Università degli Studi Roma Tre

Dipartimento di Matematica e Fisica
Doctoral program in Physics – XXXIV cycle

A thesis presented for the degree of
Doctor of Philosophy in Physics

Dynamics and structure of supercooled protein hydration water by molecular dynamics simulation

Author:
Lorenzo Tenuzzo

Supervisors:
Prof.ssa **Paola Gallo**
Dr.ssa **Gaia Camisasca**

PhD Program Coordinator:
Prof. **Giuseppe Degrassi**

january 2022

Contents

Introduction

Water is the most common liquid on Earth and it is fundamental to everyday life. We are actually made by circa 70% of water. Water is so familiar to us that one can think we already know everything about this small molecule. Nevertheless, water shows a puzzling behaviour and presents many anomalies expecially when cooled [1, 2].

Under particular conditions, water can be supercooled, i.e. it can be cooled below the melting temperature preserving its liquid state. This *supercooled water* is a metastable state and its dynamics becomes very slow, eventually not even comparable with the time scale of the experiment. In other words, supercooled water equilibrates in a very long time and this kind of system is said to undergo a glass transition [3].

As we have mentioned above, supercooled water has some unusual properties whose behaviour differs from that most of the other liquids. Thermodynamics response functions such as the isothermal compressibility and the isobaric heat capacity as well as the density are examples. In the so called glassy region, at very low temperatures, water exhibits polymorphism and is characterized by two different amourphous states. An important role in water's peculiar properties, that include also dynamical anomalies is played by the network of hydrogen bonds which characterizes the structure of water in the liquid and solid states[4–6].

Mode Coupling Theory (MCT) is the most known microscopic theory which is able to explain the behaviour of supercooled water. MCT shows that the glassy dynamics of water is due to the temporary confinement of a molecule trapped by other neighbors molecules like in a cage [6–8].

Studying supercooled water in experiments is very hard, because the crystallization time is much shorter than the usual experimental time. Some experimental methods involve the use of acqueous solutions or confinement [1]. However, computer simulations allow to study supercooled water avoiding crystallization, thanks to the tunable cooling rate and pure samples. Molecular dynamics simulations can be used to study supercooled systems, their structures and dynamics.

The aim of this thesis is to study the dynamics and the structure of the water around biomaterial, called hydration water. Here, we study the case of water surrounding a protein. This kind of water may have many dynamical and structural differences with respect to its bulk-like counterpart, due to the interaction with the protein. Understanding the properties of hydration water is particularly important in some physical applications such as cryopreservation, the technology field of preserving biomaterial for a very long time without damaging it.

The results of this work are divided in two parts. The first part focuses on the translational dynamics of lysozyme hydration water. Lysozyme is a very common globular protein. We studied the dynamics of its hydration water upon supercooling. The second part focuses on the structure of protein hydration water characterized through the calculation of several structural indicators. These results together are necessary to obtain a global pictures of the main dynamical and structural features of hydration water upon cooling.

The structure of the thesis is the following.
The **Chapter 1** is an introduction to water from the physical point of view. We present the water structure, its most known anomalies, the phase diagram and the hydrogen bond network. We give also an introduction on supercooled water. At the end of the chapter we introduce hydration water and its importance in the biological context.
The **Chapter 2** is an introduction to the Mode Coupling Theory. This is the theoretical framework in which our results on dynamics are discussed. We give a brief mathematical derivation of the main ideas of the theory with its physical interpretation in the context of the glassy dynamics of supercooled liquids.
Chapter 3 is devoted to the methods used in this work. We present the principal techniques of the Molecular Dynamics simulations: how to compute the equation of motions, how to handle pressure and temperature during a simulation and how to manage the potentials and the long distance force fields between the elements of the system. At the end of the chapter we introduce the physical observables we used to analyze the generated trajectories to study the dynamics and the structure of hydration water.
In **Chapter 4** we present our results on the translational dynamics of hydration water. We run long simulations in order to calculate self van Hove scattering functions and the mean square displacements. At the end of the chapter, we draw our conclusions on this study.
The **Chapter 5** is dedicated to the investigation of the structure of supercooled hydration water with respect to the bulk. We collect information from several structural indicators in order to characterize hydration water local order. We calculated also inherent structures, which supress thermal fluctuations, to investigate deeper in the structural conformation of hydration water with respect to the real dynamics.
In **Chapter 6** we draw our general conclusion.
Appendix A contains a list of activities, publications, schools and conferences attended during the PhD course.

Chapter 1

Liquid water

Despite the fact that water is the most common liquid on Earth, it is not a typical liquid. Water has anoumalous thermodynamical and dynamical properties, especially when supercooled. To understand water's properties it is mandatory to study its structure and its phase diagram. In this chapter we present the structure of the water molecule and the structure of the liquid water's network. The phase diagram of water is discussed and some of the most peculiar properties of water are introduced. Then, the concept of a metastable state is presented to enter in the region of the supercooled water and its phase diagram. In the end, a brief introduction to hydration water and supercooled hydration water is given.

1.1 Water structure

Water is a molecule made of three atoms: two hydrogens and one oxygen. Hydrogen atoms are covalently bonded to the oxygen in a typical V-shape molecule. The water molecule is shown in figure 1.1a: the $O-H$ bond lenght is 0.958 Å and the $H-O-H$ angle is 104.5°[9, 10]. When the water molecule is isolated, the valence electrons on the oxygen form two lone pairs that form a nearly tetrahedral structure with the two O-H bonds. Due to the difference of elettronegativity between oxygen and hydrogen, electrons are almost entirely dislocated on the oxygen, so that hydrogen atoms are strongly attracted by the oxygen and their protons are left partially unscreened. For this reason, water is a polar molecule with a permanent dipole of 1.85 D directed from the center of the negative charge to the center of the positive charge [11].
When a water molecule is sourrounded by other water molecules, its hydrogens are attracted by the partial negative charge on the other oxygens forming the so-called hydrogen bonds (HBs). This bond, usually depicted with dots, is an attractive force on the axis $O-H\cdots O$ with an energy circa of 20 kJ/mol at ambient condition. The typical tetrahedral structure is shown in figure 1.1b. The central water molecule is able to form four HBs with the other water molecules: two of them are electron-donors and the other two are electron-acceptors. The formation of HBs partially compensates the charge displacement on the $O-H$ covalent bond that is said to have a ionic character. HBs have an high degree of directionality and their energy is high compared to other intramolecular inter-

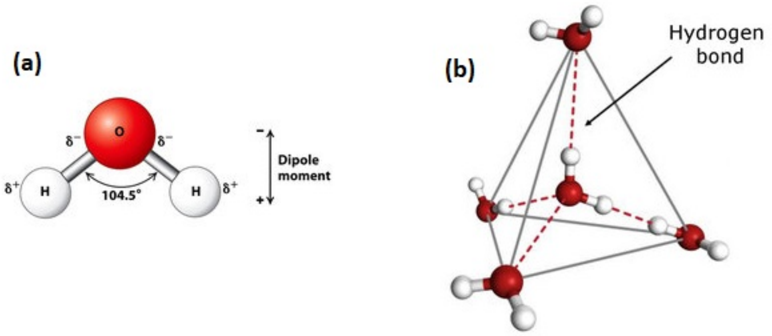

Figure 1.1: a) the structure of a water molecule and its charge distribution, b) the tetrahedral water structure made by five molecules connected by four HBs.

action like dipole-dipole van der Waals (1 kJ/mol), so water in the liquid phase is a network of HBs and most of the properties of liquid water arise from this peculiarity.

1.2 The phase diagram of water

The best way to introduce water peculiarities is to start from its phase diagram shown in figure 1.2. At ambient condition ($T \approx 298,15$ K and $P = 1$ bar) water is in the liquid phase. If we move on the horizontal line of the ambient pressure, above 373 K water becomes vapour because its stable phase is the gas phase. Below 273 K, water becomes ice because its stable phase is the solid phase. If we move to more extreme temperatures or pressures (note that the pressure scale is logarithmic) the phase diagram of stable water becomes more complex. There are, known to date, sixteen crystalline stable phases and the most common ice (hexagonal ice I_h) is just one of them. Due to this polymorphism, there are several triple points besides the well known solid-liquid-vapor triple point.
The liquid-vapor line continues above the liquid-gas critical point separating a more gas-like form from a more liquid-like form, this line is named Widom line [12–15]. The liquid-solid line has a negative slope with great implication for our lives. The Clasius Clayperon law implies therefore that when liquid water becomes ice, its volume increases. This law is defined by

$$\frac{dP}{dT} = \frac{1}{T}\frac{\lambda}{\Delta V} \tag{1.1}$$

where λ is the latent heat of fusion. Instead, most of the substances become denser when they solidify. This is called the *density anomaly* and it is one of the water anomalies that we will discuss later.

1.2.1 Metastable states

The phases depicted in the water's phase diagram are stable states, i.e. the most energetic favorable states of aggregation. A *metastable state* is a state with a finite lifetime and a higher energy with respect to the stable state: for example

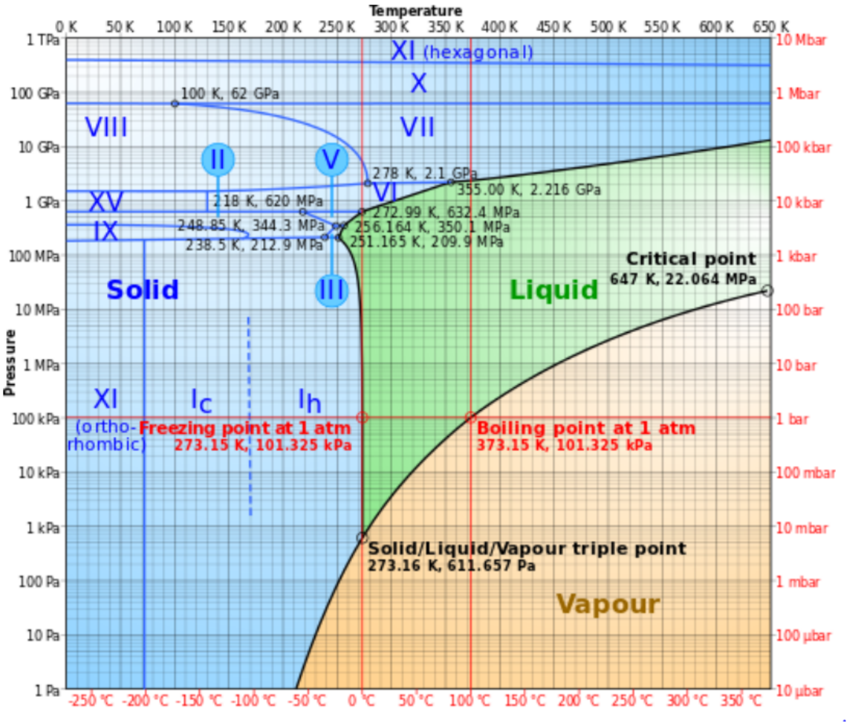

Figure 1.2: Water phase diagram in P-T plane take from ref. [16]

liquid water below the freezing point. Metastable states corresponds to a local minima in the appropriate energy landscape and the slightest pertubations can push the system towards the minimum, involving a phase change.

Water presents metastable states in the liquid phase like supercooled water, but it has also metastable ices [17, 18] and metastable amorphous solid states [19]. The phase diagram of supercooled water towards its glassy state is shown in figure 1.3. We can obtain supercooled water by cooling rapidly water under the melting temperature T_M. If we cool it down again under the homogeous nucleation temperature T_H, supercooled water quickly freeze by spontaneous thermal fluctuations. If we have a glassy state and we move from below of the diagram, heating glassy water over the glass transition temperature T_g, it enters in the ultraviscous liquid metastable phase. Heating up further leads to reach the spontaneous crystallization temperature T_x which is different from T_H. In the lowest region of the phase diagram there are three of amourphous ices: High Density Amorphous (HDA) and Low Density Amorphous (LDA) separated by a first order transition line and the Very High Density Amorphous (VHDA) [20]. Between T_x and T_H there is a region called *no man's land*, that is very hard to reach in experiments because of the fast nucleation rate. In other words, it is difficult to keep water in a metastable liquid form under T_H because the lifetime of this metastable state, not forbibben in theory, is too short to be observed experimentally. Molecular Dynamics (MD) simulations, thanks to the fast cooling rates, allows us to enter in the no man's land.

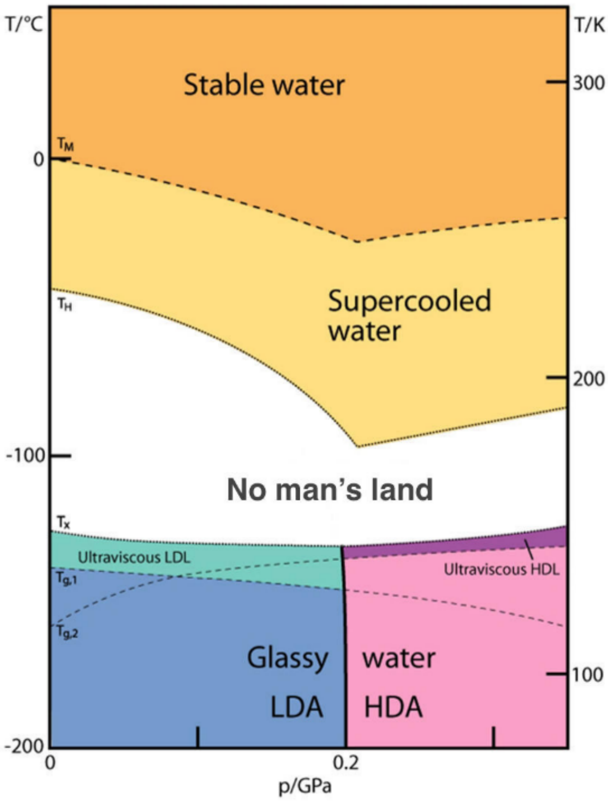

Figure 1.3: Metastable phase diagram in P-T plane of supercooled and glassy water, adapted from ref. [1]

1.2.2 Unusual properties of water

Due to the hydrogen bonds network, water shows more than 70 anomalies respect to a simple liquid [21–23]. In figure 1.4 are shown four examples of physical quantities in which water behaves differently than a simple liquid. The figure shows the temperature dependance at ambient pressure of density, thermal expansion coefficient, isothermal compressibility and isobaric specific heat. Each quantity starts at the highest temperatures like a simple liquid but it deviates from this behaviour upon cooling. Lower the temperature, greater the deviation. The three thermodynamic response functions are the thermal expansion coefficient α_P, the isothermal compressibility κ_T and the isobaric speficic heat c_P, repectively definied by

$$\begin{aligned} \alpha_P &= \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P \\ \kappa_T &= -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \\ c_P &= \frac{T}{N}\left(\frac{\partial V}{\partial T}\right)_P \end{aligned} \tag{1.2}$$

These quantities are connected to the thermodynamic fluctuations of the liquid, in particular with the fluctuations of volume and entropy

$$\begin{aligned}\langle(\delta S \delta V)\rangle &= V k_B T \alpha_P \\ \langle(\delta V^2)\rangle &= V k_B T \kappa_T \\ \langle(\delta S^2)\rangle &= N k_B c_P\end{aligned} \tag{1.3}$$

In simple liquids, lower temperatures mean smaller fluctuations, but in water the increase of the isothermal compressibility and the isobaric speficic heat correspond to an increase of the self correlation of the liquid upon cooling. In a simple liquid, the isothermal compressibility κ_T and the isobaric specific heat c_P decrease monotonically upon cooling, whereas they have a minimum in water respectively at $T = 319$ K and $T = 308$ K. Under these temperatures they show a rapid increase as seen from the figure 1.4c-d. The most known water anomaly is probably the density ρ. A simple liquid is expected to contracts when cooled and to increase monotonically its density, but water has a density maximum at T=277 K at ambient pressure. When the pressure is increased this maximum moves towads lower temperatures (see fig. 1.4a).

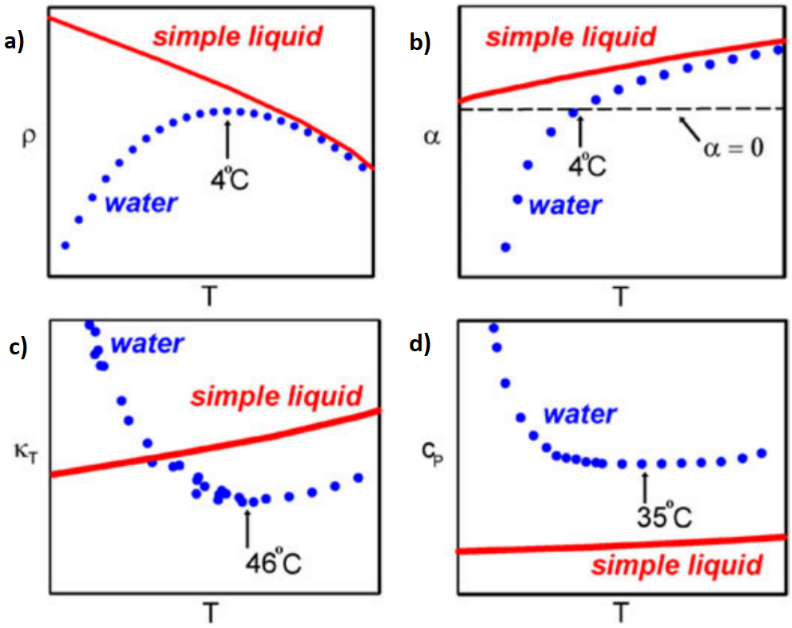

Figure 1.4: Some water anomalies compared with a simple liquid behaviour (red lines) versus the temperature at ambient pressure. a) density ρ, b) thermal expansion coefficient α_P, c) isothermal compressibility κ_T, d) isobaric specific heat c_P. Figure from ref. [1]

If we plot the density maxima versus the temperature we obtain the Temperature of Maximum Density (TMD). The isobaric expansion coefficient α_P in a simple liquid decrease monotonically while lowering the temperature but it is always positive. Instead, in water it becomes zero in correpsondance with the TMD and then it becomes negative and decreases. When we cool down

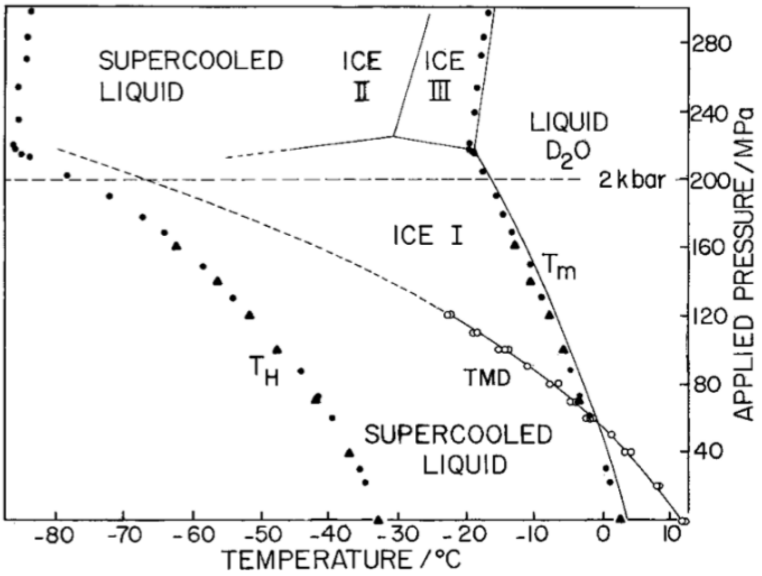

Figure 1.5: Temperature of Maximum Density for D_2O in the P-T plane. Circles are taken from experiments and the dashed line is a low temperature extrapolation, while the horizontal dashed line is the melting temperature T_M. Figure from ref. [26]

water under 278 K there is a loss of entropy accompanied by an increase of the volume (the density decreases), in opposition of what happens in simple liquids. Volume and entropy are anti-correlated and this is due to the formations of a network of hydrogen bonds that cause an increase of volume but a decrease of the orientational entropy.
The three response functions α_P, κ_T and c_P seem to diverge from the simple liquids behaviour following a power law [24, 25]

$$X \sim \left(\frac{T}{T_S} - 1\right)^{\lambda_X} \tag{1.4}$$

where X is the response function, T_S the temperature of divergence and λ_X the exponent associated with X. At ambient pressure, the estimated temperature of divergence is T_S=228 K for all the response functions. Already in 1996, it was shown by MD that T_S coincides with the MCT T_c [7]. Since $T_S < T_H$ the divergences of the response functions should be located in the no man's land. For the difficulty to reach the no man's land, these divergences are called *apparent*, simulations have shown that they are related to the Widom line [13–15] and to the fragile-to-strong crossover in dynamics [7, 8]. In figure 1.5 it is shown the experimental TMD of D_2O. This anomaly is connnected with the already mentioned negative slope of the solid-liquid coexistence line. When we cool down under the TMD, water density must show a minimum if there is no phase change. We can say that supercooled water has a density minimum because deeply supercoled water has a lower density but it increases if we cool down more the temperature. This minimum is exstimated at 203 K, located in

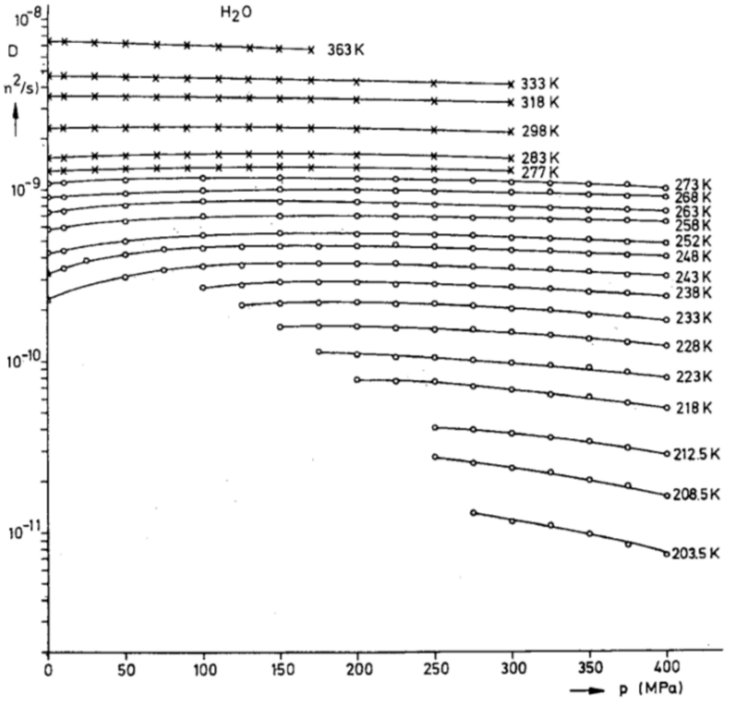

Figure 1.6: Different isotherms of experimentally measured self-diffusion coefficient D of water versus pressure. Figure from ref. [23].

the no man's land [27]. This minimum has been observed in deep supercooled water through computer simulation [**brovchenko2003multiple**], but it was observed experimentally only in confined water [28].

Another anomaly of water respect to simple liquids concerns its dynamics, in particular the diffusion coefficient D. This coefficient is defined for spherical particles by the Einstein–Smoluchowski equation

$$D = \mu k_B T \tag{1.5}$$

where $\mu = \frac{1}{6\pi\eta r}$ is the mobility of the particle, r its radius and η is the viscosity coefficient. In simple liquids, the compression is usually related with an increase of viscosity and consequently to the loss of mobility: D should decreases when increasing the pressure. In figure 1.6 the self-diffusion coeffcient of water, measured at $T \sim 283$ K, increases with the pressure until $P \approx 150$ MPa where it has a maximum. This anomaly is also related with the hydrogen bonds network because when the pressure is higher the number of hydrogen bonds decreases and the total mobility of particles raises.

The dynamic and thermodynamic anomalies just described are related to some structural anomalies often investigated in theory and in simulations [29–31]. Two order parameters are taken into account to describe the structural anomalies: the tetrahedral order parameter q, which measures the degree of tetrahedrality of the four first neighbors of a water molecule and the translational parameter t which measures the prevalent separation between two water molecules. These two parameters decrease upon compression on the contrary to what happens in normal liquids where the order increases upon compression. Firstly, the relationship between dynamic, thermodynamic and structural anomalies was studied by Errington and Debenedetti [29]. In figure 1.7 is shown the relation between water anomalies in two simulation potential: a) SPC/E in $\rho - T$ plane and b) TIP4P/2005 in p-T plane. In figure 1.7a, the structural anomalous region is the lighest grey region, bounded by q maxima and t minima. In this region, water becomes more disordered when compressed. Inside this region, there is the dynamical anomalous region (middle grey) bounded by diffusivity minima and maxima. In this region, the diffusion coefficient increases when density increases. Finally, the inner region (darkest grey) is the themodynamical anomalous region, bounded by the TMD. In this region, the density decreases upon cooling at constant pressure. The water anomalies involve each other in a domino effect that is more pronounced in the deep supercooled region.

1.3 Supercooled water

The anomalies we described earlier are even more pronounced when water is supercooled. As already described, supercooled water is a metastable state where water is in its liquid phase under the melting temperature T_M. This state has a finite lifetime because it is represented by an out-of-equilibrium local minimum in the energy where the system is trapped for a certain amount of time.

Nonetheless, we can consider a metastable state stable when its life time τ_{life} is greater than the time of observation τ_{obs}. In this way we are able to probe its properties. Moreover, this time should be much greater than the relaxation time τ_{rel}, i.e. the time the system needs to equilibrate and to rearrange. So, we can reassume

$$\tau_{rel} \ll \tau_{obs} < \tau_{life}.$$

Experiments with supercooled water are particularly challenging because the crystallization problem that must be avoided. To extend the lifetime of the supercooled liquid, the sample should be pure and there should be no mechanical or thermal pertubations. For example, the sample can be put in a smooth-wall container.

However, after τ_{life} the liquid will crystallize in a stable solid form or it will go towards a glassification process. Which destiny, depends on the nucleation time τ_1 and the relaxation time τ_2. The nucleation time is the time a certain amount of volume needs to crystallize.

These times are depicted in function of the temperature in figures 1.8 and 1.9. The nucleation time has a minimum between the two divergence temperature T_M and T_0 ($T = 0$ K) while τ_2 is a monotonic function of temperature and it diverges at T_g. When τ_2 is similar to τ_{obs} the system seems stuck . It should be chosen a cooling schedule with a right cooling rate to avoid both crystallization and glass formation. If the cooling rate is too low the system will crystallize, while if it is too high the system will end into the glassy state.

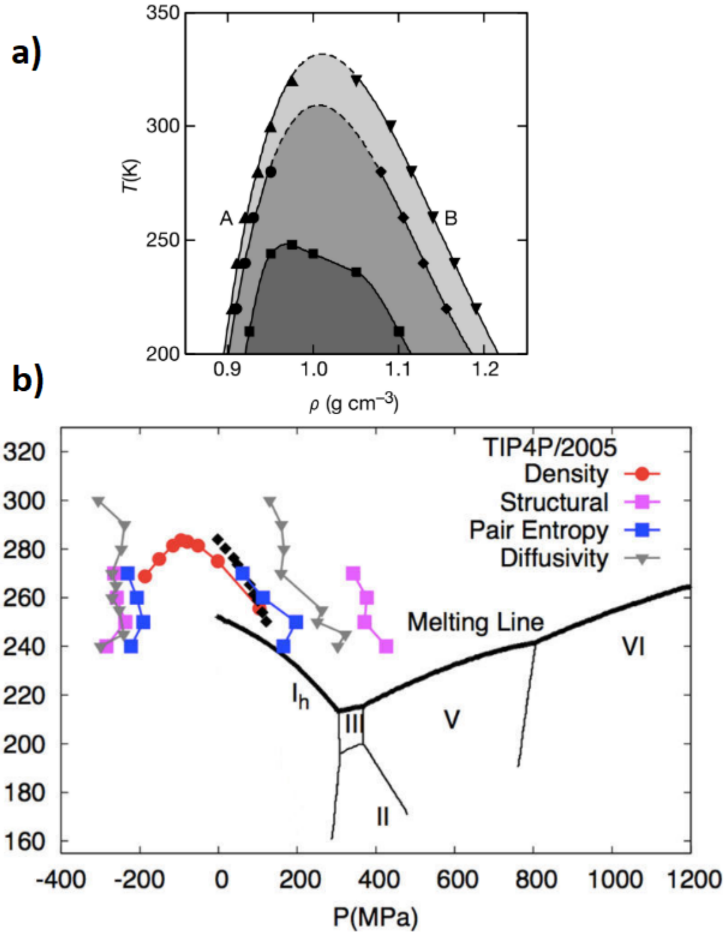

Figure 1.7: Relationship between water anomalies. a) SPC/E water in T versus density plane and b) TIP4P/2005 water in T versus pressure plane. The structural anomalous region is limited by q maxima on the left by upward triangles (a) and pink squares (b) and by t minima on the right by downward triangles (a) and pink squares (b). The dynamical anomalous region is limited by diffusivity minima on the left by circles (a) and grey downward triangles (b) and by diffusivity maxima on the right by diamonds (a) and grey downward triangles (b). The thermodynamical anomalous region is limited by the TMD squares (a) and red circles (b). Figures from ref. [29].

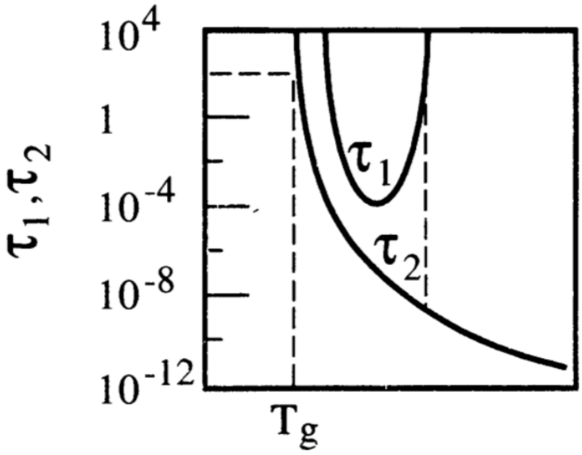

Figure 1.8: Nucleation time τ_1 and relaxation time τ_2 versus temperature. T_g is the glass transition temperature. Figure from ref. [22].

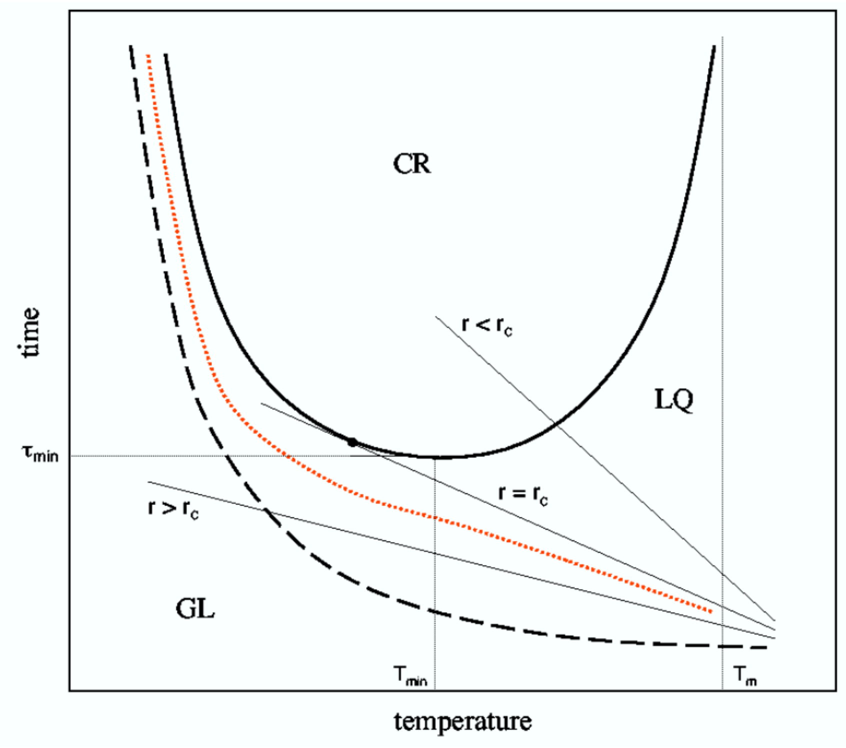

Figure 1.9: Nucleation time τ_1 (continuous line) and relaxation time τ_2 (dashed line) versus temperature. T_m is the melting temperature and T_{min} is the temperature of the minimum in τ_1. Straigh lines are cooling schedules with different cooling rates and dotted red line is a non linear cooling schedule. CR means crystal, LQ means liquid and GL is for glass. Figure from ref. [32].

1.3.1 The Liquid-Liquid Critical Point

The glassification schedule is called *quenching*. The value of T_g is still under debate: it is defined as the temperature where the system reaches a viscosity degree of $\eta = 10^{13}$ Poise, but it varies slighly using different cooling procedures. it is commonly accepted that water glassifies at 136 K [33, 34], but more recent researchs hypothesized 165 K [35, 36].

We already mention the polymorphic nature of the lowest temperature region of the phase plot with the three amorphous ices LDA (for low pressure), HDA (for high pressure) and VHDA (for very high pressure). LDA and HDA ices are separated by a first order phase transition line discovered in 1985 by Mishima and collegues [37] and confirmed later [38, 39]. Several theories have tried to explain the cohexistence of two amorphous phases in the glassy state region and the anomalies of water. The most known are the *stability limit conjecture* [24], the *Liquid-Liquid Critical Point* (LLCP) scenario [40], the *critical point free* scenario [5], *the singularity free* scenario [41]. Here, we focus on the LLCP, since evidences have accumaletd pointing to this scenario [42, 43].

The existence of a LLCP was presented in the molecular dynamics simulation work [40] by Poole *et al.* where water was studied with a five sites rigid potential called ST2. The proofs of the existence of a LLCP came from four observations:

i low-temperature isotherms have an inflection point at $\rho_C \approx 1\ g/cm^3$;

ii two distinct phases were observed below the LLCP, one similar to the LDA and the other similar to HDA;

iii simulation of a isothermal compression showed the phase transition HDL-LDL and validation of the ST2 model with the transition HDA-LDA at 77 K;

iv ST2 water's model was validated with experimental results above the no man's land

The LLCP scenario is summarized in figure 1.10. In this scenario, water has two liquid phases in the deep supercooled region and they are called respectively Low Density Liquid (LDL) and High Density Liquid (HDL). they are the corresponding higher temperatures states of the amourphous crystalline phases LDA and HDA. They differ in the local struture of the HBs network. LDL has a lower density because the water molecules form a well defined tetrahedral cell made by five molecules and four HBs. HDL has an higher density because a sixth molecule approaches the tetrahedral cell resulting in a more distorted local structure. Like their crystalline counterparts, these two liquid phases are separated by a first order phase transition line. The LLCP is out the end of this coexistence line and it is located in the no man's land. Subsequent works of Poole *et al.* locate the LLCP in ST2 water at $T_C \approx 235$ K, $P_C \approx 200$ MPa, $\rho_C = 1$ g/cm^3. In analogy with the Widom line emanated from the liquid-gas critical point, a Widom line is emanated from the LLCP towards the one-phase regione. This is a line of maxima for the thermodynamic response functions [13]. In this scheme, the anomalous properties of water are explained by the presence of a LLCP. The (apparent) divergences in the thermodynamic response functions can be explained by the increase in the correlation lenght when approaching the Widom line. At higher temperatures, anomalous properties of

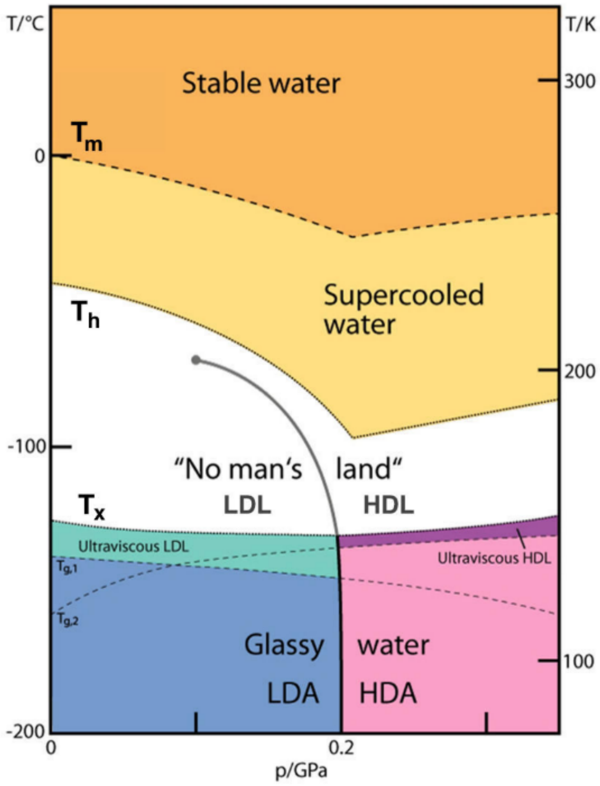

Figure 1.10: Phase diagram of water in its liquid phase showing the LLCP and the first order transition line between LDA and HDA. This line continues at higher temperatures and it separates LDL and HDL in the no man's land. The slope of this line is negative so LDL is expected to be less denser than HDL. T_H line is the homogeneous ice nucleation temperature (from above) and T_X is the spontaneous crystallization temperature (from below). Figure adapted from ref. [1].

water can be explained with the print of the LLCP in long range fluctuations. Despite the difficulties to experimentally reach the no man's land and probe the existence of the LLCP, a decisive work was performed by Mishima and Stanley [44]. They studied the difference between melting in different crystalline water phases by decompression and they located the LLCP at $T_C = 220$ K, $P_C = 100$ MPa, close to the top bound of the no man's land as showed in figure 1.10. This is considered the best estimated position of the LLCP in bulk water and was subsequently confirmed by other experimental works [42, 43, 45].

The implications of a LLCP carried by the scientific community were reviewed by Gallo et al. [1, 46, 47]. It rose clear the importance of computer simulation in this field of study. The LLCP was hypothesized in a computational work, and molecular dynamic simulations overcome the hampering of the nuclueation in experiments and they allow to enter in the no man's land. The LLCP most recent location through computation is in the supercooled region of the phase diagram.

Water was simulated with different potentials including TIP4P and TIP5P (variants of ST2) and SPC/E. Simulations rigorously proved the existence of a LLCP in ST2, TIP4P/2005 and SPC/E [48–52]. Also experimental works are being carried considering previous computational analysis, especially to detect the liquid-liquid transition. Following the example of molecular dynamics, several experiments were performed on confined water, nano droplets, water solutions of salt and water at negative pressure. Detecting the Widom line is important because when cool down to the no man's land upper bound, the Widom line is the first feature one encounter that attest the presence of the LLCP.

1.3.2 Supercooled hydration water

Water is essential for life; it is contained in living cells and it is the natural matrix for all biosystems. Water plays a key role in the exixtence and in the correct functioning of biomolecules [53]. We are interested in supercooled water in contact with biomolecules, like proteins. This kind of water is called Hydration Water (HW) and it is fundamental for the correct functioning of the protein at low temperature. These water molecules directly mediate interactions among biomolecules [54–56].

It has been found that the structure of the hydration water could be strongly modified by the interaction with the substrate due to the formation of hydrogen bonds with the hydrophilic part of the macromolecule [57]. Also, the dynamical properties are affected by the presence of the biomaterial as shown in experimental studies [58, 59] and MD simulations [60]. From a dynamical point of view, the main result is that water in bio-systems has two relaxation mechanisms, happening on two different time-scales [61]. The slower process is missing in bulk water and was consequently ascribed to the hydration water only. We will discussin details the the dynamics of HW in the fourth chapter, before presenting our results.

The protein has different conformations and sometimes has to change its shape to assolve its biological role. This dynamical ability is connected to the coupling with its hydration water. Infact, the protein flexibility is strongly suppressed at low temperatures. This dynamical properties concern only hydrated proteins and it is called Protein Dynamical Transition (PDT). This feature is not present in dried protein, so it is totally lead by hydration water.

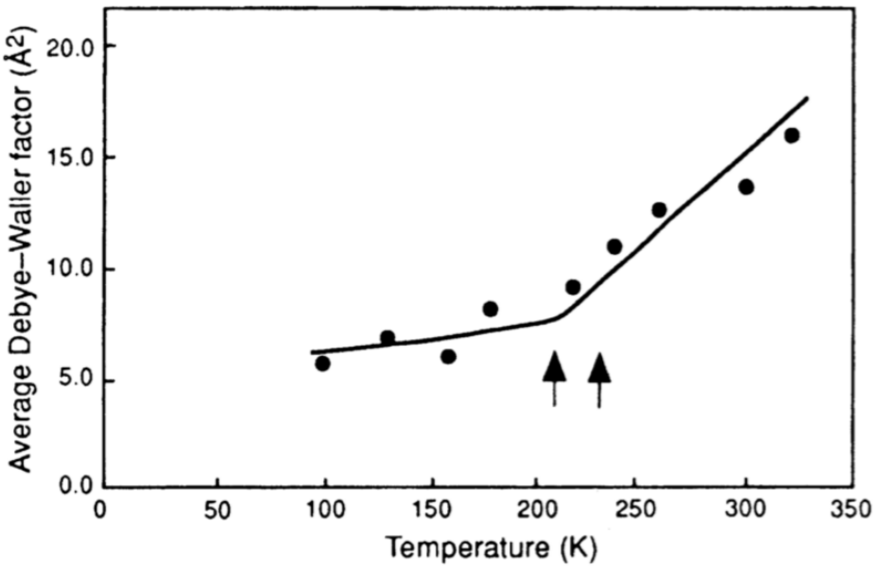

Figure 1.11: Average atomic Mean Square Factor versus temperature in the protein ribonuclease A. Figure from ref. [63].

This feature can be measured monitoring the mean square displacement of the protein atoms $\langle x^2(T)\rangle$. This quantity can be written as a sum of a vibrational part and a comformational part

$$\langle x^2(T)\rangle = \langle x^2(T)\rangle_v + \langle x^2(T)\rangle_c. \tag{1.6}$$

The vibrational part $\langle x^2(T)\rangle_v$ is dominant below the PDT and it is linear with the temperature. This part is also dominant for small biomolecules and harmonic solids [62]. The conformational contribution is observed only when the protein is hydrated and $\langle x^2(T)\rangle$ deviates from its linear behaviour. It is thought that the activation of the protein surface contributes with the conformational part of the atomic mean square displacement giving to the protein an extra mobility. In figure 1.11 is shown the fenomenon of the PDT for the ribonuclease A with the low and high temperture regimes. The PDT is close to 220 K. Above this temperature, the protein dynamics is dominated by collective anharmonic motions [63]. Moreover, above the PDT many proteins recover their biochemical activity. For examples, enzymes quickly bind substrates or inhibitors [63]. Many other experiments have observed the PDT with different techniques like the terahertz dielectric response [64], neutron scattering [65–67] and MD simulations [66, 68–71]. PDT was found in the interval 200-240 K for different biological macromolecules like membranes, DNA, RNA, etc.

Studying supercooled water at contact with biomolecules are of great relevance for cryoprotection [72, 73]. To cryopreservate a biological system means cooling it and slowing down its biological process avoiding damages due to the ice formation. For this purpose, water is used in solutions with cryoprotectors. One of the most used cryoportectors are the disaccharides, widely used as excipients because they help the stabilization of protein and cells in water solution

[74–77]. Trehalose is a non-reducing sugar and it is probably the most effective disaccharide used as a cryoprotector [78–80]. Its utility as a cryoprotector has been investigated in many studies [58, 81–84] especially in comparison to other disaccharides [60, 85–90]. The effectiveness of trehalose as a cryoprotector is mostly due to its ability to destroy the water network of hydrogen bonds. This action prevent the formation of ice bonding with water hydrogen. Moreover, trehalose slow down the water dynamics more than other sugars.

Chapter 2

The slow dynamics of supercooled water

Supercooled and glassy liquids have a very slow dynamics and this is probably their most impressive characteristic. The glass transition is indeed a dynamical feature composed by two slow relaxations with different timescales [32, 91, 92]. The relaxations can be described by correlators which stretch over several orders of magnitude of time. This chapter is devoted to introduce the Mode Coupling Theory (MCT) of the glassy dynamics [93]. A theory able to predict the dynamical behaviour of most glass formers in the region of mild supercooling. At the end of this chapter we introduce the problem of the fragile-to-strong crossover in supercooled liquids.

2.1 The Mode Coupling Theory

The MCT for glass forming liquids was formulated and devoleped by Göetze [93] and it is based on the projection formalism of Mori and Zwanzig [91, 92]. It affirms that the single particle interact with its surrounding environment and this interaction is mutual and non-linear. The consequent feedback mechanism leads to an increase of the time scale of the structural relaxation of the system. As the temperature decrease, this mechanism becomes stronger and stronger, eventually causing structural arrest and the system undergoes a glass transition. We summarize here a series of results of the MCT. Let us start considering a function $A(t)$ of positions $\bar{r}_i$ and momenta $\bar{p}_i$. The system is composed by N particles. The time evolution of $A(t)$ is given by

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \{A(t), \mathcal{H}\} \equiv i\mathcal{L}A(t) \tag{2.1}$$

where $\mathcal{L}$ and $\mathcal{H}$ are respectively the lagrangian and the hamiltonian of the system. The projection operator $\mathcal{P}$ is defined as

$$\mathcal{P} \equiv (A, ...)(A, A)^{-1}A \tag{2.2}$$

where $(A, B) \equiv \langle BA^{*}\rangle$ is the equilibrium correlation function. This operator project a generic quantity in the A space, i.e. it extracts the A component from

a variable.
For example, if A is a variable that varies slow in time, the $\mathcal{P}$ projector applied on B extract only the slow component from B. We can rewrite the eq. 2.1 as

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t} = i\Omega \cdot A(t) - \int_0^t K(\tau) \cdot A(t-\tau) d\tau + f(t) \tag{2.3}$$

where $f(t)$ is the *fluctuating random force*, defined as

$$f(t) \equiv e^{i(1-\mathcal{P})\mathcal{L}t} i(1-\mathcal{P})\mathcal{L}A \tag{2.4}$$

and K is the *memory function matrix*

$$K \equiv (f, f(t)) \cdot (A, A)^{-1} \tag{2.5}$$

and at the end $i\Omega$ is the *characteristic frequency matrix*

$$i\Omega \equiv (A, i\mathcal{L}A) \cdot (A, A)^{-1} \tag{2.6}$$

The initial values of $A(t)$ and $f(t)$ are $A = A(0)$ and $f = f(0)$.
The eq. 2.3 is an exact equation for the time evolution of $A(t)$ and it is called the *Langevin equation*. It is composed by two part: a slow part and a fast part. The fast part of $\dot{A}$ is the fluctuating random force $f(t)$. It is the time evolution of A in fast subspace orthogonal to A. The slow part of $\dot{A}$ is composed by the memory function K, given by the autocorrelation of $f(t)$, and by the frequency matrix. These components are the the part of $\dot{A}$ in the slow varying subspace.
The correlation matrix is defined as

$$C(t) \equiv \langle A(t) A^* \rangle = (A, A(t)) \tag{2.7}$$

We can put the last equation in the eq. 2.3 to obtain the equation of motion for the correlation matrix

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = i\Omega \cdot C(t) - \int_0^t K(\tau) \cdot C(t-\tau) d\tau \tag{2.8}$$

the term with the fluctuating force $f(t)$ disappeared because the orthogonality $(A, f(t)) = 0$. In eq. 2.8 the memory function K is a non Markovian term, non local in time and it takes into account the history of the particle trajectory. The eq. 2.8 is exact for a generic correlation function.
We can chose the variable A as

$$A = \begin{bmatrix} \delta\rho_q \\ j_q^L \end{bmatrix} \tag{2.9}$$

where $\delta\rho_q$ is the Fourier Transform of the density defined as

$$\delta\rho_q = \sum_i^N e^{i\overline{q} \cdot \overline{r}_i} - (2\pi)^3 \rho \delta(\overline{q}) \tag{2.10}$$

and $\rho = N/V$ is the uniform density. j_q^L is the longitudinal component of the associated current

$$j_q^L = \frac{1}{m} \sum_i^N (\overline{q} \cdot \overline{p}_i) e^{i\overline{q} \cdot \overline{r}_i} \tag{2.11}$$

where m is the particle mass. With this defitinition of A, the correlation matrix becomes

$$C(t) = \begin{bmatrix} \langle \delta\rho_{-q}\delta\rho_q(t)\rangle & \langle \delta\rho_{-q} j_q^L(t)\rangle \\ \langle j_{-q}^L \delta\rho_q(t)\rangle & \langle j_{-q}^L j_q^L(t)\rangle \end{bmatrix} \tag{2.12}$$

and the frequency matrix becomes

$$i\Omega = \begin{bmatrix} 0 & iq \\ i\frac{qk_BT}{mS(q)} & 0 \end{bmatrix} \tag{2.13}$$

where $S(q)$ is the static structure factor. The memory function becomes

$$K(\overline{q}, t) = \begin{bmatrix} 0 & 0 \\ 0 & \frac{m\langle R_{-q}R_q(t)\rangle}{Nk_BT} \end{bmatrix} \tag{2.14}$$

with

$$R_q = \frac{\mathrm{d}j_q^L}{\mathrm{d}t} - i\frac{qk_BT}{mS(q)}\delta\rho_q \tag{2.15}$$

If we take in the eq. 2.8 the lower left components of the matrices, we obtain the equation of motion for the Intermediate Scattering Function $F(\overline{q}, t)$, i.e. the density-density correlation function defined as

$$F(\overline{q}, t) = \frac{1}{N}\langle \delta\rho_q(t)\delta\rho_{-q}\rangle. \tag{2.16}$$

The equation of motion is

$$\frac{\mathrm{d}^2F(\overline{q}, t)}{\mathrm{d}t^2} + \frac{q^2k_BT}{mS(q)}F(\overline{q}, t) + \int_0^t K(\overline{q}, t-\tau)\frac{d}{dt}F(\overline{q}, t-\tau)d\tau = 0 \tag{2.17}$$

where $S(q)$ is the Static Structure factor.
Eq. 2.17 has the form of an harmonic oscillator with frequency $\Omega = \frac{q^2k_BT}{mS(q)}$ dumped by a frictional force proportional to the memory function $K(\overline{q}, t-\tau)$ and time-retarded. We don't show it here, but in the memory functions are products of the density ρ_q [94]. Equation 2.17 is an exact equation of motion but we need to do some semplification in order to solve it analytically.
The MCT is based on two approximation [95, 96]:

- the memory function can be approximated into a four-point density correlation function through the projection of the random fluctuating force onto a subspace with a new basis made by two density modes;
- Four-point correlation functions are factorized into two-point correlation function to have a simpler memory function that is not zero.

The resulting memory function is

$$K(\overline{q}, t) = \frac{\rho^2k_BT}{2Nm}\sum_K |V_{q-k,k}|^2 F(\overline{k}, t)F(\overline{q}-\overline{k}, t) \tag{2.18}$$

the $V_{q-k,k}$ are called *vertices* and they are defined as

$$V_{q-k,k} \equiv (\hat{\overline{q}}\cdot\overline{k})c(\overline{k}) + \hat{\overline{q}}\cdot(\overline{q}-\overline{k})c(|\overline{q}-\overline{k}|) \tag{2.19}$$

where $c(k)$ is defined as [97]

$$c(k) = \frac{1}{\rho}\left(1 - \frac{1}{S(k)}\right). \tag{2.20}$$

Equation 2.18 is the long time contribution to the total memory function, the most important one in the supercooled region.
The instantaneous collision between particles gives rise to the short time contribution

$$K^{(0)}(\overline{q}, t) = \gamma_q \delta(t). \tag{2.21}$$

The last two pieces can be put together

$$K(\overline{q}, t) = \gamma_q \delta(t) + \Omega_q^2 K^L(\overline{q}, t) \tag{2.22}$$

where $K^L(\overline{q}, t)$ is given by eq.2.18.
The final equation of motion for the correlator $F(\overline{q}, t)$ is

$$\frac{d^2 F(\overline{q}, t)}{dt^2} + \Omega_q^2 F(\overline{q}, t) + \gamma_q \frac{d}{dt} F(\overline{q}, t) + \Omega_q^2 \int_0^t K^L(\overline{q}, t-\tau) \frac{d}{dt} F(\overline{q}, t-\tau) d\tau = 0 \tag{2.23}$$

We can calculate and discuss the evolution of the correlator $F(\overline{q}, t)$ from the set of coupled equations 2.18 and 2.23. Before introducing the solutions we note that:

i MCT can be applied to all those correlators with an overlap with the density ρ;

ii the only input in the MCT equations is the Static Structure Factor (SSF) $S(\overline{q})$. Small changes in this quantity produce big changes in the system's dynamics.

iii the temperature is included in the SSF;

iv the main contribution to the slow dynamics comes from the non linear feedback in the memory function.

The main contribution to the vertex in the eq.2.19 comes from the SSF for $\overline{q} = \overline{q}_{max}$ [98]. It follows that another approximation can be done by replacing the sum on the wavevector space with the single value $\overline{q}_{max}$

$$\frac{d^2 F(t)}{dt^2} + \Omega^2 F(t) + \gamma \frac{d}{dt} F(t) + \Omega^2 \lambda \int_0^t F^2(t-\tau) \frac{d}{dt} F(t-\tau) d\tau = 0 \tag{2.24}$$

where λ is the exponent parameter, $F(t) = F(\overline{q}_{max}, t)$, λ and Ω play respectively the role of the density ρ and the temperature T in the full MCT. With this approach we are in the *schematic* MCT. This is a useful simplified version of the MCT because it preserves all the important features of the full MCT. Schematic MCT is able to predict different scenarios: given some values of λ and Ω the system will pass to a non ergodic phase and it won't relax anymore (fig. 2.1 curve C).In figure 2.1 is shown a schematic representation of the time evolution of the correlator for different values of the exponent parameter.
When the exponent parameter is smaller than a critical value $\lambda < \lambda_C$, the

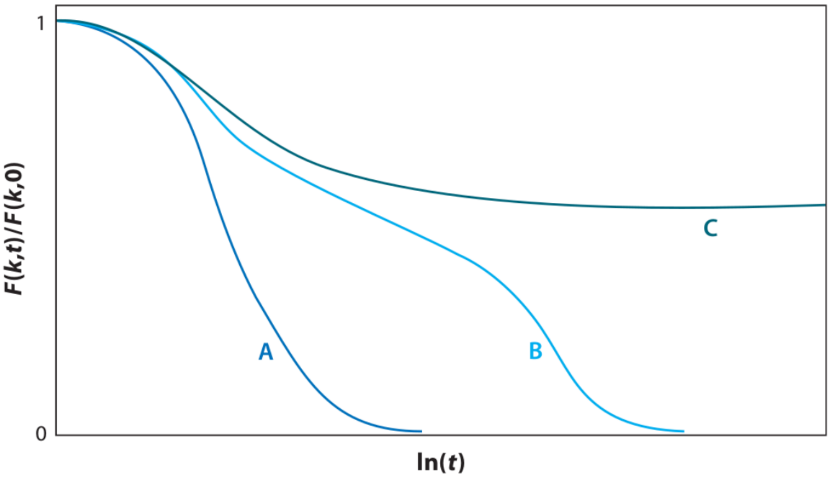

Figure 2.1: Sketch of a correlator time evolution as predicted by the Mode Coupling Theory. Curve A is the decay of the correlator in a hot liquid, curve B is the stretched dynamics of a mild supercooled liquid and curve C represents a non ergodic phase where the system isn't able to relax anymore. Figure from ref. [99].

correlators decay to zero as in the curves A and B of figure . When λ is close to λ_C the dynamics slow down and the correlator stretches in time and taking the peculiar shape of the glass transition with two relaxations. The curve B refers to this case: the first decay is called β-relaxation, after that there is a plateau and then a second long time decay (note that the time scale is logarithmic) called α-relaxation. In the curve C we have that $\lambda \geq \lambda_C$ and the correlator no longer decays. The system has lost egodicity and the final phase is the completely arrested phase, the glass. The value of the plateau is called *non ergodicity parameter*, defined by

$$\lim_{t\to\infty} F(t) = f. \tag{2.25}$$

In the full MCT, we can write the plateu value as f_q because the $\overline{q}$ dependance is restored. MCT is able to predict the time evolution of the correlator and its decays in a quantitive way: in the β-relaxation regime, when the correlator approach f_q, the time scale will behave like

$$t_\epsilon = t_0 |\epsilon|^{\frac{1}{2a}} \tag{2.26}$$

where t_0 is a microscopic time and ϵ is the distance parameter defined as

$$\epsilon = \frac{T - T_{MCT}}{T_{MCT}} \tag{2.27}$$

with T_{MCT} temperature at which the system will no longer relax. In the first part of the β regime, the correlator follows a power law behaviour

$$F(\overline{q}, t) - f_q \propto t^{-a} \tag{2.28}$$

and also at the end of the β regime it follows a power law

$$f_q - F(\overline{q}, t) \propto t^b. \tag{2.29}$$

Instead, the α-relaxation regime, that starts from the plateau region and ends to zero, can be described with a stretched exponential called Kohlrausch-Williams-Watts function

$$e^{-\left(\frac{t}{\tau_\alpha}\right)^\beta} \tag{2.30}$$

where β is the stretching parameter and τ_α is the characteristic time of the α-relaxation. According to the MCT predictions, τ_α behave like

$$\tau_\alpha = t_0 |\epsilon|^{-\gamma} \tag{2.31}$$

and it diverges at the T_{MCT}. The exponents a, b and γ are related by the relations

$$\frac{\Gamma^2(1+b)}{\Gamma(1+2b)} = \frac{\Gamma^2(1-a)}{\Gamma(1-2a)} \tag{2.32}$$

with Γ gamma function and

$$\gamma = \frac{1}{2a} + \frac{1}{2b} \tag{2.33}$$

The exponents a and b are also limited in the range $0 < a < 0.5$ and $0 < b < 1$. If we scale the correlator for the α characteristic time $\tau_\alpha(T)$ (it is T dependent), the correlator collapse in last part of the α decay into a master curve Φ

$$\Phi(t) = F\left(\frac{t}{\tau_\alpha}\right) \tag{2.34}$$

and from Φ we can extract the exponent b. This is the *Time-Temperature Superposition Principle.*
Now, we summarize the typical stages of the double relaxations mechanism shown in figure 2.2 and we introduce the concept of the *cage effect*:

1 at the very beginning, at the start of the β-relaxation, particles behave like free particles. In this regime,a particle doeas not interact with any other particle, so this is called *ballistic regime*;

2 As time passes, the particle explores more space and start feeling the environments. At a certain point, if the temperature is low enough, it will be surrounded and blocked in its course by its nearest neighbors. The particle will spend a certain amount of time trapped in this 'cage'. This is the cage effect and corresponds to the plateau region. This phase is called *cage regime*;

2a In the previous phase there is a collective rearragement of all the particles. If the temperature is decreased again, the dynamics is slowered and the time that the particle spend in the cage will increase, so the plauteu regione will be longer in time;

3 for longer times, the cage breaks and the α-relaxation can take place. Then, the usual diffusive regime is restored. This is called the *alpha relaxation regime.*

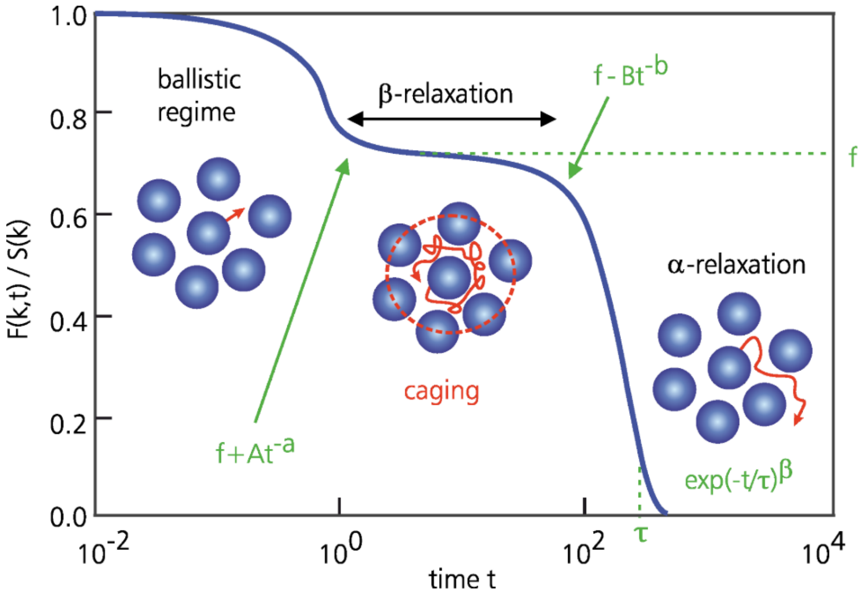

Figure 2.2: Double relaxation of the correlator predicted by the MCT for a supercooled liquid with the sketches of the different regimes. Figure from ref. [95].

What we just described is the microscopic mechanism of vitrification. MCT equations describe this mechanism with the approximations we have done, i.e. that most of the contribution to the memory function by the Static Structure Factor comes from the first peak at $q_{max} \sim 2.25$ Å^{-1} (about the cage's size). The interactions of a particle with its first neighbors (inter cage interactions) are the local interactions and they are the most important for a global structural arrest.

The MCT have been tested for many systems, as hard spheres [93] and simple liquids [100, 101], but it works well also with water [7, 8, 102]. The main features of mild supercooled water are well described by MCT. One of the most used correlator to describe the time evolution of a system is the Self Intermediate Scattering Function (SISF). This is the part of the density correlator introduced in eq. 2.16 that focuses on the single particle dynamics and what we will describe in details in the next chapter with the others dynamical and structural quantities of interest for this thesis.

The SISF is the spatial Fourie transform of the single particle density-density autocorrelation function and it can be fitted with a two exponential function [7, 8]

$$F(q,t) = (1-f_\alpha)e^{-\left(\frac{t}{\tau_{short}}\right)^2} + f_\alpha e^{-\left(\frac{t}{\tau_\alpha}\right)^{\beta_\alpha}} \tag{2.35}$$

where the first term with τ_{short} is gaussian and describe the ballistic regime, the second term with the stretched exponential account for the α-relaxation. In figure 2.3 the oxygen-oxygen SISF is shown for simulated SPC/E water. The two steps relaxation correlator are fitted with the eq. 2.35 (lines) and the first test of MCT behaviour for supercooled water it can be seen that the fuction 2.35

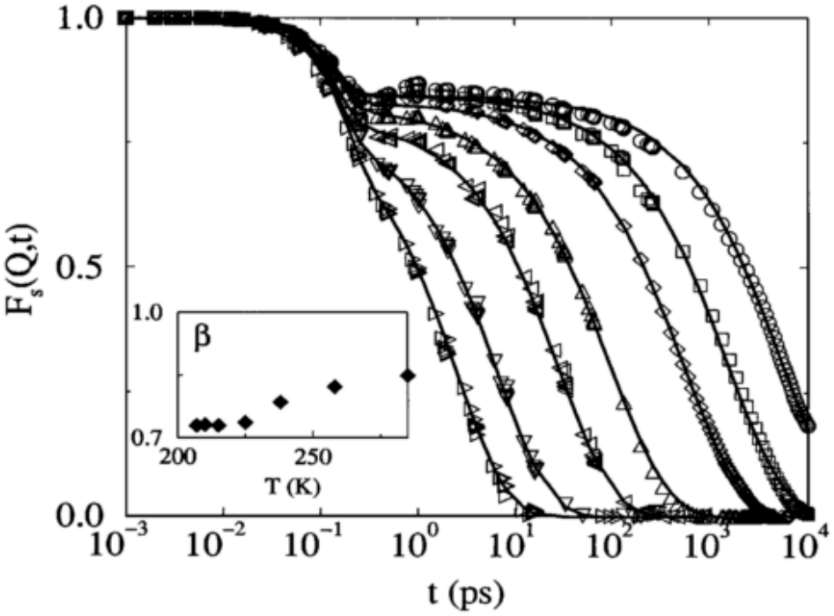

Figure 2.3: Self Intermediate Scatteriing Function of oxygen-oxygen calculated for SPC/E water. Black lines are the best fits of the data made with eq. 2.35. The inset shows the values of the stretching parameter β. Figure from ref. [7].

describes well the data. The theory predicts a complete arrest of the system with a characteristic relaxation time which diverges as a power law, written in eq. 2.31 for a certain critical value of the temperature T_{MCT}. When the system transits to a glassy state, the correlator no longer decays and mantains the plateau value.
In most glass formers that the correlator decays to zero below T_{MCT}. There is a mechanism not taken into account by the schemtic MCT that restores the ergodicity and avoid the complete structural arrest. These microscopic mechanisms are called activated processes and within the supercooling community are called *hopping* processes. The particle is allowed to escape the cage of the first nighbors by hopping to some preferred positions. This mechanism is connected to another phenomenon known as *fragile-to-strong crossover* that will be discussed in the next section.

2.2 Strong and fragile liquids

Angell proposed a classification of supercooled liquids in two groups: strong liquids and fragile liquids [103, 104]. He plotted the viscosity (or the logarithm of the relaxation time) versus the temperature scaled for the glass transition temperature T_g and two different behaviour came out. In the Angell's plot (fig. 2.4), associated to the activated processes (straight lines) are called *strong* liquids following Arrhenius behaviour. The liquids that deviate from the Arrhenius behaviour are called the *fragile* liquids. The dynamics in strong liquids is dominated by hopping phenomena. Their relaxation time (or viscosity) is

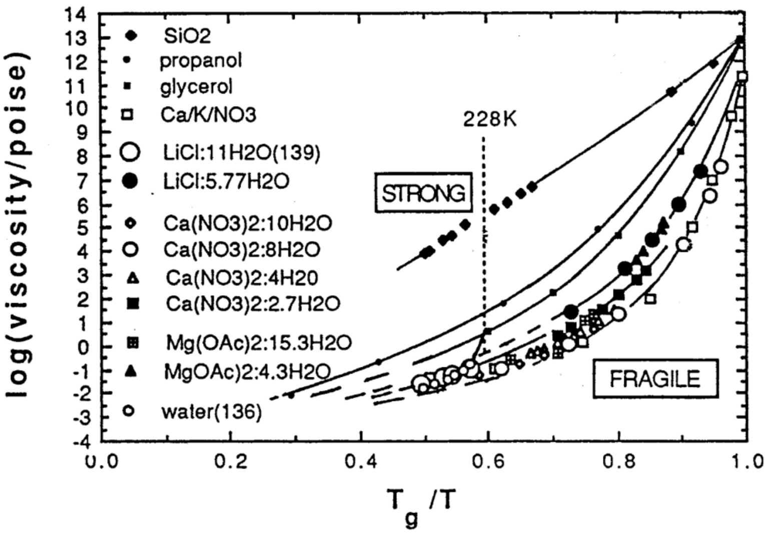

Figure 2.4: Angell plot of viscosity for many glassy liquids. Figure from ref. [36].

described by the Arrhenius law

$$\tau_\alpha = \tau_0 e^{\frac{E_A}{k_B T}} \tag{2.36}$$

where E_A is the activation energy.
Instead, relaxation times transport in fragile liquids are described by the power law

$$\tau_\alpha = \tau_0 \frac{1}{(T - T_{MCT})^\gamma} \tag{2.37}$$

following the MCT. Fragile liquids can be described also by the phenomenologically Vogel-Fulcher-Tamman law

$$\tau_\alpha = \tau_0 e^{\frac{DT_0}{T - T_0}} \tag{2.38}$$

where D is the fragility parameter and T_0 is the ideal glass transition temperature. As suggested by the name, this temperature is different from an experimentally measured real glass temperature.
All these temperatures we mentioned are not coincident. they are the MCT critical temperature T_{MCT}, the ideal glass transition temperature T_0, the real glass transition temperature T_g respect the following order [105]

$$T_0 < T_g < T_{MCT} < T_{melting}.$$

Fragile-to-Strong Crossover (FSC) was first found in supercooled water in an MD study[106] and confirmed in experiments and simulation in confinement [107] and in simulation in bulk [102] and in solutions [108]. This crossover is due to the onset of hopping process upon reaching T_{MCT} and below. In general, the presence of the FSC depends on the thermodynamic path (pressure, density) followed upon cooling.

It has been proved by simulations with different water models [7, 13, 102, 109], experiments and simulation in confined water [107, 110] that the FSC happens when the Widom line from the LLCP is crossed by the thermodynamic path followed upon cooling. In figure 2.5 it is shown the case of water TIP4P/2005, and it can be seen the crossover from a power law to an Arrhenius law (fig. 2.5a) only along density paths. The temperature of the crossing at that density corresponds with the Widom line (fig. 2.5b). The fourth path does not cross the Widom line and the relaxation time does not cross to a strong behaviour.

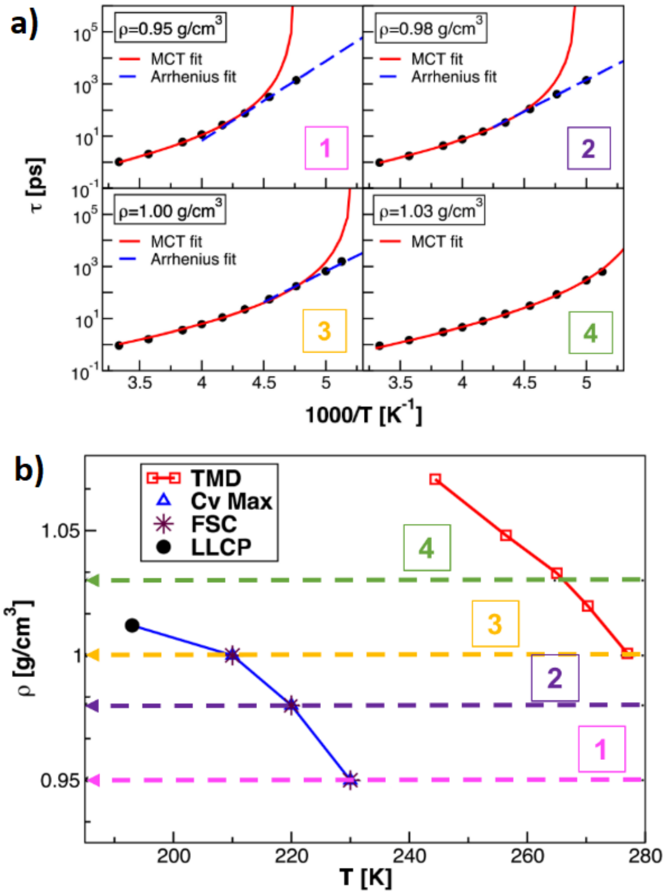

Figure 2.5: Study of the FSC in bulk water TIP4P/2005. a) Behavior of the α relaxation time τ versus the inverse of the temperature for four different densities. The red lines are the fits with the mode coupling power law eq.2.37. The blue lines are the fit with the Arrhenius function eq.2.36. b) The FSC occourrence (stars) in the ρ-T plane. The LLCP is the black circle and the Widom line pass through the spefic heat maxima c_v (empty triangles). Straight dashed lines represent the isotherm followed for every density. The empty squares are temperature of maximum density (TMD). Figure adapted from ref. [109].

Chapter 3

Methods

In this chapter we introduce the basics of molecular dynamics (MD) simulations and the techniques used in this thesis. The aim of MD simulations is to produce trajectories that can be exploited to study a particular physical system. The equations of motion and the thermodynamic environment need algorithms to be managed. We also present the water model, the force field, the dynamical and structural quantities used to study our system.

3.1 Classical Molecular Dynamics

MD is a computational technique developed at the end of 50s [111, 112]. Nowadays, MD is used in a lot of fields besides physics like chemical physics, material science, biophysics etc.

The foundations of MD is the numerical integration of the second Newton's law. MD simulations are widely used in the field of liquid state physics because such complex systems are often impossible to treat analitically. The aim is to extract a trajectory from the system. In other words, a trajectory is a sequence of points in the phase space of the system.

If we have a system composed by N particles, we need to know the positions and the velocities of every particles in order to investigate the time evolution of the system. The arrays that contain time steps, positions and velocities of every particles make up the trajectory of the system.

For a system of N particle we start with the hamiltonian $\mathcal{H}$, that can be written as a sum of two terms

$$\mathcal{H}\left(\{\overline{r}_i\}_N, \{\overline{p}_i\}_N\right) = \mathcal{K}\left(\{\overline{p}_i\}_N\right) + \mathcal{U}\left(\overline{r}_i\}_N\right) \tag{3.1}$$

where $\mathcal{K}$ and $\mathcal{U}$ are respectively the kinetic energy and the many-body interaction potential, while $\overline{r}$ and $\overline{p}$ are respectively the position and the impulse of the particle. We can approximate the many-body term with the sum of all the radial potential between pairs of particles u

$$\mathcal{U}\left(\{\overline{r}_i\}_N\right) = \frac{1}{2}\sum_i \sum_{j\neq i} u_{ij}(|\overline{r}_i - \overline{r}_j|). \tag{3.2}$$

The trajectory is obtained integrating the differential equations of the second Newton's law, that lead for the i-th particle:

$$m_i \ddot{\overline{r}}_i = \dot{\overline{p}}_i = \overline{F}_i = -\nabla_{\overline{r}_i} \mathcal{U} \tag{3.3}$$

$$u_i = \dot{\overline{r}}_i = \frac{\overline{p}_i}{m_i} \tag{3.4}$$

where m_i is the mass of the i-th particle.
The time axis is discretized in a series of time step δt and the previous equations are integrated with the finite difference method. The positions and the velocities of the particles at time $t + \delta t$ are calculated in a deterministic way, i.e. known the positions and the velocities of the particles at the previous time t.
Usually the particles are placed at the vertices of the standard lattice. In a more complex system like solutions containing biomolecules usually are employed crystallographic data.
It is known that temperature and kinetic energy of the particles are stricly connected, a method to assign a realistic initial kinetic energy to every particle is to randomly extract values from the Maxwell-Boltzmann distribution. For example, if the temperature is T, the probability to find the ith particle with mass m_i and velocity v_{ix} along the x-axis is

$$P(v_{ix}) = \sqrt{\frac{m_i}{2\pi k_B T}} e^{-\frac{m_i v_{ix}^2}{2k_B T}}. \tag{3.5}$$

To run a MD simulation different steps must be done.
The first step is the *equilibration* of the system: given the initial configuration and the point of the phase space of interest, the system has to reach that point and must be kept there in thermodynamic equilibrium. The equilibration is essentially a running phase long enough to give the system the time to relax to the minimum energy through a coupling to proper baths (see sec. 3.1.2). In order to make sure that we had a proper equilibration, some quantities must be monitorized, such as the total and potential energies. If these quantities are completely minimized, all the counters are set to zero we can start the real evolution of the system.
The second step is called *production* because trajectories are produced and they can be used to calculate physical quantities at equilibrium of the system. The quantities we want to obtain are several macroscopic physical properties of the system. Since a trajectory is the sequence of points which correspond to positions and velocities as a function of time, these are the microscopic quantities we use to calculate the macroscopic properties.
An average value of an observable of interest A, in statistical mechanics is defined as

$$\langle A \rangle = \iint A\left(\{d\overline{r}_i\}\{d\overline{p}_i\}\right) \rho\left(\{d\overline{r}_i\}\{d\overline{p}_i\}\right) \mathrm{d}\overline{r}\mathrm{d}\overline{p}. \tag{3.6}$$

$\langle A \rangle$ is the ensemble average and the integrations are all over the values of $\overline{r}$ and $\overline{p}$.
The quantity $\rho(\overline{r}, \overline{p})$ is the probability density of the ensemble and can be written through the partition function

$$\rho(\overline{r}, \overline{p}) = \frac{1}{\mathcal{Z}} e^{-\beta \mathcal{H}(\overline{r}, \overline{p})} \tag{3.7}$$

where $\mathcal{H}$ is the hamiltonian and $\beta = \frac{1}{k_B T}$ with k_B Boltzmann constant and T the temperature. The ergodic hypothesis implies that ρ cannot be zero in the phase space of the system.

We want to know the time evolution of the observable A, and the ergodic hyphothesis allow us to pass from the integrals over the phase space (the ensemble average) to a time-average integral

$$\langle A \rangle = \langle A \rangle_{time} \tag{3.8}$$

therefore we can write the equality between statistical mechanics and MD as the time integral over the lenght t of the simulation and the summation over M discrete the time steps

$$\langle A \rangle = \langle A \rangle_{time} = \lim_{t \to \infty} \frac{1}{t} \int_0^t A(\overline{r}, \overline{p}) d\tau \simeq \frac{1}{M} \sum_{t=1}^{M} A(\overline{r}, \overline{p}) \tag{3.9}$$

where $A(\overline{r}, \overline{p})$ is the istantanueous values of A that can be calculated from the informations stored in the trajectories. The last passage is valid if we can generate enough microstates in the phase space.

In other words, we need to sample the phase space in a reliable way. In order to do that, the simulation time should be much longer than the characteristic time of the molecular motions. This is the reason why there are two steps in obtaining a trajectory: the equilibrium run in which the system is evolved until the equilibrium and the production run in which the system can evolve from the equilibrium. In this case the ergodic hyphotesis is valid and the eq.3.9 can be used to calculate the observables.

Here, we list the ensembles that can be used to perform MD simulations depending on the thermodynamic state of the system:

NVE the Microcanonical Ensemble. When the thermodynamic state has fixed quantities in N particles, volume V and total energy E;

NVT the Canonical Ensemble. When the thermodynamic state has fixed quantities in N particles, volume V and temperature T;

NPT the Isobraic-Isothermal Ensemble. When the thermodynamic state has fixed quantities in N particles, pressure P and temperature T;

μVT the Grandcanonical Ensemble. When the thermodynamic state has fixed quantities in chemical potentialμ, volume V and temperature T.

The temperature of the system can be calculated through the equipartition theorem, that relates the temperature with the average kinetic energies of the system

$$T = \frac{2\langle \mathcal{K} \rangle}{(3N - N_C) k_B} \tag{3.10}$$

where N_C is the number of constraints. The pressure can be calculated via the virial theorem, which relates the time average of the potential and kinetic energy to the temperature

$$P = \frac{N k_B - \frac{1 \langle \mathcal{V} \rangle}{3}}{V} \tag{3.11}$$

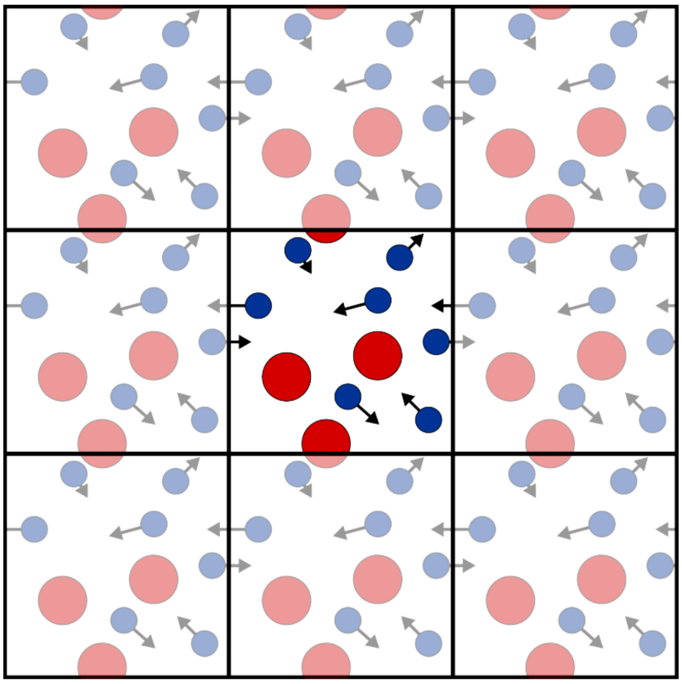

Figure 3.1: This sketch illustrates how periodic boundaries conditions work.

where $\mathcal{V} = \sum_{i=1}^{N} \overline{r}_i \cdot \overline{F}_i$ is the viral. In a MD simulation we simulate a finite system: for example the simplest is a simulation cubic box, so the way we treat boundaries is particularly relevant in this technique. Periodic Boundaries Conditions (PBC) are usually applied to the box. The finiteness of the box can introduce artifacts and it can act as an external field. The PBC avoid these problems and from a computational point of views allow to calculate bulk properties of gases, liquids or crystals with a smaller number of molecules. PBC in 2D are shown in figure 3.1: the box in the middle is replicated in all directions which creates a periodic lattice. The total number of particles in the box are always the same because if a particle leave the box on the left with velocity $\overline{v}$, an identycal particle with the same velocity will enter the box from the right.

3.1.1 The equations of motion in MD

As we already mentioned, the trajectories are calculated by integrating numerically the equations of motion. The time step is chosen in order to be much shorter than the time of a molecule needs to cover its own lenght: usually $\delta t \sim 10^{-15} s = 1$ fs is a good choice to study phenomena in the picoseconds or nanoseconds scales.

The leap-frog algorithm

In the case of rigid molecules the motion can be rotational and translational. The *leap-frog* algorithm is one of the most used algorithms to integrate the translational motion. This algorithm derives from the Verlet algorithm [113].

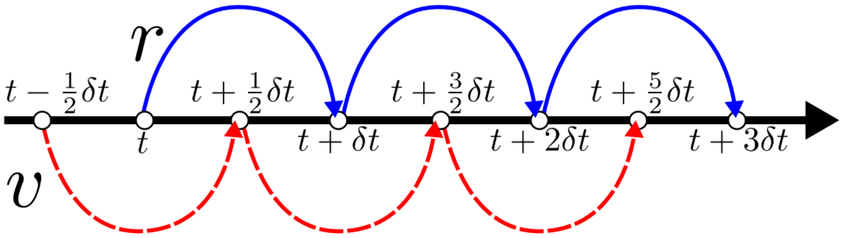

Figure 3.2: Sketch of the functioning of the leap-frog algorithm. The step i in eq. 3.18 are the dashed-red lines and they represent the calculation of velocities. The solid-blue lines are the step ii in eq. 3.19 and they represent the calculation of the positions.

It considers the Taylor expansion for position, velocity and acceleration of a molecule i:

$$\overline{r}_i(t+\delta t) = \overline{r}_i(t) + \overline{v}_i(t)\delta t + \frac{1}{2}\overline{a}_i(t)\delta t^2 + o(\delta t^3) \tag{3.12}$$

$$\overline{v}_i(t+\delta t) = \overline{v}_i(t) + \overline{a}_i(t)\delta t + \frac{1}{2}\overline{b}_i(t)\delta t^2 + o(\delta t^3) \tag{3.13}$$

$$\overline{a}_i(t+\delta t) = \overline{r}_i(t) + \overline{b}_i(t)\delta t + o(\delta t^2). \tag{3.14}$$

If we consider the following

$$\overline{r}_i(t-\delta t) = \overline{r}_i(t) - \overline{v}_i(t)\delta t + \frac{1}{2}\overline{a}_i(t)\delta t^2 + o(\delta t^3) \tag{3.15}$$

and we sum it with eq.3.12 we can simplify the previous taylor expansions

$$\overline{r}_i(t+\delta t) = 2\overline{r}_i(t) - \overline{r}_i(t-\delta t) + \overline{a}_i(t)\delta t^2 + o(\delta t^4). \tag{3.16}$$

We notice that the velocity term has been eliminated, so the algorithm becomes faster. In the Verlet's algorithm, the velocity can be calculated via

$$\overline{v}_i(t) = \frac{\overline{r}_i(t+\delta t) - \overline{r}_i(t-\delta t)}{2\delta t} \tag{3.17}$$

Even throught the Verlet's alogorithm is robust, the velocity calculation is not accurate. The leap-frog algorithm introduce some modifications to the previous procedure and allows a more accurate calculation of the velocities. The new procedure is a two steps iteration: given $\overline{r}_i(t)$, $\overline{a}_i(t)$ and $\overline{v}_i(t-\frac{1}{2}\delta t)$, first the velocities are updated and then the positions

$$i) \qquad \overline{v}_i(t+\frac{1}{2}\delta t) = \overline{v}_i(t-\frac{1}{2}\delta t) + \overline{a}_i(t)\delta t \tag{3.18}$$

$$ii) \qquad \overline{r}_i(t+\delta t) = \overline{r}_i(t) + \overline{v}_i(t+\frac{1}{2}\delta t) \tag{3.19}$$

at the end of step ii the forces can be calculated and the new accelerations can be put in step i to start again. In figure 3.2 it can be seen how the leap-frog algorithm works. The positions in the leap-forg algorithm are calculated using velocities at a time closer to $t+\delta t$, so they are more accurate than the

Verlet's algorithm. Calculating the derivative of eq. 3.19, i.e. the distribution of particles, the potential and the forces can be obtained.
The velocities at time t are calculated through the simple relation

$$\overline{v}_i(t) = \frac{\overline{v}_i(t + \frac{1}{2}\delta t) + \overline{v}_i(t - \frac{1}{2}\delta t)}{2}. \tag{3.20}$$

The intergration of the rotational motion is in general more complex. We can define the torque with respect the center of mass of a molecule in position $\overline{r}_i$, as a vector $\overline{\tau}_i$

$$\overline{\tau}_i = \sum_a (\overline{r}_{ia} - \overline{r}_i) \times \overline{F}_{ia} = \sum_a \overline{d}_i \times \overline{F}_{ia} \tag{3.21}$$

where the subscript d_{ia} is the position of the atom a in the molecule i with respect its center of mass.
The orientation of a rigid body can be defined by a matrix that allows a rotation from the frame of reference of the laboratory to the frame of reference of the center of mass. The rotation matrix is defined through the Euler's angles (ϕ, θ, ψ). A quadri-dimensional vector, called *quaternion*, with unitary norm is defined as a function of the Euler's angles. This vector consents to have convergent equations of motion

$$\overline{q} = (q_0, q_1, q_2, q_3) \tag{3.22}$$

and its norm is $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. The rotation matrix can be write in term of the quatertion

$$\mathcal{R} = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix}. \tag{3.23}$$

If $\hat{\overline{d}}_{ia}$ is the position of the atom a of the molecule i in the frame of its center of mass, its position in the frame of reference of the laboratory is $\overline{d}_{ia} = \mathcal{R}^T \hat{\overline{d}}_{ia}$.
The equations of motion are satisfied by the quaternion

$$\begin{pmatrix} \dot{q}_0 \\ \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{pmatrix} = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \tag{3.24}$$

where $\overline{\omega} = (0, \omega_x, \omega_y, \omega_z)$ is the angular velocity vector in the frame of reference of the center of mass.
The aim is to solve the rotational motion in the same way we did for the translational one. The system of equations 3.24 does not have singularities and can be solved applying the following equation

$$\frac{\mathrm{d}\overline{J}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(I\overline{\omega}) = \overline{\tau} \tag{3.25}$$

where $\overline{\tau}$ is the usual torque (we droppep the subscript i for convenience), I is the moment of inertia matrix and $\overline{J}$ is total angular momentum of the molecule. Eq. 3.24 and 3.25 are coupled. Having in memory $J(t - \delta t)$, $\overline{q}(t)$ and $\overline{\tau}(t)$, the

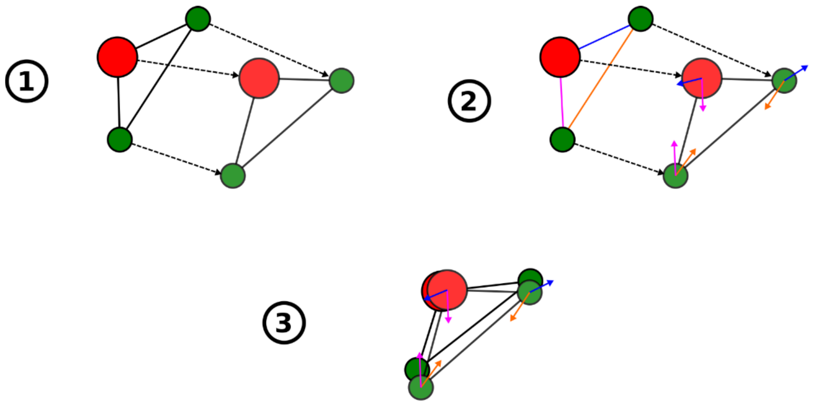

Figure 3.3: Sketch of the three steps of the constraint algorithm.

equations of rotational motion can be solved iteratively. First of all, the angular momentum has to be updated with

$$\overline{J} = \overline{J}\left(t - \frac{1}{2}\delta t\right) + \overline{\tau}(t)\frac{1}{2}\delta t \tag{3.26}$$

From eq.3.24, $\dot{\overline{q}}$ can be calculated and so an estimation of $\overline{q}$ at the half-step time

$$\overline{q}\left(t + \frac{1}{2}\delta t\right) = \overline{q}(t) + \dot{\overline{q}}(t)\frac{1}{2}\delta t. \tag{3.27}$$

Now, the new $\overline{J}$ and $\overline{q}$ can be calculated with the following two equations

$$\overline{J}\left(t + \frac{1}{2}\delta t\right) = \overline{J}\left(t - \frac{1}{2}\delta t\right)\overline{\tau}(t)\delta t \tag{3.28}$$

$$\overline{q}(t + \delta t) = \overline{q}(t) + \dot{\overline{q}}\left(t + \frac{1}{2}\delta t\right)\delta t \tag{3.29}$$

The constraint algorithm

The constraint algorithm is another way to reproduce the motion of a rigid body. In this method, a constraint is inserted in the distance between atoms and it follows three steps, summarized in figure 3.3:

i the particles of the same molecules can move separately;

ii explicit constraint forces are introduced;

iii these forces are minimized using the Lagrangian multipliers technique or another method.

The method of Lagrangian multipliers is one of the most common in MD to minimize the force of the constraints. First al all, a set of n holonomic linear constraints at time t are introduced

$$\sigma_k(t) = \|\overline{x}_{k\alpha}(t) - \overline{x}_{k\beta}(t)\|^2 - d_k^2 = 0, \text{ with } k = 1...n \tag{3.30}$$

where d_k is the intramolecular distance fixed by the constraint while $\overline{x}_{k\alpha}(t)$ and $\overline{x}_{k\beta}(t)$ are the positions of the two particles bonded with the k-constraint.
The total work done by the constraint forces is zero, because they act in the opposite direction for the two particles, so the total energy does not change. For each particles, the force is

$$\frac{\partial \overline{x}_i(t)}{\partial t^2} m_i = -\frac{\partial}{\partial \overline{x}_i}\left[V(\overline{x}_i(t)) + \sum_{k=1}^{n} \lambda_k \sigma_k(t)\right] \text{ with } i = 1...N \tag{3.31}$$

where N is the total number of particles. By integrating both sides of the last equation with respect to the time, the new constrained coordinates at time $t+\delta t$ are obtained

$$\overline{x}_i(t+\delta t) = \overline{x}_i^f(t+\delta t) + \sum_{k=1}^{n} \lambda_k \frac{\partial \sigma_k(t)}{\partial \overline{x}_i}(\delta t)^2 \frac{1}{m_i} \text{ with } i = 1...N \tag{3.32}$$

where $\overline{x}_i^f(t+\delta t)$ is the position of the particle after integrating the equations of motion without constraint, so basically it's the 'free' position of the particle. The constraints should be satisfied in the next time-step, so they should respond to the equation 3.30 of the Lagrangian multipliers

$$\sigma_k(t+\delta t) = \|\overline{x}_{k\alpha}(t+\delta t) - \overline{x}_{k\beta}(t+\delta t)\|^2 - d_k^2 = 0 \tag{3.33}$$

A system of n non-linear equations should be solved for the n unknown Lagrangian multipliers λ_k.

3.1.2 Barostats and thermostats

In a simulation it is often necessary to fix the pressure or the temperature, or both these quantities to simulate, for example, the real conditions of an experiment.
Even for the microcanical ensemble NVE, it is needed to regulate the temperature during the equilibration phase. We already know that temperature and kinetic energy are connected by the equation 3.10, therefore controlling the velocities of the particles means to control the temperature. If we multiply the velocities by a factor λ at time t, the variation of the temperature of the uncostrained system is

$$\Delta T = \frac{1}{2}\sum_{i=0}^{N} \frac{2m_i(\lambda \cdot v_i)^2}{3Nk_B} - \frac{1}{2}\sum_{i=0}^{N} \frac{2m_i\lambda \cdot v_i^2}{3Nk_B} = (\lambda^2 - 1)T(t) \tag{3.34}$$

The system is at temperature $T(t)$ and we want to take it at the desired temperature T_D. What we need to do is to multiply velocities at every time step by a factor

$$\lambda = \sqrt{\frac{T_D}{T(t)}} \tag{3.35}$$

The Berendsen thermostat

The Berendsen thermostats is one of the oldest and it was proposed by Berendsen *et al.* in 1984 [114]. This method provides a new way to control the temperature, adding to the system an external heat bath at a tunable temperature. The external bath add or remove heat form the system in order to reach and keep the desired temperature. If there is a temperature shift between two time steps, the change is given by

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_B - T(t)) \rightarrow \delta T = \frac{1}{\tau}(T_B - T(t))\delta t \tag{3.36}$$

where T_B is the temperature of the external bath and τ is coupling parameter between the bath and the system.

The parameter λ that rescale velocities is proportional to the difference between the bath temperature and the system temperature

$$\lambda = \sqrt{1 + \frac{\delta t}{\tau}\left(\frac{T_B}{T(t) - 1}\right)} \tag{3.37}$$

It has been shown that the quantities calculated through this method differ from the canonical ensemble by $\mathcal{O}(1/N)$ [113]. In most of the cases, the Berendsen method is accurate enough for several applications used in MD.

The method discussed until now, can be extended to treat barostats. A 'Berendsen barostat', can be imagined if to reach the desired temperature we act on velocities, to reach the desired pressure we rescale the volume of the system. The λ factor, in this case, is multiplied by the volume

$$r_i' = \sqrt[3]{\lambda} r_i \tag{3.38}$$

Similarly to what we did with the temperature bath, we can couple the system with a 'pressure bath' and the change in pressure can be calculated by the following expression

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P - P_B) \tag{3.39}$$

where τ_P is the coupling pressure constant, P_B is the pressure of the bath and $P(t)$ is the pressure of the system at time t.

The λ factor determines the rescaling of the volume as

$$\lambda = 1 - \kappa_T \frac{1}{\tau_P}(P - P_B)\delta t \tag{3.40}$$

where κ_T is the isothermal compressibility, which describes the change of volume responding to the change of the pressure at fixed temperature and it is defined as

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{3.41}$$

3.2 Potentials and force fields

In this section we briefly describe how potentials are treated in MD simulations. Potential energy can be split in two terms: one is relative to the bonded interactions (intramolecular) and the other one is relative to the non-bonded interactions (intermolecular)

$$\mathcal{U} = \mathcal{U}_{bonded} + \mathcal{U}_{non-bonded} \tag{3.42}$$

The intramolecular part $\mathcal{U}_{bonded}$ could be divided into three terms relative to the basics movement of the molecule: stretching, bending and torsion

$$\mathcal{U}_{bonded} = \mathcal{U}_{stretching} + \mathcal{U}_{bending} + \mathcal{U}_{torsion}. \tag{3.43}$$

$\mathcal{U}_{stretching}$ describes the oscillations of the bond lenght around the equilibrium, $\mathcal{U}_{bending}$ describes the oscillations of the angle between three atoms around the equilibrium angle and $\mathcal{U}_{torsion}$ describes the torsional rotation of four atoms around a central bond.

The intermolecular part $\mathcal{U}_{non-bonded}$ can be written as a sum of two terms

$$\mathcal{U}_{non-bonded} = \mathcal{U}_{LJ} + \mathcal{U}_{Coulomb} \tag{3.44}$$

where $\mathcal{U}_{LJ}$ is the Lennard-Jones potenatial, defined as

$$\mathcal{U}_{LJ} = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{3.45}$$

and $\mathcal{U}_{Coulomb}$ represents the Coulombian interactions

$$\mathcal{U}_{Coulomb} = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r}. \tag{3.46}$$

The most expensive part in term of computational power are the non-bonded interactions, which are pairwise. The way to deal with intermolecular interaction involves a non-bonded cutoff lenght and the minimum image convention.

The cutoff works by setting to zero all the pairwise interactions between particle i and j distant more than the cutoff lenght, i.e. the interactions are only calculated when their distance is $r = |r_i - r_j| \leq r_{cutoff} < L/2$ where L is the linear dimension of the box. This upper limit is introduced to avoid considering the same particles twice, due to the Periodic Boundary Conditions.

The image convention provides that each particles in the simulation box interacts only with the closest images of all the other particles. The introduction of the cutoff makes it necessary to introduce corrections for non-bonded interactions [113].

Corrections for short range potentials like the Lennard-Jones is typically performed in the real space, while long range potentials like the Coulombian are handled in the reciprocal space (Fourier space) with, for example the Particle Mesh Ewald method.

3.2.1 Particle Mesh Ewald method

As we already mentioned, it is delicate to deal with truncation of long range potentials, and often corrections are needed. If long range interactions are not adequately treated, some severe artifacts can arise [52]. The Particle Mesh Ewald method, or Ewald summation method is one of the most used technique to deal with corrections of long range potentials.

Taking into account the PBC, the potential of electrostatic interactions can be written as

$$\mathcal{U}_E = \frac{1}{2} \sum_{i=1}^{N} q_i \phi_E(\overline{r}_i) \tag{3.47}$$

where ϕ_E is given by

$$\phi_E = \sum_{j=1}^{N} \sum_{\overline{R}} \frac{q_j}{|\overline{r}_{ij} - \overline{R}|} \tag{3.48}$$

and $\overline{R} = \overline{n} \cdot L$, with $\overline{n} = (n_x, n_y, n_z)$ and $n_x, n_y, n_z \in \mathbf{Z}$. In the sum there is ad hoc convention that $\overline{R} = 0$ when $i = j$. Screening charges can be added and subtracted in order to split the sum in two terms $\mathcal{U}_E = \mathcal{U}_{SR} + \mathcal{U}_{LR}$ where SR stands for short ranged and LR for long ranged. The first term $\mathcal{U}_{SR}$ is composed by the original charges screened with clouds of opposite charge; this term is short ranged and can be evaluated in the real space. The second term $\mathcal{U}_{LR}$ is evaluated in the Fourier space and it is made by screening charge densities used to compensate the screening charges of the first term. Screening charges are usually distribuited as Gaussian

$$\rho_s(r) = q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha r^2}. \tag{3.49}$$

With this choice, the compensating charges distribution in the second term can be written as

$$\rho_c(\overline{r}) = \sum_{j=1}^{N} \sum_{\overline{R}} q_{ij} \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha|\overline{r}-(\overline{r}_j+\overline{R})|^2} \tag{3.50}$$

and its Fourier transform into the reciprocal space is

$$\tilde{\rho}_c(\overline{k}) = \int_V d\overline{r}\, e^{-i\overline{k}\cdot\overline{r}} \rho_c(\overline{r}) = \sum_{j=1}^{N} q_j e^{-i\overline{k}\cdot\overline{r}_j} e^{-k^2/4\alpha} \tag{3.51}$$

This charge distribution generates a field that can be calculated by the Poisson's equation in the reciprocal space:

$$k^2 \tilde{\phi}(\overline{k}) = 4\pi\tilde{\rho}(\overline{k}), \tag{3.52}$$

and then antitrasformated it into the real space. In this way, the long range sum becomes

$$\mathcal{U}_{LR} = \frac{1}{2} \sum_{i=1}^{N} q_i \phi_c(\overline{r}_i) = \frac{1}{2V} \sum_{k \neq 0} \frac{a\pi}{k^2} |\tilde{\rho}(\overline{k})|^2 e^{-k^2/4\alpha} \tag{3.53}$$

with $\tilde{\rho}(\overline{k}) = \sum_{i=1}^{N} q_i e^{-i\overline{k}\cdot\overline{r}_i}$. The point charges $\overline{r}_i$ interact with the the compensating charges also in $\overline{r}_i$, therefore this part of potential in equation 3.53 contains some self-interactions terms. To restore the correct electrostatic energy, these terms should be eliminated from the final expressions. They represent the potential at the center of Gaussian charge distribution.
Now, we calculate the short range term $\mathcal{U}_{SR}$. By means of Poisson's equation in the real space

$$-\nabla^2 \phi(\overline{r}) = 4\pi\rho(\overline{r}) \tag{3.54}$$

it is possibile to show that the screening potential can be written as

$$\phi_s(r) = \frac{q_i f_{err}(\sqrt{\alpha} r)}{r} \tag{3.55}$$

where f_{err} is the error function, defined by

$$f_{err} = \frac{2}{\sqrt{\pi}} \int_0^x dt\, e^{-t^2}. \tag{3.56}$$

The self-interactions term is

$$\phi_{self} = \phi_s(0) = 2\sqrt{\frac{\alpha}{\pi}} q_i \tag{3.57}$$

and the total self-interaction energy becomes

$$\mathcal{U}_{self} = \sqrt{\frac{\alpha}{\pi}} \sum_{i=1}^{N} q_i \tag{3.58}$$

this last term must be subtracted to the total electrostatic energy in order to restore the correct Coulombian field. The short range term of the potential energy can be written as

$$\mathcal{U}_{SR} = \sum_{\overline{R}} \sum_{i \neq j}^{N} q_i q_j \frac{1 - f_{err}\sqrt{\alpha}|\overline{r}_{ij} + \overline{R}|}{|\overline{r}_{ij} + \overline{R}|} = \sum_{\overline{R}} \sum_{i \neq j}^{N} q_i q_j \frac{f^c_{err}\sqrt{\alpha}|\overline{r}_{ij} + \overline{R}|}{|\overline{r}_{ij} + \overline{R}|} \tag{3.59}$$

where $f^c_{err}(x) = 1 - f_{err}(x)$ is the complementary error funtion deined by

$$f^c_{err}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} dt\, e^{-t^2}. \tag{3.60}$$

Finally, by putting together equations 3.59, 3.53 and 3.58, the total electrostatic energy can be written

$$\begin{aligned} \mathcal{U}_E = \mathcal{U}_{SR} + \mathcal{U}_{LR} - \mathcal{U}_{self} = & \sum_{\overline{R}} \sum_{i \neq j}^{N} q_i q_j \frac{f^c_{err}\sqrt{\alpha}|\overline{r}_{ij} + \overline{R}|}{|\overline{r}_{ij} + \overline{R}|} + \\ & + \frac{1}{2V} \sum_{k \neq 0} \frac{a\pi}{k^2} |\tilde{\rho}(\overline{k})|^2 e^{-k^2/4\alpha} - \sqrt{\frac{\alpha}{\pi}} \sum_{i=1}^{N} q_i. \end{aligned} \tag{3.61}$$

As can we see, there are different sums in the last formula. The short range sum is performed into the real space, while the long range sum is in the Fourier space. Moreover, the α parameter is tunable and particularly important for the convergence of the integrals: to obtain a fast convergence, α should be large in the real space and small in the reciprocal space. Usually, α is set at the order of $\frac{5}{L}$ [113], taking an order of ten wave vector $\overline{k}$ for the computation of the sum into the reciprocal space.

3.2.2 SPC/E potential for water

There are several rigid potential used to model water as SPC/E and TIP4P. These are rigid potentials which constrain and simplify the internal motion of the atoms composing molecule. Among of them, one of the the most popular is probably the *simple point charge extended* or SPC/E potential. It was developed in 1987 by Berendsen *et al.* [115].

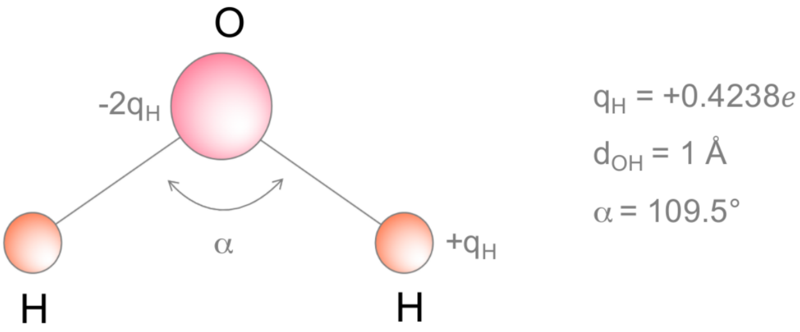

Figure 3.4: Sketch of the SPC/E model of the water molecule.

This potential is an improvement of the already existing SPC (*simple point charge*) potential that consists in a correction of the polarization and a consequent reparametrization of all the older SPC parameters. This correction was introduced to better characterize the properties of polar liquids. The model of the water molecule offered by the SPC/E potential is shown in figure 3.4: the O-H distance is 1 Å and the H-$\hat{O}$-H angle is 109.5°. The charges on the hydrogen atoms are $q_H = 0.4238e$ while the charge on the oxygen atom is $q_O = -2q_H = -0.8476e$ On the oxygen atom there is also the only Lennard-Jones interaction site with parameters

$$\epsilon_{OO} = 0.650\ kJ/mol \qquad\qquad \sigma_{OO} = 3.166\ \text{Å}.$$

The total interaction potential of the i^{th} water molecule is given by

$$u(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_i - r_j}\right)^{12} - \left(\frac{\sigma_{ij}}{r_i - r_j}\right)^{6}\right] + \frac{e^2}{4\pi\epsilon_0}\frac{q_i q_j}{|r_i - r_j|} \tag{3.62}$$

3.2.3 The CHARMM force field for biosystem

In this paragraph we describe the CHARMM force field, used in this work for the modellization of the lysozime protein. CHARMM is the acronym for *Chermistry at Harvard Macromolecular Mechanics* and represents a set of force fields widely used in MD to modellize many biomolecules like saccharides, proteins, lipids, nucleid acids etc. [116–121]. As we have already seen in the geneal discussion of section 3.2, we can divide the CHARMM potential into two terms

$$\mathcal{U} = \mathcal{U}_{bonded} + \mathcal{U}_{non-bonded}. \tag{3.63}$$

The first term takes into account all the intramolecular interactions between neighbors atoms chemically bonded like stretching, bending, proper and improper dihedral energy

$$\mathcal{U}_{bonded} = \mathcal{U}_{stretching} + \mathcal{U}_{bending} + \mathcal{U}_{dihedral} + \mathcal{U}_{impr-dihedral} + \mathcal{U}_{UB} + \mathcal{U}_{CMAP} \tag{3.64}$$

where

$$\mathcal{U}_{streching} = \sum_{bonds} K_b(b - b_0)^2 \tag{3.65}$$

is the energy function account for the bond stretches and where K_b is the bond force constant and $b - b_0$ is the lenght of atom's shift from the equilibrium.
The second term is

$$\mathcal{U}_{bending} = \sum_{angles} K_\theta(\theta - \theta_0)^2 \tag{3.66}$$

and it represents the bending energy of every angle: K_θ is the angle force constant and $\theta - \theta_0$ is the angle's shift from the equilibrium.
The third term is

$$\mathcal{U}_{dihedral} = \sum_{dihed.} K_\phi(1 + \cos(n\phi - \delta)) \tag{3.67}$$

and it represents the torsion angles (dihedrals). K_ϕ is the dihedral force constant, n is the multiplicityof the function, ϕ is the dihedral angle and δ is the phase shift.
The fourth term is

$$\mathcal{U}_{impr-dihedral} = \sum_{imp.dihed.} K_\omega(\omega - \omega_0)^2 \tag{3.68}$$

and it accounts for the out-of-the-plane bending: K_ω is the force constant and $\omega - \omega_0$ is the out-of-plane angle's shift from equilibrium.
The last two terms were introduced by CHARMM. The fifth term is called Urey-Bradley term and it is

$$\mathcal{U}_{UB} = \sum_{UB} K_{UB}(b^{(1;3)} - b_0^{(1;3)})^2. \tag{3.69}$$

This is an harmonic term in the distance between atoms 1 and 3 that forms an angle θ. This term is an improvement for in-plane bending and vibrational stretching because it separates symmetric and non-symmetric stretching modes [116]. In the end, the last and sixth term is the CMAP term

$$\mathcal{U}_{CMAP} = \sum_{residues} u_{CMAP}(\phi, \psi) \tag{3.70}$$

and it is a cross term for backbone dihedral angles ϕ and ψ realized by grid based correction maps [117]. This is used as an improvement for dihedral angles only in proteins backbones.
As already discussed, the $\mathcal{U}_{non-bonded}$ part is made by the usual two contributions by Lennard-Jones and Coulombic interactions

$$\mathcal{U}_{non-bonded} = \mathcal{U}_{LJ} + \mathcal{U}_{Coulomb} \tag{3.71}$$

3.3 Observables

In this last section we discuss the MD observables used in this study to characterize our system. They are divided in two kinds: dynamical quantities and structural quantities of our interest. The dynamical quantities are the Self Van-Hove correlation function (SVHS), the Self Intermediate Scattering Function (SISF) and the Mean Square Displacement (MSD). The structural quantities are the Radial Distribution Function (RDF), the Local Structure Index (LSI), the Orientational Tetrahedral Order Parameter q and the V_4 parameter [122].

3.3.1 Dynamical quantities

To characterize the dynamics of a liquid, correlation functions are powerful tools [97].
The correlation function between two observables $A(t_1)$ and $B(t_2)$, that are functions of the coordinates $\overline{r}_i(t)$ and the momenta $\overline{p}_i(t)$, is give by

$$C_{AB}(t_1, t_2) = \langle A(t_1)B(t_2)\rangle = \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau A(t_1 + t)B(t_2 + t)dt \tag{3.72}$$

with the convention of $t_1 \geq t_2$. The correlation is invariant under time traslation. At equilibrium, therefore we can define $t = t_1 - t_2$ and write

$$C_{AB}(t_1, t_2) = \langle A(t_1)B(t_2)\rangle = \langle A(t)B(0)\rangle. \tag{3.73}$$

The two limit cases for $t \to 0$ and $t \to \infty$ are

$$\lim_{t\to 0} C_{AB}(t) = \langle A(0)B(0)\rangle = \langle AB\rangle \tag{3.74}$$

$$\lim_{t\to\infty} C_{AB}(t) = \langle A\rangle\langle B\rangle \tag{3.75}$$

i.e. in the second case A and B become uncorrelated.
Correlation functions are often normalized by thei value at initial time

$$\hat{C}_{AB}(t) = \frac{\langle A(t)B(0)\rangle}{\langle AB\rangle} \tag{3.76}$$

When the observables are the same, i.e. $A(t) = B(t)$, we call C_{AA} an *autocorrelation function* defined by

$$C_{AA}(t) = \langle A(t)A(0)\rangle \tag{3.77}$$

Self Van-Hove correlation function

The Self van Hove Function (SVHF) is the single particle density-density correlation function in the real space. It is defined as [97]

$$G_s(\overline{r}, t) = \frac{1}{N}\langle\sum_i^N \delta(\overline{r} - \overline{r}_i(0) - \overline{r}_i(t))\rangle \tag{3.78}$$

It represents the probability density of finding a particle i in the position $\overline{r}$ at time t when the particle was located in the origin at time 0. Usually, we display the radial part of the G_{self} calculated as $4\pi r^2 G_s(r, t)$ which is the probability density of a displacement $\overline{r}$ in a time interval t. In figure 3.5 is shown the SVHFs at different times for bulk water. The function starts as a sharp gaussian, but as the particle diffuses, the function becomes lower and wider over longer lenghts.

Self Intermediate Scattering Function

We already introduced the SISF as the autocorrelation function of the Fourier components of the local density in the reciprocal space. Let us rewrite here the definition:

$$F_s(\overline{q}, t) = \frac{1}{N}\langle\rho_{\overline{q}}(t)\rho_{-\overline{q}}(0)\rangle = \left\langle \frac{1}{N}\sum_{i=1}^N e^{i\overline{q}\,\cdot\,(\overline{r}_i(t) - \overline{r}_i(0))}\right\rangle \tag{3.79}$$

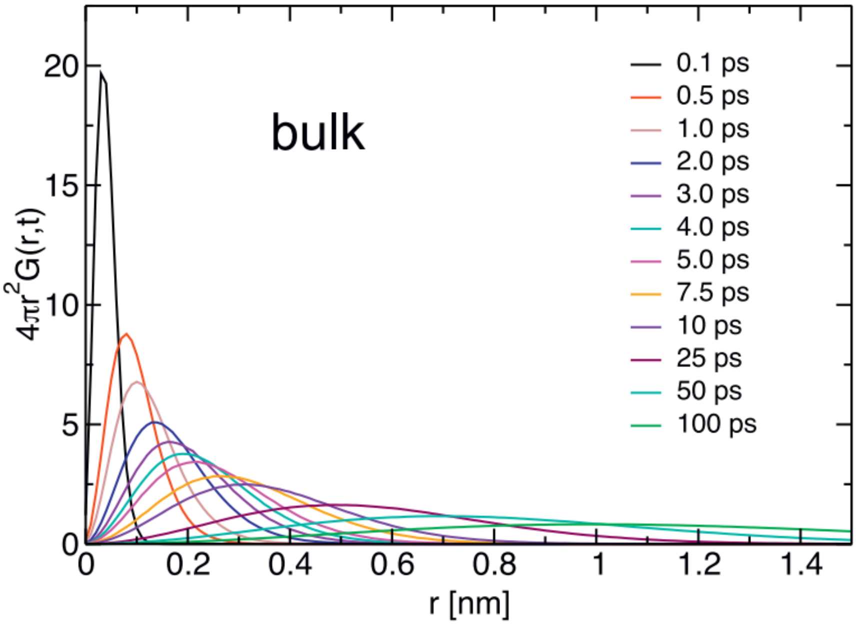

Figure 3.5: Oxygen SVHFs of SPC/E bulk water at different time interval. Figure from ref. [123].

where N is the number of particles and $\overline{r}_i(t)$ is the position of the i-th particle at time t. Figure 3.6 shows different SISF calculated in a range of rescaled temperatures from 0.7 to 0.46 for a dense polymer liquid model.
As we already described in section 2.1, the SISF probes the translational dynamics of the single particle and it is useful to characterized different regimes in the diffusion process. In figure 3.6, the functions at higher temperatures (on the left) exhibit a rapid diffusion. As T decreases, the cage regime shows as a plateau followed by the α-relaxation.

Mean Square Displacement

The Mean Square Displacement (MSD) is defined as

$$\langle \Delta r^2(t) \rangle = \left\langle \frac{1}{N} \sum_{i=1}^{N} (\overline{r}_i(t) - \overline{r}_i(0))^2 \right\rangle \tag{3.80}$$

with N is total number of particle, $\overline{r}_i(t)$ is the position of the i-th particle at time t. MSD is connected with the autocorrelation function of velocity [97]. The way to measure experimentally the MSD are the neutron scattering and the photon correlation spectroscopy.
Towards the MSD the diffusive regime of the system under investigation can be studied. If we write the MSD as a power law

$$\langle \Delta r^2(t) \rangle \propto t^{\alpha} \tag{3.81}$$

we can catalogate the different regimes depending on the exponent α [125]:

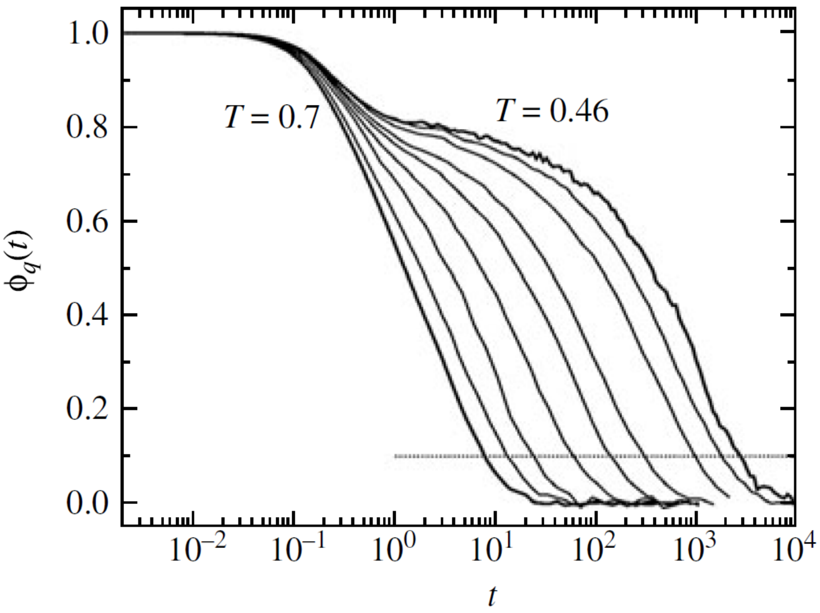

Figure 3.6: SISF $\phi_q(t)$ for wave number q=6.9 obtained by molecular-dynamics simulations for a dense-polymer liquid model. Temperatures are in ascending order from left to right. Figure from ref. [124].

- $0 < \alpha < 1$ → sub-diffusive regime;
- $\alpha = 1$ → diffusive regime;
- $1 < \alpha < 2$ → super diffusive regime;
- $\alpha = 2$ → ballistic regime;

In figure 3.7 is shown the MSD for a Lennard-Jones binary mixture. The curves at higher temperatures (on the left) exihibit the ballistic regime followed by a diffusive regime. A sub-diffusive regime appears upon cooling ending in a plateau, that corresponds to the cage regime of the SISF.

3.3.2 Structural quantities

Radial Distribution Function

The Radial Distribution Function (RDF) is written $g(r)$ and corresponds to the probability of finding a pair of particle at distance r, relative to the probability of a completely random density distribution. The definition of the RDF is [97]

$$g(r) = \frac{V}{N^2}\left\langle \sum_i \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle \tag{3.82}$$

where N is the total number of particle, V is the volume of the system. The last definition is an ensemble average over pairs and it is useful to investigate

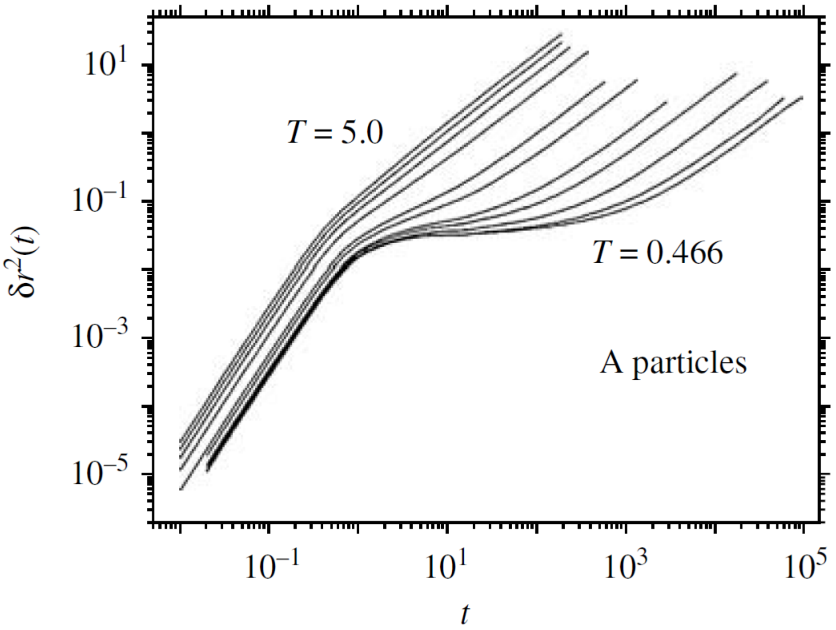

Figure 3.7: Mean-squared displacements $\delta r^2(t)$ obtained by molecular-dynamics simulations for the A particles of the Lennard-Jones binary mixture. Temperatures are in ascending order from left to right. Figure from ref. [100].

the internal structure of the system. The RDF is a sort of 'static' Van Hove Correlation function and shows picture of the density distribution of the system. The RDF points out the differences in the system's structure depending on the dynamical phase as it is shown in figure 3.8. In a normal liquid (top) it is possible to recognize three maim coordination shells. As we drift away from the tagged particle, the density fluctuations approach the unity. The amourphous solid (middle) is an intermediate system between liquid and solid: the $g(r)$ shows three main peaks as for the normal liquid, but they are sharper and better defined. Minor peaks corresponding to more fine structures can be seen in the second shell. A solid (bottom) has a regular structure and the RDF reflects this arrangement: well defined isolated peaks describe the lattice regularity.

Local Structure Index

The Local Structure Index (LSI) is a used stuctural parameter that distinguishes molecules with well separated first and second coordination shells from molecules surrounded by a disordered environment.
A molecule i is chosen as the reference molecule and j are its neighbors ordered according to increasing distance from i. We define $n(i)$ as the number of molecules within 3.7 Å from the i-th molecule. Therefore $r_{n(i)} < 3.7\,\text{Å} < r_{n(i)+1}$. The LSI is defined as

$$I(i) = \frac{1}{n(i)} \sum_{j=1}^{n(i)} \left[\Delta(j;i) - \bar{\Delta}(i)\right]^2 \tag{3.83}$$

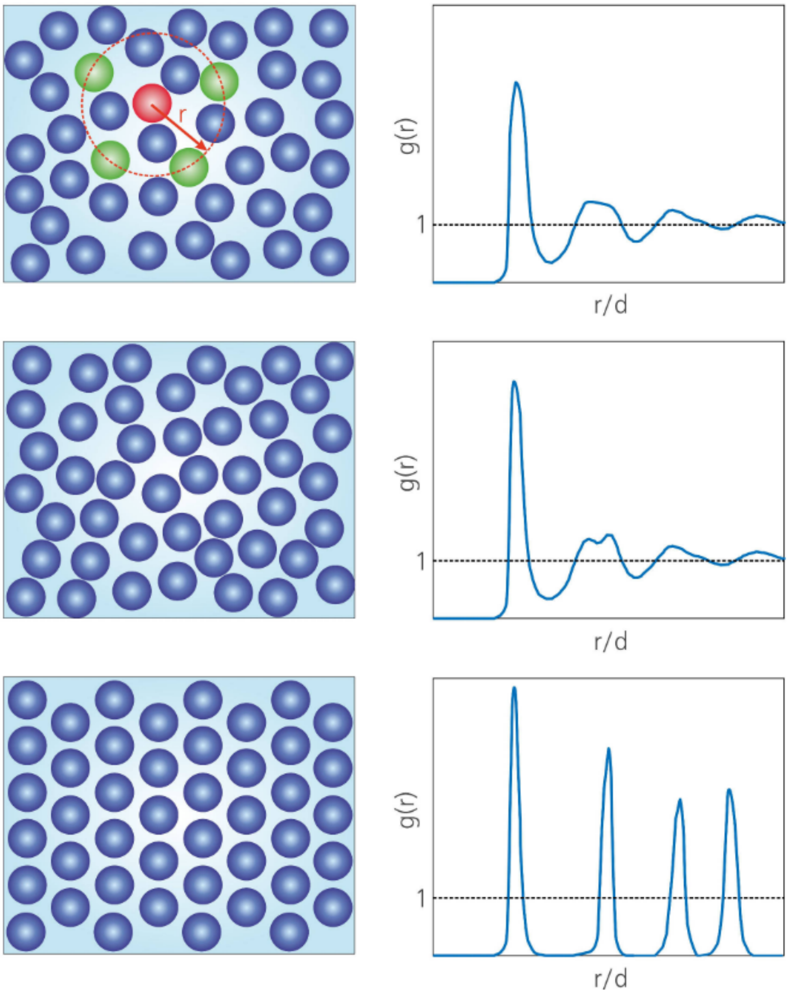

Figure 3.8: Structures of different systems and the associated $g(r)$: normal liquid (top), amrphous solid (middle), crystalline solid (bottom). The dashed horizontal lines refers on a ideal gas. Figure adapted from ref. [95].

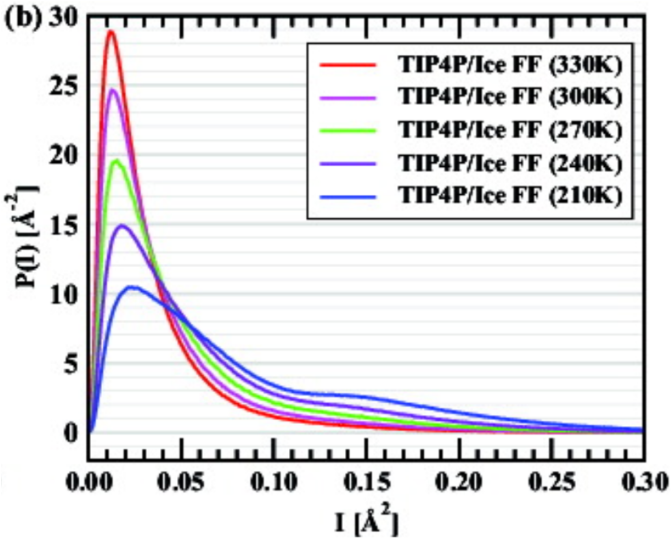

Figure 3.9: Probability density distributions of the LSI (P(I)) in liquid water obatained from classical force field (TIP4P/Ice) based simulations. Figure adapted from ref. [126].

where $\Delta(j;i) = r_{j+1} - r_j$ and $\bar{\Delta}(i)$ is the average of $\Delta(j;i)$ over all neighbors j of molecule i within the cutoff distance. The LSI is a radial parameter sensitive to local environments, including interstitial water between first and second neighbors shells. In figure 3.9 is shown the LSI for liquid water simulated with TIP4P/Ice. It can be observed a decrease of the first peak upon cooling. This means a decrease in the relative population of water molecules with lower LSI values (i.e., molecules in locally disordered environments) coupled with an increase in the relative population of water molecules with higher LSI values (i.e., molecules in locally ordered environments).

Orientational Tetrahedral Order Parameter

The orientational tetrahedral order parameter q is defined as [29]

$$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos(\gamma_{ij}) + \frac{1}{3}\right)^2 \tag{3.84}$$

where γ_{ij} is the angle formed by the lines joining the oxygen atom of a given molecule and those of its first four nearest neighbours j and k. This parameter is used to quantify the degree of tetrahedrality of the local structure of a system. The average value of q varies between 0 for an ideal gas and 1 for a system with perfect tetrahedral arrangement. In figure 3.10 is shown the q parameter for liquid water for three temperatures: 320 K (dashed line), 280 K (solid line) and 240 K (dot-dashed line). The highest and the lowest temperatures exhibits a unimodal behaviour respectively with a low-q peak and a high-q peak. The bimodal distribution at the intermediate temperature, suggests an arrangement that can be described as 'more tetrahedral-ordered' (high-q peak) or 'less tetrahedral-disordered' (low-q peak).

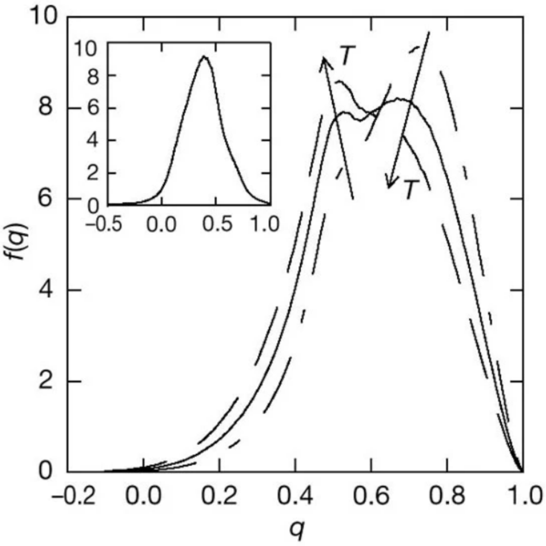

Figure 3.10: Probability distributions of the orientational tetrahedral order parameter for liquid SPC/E water. Three temperatures are considered: 320 K (dashed line), 280 K (solid line) and 240 K (dot-dashed line).Arrows indicate the effect of increasing temperature and the area under each curve is normalized to unity. Figure adapted from ref. [29].

V_4 **Parameter**

V_4 parameter was recently proposed by Montes de Oca *et al.* [122] as a new structural parameters, in order to distinguish better different local structures in liquid water already at ambient conditions.
V_4 parameter is build upon energy consideration: for each water molecule i, all pairwise interactions V_{ij}, with $j \neq i$, are evaluated and sorted according to their intensity, from the smallest to the largest value. V_4 is defined as the forth V_{ij}. If a molecule forms four strong linear hydrongen bonds, the fourth is expected to be the order of a linear hydrogen bond. When the structure of the elementary water shell is perturbed, then al least one of the four hydrogen bond is distorted. A distortion on the tetrahedral water arrangement results on a larger V_4 value. The minimum in V_4 probability distributions is at -12 kJ/mol and this value is used as a threshold which separates interstitial-distorted water from more ordered water (lower V_4 values). In figure 3.11, V_4 probability distributions in SPC/E bulk water upon cooling are shown. The dashed line are the equilibrium configurations or Real Dynamics (RD) and the full lines are the Inherent Structures (IS) that eventually reveal underlying structure hidden by thermal fluctuations. We point out that IS distributions shown a bimodal shape already at ambient condition.

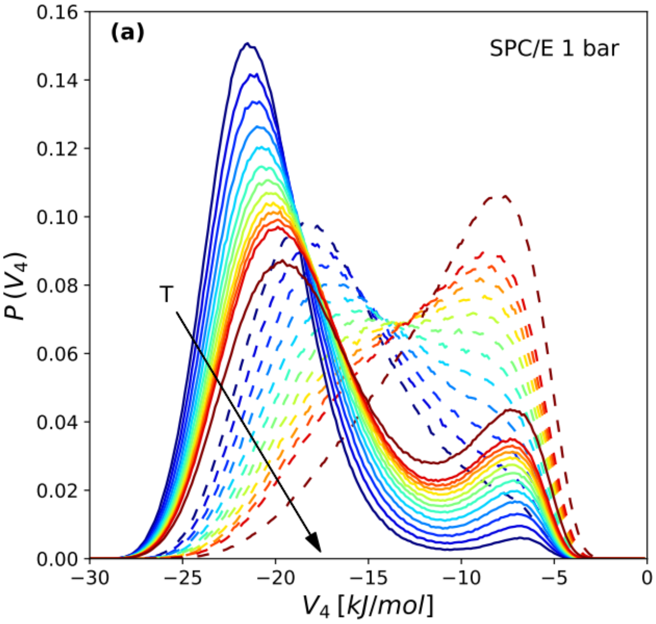

Figure 3.11: V_4 V4 for SPC/E water for equilibrium configurations or real dynamics (RD, dashed lines) and for inherent-structures (IS, full lines) at 1 bar. The temperature ranges are as follows from 190 K to 300 K in 10 K steps. Figure from ref. [122]

Chapter 4

Dynamics of protein hydration water

In this chapter we present some original results in the study of the dynamics of lysozyme hydration water obtained using MD simulations. In particular, we characterized the dynamics of hydration water through the Self Van Hove correlation function and the Mean Square Displacement. Before doing that, in the first section we give details on the systems simulated and in the second we review previous results on protein hydration water obtained by Gallo's research group through the Self Intermediate Scattering Function and the extraction of the relaxations times.

4.1 System and simulation details

Our system consists on a lysozyme protein immersed in 13982 water molecules. The system is shown in figure 4.1 and it includes also 8 chlorine ions added to the liquid bulk to neutralize the positive charges of the residues of the protein. In this way, the net charge of the system is zero.

The system was run with classical all-atom MD simulations performed with the GROMACS 4.5.5 package [127]. The aim was studying the dynamics of water proximal to the protein surface. Usually, hydration water is defined in a shell of 4-6 Å surrounding the biomolecules [60, 128–130]. In this thesis, protein hydration water refers to water molecules at a distance minor or equal to 6 Å from any lysozyme atoms (blue sticks in figure 4.1).

We took already equilibrated trajectories from previous studies [61, 130] and we simulated them up to 100 ns for each temperature. In a second moment, the 200 K trajectory was elongated up to 200 ns. The pressure was set at $p = 1$ bar and the temperature spanned from 300 K to 200 K. Both pressure and temperature were handled with the Berendsen method [114] during the equilibration phase (NPT). It was used the CHARMM force field [116, 117], described in section 3.2.3, to model the lysozyme protein and the Cl^- ions, while the SPC/E potential [114] (sec. 3.2.2) was used for the water. The cut-off value for the non-bonded interactions was set at 10 Å and the electrostatic force field was managed with the particle mesh Ewald method. The equations of motion were integrated with the Verlet leap-frog algorithm, with a time step

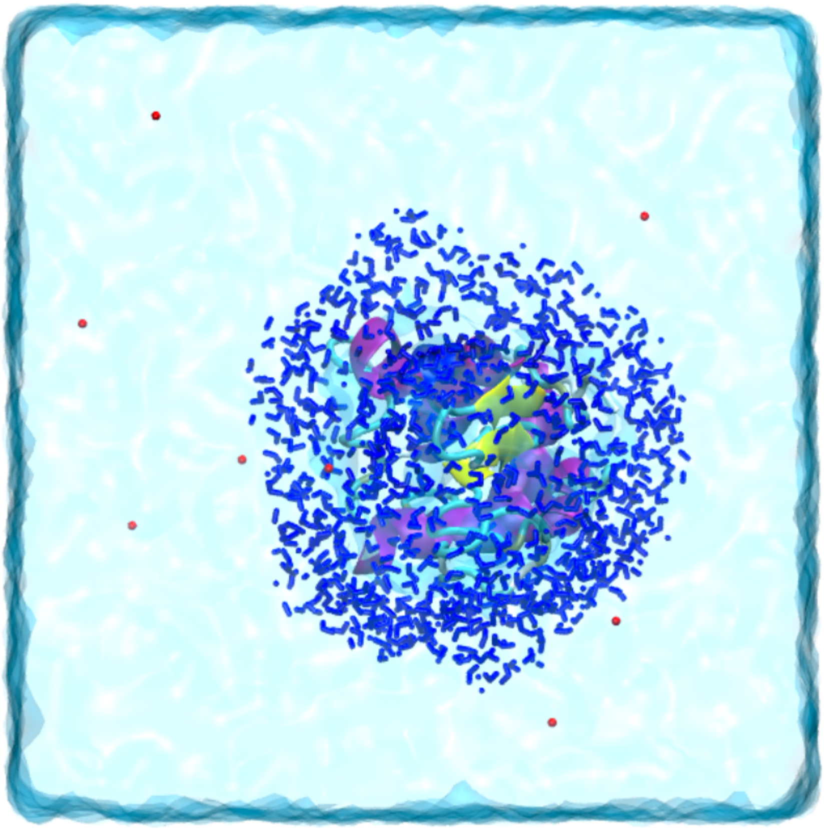

Figure 4.1: The system in a snapshot taken at 280 K. The blue stick are the protein hydration water, surrounding the lysozyme coloured according its secondary structure. The red balls are the chlorine atoms while the rest of the water is depicted as a transparent surface.

Table 4.1: Details of the simulated trajectories for the system of lysozyme and hydration water. For every tempeature, we show density, old equilibration runs lenght t_{eq} and the lenght of the new production runs performed for this thesis t_{prod}.

T(K)	ρ (kg/m^3)	t_{eq} (ns)	t_{prod} (ns)
300	1014.38	10	100
280	1023.28	10	100
260	1029.12	10	100
250	1030.28	30	100
240	1030.22	30	100
230	1028.63	30	100
220	1025.16	30	100
210	1019.97	30	100
200	1013.33	60	200

of 1 fs.
The equilibration lenght ranges from a minimum of 10 ns at high temperatures to a maximum of 90 ns at low temperatures. The production run led every trajctories up to 100 ns and up to 200 ns for the lowest temperature (200 K). During every simulations, we monitored the thermodynamic variables as temperature, total energy and pressure. In figure 4.2 we show that the system has reach a stable temperature (a), a stable total energy (b) for the full lenght of each trajectories. In the panel c) is shown the representative pressure check for T=300 K. Details on the protein hydration water trajectories can be found in table 4.1.

To better understand the concept of hydration water, we show in figure 4.3 the RDFs calculated between two specific atoms on the protein (NZ and N) and the oxygen (and the hydrogen) atoms of the hydration. NZ and N refer to two nitrogen atoms located on the 13^{th} residue (a residue is a part in which is divided the protein's chain). The curves show two behaviours:

NZ the hydrated site shows two peak within 6 Å;

N the not-hydrated site does not show any peaks.

Simulations were conducted on the INFN-Grid Roma Tre cluster. In the next section, a brief summary on previous findinds on the dynamics of hydration water is given, in order to collocate this thesis in the right framework.

4.2 The second slow relaxation of hydration water

We already introduced the role and the importance of hydration water (HW) for biomolecule systems in section 1.3.2 and we describe the dynamics of supercooled water in the second chapter. Here, we focus on the dynamics of supercooled HW.
In the last years, many experiments have been perfomed by depolarized light scattering [59, 131] on dilute water solutions of protein and other biomolecules. These works showed that HW has two slow relaxations, while bulk water has only one. The first one is the well known α-relaxation: the structural relaxation of bulk water upon cooling that we described in details in chapter 2. The second one is slower and it is associated with the HW only, because its timescale is related with the complexity of the biomolecule [59, 132].
These two relaxations were also detected in sugar solutions. The second one was assigned to the coupling of water with the solute dynamics [60] and measurements by means of broadband dielectric spectroscopy have confirmed the simulations [133]. However, these experiments are indirect evidences because they average over HW and bulk water contained in the systems. There were also exeperiments being able to focus only on the dynamics of HW performed with intrincinc fluorescent probes [134, 135] or 2D infrared spectroscopy [136, 137]. They also found two slow relaxations belonging to HW, with the time scale depending on the protein site. The different residence time of water depending on the site of the protein-water interface is in agreemeent with other studies [129, 138, 139].
When the HW is studied in the supercooled regime, new dynamical features

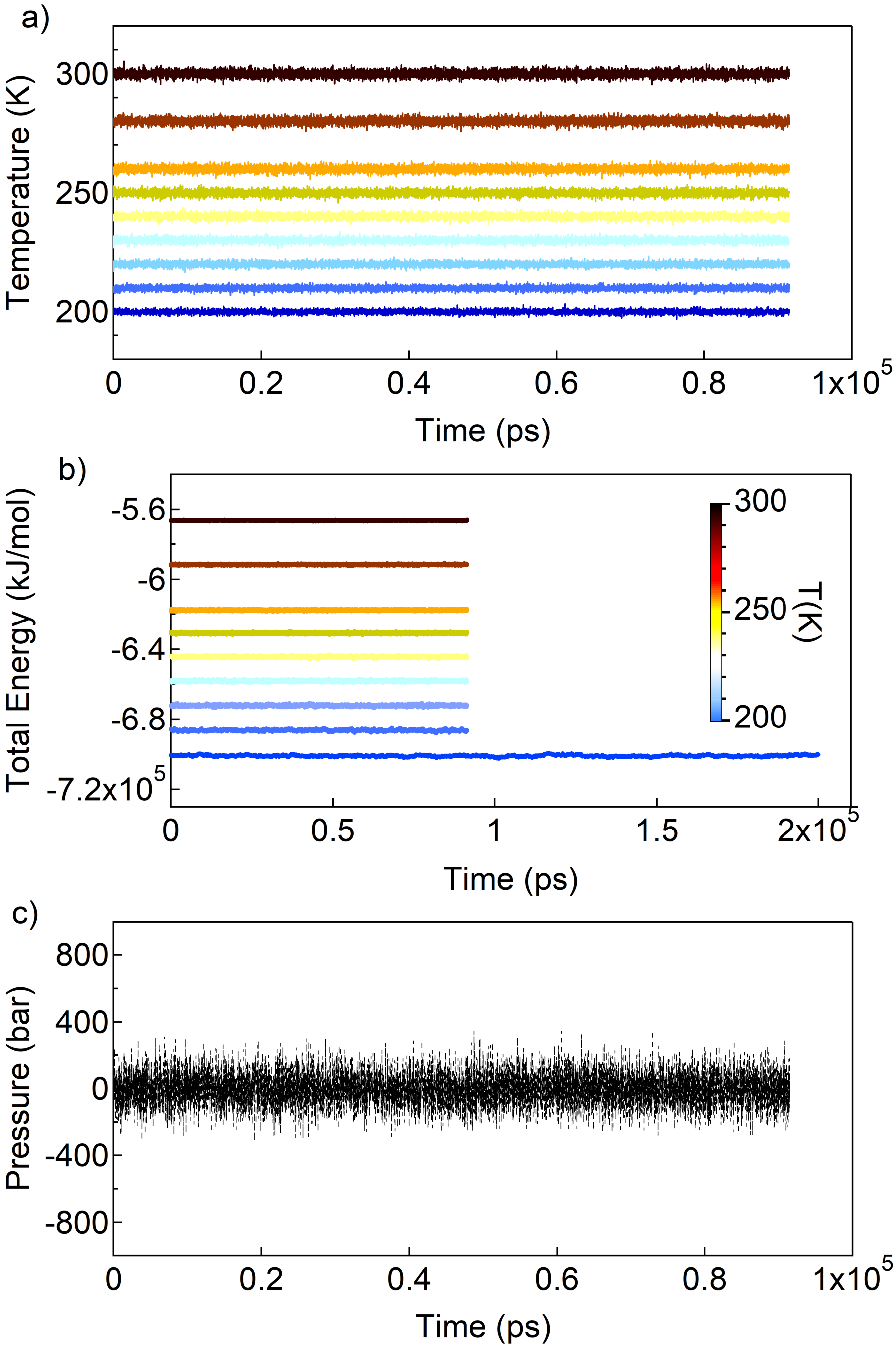

Figure 4.2: a) Temperature and b) total energy of the system of protein and hydration water on for mthe full lenght of every trajectory; c) pressure check at T=300 K. Color coding is indicated in the panel b.

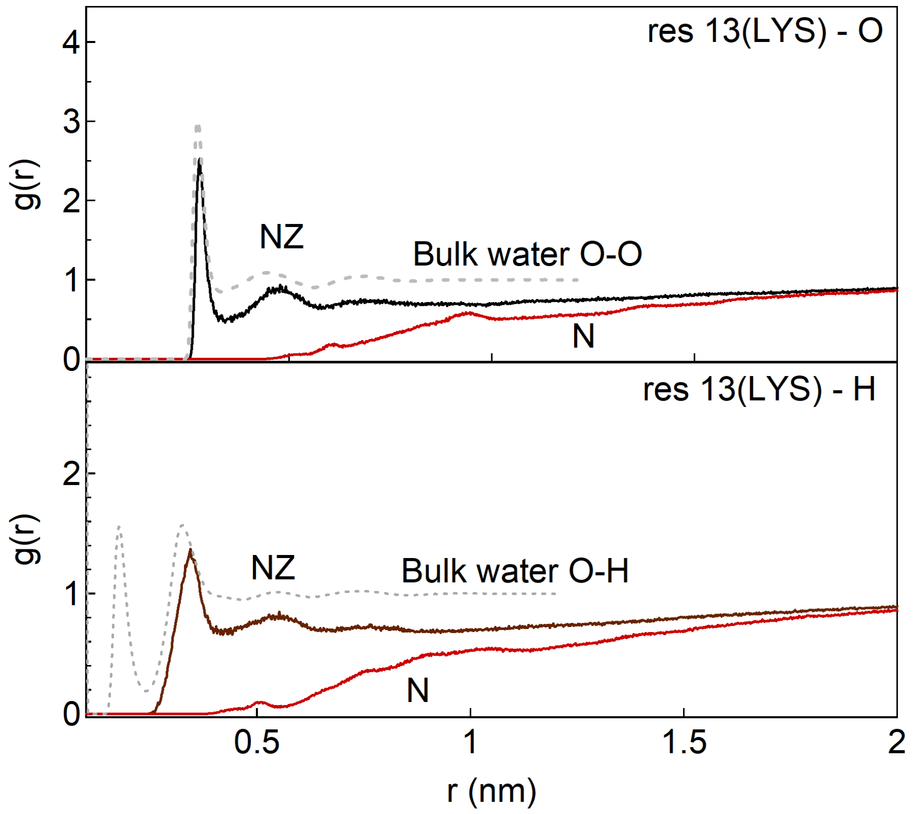

Figure 4.3: Radial distribution functions at 300 K between two lysozyme sites and hydration water. The curves are site-oxygen (top) and site-hydrogen (bottom). The sites are two nitrogen atoms labeled NZ and N both located on the 13th residue. Grey dashed curves are the oxygen-oxygen (top) and oxygen-hydrogen (bottom) of bulk water.

emerge both in protein and HW. We already mention in section 1.3.2 the Protein Dynamical Transition (PDT), an abrubt increase of the protein's flexibility over a certain temperature that happen only when the protein is minimally hydrated [140]. Experiments performed through quasi-elastic neutron scattering and MD simulations on poorly hydrated proteins [65, 66, 70, 141] showed the presence of a dynamic crossover in the α-relaxation in coincidence of the PDT temperature.

MD simulations with density-density autocorrelation functions enables the direct study the traslational dynamics of HW alone [61, 130] in a range of hydration levels and with different biological evironments. In these studies, the two slow relxations in supercooled HW emerge clearly. Comparing the time scale and the temperatures of these two decays of the density correlators, the main results were:

- the first relaxation is faster and it is the analogous of the α relaxation observed in bulk water. Consequently it was called α relaxation;
- the second relaxation is slower than the first α and it exhibits a different temperature behaviour with respect to the α decay. It is termed *long* relaxation. The long relaxation was show to be sensitive to the PDT: the strong loss of the protein flexibility directly contribuites to the change in the temperature relaxation time behaviour of the correlator and this change corresponds to the PDT transition.

Despite these findings, much of the work in order to characterize and understand the nature of the long relaxation has still to be done.

4.2.1 Slow dynamics of hydration water by SISF

After introducing the second long relxation, we can show the main findings on the translation dynamics of hydration water upon cooling [60, 61, 123, 130, 142]. The SISFs was defined in section 3.3. Analysing the correlator is particularly useful to characterize the translational dynamics of the system: different microscopic mechanisms in the diffusion can be tracked in the correlator shape.

The SISFs of hydration water have been calculates from oxygen trajectories obtained by MD simulations in the temperature range 300-200 K. In figure 4.4 is shown the SISF calculated at 220 K and at the maximum value of the structure factor for water, i.e. $q = q_{max} = 22.5$ nm^{-1}, where the features of the slow dynamics in the supercooled regime are best detectable [143]. We can summarize the the dynamics emerging from figure 4.4 in the following way:

- Ballistic regime: the first decay of the curve it happened at short time. In this short time the tagged molecule is far enough from its neighbors and moves as a free particle;
- Cage regime: as the time increases, the tagged molecule starts to feel the neighbors interactions and it gets trapped in a sort of cage. The motion here is a rattling inside the cage and this situation is placed on intermediate times (~ 1 ps), and it corresponds to the plateau region in figure 4.4. The value of the plateau is connected to the cage dimension and the lenght of the plateau depens strongly on the temperature. Infact, the cage regime is absent at high temperatures and becomes longer and longer

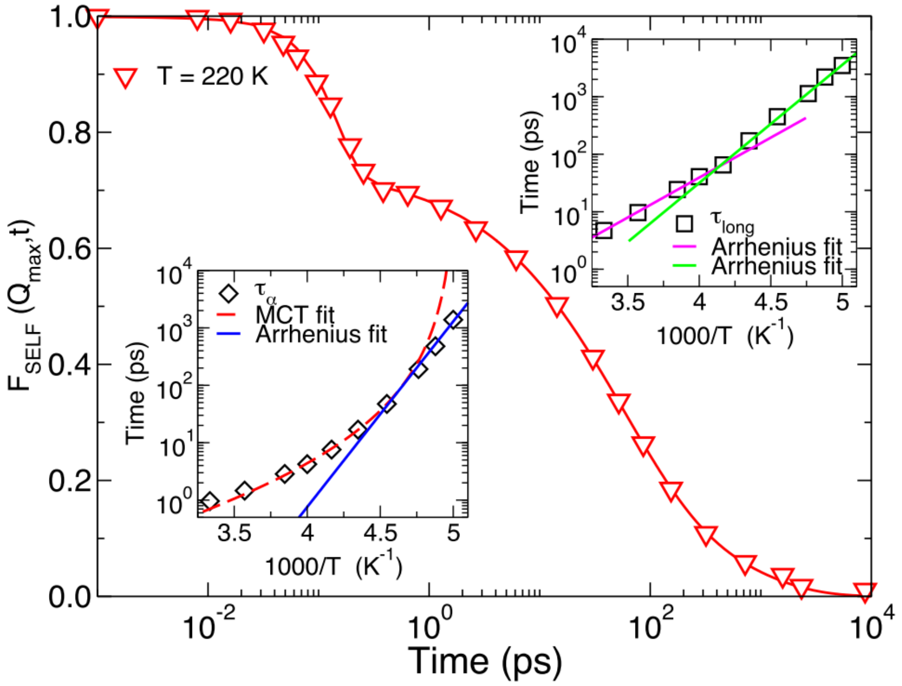

Figure 4.4: $F_{self}(\overline{q}, t)$ of lysozyme hydration water oxygen at 220 K. The continuous line is the fit to equation 2.35. Left inset: time of the α relaxation τ_α as a function of the temperature. It has a fragile behaviour at high temperature fitting the MCT (dashed red line), while it switches to a strong behaviour at low temperature fitting the Arrhenius law (continuous blue line). The fragile-to-strong crossover is observed at 215 K. Right inset: time of the *long* relxation τ_{long} as a function of the temperature. It fits two Arrhenius laws with two activation energies at low and high temperatures. This strong-to-strong crossover and is observed at 240 K, where the lysozyme shown the PDT. Figure from ref. [123].

(several order of magnitude) when the system approaches the glassy state [143];

- α-relaxation: when in the liquid state, the cage eventually will relax and the tagged particle is again free to explore othe locations. In pure bulk water, the α-relaxation leads the correlator to zero. This restored diffusive regime is described by MCT in the mid-supercooled region for many glass formers including water. For HW, a sub-diffusive regime is reached in this phase [80, 123];
- *long* relaxation: the long decay to zero of the correlator is not due only to the α-relaxation. It is a second slower relaxation arises from the analysis of the correlator. This decay is due to the coupling of the HW with the protein [61].

Already in a study of hydration water of trehalose [60], it was found that hydration water SISF could not be fitted with a Gaussian plus a stretched exponential, as shown in the formula 2.35, but two stretched exponential were needed. Two times, τ_α and τ_{long}, can be extracted by a double-stretched exponential function [60, 130]. The formula is

$$F_{self}(\overline{q},t) = (1 - f_\alpha - f_{long})\, e^{-\left(\frac{t}{\tau_s}\right)^2} + f_\alpha\, e^{-\left(\frac{t}{\tau_\alpha}\right)^{\beta_\alpha}} + f_{long}\, e^{-\left(\frac{t}{\tau_{long}}\right)^{\beta_{long}}}. \quad (4.1)$$

The first term is Gaussian, related to the ballistic regime with characteristic time τ_s (short time), while the two stretched exponential functions fit respectively the α and the *long* relaxations. The *long* term is essential to fit the tail of the correlator.

Both structural relaxations have been charachterized in temperature as shown in the insets of figure 4.4. The behaviours in temperature of τ_α and τ_{long} are shown in two Arrhenius plots:

τ_α can be described with a power law upon cooling, in the form of

$$\tau_\alpha \sim \frac{1}{(T - T_C)^\gamma} \quad (4.2)$$

with $T_C = 199$ K and $\gamma = 2.68$. This is the typical behaviour of the α-relaxation time of a fragile liquid as predicted by the MCT. Instead, the behaviour of this characteristic time at low temperature can be described by the Arrhenius law

$$\tau_\alpha \sim e^{\frac{E_A}{k_B T}} \quad (4.3)$$

with activation energy $E_A = 61.9$ kJ/mol. This is the energy that characterized the activation of hopping processes. These processes start when the cage is frozen and they are not taken into account by the MCT. Systems with α-relaxation times which follow the Arrhenius law are classified as strong liquids. A dynamic crossover fragile-to-strong is therefore detected at 215 K for HW of the lysozyme protein. For comparison, the fragile-to-strong crossover in SPC/E bulk water in the α-relaxation has been detected at 210 K [106]. The fragile-to-strong crossover in bulk water is related to the crossing of the Widom line, the line of maxima response functions emanated by the LLCP in the no man's land [1, 13, 14, 45] as discussed in 1.2;

τ_{long} is the charatertistic time of the *long* relaxation that is missing in bulk water. This time has a different temperature behaviour with respect to τ_α. It can be described with two Arrhenius laws (see insets in figure 4.4) with two activation energies $E_{A1} = 26.6$ kJ/mol at high temperatures and $E_{A2} = 39.3$ kJ/mol at low temperatures. τ_{long} show a strong-to-strong crossover at 240 K. This relaxation happens over longer timescales with respect to the α one and it is coupled with the protein dynamics, as it was shown in ref. [61]. The MSD of the hydrated protein fluctuations changes in the slope at the PDT (240 K), therefore in coincidence with the strong-to-strong crossover of the τ_{long}.

4.3 Results and discussion

We now show the results obtained in this thesis on the dynamics of hydration water by analyzing the Self Van Hove correlation Function and by the Mean Square Displacements.

The SVHF was defined in section 3.3. In this study we calculated the radial part of the SVHF, i.e. $4\pi G_{self}(r,t)$ which is the probability density of a displacement $\overline{r}$ at time t. By the radial part of the equation 3.78 we can describe the average evolution in time and space of hydration water molecules by considering the positions of oxygen atoms moving in the lysozyme hydration water shell. The SISF and the SVHF are Fourier transform pairs, therefore the study of the SVHF gives complementary information about the relaxation mechanism of supercooled hydration water. These functions have been calculated for water simulated with different potentials [8, 144–146] and recently measured with X-ray scattering experiments [147].

The SVHF is usually unimodal when the dynamics follows the MCT: the probability distribution has a single peak which moves towards larger distances while the distribution spreads and becomes lower. We already described the hopping phenomena (or activated processes): in the cage regime, when the molecules are trapped by their frozen neighbors, some molecules are allowed to hop toward to some energetically favoured positions. When these processes become relevant in the dynamics, the shape of the SVHF become multimodal, i.e. it can develop multi-peaks or shoulders, meaning at a given time different groups of molecules have travelled (on average) different distances. The hopping phenomena and the cage regime have been already observed through SVHF in many glass formers [148–150], but also in bulk water [8, 145] and in sugar-water solutions [151].

The MSD was used to better characterized the translational motion of hydration water. The MSD is defined as

$$MSD = \frac{1}{N}\left\langle \sum_i |\overline{r}_i(t) - \overline{r}_i(0)|^2 \right\rangle \tag{4.4}$$

and we calculated this quantity considering the positions of the oxygen atoms of HW only.

4.3.1 SVHFs of hydration water for the short time dynamics and α slow dynamics

In a previous work [123], the study of hydration water through SVHF was at 300 K on trajectories of 20 ns. There, was observed a new hight temperature hopping regime not connected to the activated processes of the glassy dynamics, but rather on the protein rearrangements. Here we have added this study in order to better undestand these mechanisms and how and if they are present at low T.

In this section we present many SVHFs calculated at different correlation time in the temperature range 300-200 K. In figure 4.5 are shown the SVHFs of hydration for three temperatures (200, 240 and 280 K) with correlation times spanning over six orders of magnitude (from 56 fs to 15 ns). At short time (blue curves) we observe the ballistic regime: the shape of the probability distribution is narrow and highly peaked at short distances. Their maximum shifts to longer distances and the tagged molecules moves as a free particle up to 240 fs. This regime is not strongly affected by the temperature, as we can see by looking the red curves of the three panels.

After the ballistic regime, the structural relaxation stretches as the cooling proceeds. Correspondingly, the curves continue to shift to longer distances and spreading.

At 240 and 200 K, orange curves are used to highlight an intermediate regime of the dynamics: we observe that the orange curves group together for several picoseconds. This is the cage regime induced by the supercooling. As the time passes, these curves look overimposed without showing any temporal evolution. The particles are trapped by the caged formed by their neighbors at a distance of $\sim$ 0.05 nm from the original point, which is well inside the position of the first peak ($\approx$ 0.28 nm) of the radial distribution function. This estimate is also in agreement with the cage radius measurement extracted from the SISFs of hydration water, which is

$$a \sim \frac{-3\sqrt{\ln(\phi)}}{q_{max}^2} = 0.046\,\mathrm{nm}$$

where $\phi \sim 0.7$ is the plateau value of the SISFs. The same cage radius was also determined for bulk water [8].

The cage relaxation permits to the molecules to move again and to the SVHFs to evolve and spread over larger distances. In figure 4.6 we show the longest correlation times from 100 ps to 15 ns in order to show the long time correlation regime. All the curves except the ones for $t > 5$ ns at high T are unimodal distributions as the predicted by the MCT. In the next section we will focus on the multimodal distributions at high temperatures, that are not related with the bulk water glassy behaviour.

From the comparison among the panels in figures 4.5 and 4.6 it is also possible to appreciate the slowing down of the motion of hydration water molecules upon supercooling: the particles move less at fixed correlation time when lowering the temperature. We now focus on the lowest temperatures, below the MCT crossover, at T=200 K where the cages are frozen. In this situation, hopping phenomena restore ergodicity. In figure 4.7 we show the SVHF at 200 K for the extremely long correlation time of t=50 ns. In the same figure we reported the radial distribution function (RDF) of SPC/E bulk water at the same

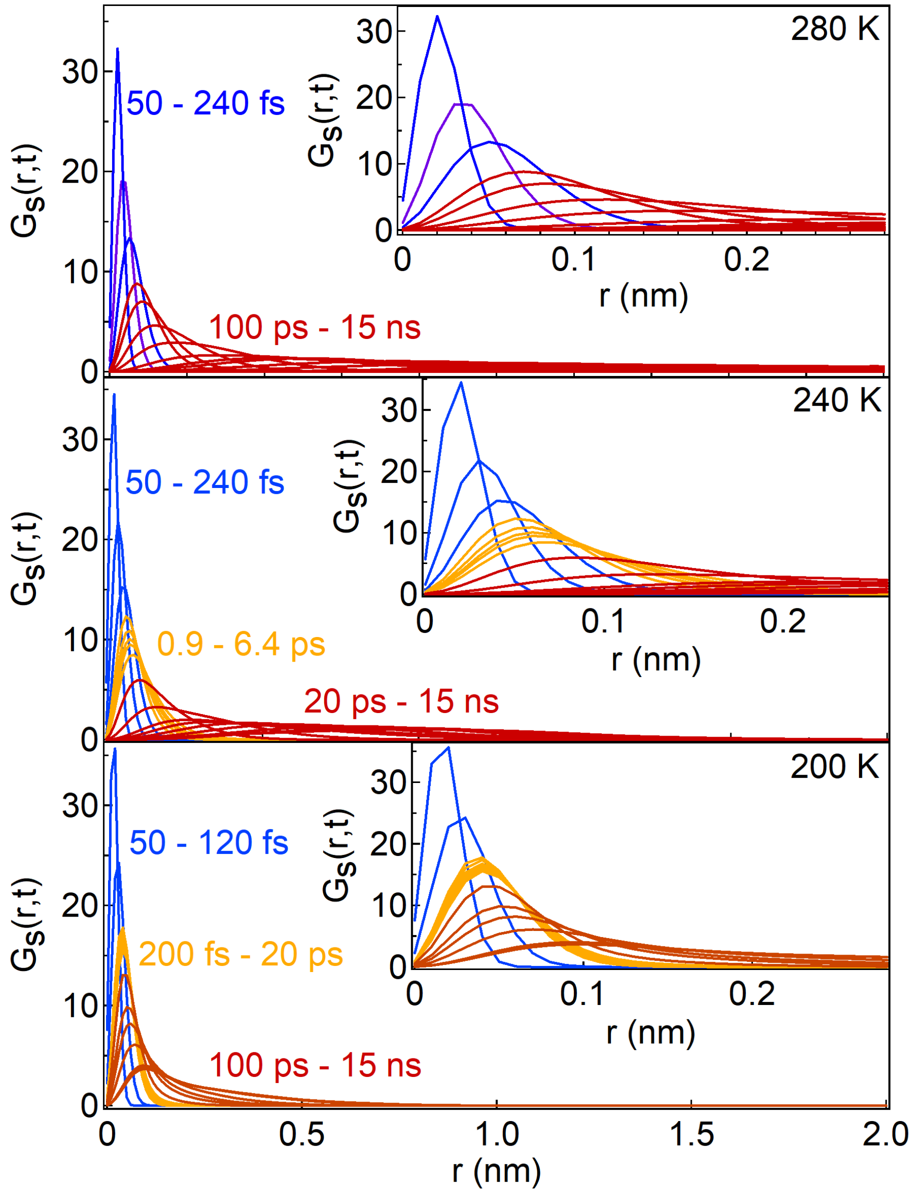

Figure 4.5: Self Van Hove correlation functions (SVHFs) calculated at 280 K (top panel), 240 K (middle panel) and 200 K (bottom panel). The different colours are used to better highlight the different dynamics regimes. At short time the particle moves as a free particle in the so called ballistic regime (blue) that is found for all temperatures. At intermediate times and enough low temperature water is the transient (or cage) regime (orange) where every molecule is trapped in cages made by its neighbors. We can observe how the orange curves peaks mantain its position over time. Finally, at long times (red), the probability density curves show an α-relaxation process similar to bulk water. This process is well described by the MCT. In the insets of the figure we show a blow-up at short distances to best underline the cage effect.

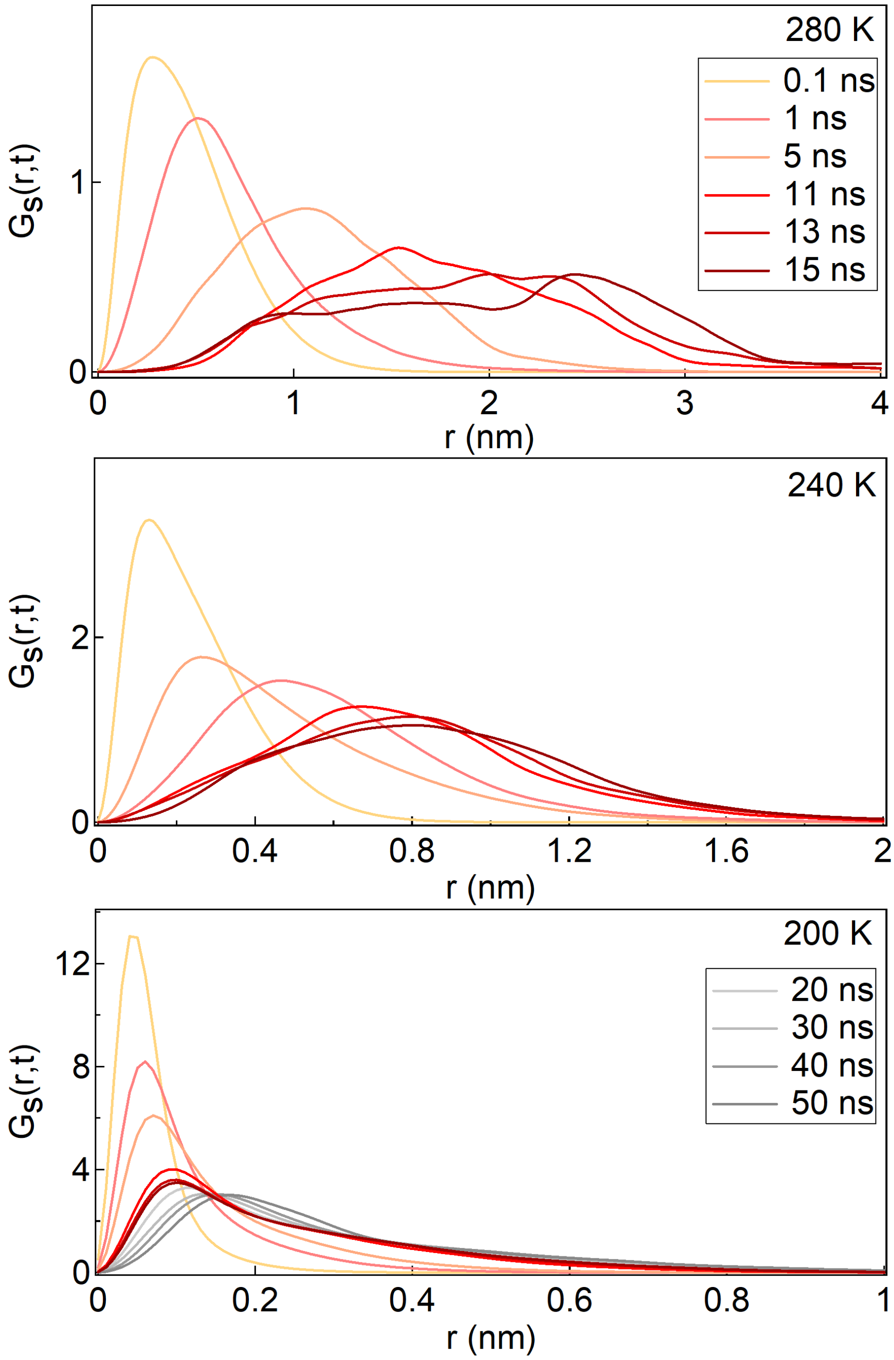

Figure 4.6: A blow up of the SVHFs curves showed in figure 4.5 for long correlation times: from 100 ps to 15 ns. At 280 K (top panel), the distributions start to assume a multimodal shape, sign of some new hopping phenomena. At 240 K (middle panel) the SVHFs are more regular and they re-gain a unimodal shape like the α-relaxation should appear as predicted by the MCT. At 200 K, the lowest temperature (bottom panel), the trajectory was elongated up to 200 ns, so new SVHFs from 20 to 50 ns were calculated (grey curves). Notice the scale changes in both axes for the three panels.

temperature and pressure. The peaks of the RDF permits to easily identify the first three coordination shells around the central water molecule. In the blow-up (bottom panel of figure 4.7), it is possibile to appreciate the hopping phenomena not predicted by the classic MCT but included in its extended version. These phenomena appear like bumps os shoulders in the SVHF and we highlight how these bumps correspond with the second and third shells of the RDF of bulk water. When approaching the MCT crossover temperature T_C from above, thermal fluctuations become small enough to prevent the cage relaxation and the tagged particle would remain stucked in the cage causing the structural arrest of the liquid. However, we observe translational motion of the water molecules at slow temperatures due to the activation of a new relaxation mechanism called hopping. Hopping phenomena permit the structural relaxation at low temperatures and their activation results in a strong regime for the α relaxation time. In order to escape the cage, water molecules hop to new, more energetically favoured, locations which coincide with the distances of the neighbors. These phenomena were extensively studied in bulk water [145].

4.3.2 SVHFs of hydration water for the long time dynamics

In the figure 4.8 we show the temperature evolution of SVHFs of hydraiton water in the entire temperature range (300-200 K) at four long correlation times: 1, 5, 10 and 15 ns.

Comparing these curves at a fixed correlation time permits to quantify the slowing down of HW induced by the cooling. We notice that a given curve is lower and wider with respect to a curve calculated at lower temperature. Moreover, when comparing curves at the same temperature and different correlation times we notice that the system continues to evolve, expecially at high temperatures as the curves change significantly their shape. At low temperatures the curves vary much more slower with time even if we are calculating them for very long correlation times. For example, at T=200 K we found a single peak at ~ 0.2 nm both for t=5 ns and t=15 ns. We start noticing a multimodal shape in the curve calculated at 300 K for a correlation time of 5 ns (fig. 4.8 top): two shoulders at larger distance with respect the main maximum develop. This feature becomes more evident at longer correlation times. In particular, the curves at 300 and 280 K for t=10 ns and t=15 ns. These curves have completely lost their unimodal shape that they had at shorter correlation times. In figure 4.9 this fact is particularly evident. Here, we compare the SVHFs calculated at 280 K (top) and 240 K (bottom) with correlation times which span from 9 to 15 ns. We can appreciate how, at fixed correlation times, the diffusion mechamism is different at high temperature where the curves have lost their unimodal shape with respet the curves at 240 K (which coincide with the PDT temperature), where the curves show a slower bulk-like diffusion. To highlight this fact we show again the curves at T=280 K (figure 4.10) and T=240 K (figure 4.11) with correlation times from t=240 fs to t=15 ns. In the bottom panels the correlation times at which we observe the main peak and its shoulders are reported as a function of the position of these features. At T=280 K we find a single peak of the probability distribution up to 9 ns, while for longer times multiple peaks and shoulders appear. Some curves, such as the one calculated at t = 12 ns, have up to four different shoulders and therefore locations in space where

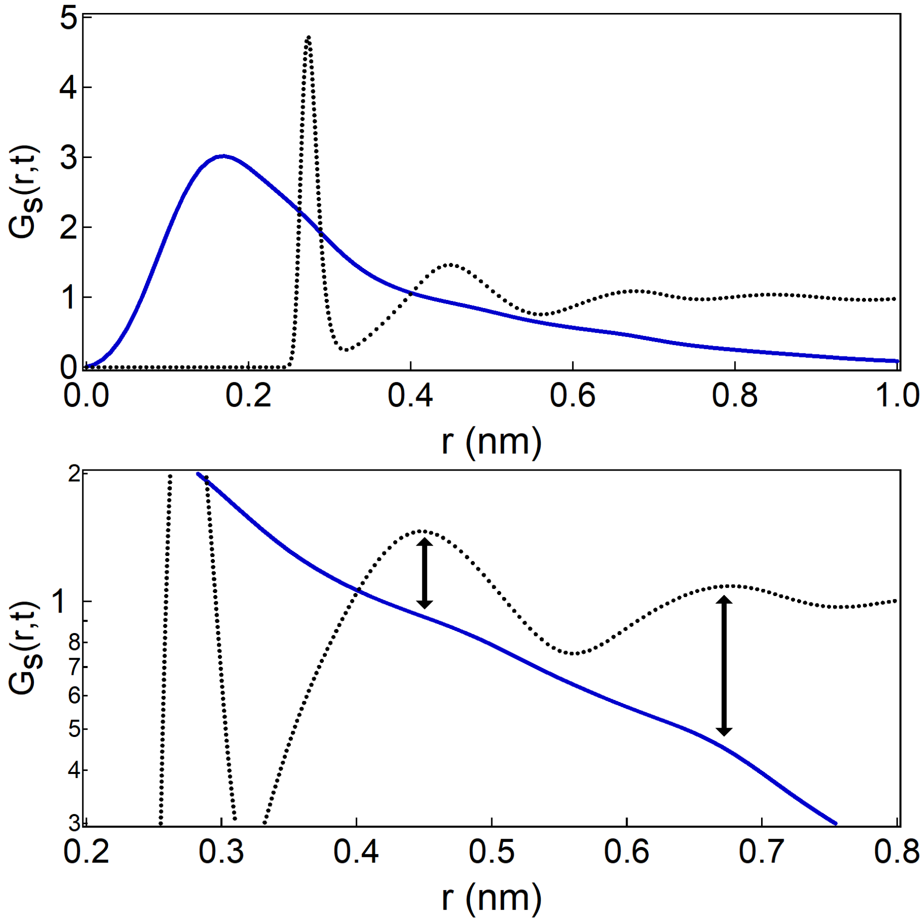

Figure 4.7: Top panel: SVHF (blue continuous line) for hydration water at 200 K for a correlation time of 50 ns. In this curve hopping phenomena are visible. The shoulders in the right tail of the curve are due to the hop of the particle out of the cage. The black dashed line is the radial distribution function of oxygen-oxygen bulk water at the same T. Bottom panel: blow-up of the same curves in semi-logarithmic scale. This plot allows us to appreciate the correspondance of the SVSF shoulders with the location in space of the second and third shells around the particle.

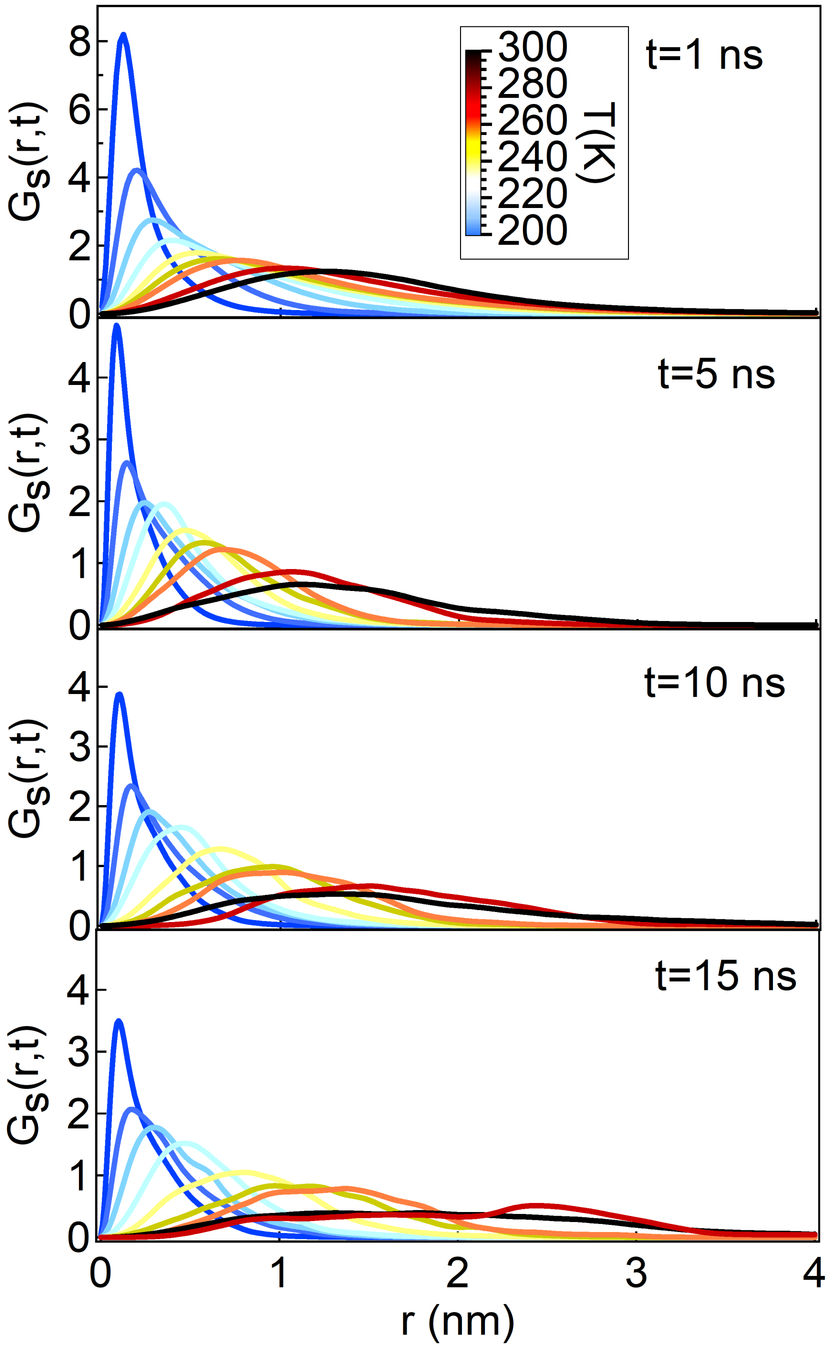

Figure 4.8: SVHFs calculated from 300 to 200 K for oxygens of hydration water with a correlation time of t=1, 5, 10 and 15 ns. For t$\geq$ 5 ns at high T we start to see hopping phenomena described in figure 4.6. These phenomena appears already at high temperatures with very well defined peaks. They appear at lower and lower temperatures as the correlation time become longer.

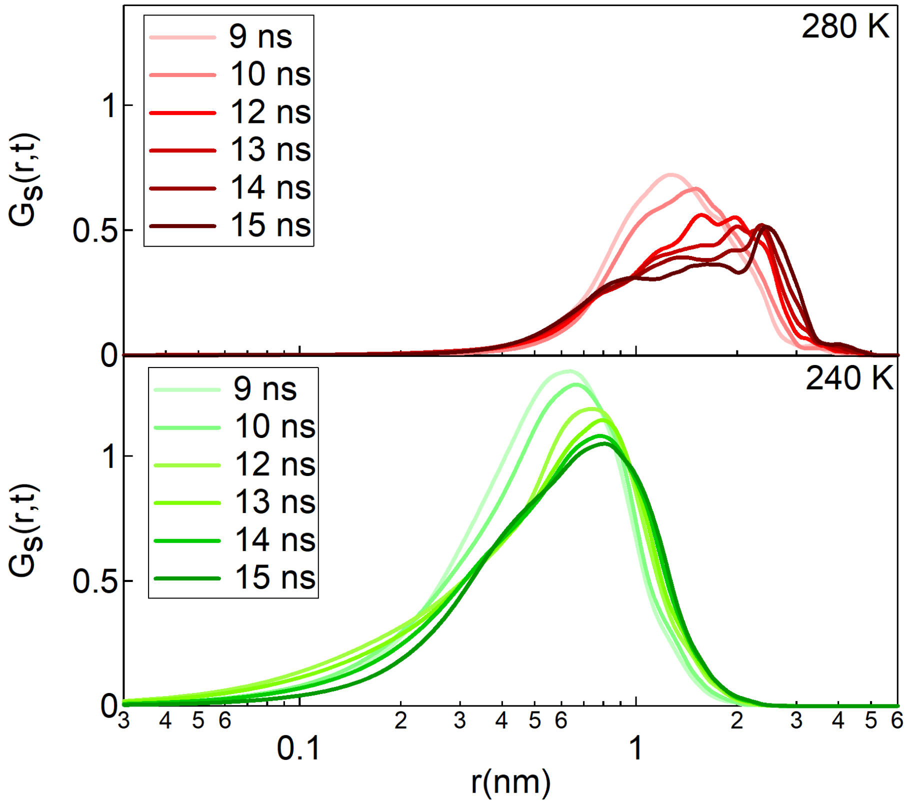

Figure 4.9: SVHFs of hydration water calculated at 280 K (top) and 240 K (bottom) for long correlation time. The curves at 240 K are still unimodal, while they are multimodal at higher temperatures due to the PDT.

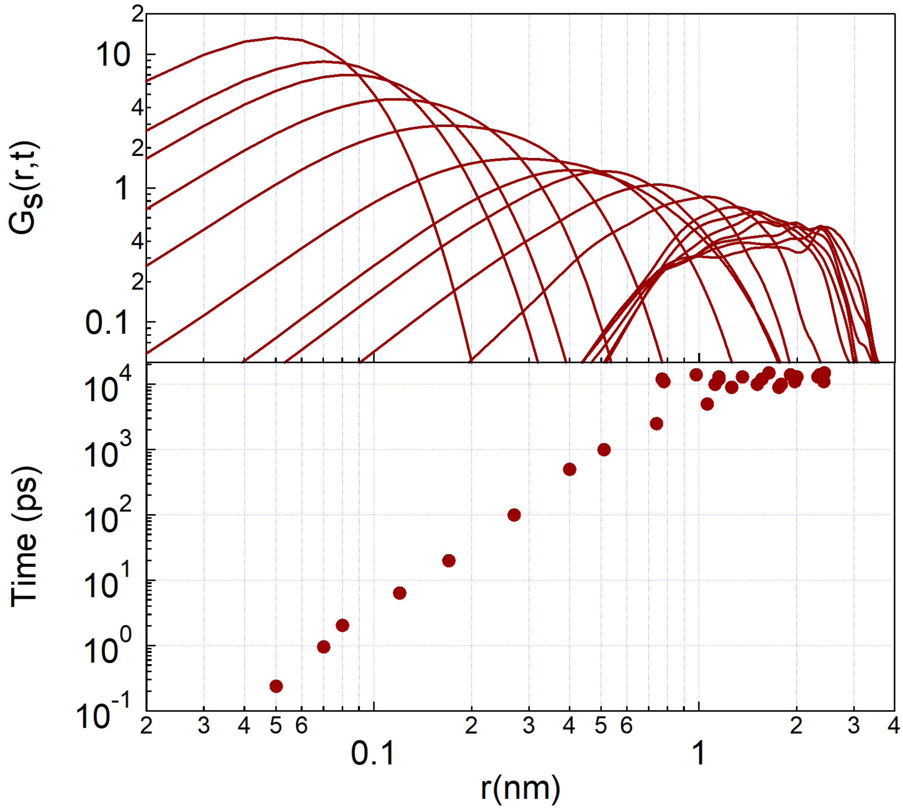

Figure 4.10: Top panel: SVHFs calculated at 280 K with correlation time from 240 fs to 15 ns. Bottom panel: peaks and shoulders of the distributions for every correlation time versus their positions. The curves are showed in double logarithmic scale to best highlight their shape features at all the correlation times.

the water molecule has a probability to have hopped. The peak at 2.5 nm becomes sharper and sharper as the time proceeds for t=13, 14 and 15 ns. These shoulders cover a wide space, form 0.8 to 3 nm. At T=240 K (fig. 4.11top) a cage regime between 2 and 6 ps is visible where peaks of the SVHFs (points in the bottom panel) are vertically aligned. After this regime we observe a single peaks moving toward larger distances. Despite the longer correlation time, the single peak shape is preserved.

We already said that the multimodal shape of the curves are a mark of the heterogeneities in the diffusion of the particles, generally called hopping phenomena. The hopping mechanim allows water molecules to escape from the frozen cage during the α-relxation in the strong regime. The mark of the hopping expected by the the extended MCT are shoulders in the SVHFs at temperature close to the fragile-to-strong crossover temperature T_C like the ones showed in figure 4.7. By the way, the hopping phenomena that we are dealing with in this case have a different nature. Indeed, we are observing:

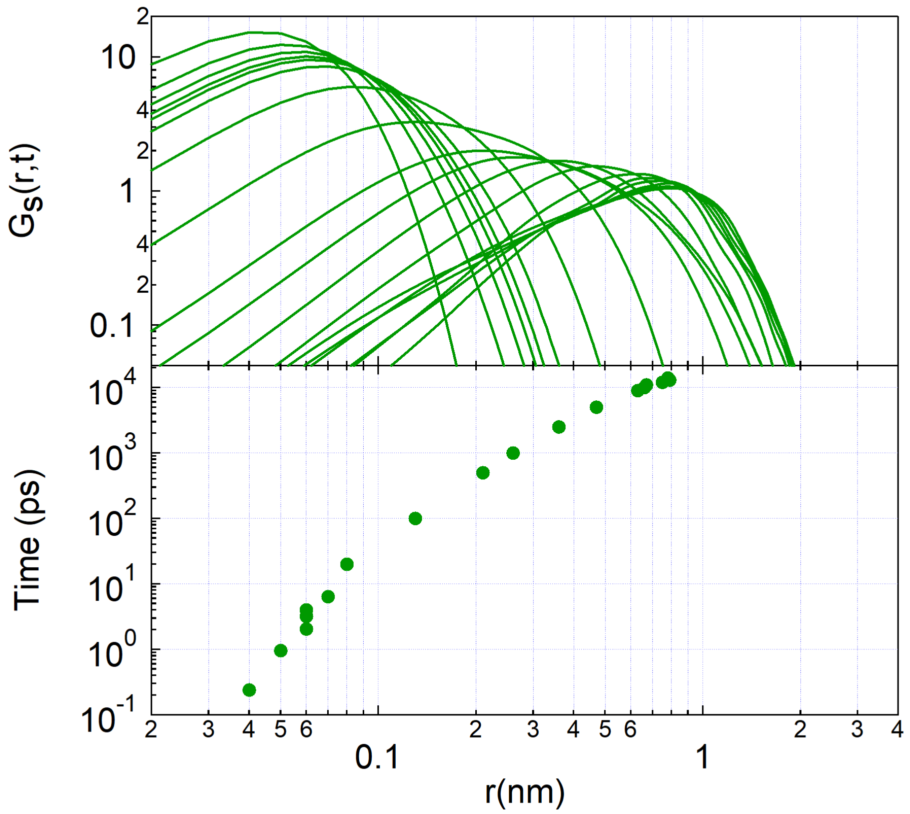

Figure 4.11: Top panel: SVHFs calculated at 240 K in double logarithmic scale with correlation time from 240 fs to 15 ns. Bottom panel: peaks and shoulders of the distributions for every correlation time versus their positions. The curves are showed in double logarithmic scale to best highlight their shape features at all the correlation times.

i the development of the shoulders in the SVHFs affects the large distances, being their position at 0.8-3 nm, where the RDFs approach unity, i.e. distances way larger the typical inter-particle lenghts of few water shells;

ii the temperature are higher than the temperature at which sets on the MCT cage effect;

iii the timescale of these phenomena is $\sim$ ns, beyond the usual α-relaxation time regime at that temperature.

We relate these long-time-high-temperatures hopping phenomena with the long time protein internal dynamics such as relative motion of different globular domains or the long relaxation of hydration water molecules positioned in different sites of the protein and with differente residence times.
Moreover, we observe these phenomena mainly from T=250 K, over the PDT temperature of our protein. Infact, below this temperature the protein movements are strongly limited.
Therefore, below the PDT, from 240 to 200 K, we are not able to follow the hopping that determines the strong behaviour of the long relaxation because below this transition the protein motion because is too slow, given the lenght of our simulations. The long relaxation has a strong-to-strong transition at the PDT and the correlation time grows very much below it. At low temperatures we are sensitive to the hopping due to the MCT cage effect and at high T to the protein-related hopping phenomena.

4.3.3 Two caging regimes and the Mean Square Displacements

We were interested to further investigate this long time diffusion mechanism only present in hydration water. In figure 4.12 we plot the value of the main maximum of SVHFs of hydration water as a function of the correlation time. Each curve corresponds to a different temperature, from 300 K to 200 K. For complete structural relaxations, the main peak of the SVHFs decays to zero, similarly to the SISFs.
We firstly observe a fast-decay at short time corresponding to the ballistic regime and at low temperatures the curves develop the typical plateau due to the cage-effect,that we observed also in the SISFs. This plateau begins around $\sim$ 250 fs and extends up to $\sim$ 10 ps at the lowest temperature.
After the ballistic regime at high T or after the cage regime at low T, the maximum of the SVHF starts to decay to zero: this corresponds to the α-structural relaxation regime. From the study of the SISFs this regime should be followed by the long relaxation regime.
Here, we observe a novel feature: a second plateau region appears in the curves at the high temperature, in constrast with to MCT cage-regime plateau that characterizes low T curves. This plateu does not appear in the SISFs calculated at a fixed wavevector q. The plateau new is observable at T=300 K in the long interval from 100 ps up to 1 ns. This plateau shifts later in time upon cooling, and we can follow it down to $\sim$ 230 K. This plateau indicates as a second caging regime, related to the protein-molecules bonds, which precedes the escape of the water molecules from the *protein cage*. This escape mechanism is lead by the high tempertures hopping phenomene we described earlier.

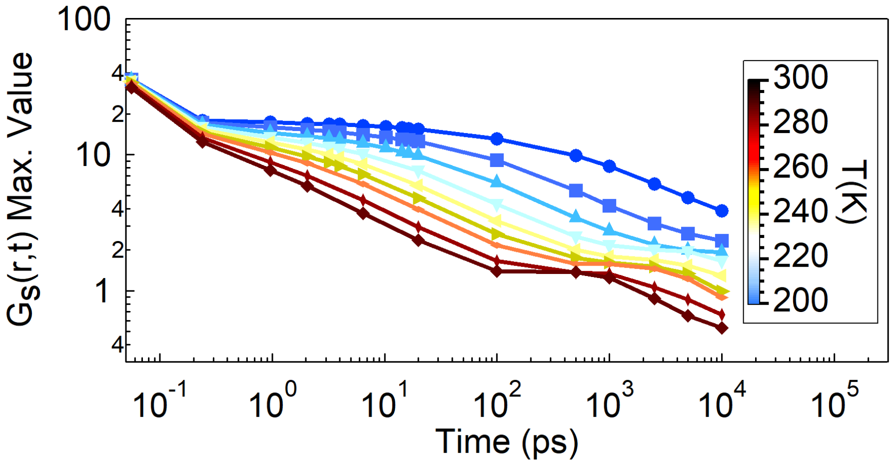

Figure 4.12: Value of the main maximum of every SVHFs over time.

To conclude our characterization, we calculated also the MSD of the protein hydration water, showed in figure 4.13. Here, we could also observe regimes of the translational dynamics of hydration water upon cooling. The ballistic regime stands up to ~ 250 fs and it depends weekly from the temperature. In this regime the MSD follows the predicted quadratic law $\sim t^2$. At the lowest temperatures the cage regime appears: here we observe a plateau region 10 ps long at 200 K, and we note that this is the same temporal window where the main maximum of the SVHFs in figure 4.12 shows its first plateau. This plateau of value $\phi^1_{MSD} \sim 0.003$ nm^2 corresponds to a trapping lenght of the water molecules of ~ 0.055 nm, a value in agreement with the cage dimensions of the water first neighbors. This estimate of the cage radius is in agreement with the one done by the position of the main maximum of the SVHFs and the plateau of the SISFs.

When the cage relaxes, or when the low temperatures hopping phenomena intervene, the MSD starts to increase again because the translational motion of the particle is restored. We measure here a sub-diffusive motion with time-dependence of the MSD of $\approx t^\delta$, with $\delta = 0.66$.

Then, we observe a new second plateau region in some of the curves. In particular this can be seen from T=300 K down to T=230 K. We note that the intervals of temperatures and the time-scale of this second plateau region is the same of the second plateau region of the main maximum of the SVHFs of hydration water in figure 4.12.

From this second plateau we measure $\phi^2_{MSD} \sim 0.6$ nm^2 at 300 K and ~ 0.3 nm^2 at 230 K. Correspondingly, the lenght scales of these traps are respectively ~ 0.77 and ~ 0.55 nm. This corresponds to the starting point of the shoulders we observe in the SVHFs in figure 4.7. Moreover, we stress out that this second trapping mechanism is temperature dependent, because the water molecules at high temperatures will escape with the long time hopping. This allow the MDS to increase again.

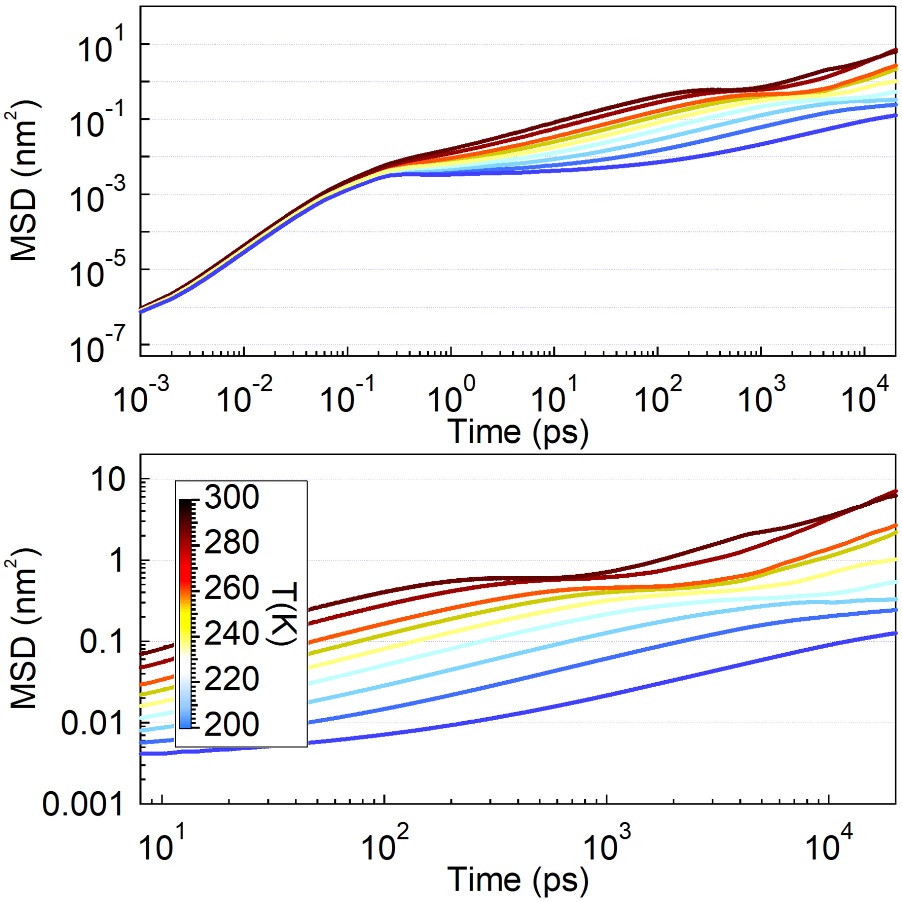

Figure 4.13: Mean Square Displacement (top) of protein hydration water. The molecules follow a sub-diffusive regime until 100 ps at 300 K, when they reach the new plateau. This feature is related to the plateau showed in figure 4.12 and it has different characteristic as compared to the plateau seen at low temperatures. Bottom: a blow-up of the MSD on the second protein related plateau.

4.4 Summary and Conclusion

In this chapter, we presented our results on the translational dynamics in the direct space of the lysozyme hydration water obtained by means of 100 ns long MD simulations. We analyzed the motion of the water molecules within a shell of thickness of 6 Å around the protein. To characterize the dynamics of these molecules we computed the self van Hove correlation functions and the Mean Square Displacements from T=300 K to the supercooled region T=200 K. We calculated the SVHFs in a wide range of correlation times: from tens of femtoseconds to tens of nanoseconds.

We found two caging regimes followed by a relaxation dominated by hopping phenomena. The two mechanisms have different origins. The first cage regime is related to the trapping cage made by the particle neighbors, a feature of the supercooled regime described by the Mode Coupling Theory for the study of glassy dynamics. This process was already seen in bulk water and we have the same timescale (tens of ns) also for HW. The lenght of the trapping region found is circa ~ 0.05 nm.

At very low temperatures, the cages should be frozen, hypothetically blocking further diffusion. Nonetheless, hopping phenomena intervene to restore ergodicity, within a lengthscale of 2-3 water shells and a time scale of tens of nanoseconds. Correspondingly, the structural α-relaxation times of hydration water shows strong behaviour, and therefore it is related to the fragile to strong transition of the α-relaxation time.

The second cage regime was observed at high temperatures, the cage radius is about 10 times larger than that charaterize the MCT cage. This lenght increases with the temperature (radius of ~ 0.77 nm at 300 K and ~ 0.55 nm at 230 K). The escape mechanism is a new high temperature hopping, which is characterised by a longer length scale (0.8–3 nm) and shorter time scale (nanoseconds) with respect to the MCT hopping phenomena characterising the low temperature region. Given the characteristics of this mechanism we relate it to the protein motions, and in particular to transient trapping of water molecules performed by the protein through hydrogen bonding. The water molecules that are bonded to the protein surface by hydrogen bonds could not move freely, but with the whole protein, following its domain rearrangements. We recall that we observe this mechanism mainly above the Protein Dynamical Transition temperature, when the protein mobility is generally enhanced.

Chapter 5

Structure of protein hydration water

This chapter is devoted to present the findings of this thesis on the structure of protein hydration water. We wanted to investigate the regularity and the distorsions of the water tetrahedral network of protein hydration water with respect to bulk water upon cooling. We calculated three different structure parameters among the most used and the radial distribution functions. Moreover, we compared the results in the equilibrated trajectories called Real Dynamics (RD), with the results in the Inherent Structures (IS), the trajectories where the thermal fluctuations are suppressed [152].

5.1 System and simulation details

We already described the system in the previous chapter. For the sake of the comparison between protein hydration water and bulk water, we exploited a system of bulk water containing 500 SPC/E molecules, at similar thermodynamic conditions as the protein plus water system [61]. We called *Shell* the system of protein hydration water, while we called *Bulk* the system of 500 SPC/E molecules. This nomenclature is used on all the figures. In order to calculate the structure parameters, that we described in section 3.3.2 we wrote custom programs.

For both systems, the lysozyme-water solution and bulk water, we calculated IS from the equilibrated MD configurations. Energy minimization was performed with the steepest descent algorithm at each configurations in GROMACS, the stopping criterion corresponds to a maximum force of 10 kJ/mol. We extracted from every RD trajectories 20 thousands frames of the system water+lysozyme. In figure 5.1 are shown the histograms of the potential energy of every minimized frame. These distributions become sharper at lower temperature.

We calculate the average energy values from each histogram and we plot them as a function of T in figure 5.2. The minima exhibit two different linear behaviour with a crossover around 240 K. This crossover coincides with the strong-to-strong crossover in the long-relaxation time that we discussed in the previous chapter as well as with the Protein Dynamical Transition temperature. We recall here that the *long* relaxation is a diffusion process connected to the proteinhyration water coupling and the PDT is a sudden lost of vibrational properties on the protein surface upon cooling. The loss of degrees of freedom below the PDT is in agreement with a deeper minimum in the energy landscape. IS reflect the sudden change in the dynamical properties of the system. Studying the

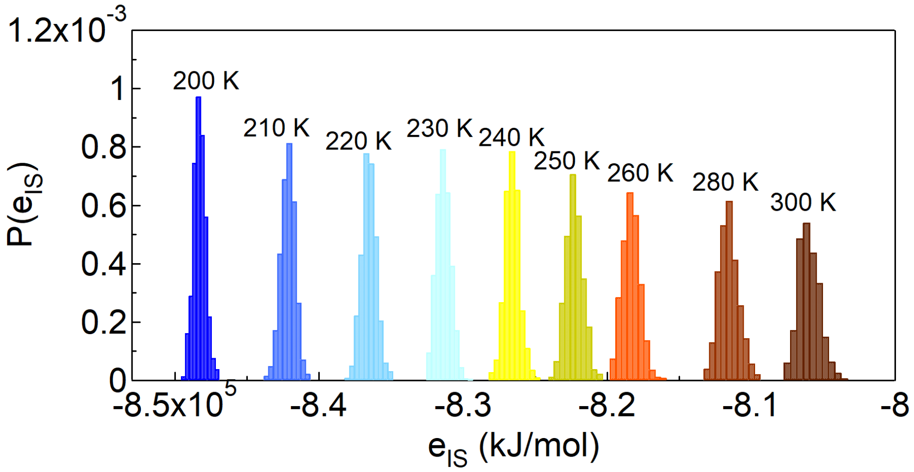

Figure 5.1: Histograms of potential energy in the inherent structures (IS) for the system made by lysozyme and water for all the temperatures.

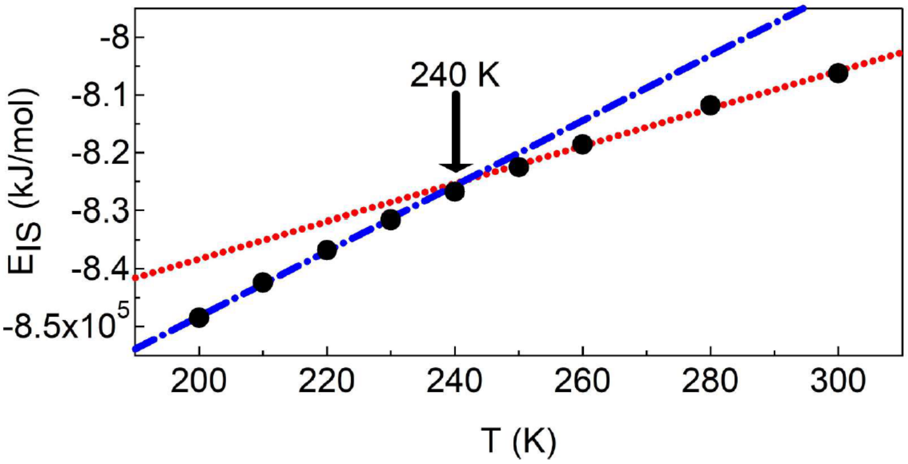

Figure 5.2: Average potential energy E_{IS} of the inherent structures versus temperature. The dotted red line and the blue dot-dashed line are linear best fit to the data points. The crossing temperature T=240 K is also indicated.

local structure of protein hydration water, the Radial Distribution Functions (RDFs) were the first observables we calculated to osbserve the internal structure of hydration water. In a liquid system, the RDF should quickly reach the unity after a lenght of 3-4 coordination shells. In a isotropic system like bulk water, the RDFs reach spontaneously the unity value, but in the shell of 6 Å

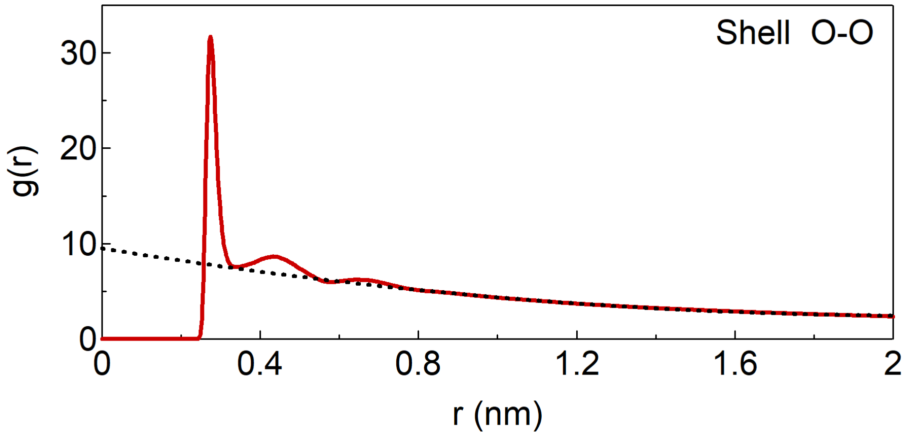

Figure 5.3: Normalization procedure of the hydration water RDFs. The red curve is the oxygen-oxygen RDF at 300 K of hydration water as calculated with GROMACS embedded tool, while the black dashed curve is the polynomial best fit calculated on the long tail of the RDF.

of hydration water around the protein, the RDFs are affected by a distortion because the system of the protein with its hydration water is anisotropic. For this reason, in order to obtain a normalized $g(r)$ of hydration water, we have divided each curve for a polynomial fit calculated on the tail of the RDF, as showed in figure 5.3.

5.2 The double local structure of water

In the first chapter, when presenting the phase diagram of supercooled water, we introduce the polymorphism of the amourphous ice in the deep supercooled region with LDA and HDA states separated by a first order transition line which ends into the Liquid Liquid Critical Point (LLCP). The continuations of this line, beyond the LLCP is the Widom line that at higher tempeatures separates two metastable liquids: HDL and LDL.

In the two phase region water molecules show a two-liquid local arrangement, studied in both experimental and simulation works, see for example [49, 79, 123, 153–160] and a pre-print [161] that appears coherent with the second critical point scenario expecially upon supercooling, and persisting also up to ambient conditions, see for example [153].

Several indicators of local structure have been used in literature to charachterized the features of bulk water as a two state liquid. Among the most used indicators are the Local Structure Index (LSI) [152, 160, 162], the orientational tetrahedral order parameter (q) [29] and the recently proposed new structural indicator called V_4 [122] based on potential energy considerations. The fraction of the two liquids changes depending on the temperature and there is no a unique parameter able to distinguish the two phases.

Structural differences between bulk water and water in biomolecules hydration layers have been found in literature, but most of these studies are limited to

ambient conditions [163, 164]. Close to a protein the hydration water network is affected not only by the temperature, like bulk water network, but also by the coupling with the protein surface [57]. Some results show that also the protein shape influences the hydration water network, e.g. globular intrinsecally disorderd proteins [165].
We expect water structure to undergo some change when a protein breaks into the water network. We tried to understand the modifications of the structure of lysozyme hydration water upon cooling and discuss them in the light of the biological role of this important kind of water.
We mainly use three structural indicators: LSI, q and V_4 both in bulk and hydration water.

5.3 Order parameters

We describe in the following our results for the selected order parameters for bulk and hydration water: Local Structure Index, Orientational Tetrahedral Order parameter and the V_4 parameter. To assess the difference between bulk and hydration water we compare the order parameters for both RD and IS. We report the probability distributions for the before mentioned parameters. The distributions are bi-modal, so their peaks can be related to a *moreLDL* behaviour or a *moreHDL* behaviour. We used the minimum in the IS probability distributions as separator the two populations upon cooling, labelled *moreLDL* or *moreHDL* as the corresponding peaks. For the parameters with better shaped bimodal distributions q and V_4, we report the positions (abscissa) of the peaks and the height of the peaks. The first gives us information about the shape of the local arrangement (more tetrahedral or more distorted) and the latter describes qualitatively the change into the amount of the two water kinds.

5.3.1 Local Structure Index

We already presented the definition of our observables of interest in section 3.3. LSI is a radial parameter that distinguishes molecules with well separated first and second coordination shells from molecules surrounded by a disordered environment where the second coordination shell collapses into the first. A high LSI corresponds to the first situation, while a low LSI corresponds to a collapsed second shell.
In fig. 5.4 we show the LSI distributions in RD for bulk water and protein hydration water. The dominant peak is the one associated to disordered water and low LSI values. In bulk RD, the LSI is not able to divide into two separated peaks. The first peak decreases upon cooling but it is still dominant at all temperatures, and there is an increase of the tails upon cooling. In both bulk and shell RD, the tails that are relative to higher LSI values, increases.
In RD the distributions are unimodal. Moreover, no particular differences between bulk and shell RD appear from the LSI distributions. It was shown for bulk water that LSI divides water into two population in the IS [126, 160]. The LSI distributions for IS trajectories are shown in figure 5.5. In this case, a clearly second peak appears around 0.17 Å^2. This second peak shifts towards higher LSI values upon cooling. The second peak, at variance to the first peak, increases its intensity while lowering the temperature. We can affirm that su-

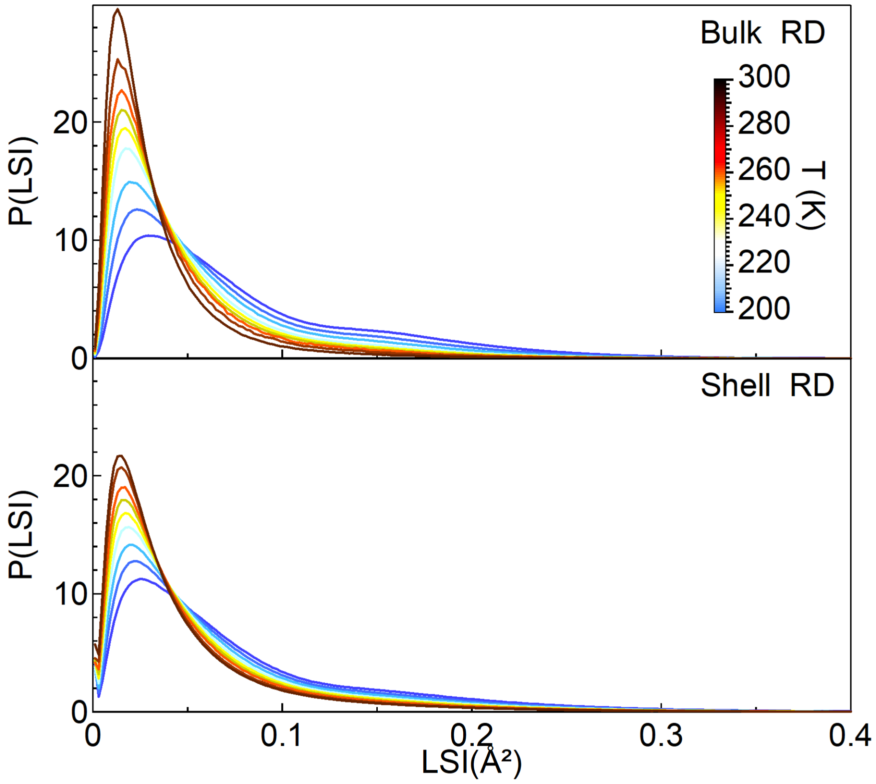

Figure 5.4: Distribution functions of the Local Structure Index (LSI) in real dynamics (RD) for bulk water (top panel) and for protein hydration water (bottom panel) for all the temperatures investigated. Colorscale shows the color coding of the curves: hightest temperature corresponds to the black line and lowest temperature corresponds to the dark blue line.

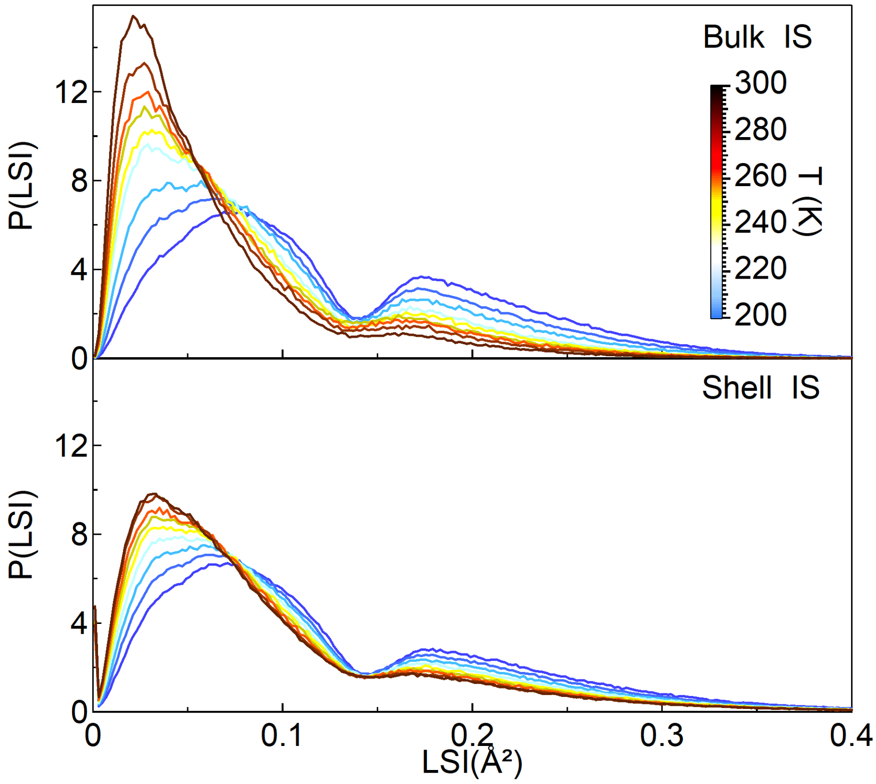

Figure 5.5: Distribution functions of the Local Structure Index (LSI) in inherent structures (IS) for bulk water (top panel) and for protein hydration water (bottom panel) for all the temperatures investigated. Colorscale shows the color coding of the curves: hightest temperature corresponds to the black line and lowest temperature corresponds to the dark blue line.

percooled water experiences a rearrangement in its network upon cooling, from a more disordered-interstitial water to a more well separated shells of first and second neighbors. The distributions have a minimum at 0.13 Å^2 that can be taken as a value to divide the populations into moreHDL (LSI<0.13 Å^2) or moreLDL (LSI>0.13 Å^2). We point out also a feature also presents in the next parameters: the shell case is less affected by the temperature than the bulk case.

5.3.2 Orientational Tetrahedral Order parameter q

The q parameter is used to quantify the degree of tetrahedrality of the local structure of a system. The average value of q varies between 0 for an ideal gas and 1 for a system with perfect tetrahedral arrangement. The q parameter is sensitive to the first neighbors angular distribution, so it is not uniquely radial as the LSI parameter or the RDFs.
In figure 5.6 we show the q probability distributions for the bulk case (top) and

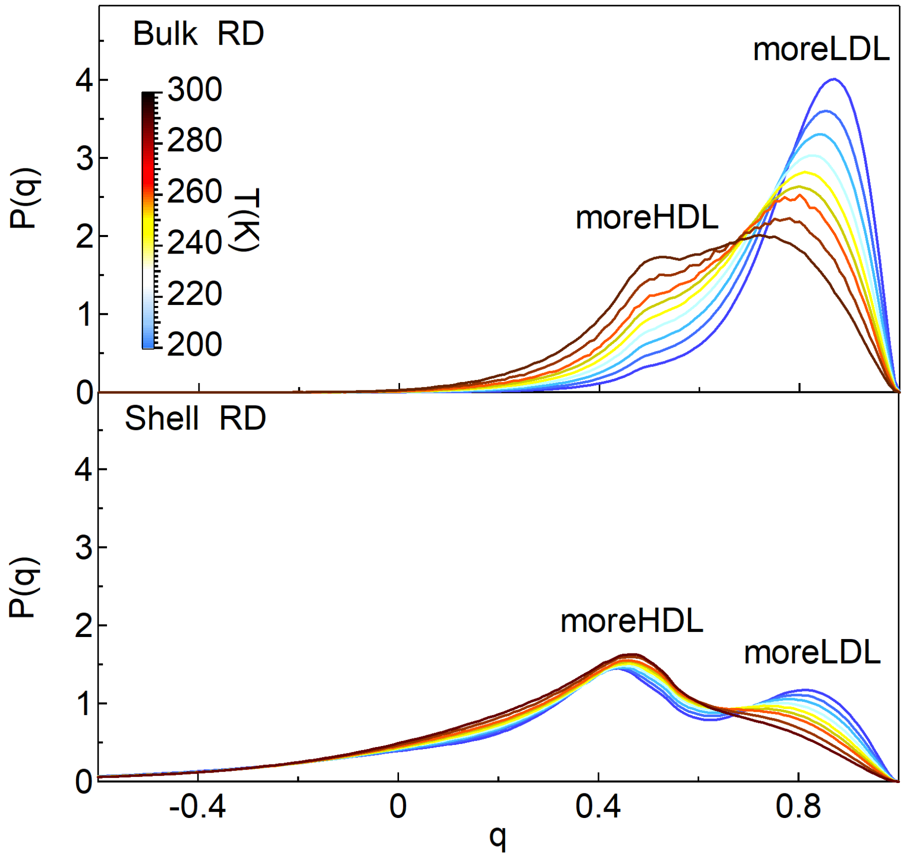

Figure 5.6: Distribution functions of the Orientational Tetrahedral Order Parameter in real dynamics (RD) for bulk water (top panel) and for protein hydration water (bottom panel) for all the temperatures investigated. Colorscale shows the color coding of the curves: the hightest temperature corresponds to the black line and the lowest temperature corresponds to the dark blue line.

the shell case (bottom) in RD. Both cases show bi-modal distributions already at ambient conditions. The right peak is related to higher q values, it is associated to water molecules in a more tetrahedral and therefore more ordered arrangement. We labelled the right peak with *moreLDL*. Instead, the left peak corresponds to lower q values, it is related to a more distorted local arrangement and we labelled it with *moreHDL*. The peaks heights give us qualitatevely information about the amount of the most probable LDL-HDL enviroments. A higher peak corresponds to an increase in the amount of water which has that most common q value.

In the bulk RD case, the moreLDL peak is higher than the moreHDL peak already at ambient conditions (black line). The moreHDL peak decreases upon cooling while the moreLDL peak increases. This process is particularly evident in the bulk case: in fact at 200 K the moreHDL peaks is almost vanished while the moreLDL peak reach the highest level.

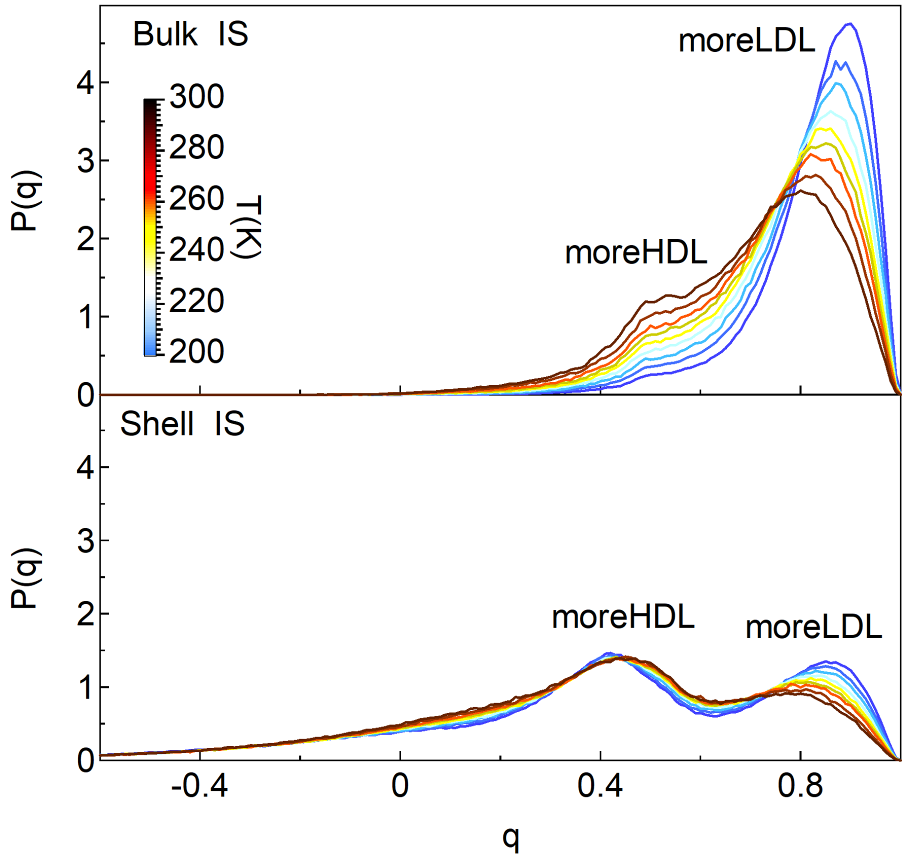

Figure 5.7: Distribution functions of the Orientational Tetrahedral Order Parameter in inherent structures (IS) for bulk water (top panel) and for protein hydration water (bottom panel) for all the temperatures investigated. Colorscale shows the color coding of the curves: the hightest temperature corresponds to the black line and the lowest temperature corresponds to the dark blue line.

In the shell RD case, the moreHDL peak is higher than the moreLDL at every temperature. Moreover, the moreHDL peak is almost stable in temperature, i.e. it is not affected by the cooling while the moreLDL peak is growing up. Protein hydration water appears to have a higher concentration of HDL-like molecules and at the same time these configurations are less affected by the cooling. The former is due to the presence of the protein that affect the tetrahedral local arrangement of hydration water. The water molecules directly bonded with the protein surface can not form a regular tetrahedral network with the other water molecules because the presence of protein atoms.

The q distributions in the IS in figure 5.7 supports these interpretation. The shape of the distributions are very similar to the RD case, with only slightly lower moreHDL peaks and slightly higher moreLDL peaks. The energy minimization, as seen with the LSI, tends to to stabilize on LDL-like structures. Nonetheless, in the shell case the effect appears much smaller as compared to the case of bulk water.

In order to quantify the differences between bulk and shell cases, we display in figure 5.8 the moreHDL and the moreLDL populations percentage. As for the LSI, we took the minimum in the q distributions in the IS as separator for the two populations. The integration of the resulting areas gives us the fraction of each population.

In the upper panel we show the RD case: at 300 K, in bulk water there are 52% moreLDL and 48% moreHDL molecules, while in shell water there are 19% moreLDL and 81% moreHDL molecules. The moreLDL water concentration increase almost linearly upon cooling both for bulk and shell, but in the shell case the rate of increase is slower. At 200 K, in the bulk case we have 90% moreLDL and 10% moreHDL while the shell case we have 32% moreLDL and 68% moreHDL. In the shell RD case, at 200 K almost 70% of water remains HDL-like, while in the bulk RD case water at 200 K is almost entirely LDL-like. The exchange between moreHDL and moreLDL populations upon cooling appears more pronounced in the shell case than in the bulk case, also in the IS. Protein hydration water is more disordered but also more stable than bulk water upon cooling.

We plot in figures 5.9 and 5.10 the peaks abscissa and the peaks heights in RD (top panels) and IS (bottom panels) respectively. We remind that the peaks coordinates of the q probability distributions describe the probability of water molecules to arrange with a particular q value. In figure 5.9, the position of the moreLDL peaks in the shell RD and bulk RD cases have similar behaviour upon cooling: at 300 K, bulk and shell are respectively 0.72 and 0.69, while at 200 K they are respectively 0.87 and 0.81. The moreLDL peaks shift towards higher q values upon cooling. The bulk RD curve is shifted up by 0.03 respect to the shell RD line. Therefore, the HDL-like bulk water decreases upon cooling, while at fixed temperatur LDL-like bulk water is slightly more tetrahedral than LDL-like hydration water. About the moreHDL peaks abscissa, at 300 K, bulk RD case and shell RD case are respectively 0.52 and 0.46 while at 200 K they are respectively 0.48 and 0.44. As for moreLDL peaks, the HDL-like bulk fraction is slighly more tetrahedral than the HDL-like shell fraction at fixed temperature. These results are confirmed in the IS panel above.

In figure 5.10 we show the peaks heights versus the temperature. The peak height is the probability of the most common q value. In the RD panel, at 300 K the bulk moreLDL peak is almost 3 times higher than the shell moreLDL peak and this ratio goes up to 4 at 200 K. MoreHDL peaks in both bulk and shell RD cases have the same height, equal to 1.65. At 200 K in the bulk RD case, the moreHDL peak's height decrease to 0.33 while in the shell RD case it falls to 1.5. The moreHDL peak height in the shell RD case is therefore constant. We point out that the peak height does not correspond with the exact amount of that type of water. The HDL-LDL fractions are calculated integrating the area under the peaks and a change in the fraction could be caused by a change in the shape of the distribution. In figure 5.8, at 300 K, the percentage of moreHDL water in the bulk RD case and in the shell RD case are respectively 48% and 81%. At 200 K, these percetages fall to, respectively, 68% and 10%. Despite the same height of the moreHDL peaks, protein hydration water has almost double HDL-like fraction than bulk water. This difference becomes even wider at 200 K, where HDL-like water fraction is almost 7 times higher in the shell RD case than the bulk RD case. Infact, the moreHDL peak in the shell RD case is wider

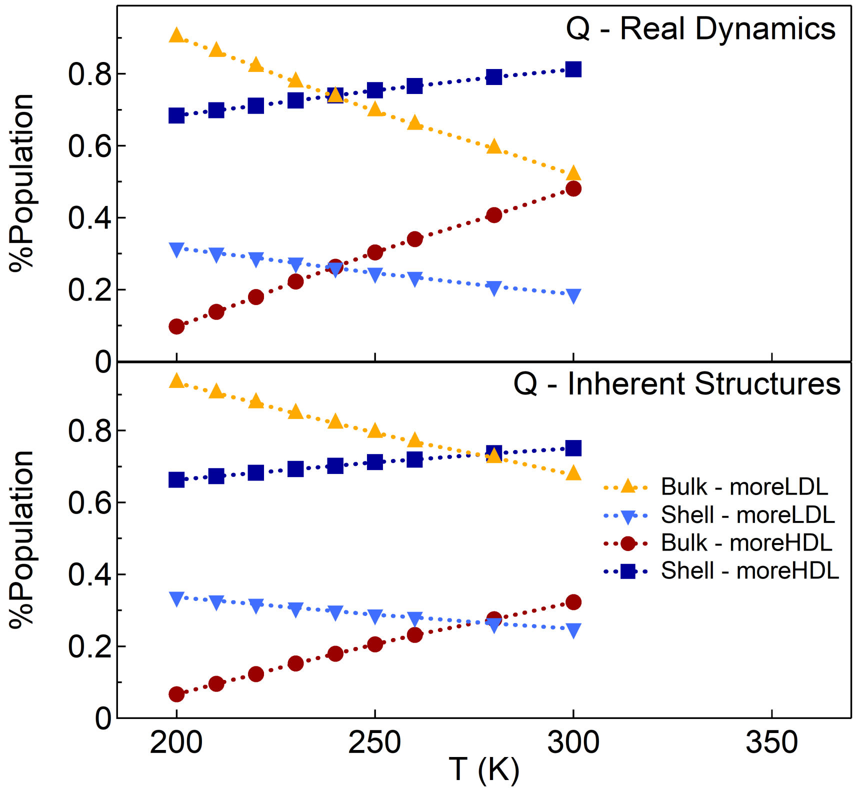

Figure 5.8: Percentage of the moreLDL and the moreHDL populations in the q distributions for bulk water (up triangles and rounds) and for protein hydration water (reverse triangles and squares) in Real Dynamics and in Inherent Structures.

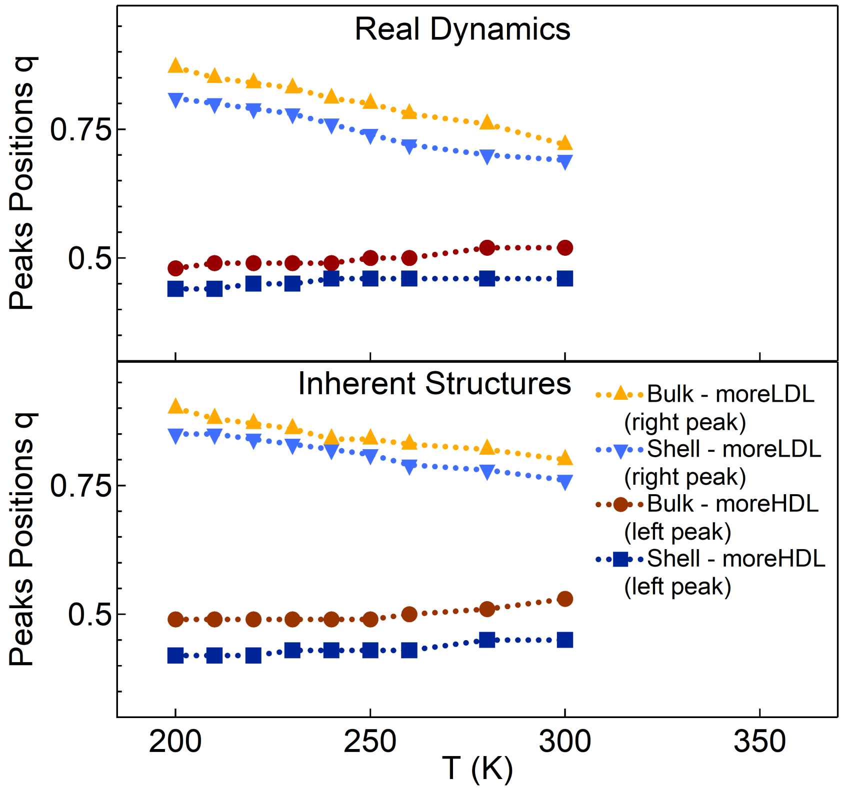

Figure 5.9: Peaks abscissa of the distributions of the orientation tetrahedral order parameter versus temperature in Real Dynamics and in Inherent Structures for both bulk water and protein hydration water.

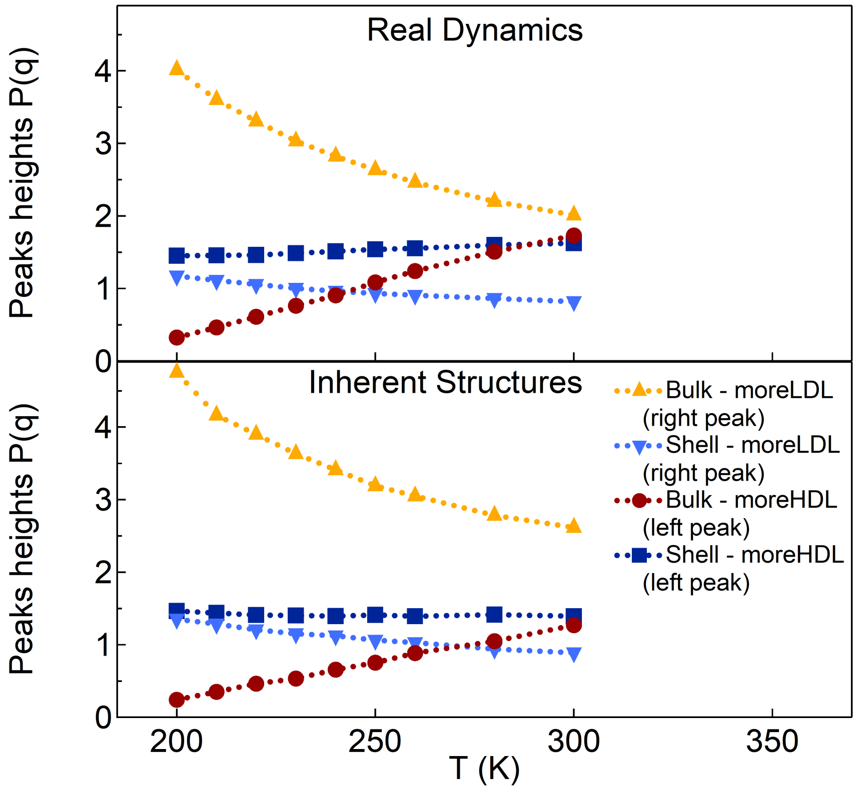

Figure 5.10: Peaks height of the distributions of the orientation tetrahedral order parameter versus temperature in Real Dynamics and in Inherent Structures for both bulk water and protein hydration water.

than in the bulk RD case, with q values which spans down to negative values. This could be explained with the presence of the water-protein interface which provide contain number of HDL-like water connected directly with the protein and less affected by the cooling as compared to 'more free' water molecules, surrounded only by other water molecules as in the bulk environment.

5.3.3 Potential Energy V_4 parameter

V4 parameter is build upon energy consideration [122]. A distortion on the tetrahedral water arrangement results on a larger V_4 value. The minimum in V_4 probability distributions in IS is at -12 kJ/mol and this value is used as a threshold which separates interstitial-distorted water which we call moreHDL (greater V_4 values) from tetrahedral-ordered water which we call moreLDL (lower V_4 values).

V_4 parameter probability distributions are bi-modal already at ambient conditions in the IS. The two peaks represent two water networks, moreHDL and moreLDL, as we already seen in the LSI and q parameter discussion. In figure 5.11 the RD is reported. In the bulk case we observe from the T behaviour, that the peak at higher V_4 values associated to moreHDL molecules vanishes upon cooling while the peak at lower V_4 values associated to moreLDL molecules increases. This indicates a transition of the HDL population to LDL upon cooling. The shell probability distribution at 300 K is similar to the bulk case, but the moreHDL peak does not vanishes upon cooling and it remains still evident at 200 K

In the IS case in figure 5.12, V_4 probability distributions are bi-modal at every temperature. In the bulk IS case, already at 300 K the moreHDL peak is lower than the moreLDL peak. The moreLDL peak becomes higher and shifts towards lower V_4 values upon cooling. In the shell IS case, at 300 K the moreHDL peak is higher than the moreLDL peak and its height is stable at every temperature. The moreLDL peak in the shell IS case behave similar to the moreLDL peak in the bulk IS case.

In figure 5.13 we show the percentage of the populations extracted from the V_4 probability distributions versus the temperature. In bulk RD, at 300 K moreLDL and moreHDL populations are almost the same percentage, 50% and 50%, while at 200 K they're respectively 83% and 17%. In shell RD, at 300 K moreLDL and moreHDL populations are respectively about 47% and 53%, at 260 K they're are the same percentage and at 200 K they're respectively 62% and 38%. In every case except shell RD, moreLDL population is higher than moreHDL population already at ambient conditions. In the bulk IS, at 300 K moreLDL and moreHDL populations are respectively 84% and 16% while at 200 K they're respectively 96% and 4%. In shell IS, at 300 K moreLDL and moreHDL populations are respectively about 59% and 41% and at 200 K they're respectively 69% and 31%. In every case, moreLDL population is higher than moreHDL population already at ambient conditions with a wider gap than in the RD. Globally, we see that the HW has higher HDL population as compare to water; this is consistent to the results of the q parameter study.

As we did for the q parameter, we show in figure 5.14 and in 5.15 respectively the position and the height of the peaks in V_4 distributions upon cooling. The observation of how peaks move upon cooling can provide a qualitatevely description of how the most probable water arrangements change upon cooling in

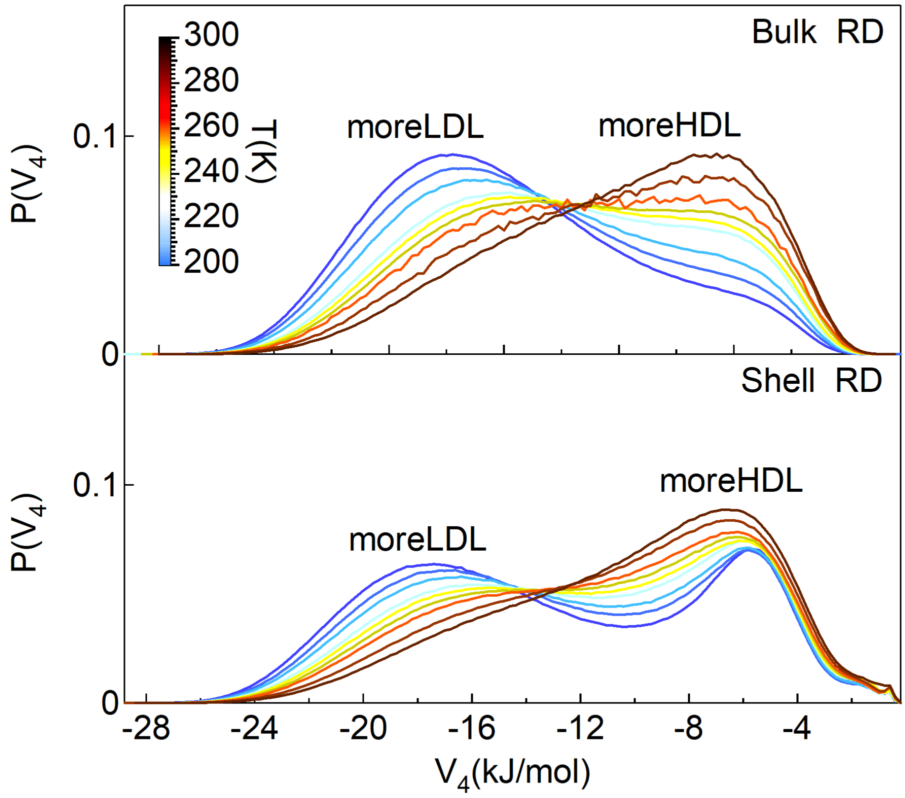

Figure 5.11: Distribution functions of the V4 structural indicator in real dynamics (RD) for bulk water (top panel) and for the hydration shell water (bottom panel) for all the temperatures investigated. Colorscale shows the color coding of the curves: the hightest temperature corresponds to the black line and the lowest temperature corresponds to the dark blue line.

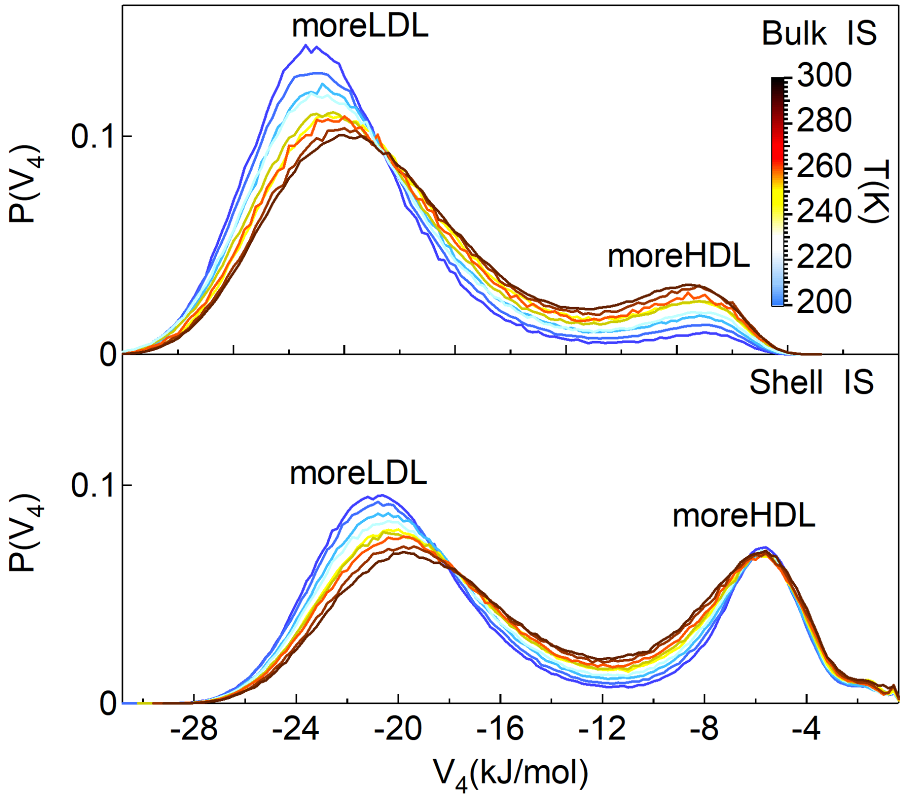

Figure 5.12: Distribution functions of the V4 structural indicator in inherent structures (IS) for bulk water (top panel) and for the hydration shell water (bottom panel). Colorscale shows the color coding of the curves: the hightest temperature corresponds to the black line and the lowest temperature corresponds to the dark blue line.

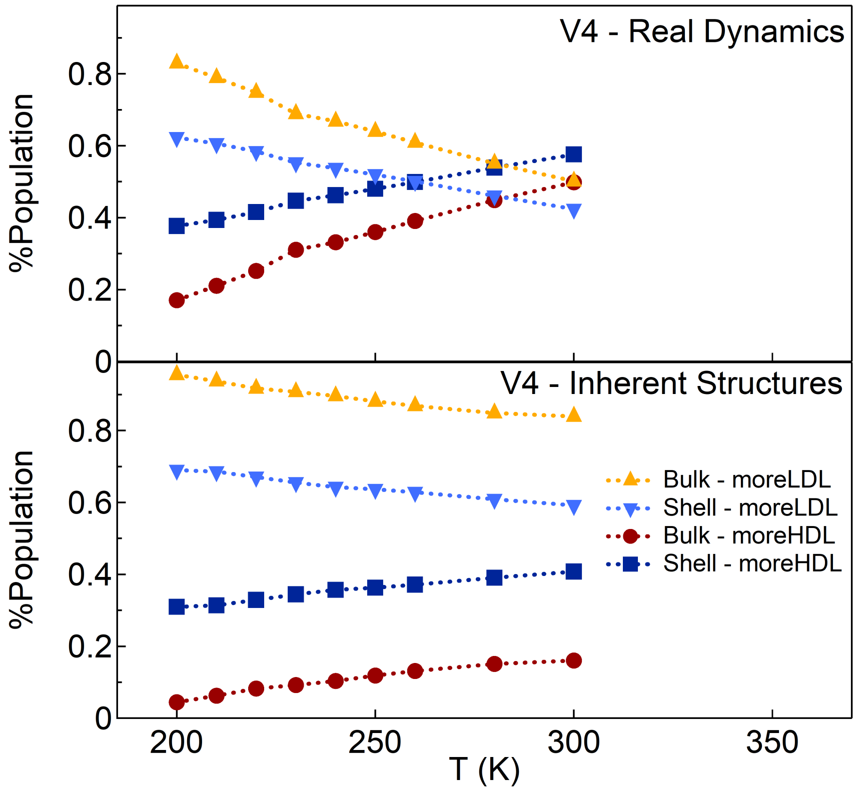

Figure 5.13: Percentage of the moreLDL and the moreHDL populations in the V4 distributions for bulk water (up triangles and rounds) and for protein hydration water (reverse triangles and squares) in Real Dynamics the Inherent Structures.

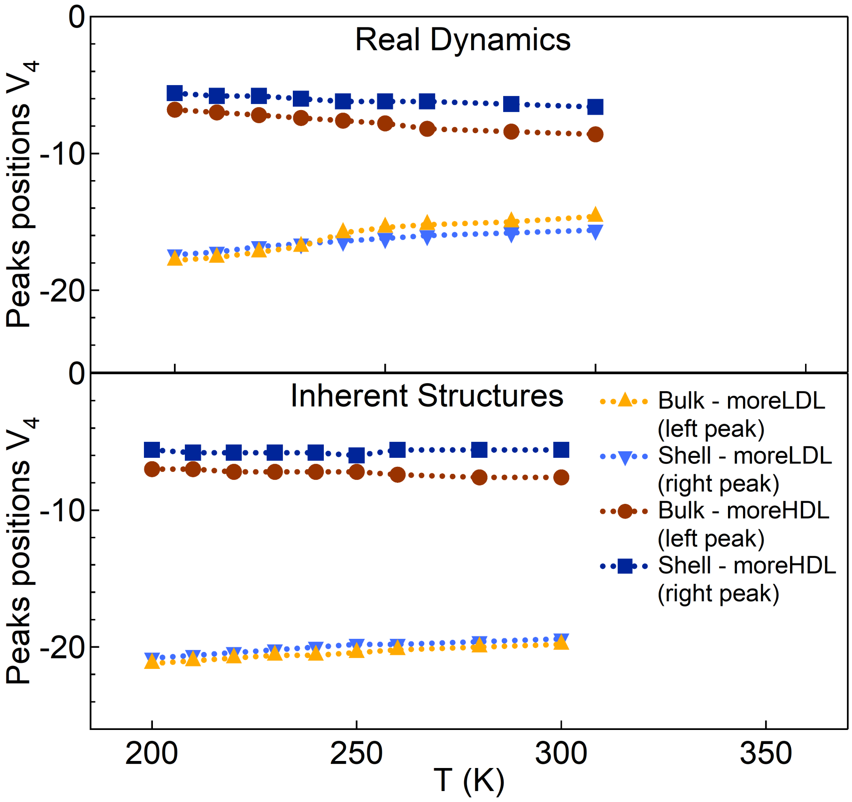

Figure 5.14: Peaks abscissa of the distributions of the orientation tetrahedral order parameter versus temperature in Real Dynamics and in Inherent Structures for both bulk water and protein hydration water.

bulk and hydration water. In figure 5.14 we plot the peaks position versus the temperature. If the V_4 peak is shifting towards higher (lower) values upon cooling it means that the corresponding water structure is becoming more distorted (tetrahedral). In the RD case, moreHDL peak for bulk and shell remain almost constant. Both moreLDL peaks in bulk and shell cases shift toward lower V_4 values almost with the same slope as T decreases. In the IS case the results are similar but the moreLDL lines are shifted down by almost 4 kJ/mol. As compared to the RD case, HDL bulk peak rearrange in a slighly distorted way upon cooling while HDL shell peaks is almost constant upon cooling. Instead, both bulk and shell LDL peaks shift towards more ordered arrangements upon cooling. In figure 5.15 we show the peaks heights versus the temperature. These lines can give us a qualitative idea of the concentration of the two most probable water networks upon cooling. In the RD case, bulk and shell moreLDL peaks grow up in the same way upon cooling, but the shell line is shifted down by a value of 0.03. Instead, moreHDL peaks decrease in a different way: they both start with a value close to 0.1 but moreHDL bulk peak fall to 0.02 while moreHDL shell peak decrase until 0.07. In the IS case, bulk and shell moreLDL peaks grow up in a way similar to the RD case but they are both shifted to

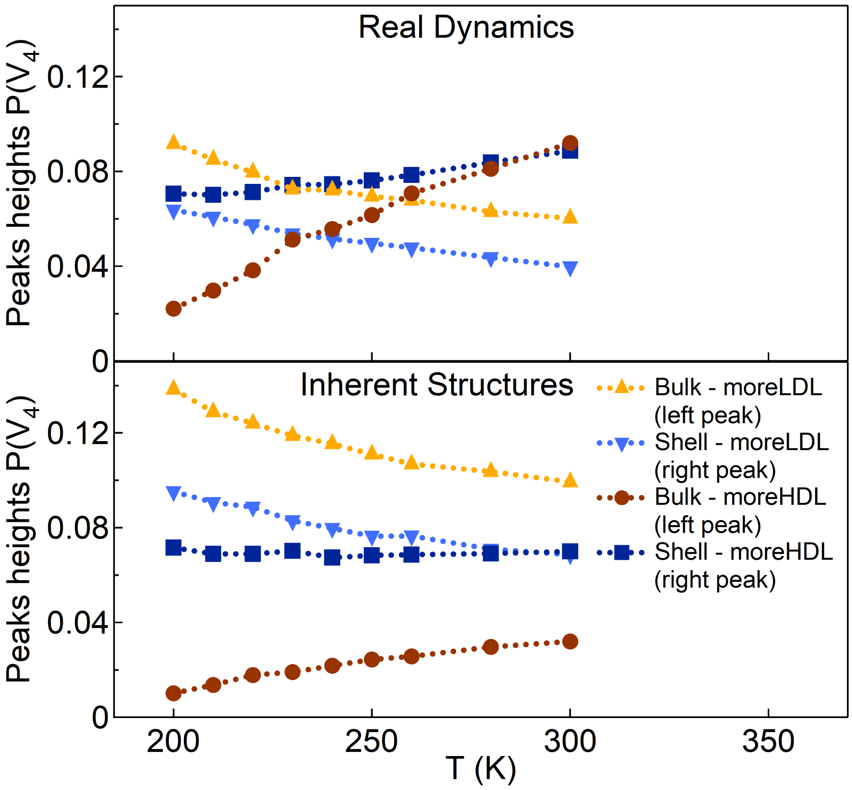

Figure 5.15: Peaks height of the distributions of the orientation tetrahedral order parameter versus temperature in Real Dynamics and in Inherent Structures for both bulk water and protein hydration water.

higher probability values: moreLDL bulk is shifted up by 0.05 while moreLDL shell is shifted up by 0.035. Instead, moreHDL bulk peak is shifted down from 0.1 to 0.03 and it goes down to 0.01 upon cooling while more HDL shell peak remains almost constant around a value of 0.075. In RD bulk water, moreHDL peak is higher than moreLDL peak at the three highest temperature while in RD hydration water, moreHDL peak is higher than moreLDL peak at every temperature. In the IS case, moreHDL shell is the only curve that is not affected by the cooling.

5.3.4 Radial Distribution Functions

The RDF, indicated by $g(r)$, corresponds to the probability of finding a pair of particle at distance r, relative to the probability of a completely random density distribution. Concerning of the study of the local structure of protein hydration water, the RDFs were the first observables we calculated to osbserve the internal structure of hydration water and comparing it with the RDFs of bulk water. Firstly, we show in figures 5.16 and 5.17 the RDFs of bulk water (left panels) and shell (right panels) calculated both in Real Dynamics (RD) and Inher-

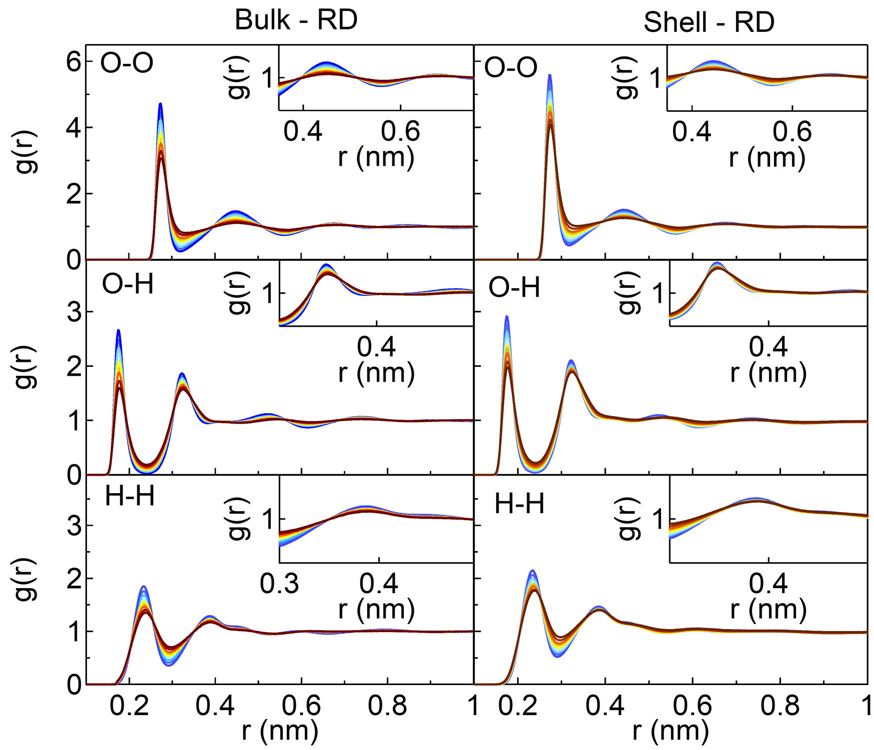

Figure 5.16: RDFs of bulk water (left panels) and hydration water labelled *shell*(right panels) in RD of oxygen-oxygen (O-O), oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H). Temperature spans from 300 K (black curve) to 200 K (dark blue).

ent Structures (IS). The RDFs are the oxygen-oxygen (O-O), oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H). As can be seen, the RDFs are similar between bulk and shell, both in RD and IS. IS distributions have sharper peaks and deeper minima with respect RD as we seen also for the others structure parameters. The RDFs calculated on the total water molecules of out systems are not enough to appreciate differences between bulk and hydration water.
To further investigate the inner structures of the two water groups, we calculated the RDFs on the populations extracted by the stucture parameters. In the figures 5.18, 5.19, 5.20 and 5.21 are shown the oxygen-oxygen RDFs of hydration water calculated on the HDL, LDL and total water molecules, as extracted respectively from the tetrahedral order parameter q and the V_4 parameter both for RD and IS. As for the totality of the water molecules, in these cases no particular differences are detectable between HDL and LDL populations extracted by the same parameter. The populations separation operated by both the q and the V_4 parameter produce two groups of molecules with no structural differences by the RDFs point of view. From the comparison between RD and IS, we record that the second coordination shell in IS shows two peaks at 0.35 and 0.43 nm while in RD the peak at 0.35 nm is hidden by thermal fluctuations which spread interstitial water in the space between the first and the second

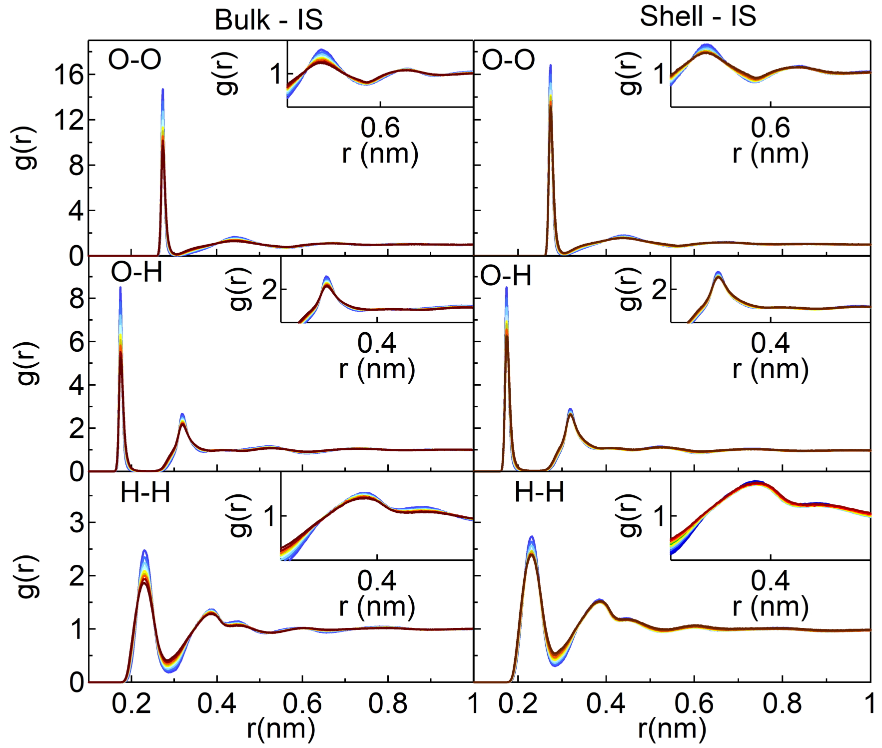

Figure 5.17: RDFs of bulk water (left panels) and hydration water labelled *shell*(right panels) in IS of oxygen-oxygen (O-O), oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H). Temperature spans from 300 K (black curve) to 200 K (dark blue).

coordination shells.

A different results is obtained with the RDFs calculated on the population extracted from the LSI. In figures 5.22 and 5.23 are shown the oxygen-oxygen RDFs of hydration water calculated on the HDL, LDL and total water molecules, as extracted from the LSI. In this case we can appreciate differences in the structure of HDL and LDL: the first shows the presence of interstitial water before the second peak at 0.43 nm, with an interstitial peak which appears at 200 K at 0.38 nm; the latter shows a wide unique second peak preceded by a deep minimum. The peculiarity of RDFs calculated on populations extracted from the LSI lies on the radial nature of the LSI. Infact, the LSI separates water moecules clusters with well separated first and second coordination shell from the others. Therefore, among our structure parameters, the LSI is the only one which select structures only in the radial direction. Due to its radial nature, the RDF is able to distinguish different structures only in the populations extracted by the LSI.

5.4 Summary and Conclusions

In this work, we studied by molecular dynamic simulations the structure of hydration water of a protein (lysozyme). The HW was defined as the water inside a shell of thickness of 6 Å around the protein. The tools we used to investigate the structure of protein hydration water were the probability distributions of three structure parameters calculated at different temperatures ranging from T=300 K upon supercooling down to T=200 K. The parameters were the Local Structure Index, the orientational tetrahedral order parameter q and the V_4 parameter, (recently proposed by ref. [122]) and the RDFs.

For every Real Dynamics (RD) simulation we calculated the Inherent Structures (IS) which suppress thermal fluctuations and allowed us a better recognition of the interstitial-disordered water molecules. Most of the probability distributions have a bi-modal shape with two peaks corresponding to two water species similar to the two liquid mixture found in the supercooled regime. A more tetrahedral-ordered liquid (LDL) similar to the low density amorphous ice and a more distorted-disordered liquid (HDL) similar to the high density amorphous ice. The minimum in the IS probability distributions was set to divide HDL and LDL areas and we calculated HDL-like and LDL-like water fractions integrating these areas.

The Local Structure Index reveals differences between water network with well separated first and second coordination shell and network richer of interstitial water. The probability distribution obtained with the LSI in RD are unimodal and very similar between bulk and hydration water. Probably due to the intrinsic nature of the LSI, the system of the irregular shell 6 Å thick around the protein does not lend itself to be investigated through this radial parameter. Infact, RDFs have the same limitation. In the IS a clear second peak arises at higher LSI values. This permits to distinguish two distinct water groups. We notice that this second peak grows less in protein hydration water than in bulk water upon cooling, as well as the first peak's decrease is less pronounced in protein hydration water than bulk water upon cooling. We record that protein hydration water is less affected by temperature variations than bulk water.

The orientational tetrahedral order parameters describes the geometry of the

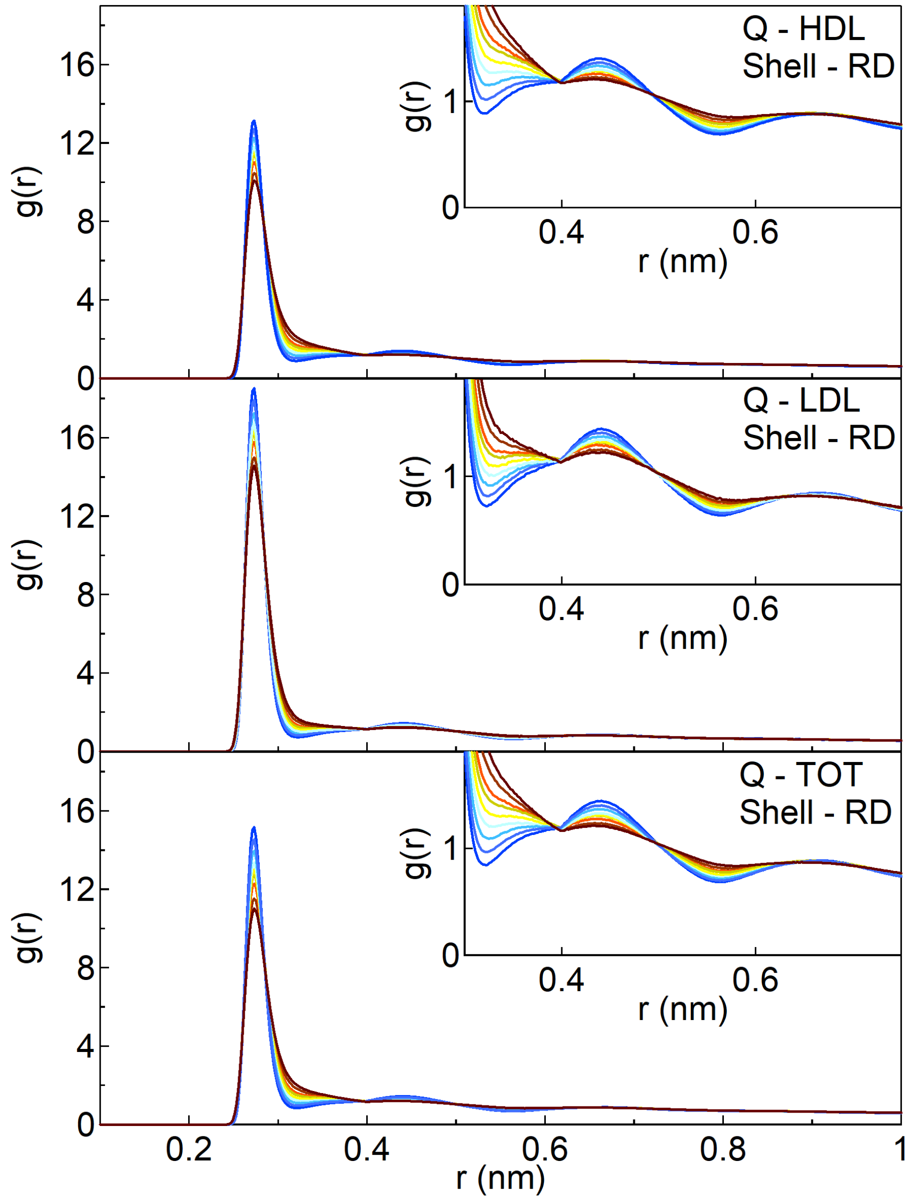

Figure 5.18: Oxygen-oxygen RDFs of protein hydration water for HDL (top panel) LDL (middle panel) and total (bottom panel) populations extracted from the q parameter in Real Dynamics. Temperature spans from 300 K (black curve) to 200 K (dark blue).

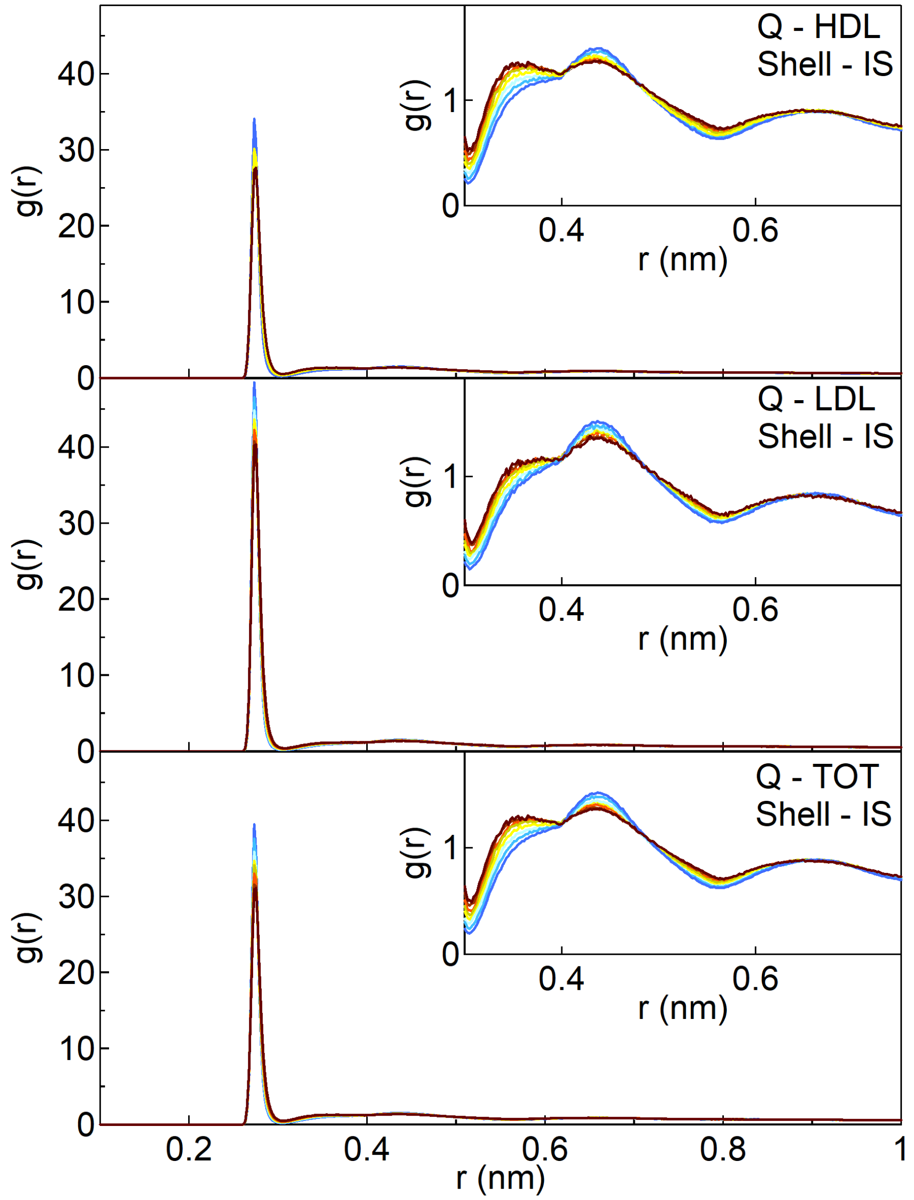

Figure 5.19: Oxygen-oxygen RDFs of protein hydration water for HDL (top panel) LDL (middle panel) and total (bottom panel) populations extracted from the q parameter in Inherent Structures. Temperature spans from 300 K (black curve) to 200 K (dark blue).

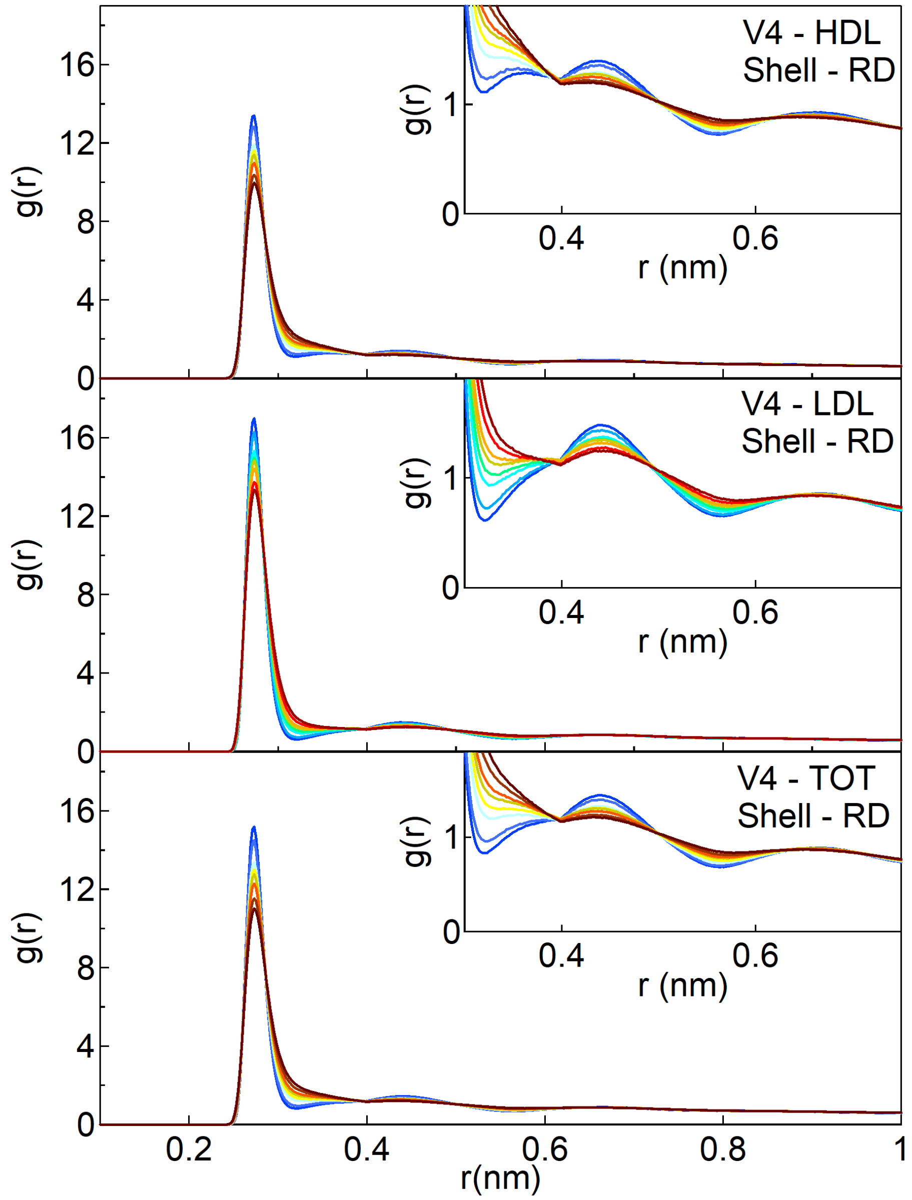

Figure 5.20: Oxygen-oxygen RDFs of protein hydration water for HDL (top panel) LDL (middle panel) and total (bottom panel) populations extracted from the V_4 parameter in Real Dynamics. Temperature spans from 300 K (black curve) to 200 K (dark blue).

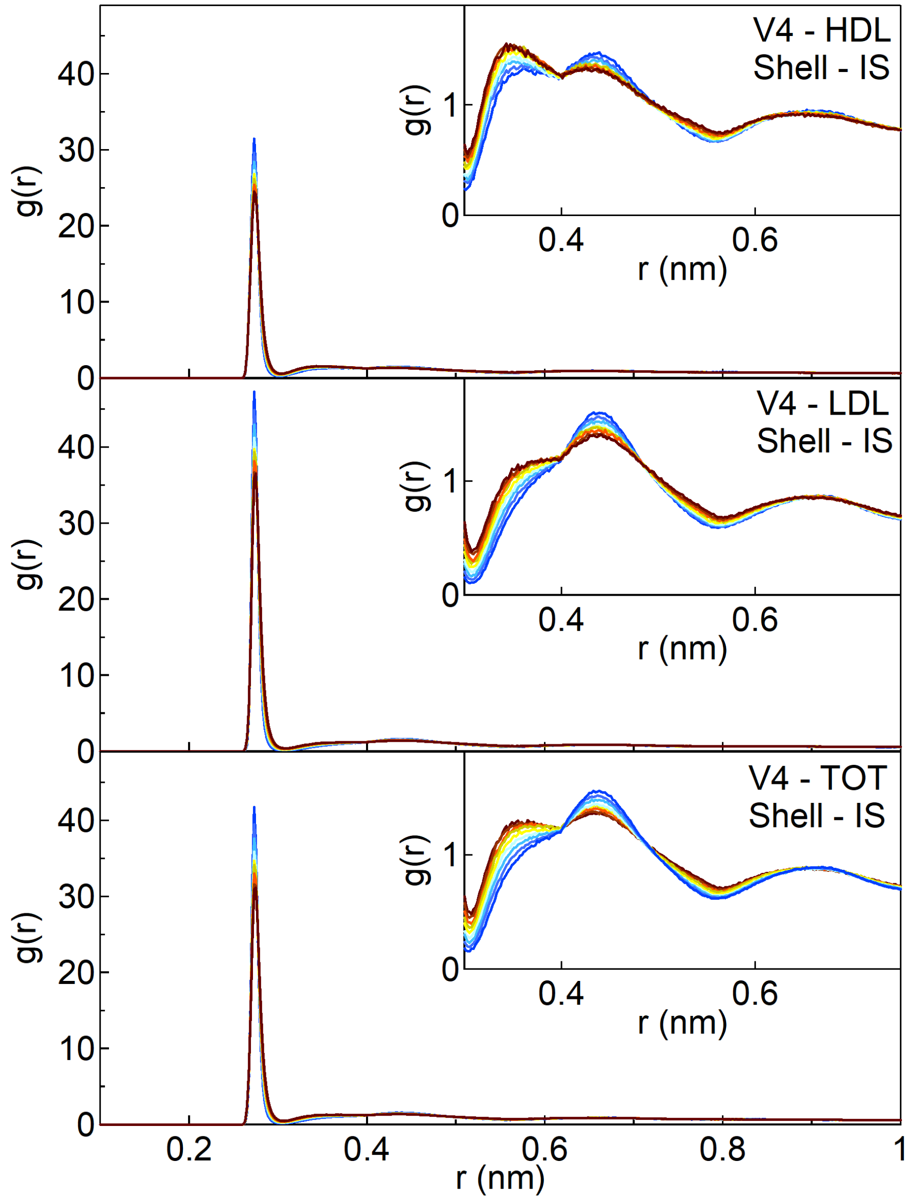

Figure 5.21: Oxygen-oxygen RDFs of protein hydration water for HDL (top panel) LDL (middle panel) and total (bottom panel) populations extracted from the V_4 parameter in Inherent Structures. Temperature spans from 300 K (black curve) to 200 K (dark blue).

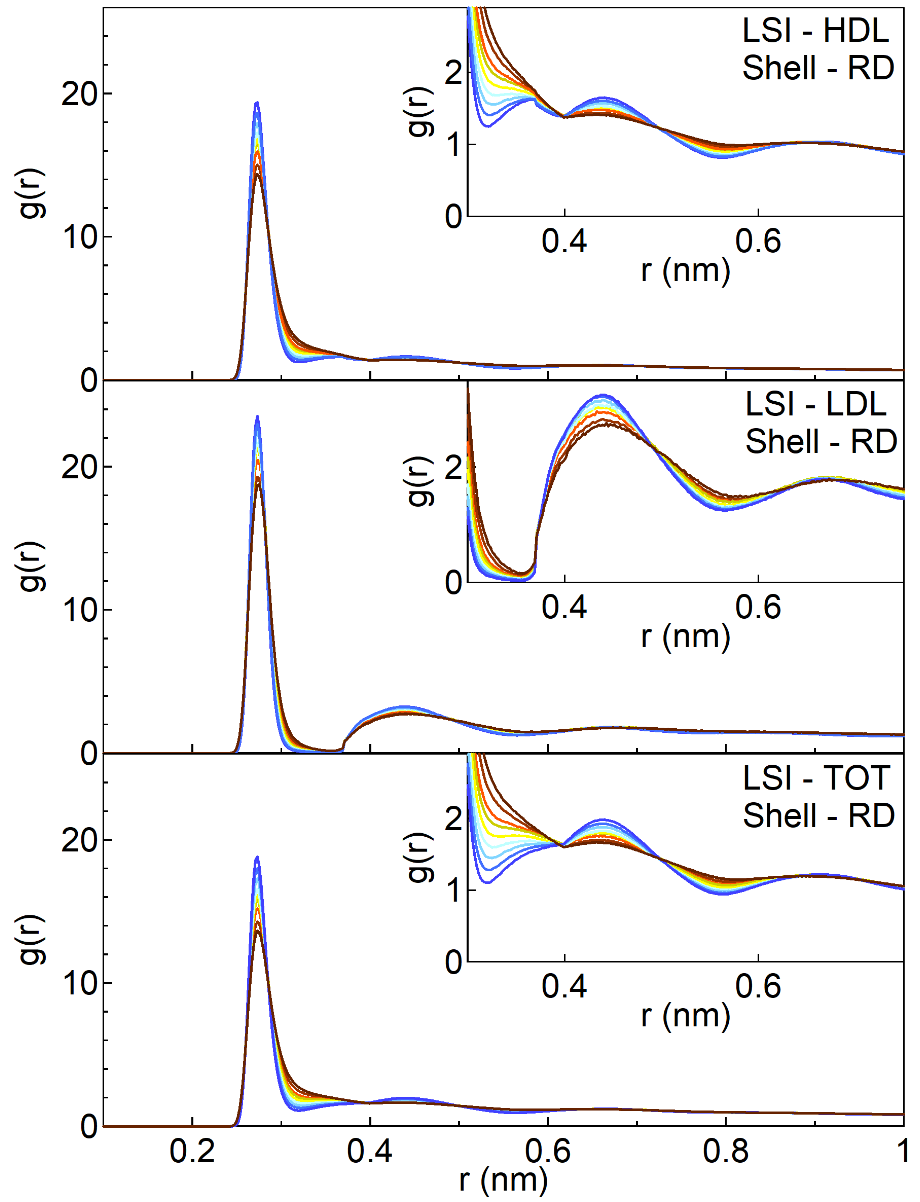

Figure 5.22: xygen-oxygen RDFs of protein hydration water for HDL (top panel) LDL (middle panel) and total (bottom panel) populations extracted from the LSI in Real Dynamics. Temperature spans from 300 K (black curve) to 200 K (dark blue).

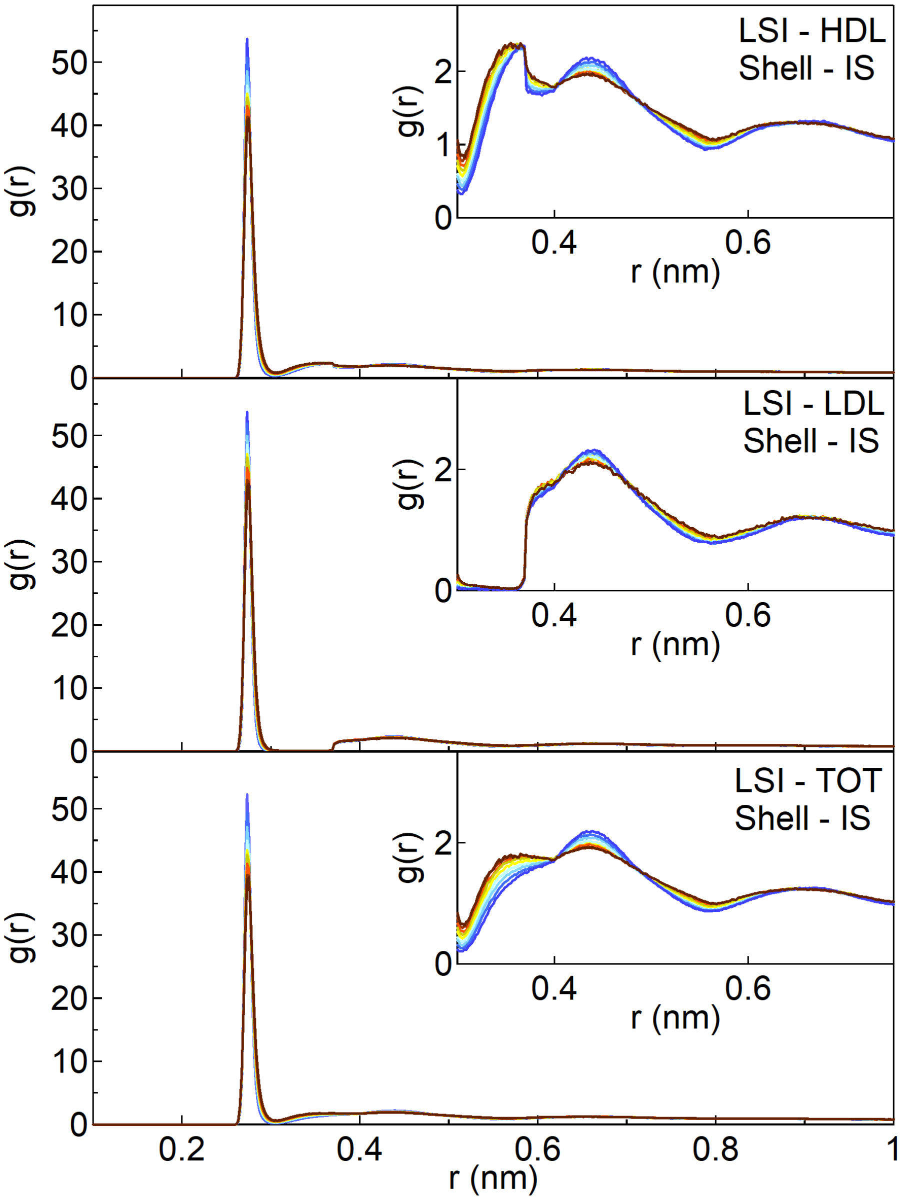

Figure 5.23: Oxygen-oxygen RDFs of protein hydration water for HDL (top panel) LDL (middle panel) and total (bottom panel) populations extracted from the LSI in Inherent Structures. Temperature spans from 300 K (black curve) to 200 K (dark blue).

Table 5.1: Percentage of HDL and LDL populations extracted from q and V_4 distributions at the highest and the lowest temperatures in RD.

	HDL		**LDL**	
Par - T(K)	Bulk RD	Shell RD	Bulk RD	Shell RD
q - 300 K	48%	81%	52%	19%
q - 200 K	19%	32%	81%	68%
V_4 - 300 K	50%	53%	50%	47%
V_4 - 200 K	17%	38%	83%	62%

first four neighbors of a water molecule taking into account angular orientation. Other parameters can, instead, deal with interaction energies between the central molecules and its neighbors [122]. In particular, V_4 is the fourth strongest energy term, which should be on the order of the HBs energy if the molecules is in tetrahedral structure or higher otherwise. The q parameter is build upon geometrical considerations and describe directly the degree of tetrahedrality of water elementary cell, while V_4 parameter is build upon potential energy consideration and describe the distortion of a typical hydrogen bond, so indirectly the changes in the tetrahedrality of a water molecule environtemt.

We found that protein hydration water has a different composition with respect to bulk water in terms of HDL/LDL populations and also behave different upon cooling. The water fractions extracted by q and V_4 parameters are summarized in the table 5.4, for the highest and the lowest temperatures investigated here. From the q probability distributions we found that protein hydration water is made for more than 80% of HDL. This percentage decrease upon cooling but slower than in bulk water. As a result, protein hydration water is made by almost 70% of HDL-like at 200 K against the 10% of HDL-like in bulk water. From V_4 probability ditributions we found that protein hydration water is made for more than 53% of HDL. As we found for q parameter, this fraction decreases upon cooling down to 38% at 200 K while in bulk water HDL fraction starts from 50% and fall down to 17% at 200 K. When we cool down the system, water molecules rearrange in a lower-energy more-tetrahedral network, i.e. they switch from HDL-like to LDL-like structures. Protein hydration water is less affected by this switching upon cooling than bulk water. In other words, we can say that protein hydration water is more stable upon cooling than bulk water, probably due to the presence of the protein-water interface.

The q and V_4 parameters have given different results because their different nature and definition. They also responded differently in the IS. The q parameter probability distributions in the IS well reproduced the results in RD.

We notice how the percentage of HDL-like water in the shell case is much higher when extracted by the q parameter than the V_4 parameter. This is due to the presence of the protein-water interface combined with the definition of the two parameters. The q parameter describe the tetrahedrality of the overall water elementary cell, therefore the water bonded with the protein interface results to be highly distorted showing low or negative q values, increasing abruptly the percentage of HDL-like water. Instead, V_4 is related to a pair of water molecules sorting the first neighbors based on the potential value of their bond and taking the fourth with the largest value. Therefore, not all the molecules close to the protein surface enter in the compute of HDL-like fraction as for the q parameter. We analyzed also the heights and the position of the LDL/HDL peaks in every distributions. The moreHDL peaks decrease their height upon cooling while moreLDL peaks, on the contrary, increase their heights. These changes in height not correspond directly in a change in the amount of the corresponding water type, but the analysis of the peaks height can give us a qualitatevely the idea of how the most common q or (V_4) value and its corresponding structure is changing upon cooling. From the q peaks analysis we found that moreHDL peak is almost constant in height while moreLDL peak decreases. Instead, in bulk water, moreHDL peak decrease in height while moreLDL increases non simmetricaly with respect to the HDL curve. Morever, moreHDL peak position (q value) decreases upon cooling while moreLDL peak abscissa increases. Bulk water peaks positions reproduce the same trend at higher q values both for HDL-like and LDL-like waters. Similar results can be obtained by V_4 distributions. Despite the HDL fraction decreases upon cooling, the corresponding peak is almost constant. The structure of protein hydration water is mostly HDL-like near the protein surface and this structure becomes even more distorted upon cooling, while the water molecules farer from the protein surface should be already LDL-like and they are free to rearrange like bulk water.

Chapter 6

General conclusions

In this thesis we analyzed supercooled hydration water of lysozyme both from a dynamical and a structural point of view. We used MD simulations to extract long trajectories (up to 200 ns) of the system composed by the lysozyme protein and more than thirteen thousands water molecules in a temperature range from 300 K down to 200 K. We focused on the hydration water (HW), a shell made by the water molecules within 6 Å from the protein surface. The observables of interest were calculated from the trajectories using mainly custom programs. The translational dynamics have been previously investigated in the reciprocal space by means of the Self Intermediate Scattering Functions (SISFs) [61, 130]. In this work we probed the dynamics in the real space by means of the Self Van Hove Functions (SVHFs), for the first time down to the supercooled regime [158]; this gives complementary informations to the SISFs, and the Mean Square Displacements (MSDs). The local structure of protein hydration water was studied by calculating three stucture parameters [161]: the Local Structure Index (LSI), the tetrahedral orientational structure parameter (q) and the recently proposed V_4 parameter [122]. Moreover, we calculated the Radial Distribution Functions of the hydration water divided into two populations defined by the parameters. All the simulations and the calculations of the obsevables were performed exploiting the multicore architecture of the machines in the Roma Tre High Performance Computing cluster.

We found many differences in the dynamical behaviour and in the local structure between protein HW and bulk water. The presence of the protein-water interface has a crucial role in affecting the dynamics and the structural composition of the shell of HW, as well as the Protein Dynamical Transition (PDT), a crossover temperature found at 240 K [61] under which the protein abruptly loses its motion capacity.
From a dynamical point of view, the motion of the water molecules around the protein reproduce the same dynamical regimes of supercooled bulk water, such as the ballistic, the MCT-cage and α structural relaxation regimes, including low T MCT hopping phenomena, but more importantly it also shows new features absent in bulk water. The translational dynamics of HW in the real space through the SVHFs was studied in a wide range of correlation times (from tens of femtoseconds to tens of nanoseconds). The new feature consists in another cage regime followed by new protein related hopping phenomena at high temper-

atures. The MCT hopping has a typical lenght scale of 2-3 neighbors distances (0.5-0.7 nm) and it only appears at very low T and long times (200 K and 50 ns correlation time). The new cage regime has a different nature with respect to the MCT one and it is observed at high temperatures with a cage radius ten times larger than the first neighbors cage. The new hopping is characterized by longer lenghts and it appears at high T and for long correlation times (240 K and correlation time >5 ns). Our physical explanation relates these phenomena to the protein motion: above the PDT at high temperatures, the protein has complete flexibility. The protein surface is able to strecth, relax, rotate etc. and the water molecules directly bonded with the protein surface are constrained to move following the protein, which has a dynamics timescale slower than the water dynamics. This is the new cage regime. In other words, we can say that the dynamics of the first layer of HW depends on the protein movements. This fact is reflected in multimodal SVHFs. The many shoulders on the distributions can be described as hopping phenomena related to the clusters of water molecules constrained in different locations on the protein.

From a structural point of view our aim was to analyze the composition of hydration water based on the LDL-HDL dichotomy. The Low Density Liquid (LDL) and the High Density Liquid (HDL) are two types of water local structure based on the polymorphism of glassy water that has two main glass states: the Low Density Amourphous and the High Density Amorphous. The LDL corresponds to a high regular tetrahedral water structure, composed by five water molecules. The HDL corresponds to a more packed water network, i.e. with interstitial water molecules that lead to a less tetrahedral and less ordered structure than the LDL liquid.

To characterize the local structure of HW we compared the calculated parameters with the ones calculated for the bulk case and the inherent structures (IS) that help to find underlying structures hidden by thermal fluctuations.

The results obtained by the q and the V_4 parameters show bimodal probability distributions, i.e. in each curve we can identify a LDL peak and a HDL peak. In the bulk case, water has a majority of HDL at high temperatures which quicly decrease upon cooling in favour of LDL. Instead, hydration water at high temperatures shows equal percentanges of LDL and HDL for V_4 parameter, while a major percentage of HDL (81 %) in the q parameter distributions. The HDL fraction decreases upon cooling less than in the bulk case. Hydration water network results to be less tetrahedral than the bulk one, with a percentage of disordered water less affected by the cooling than the bulk case. In general, inherent structures distributions have sharper peaks and a major concentration of LDL as compareed to RD. The main differences are: *(i)* the q parameter indicates and higher concentration of HDL (about 30 %) with respect to the V_4 parameter and *(ii)* the V_4 parameter is more sentive to the calculations of IS as compared to q.

The fact that different parameters detect different fractions of population with different T behaviour, is due to the fact that they essentialy probe different 'ways' of being HDL/LDL. q measures the tetrahedrality of the elementary water cell and the water molecules directly bonded with the protein lead to detect more HDL geometries. V_4 measures the fourth strongest term of the potential energy for each water molecule and it will be less sentive to other distortion of the water shell bonded to the protein.

Regarding the LSI, this parameter does not detect two populations in RD of hydration water, but only in the IS as it happens in the bulk case. Nonetheless,it is the only parameter among the ones here investigated for which LDL and HDL RDFs show significative differences in the local density in the region between the first and the second coordination shell.

We can say that the protein presence affects both the dynamics and the structure of water molecules in the hydration shell, but water features are still present. Further investigations are needed to understand the hydration water-protein mutual interaction and how the protein affects in details the structure of the water enviroments. Unveiling these mechanisms is fundamental for further developments in the field of biosolutions.

Appendix A

List of Activities

List of Publications

- L. Tenuzzo, G. Camisasca & P. Gallo, *Protein-Water and Water-Water Long-Time Relaxations in Protein Hydration Water upon Cooling—A Close Look through Density Correlation Functions.* Molecules, 25(19), 4570 (2020).
- G. Camisasca, A. Iorio, L. Tenuzzo & P. Gallo, *Chapter 2 - Slow Dynamics of Biological Water* included in the book by Bulavin, L., & Lebovka, N. (2021). *Soft Matter Systems for Biomedical Applications.*, Springer (2021).
- G. Camisasca, L. Tenuzzo, & P. Gallo, *Protein hydration water - focus on low density and high density local structures upon cooling* In preparation (2022).

Attended school and conferences

- Roma Tre Congress on *Water under Extreme Conditions 2019*, Roma, June 2019. Member of the Local Organizing Committee
- Congress *Frontiers in Water Biophysics*, Erice, July 2019.
- School *CCP5 Methods in Molecular Simulation Summer School*, Durham (UK), July 2019.

Bibliography

[1] Paola Gallo et al. “Water: A tale of two liquids”. In: *Chemical reviews* 116.13 (2016), pp. 7463–7500.

[2] Paola Gallo and Mauro Rovere. *Physics of Liquid Matter*. 2021.

[3] Charles A Angell. “Supercooled water”. In: *Annual Review of Physical Chemistry* 34.1 (1983), pp. 593–630.

[4] Maud Formanek and Fausto Martelli. “Probing the network topology in network-forming materials: The case of water”. In: *AIP Advances* 10.5 (2020), p. 055205.

[5] Peter H Poole et al. “Effect of hydrogen bonds on the thermodynamic behavior of liquid water”. In: *Physical review letters* 73.12 (1994), p. 1632.

[6] Thomas M Truskett et al. “A single-bond approach to orientation-dependent interactions and its implications for liquid water”. In: *The Journal of chemical physics* 111.6 (1999), pp. 2647–2656.

[7] P Gallo et al. “Slow dynamics of water molecules in supercooled states”. In: *Physical review letters* 76.15 (1996), p. 2730.

[8] F Sciortino et al. “Supercooled water and the kinetic glass transition”. In: *Physical Review E* 54.6 (1996), p. 6331.

[9] Attila G Csàszàr et al. “On equilibrium structures of the water molecule”. In: *The Journal of chemical physics* 122.21 (2005), p. 214305.

[10] David Eisenberg and Walter Kauzmann. *The structure and properties of water*. OUP Oxford, 2005.

[11] JB Hasted. “Liquid water: Dielectric properties”. In: *The Physics and Physical Chemistry of Water*. Springer, 1972, pp. 255–309.

[12] P Gallo, D Corradini and M Rovere. “Widom line and dynamical crossovers as routes to understand supercritical water”. In: *Nature communications* 5.1 (2014), pp. 1–6.

[13] Limei Xu et al. “Relation between the Widom line and the dynamic crossover in systems with a liquid–liquid phase transition”. In: *Proceedings of the National Academy of Sciences* 102.46 (2005), pp. 16558–16562.

[14] Giancarlo Franzese and H Eugene Stanley. “The Widom line of supercooled water”. In: *Journal of Physics: Condensed Matter* 19.20 (2007), p. 205126.

[15] GG Simeoni et al. “The Widom line as the crossover between liquid-like and gas-like behaviour in supercritical fluids”. In: *Nature Physics* 6.7 (2010), pp. 503–507.

[16] Cmglee. *Phase diagram of water as a log-lin chart with pressure from 1 Pa to 1 TPa and temperature from 0 K to 650 K.* [Online; controllata il 30-novembre-2021]. 2008. URL: `https://it.wikipedia.org/wiki/File:Phase_diagram_of_water.svg#/media/File:Phase_diagram_of_water.svg`.

[17] Leonardo del Rosso et al. "Refined structure of metastable ice XVII from neutron diffraction measurements". In: *The Journal of Physical Chemistry C* 120.47 (2016), pp. 26955–26959.

[18] John Russo, Flavio Romano and Hajime Tanaka. "New metastable form of ice and its role in the homogeneous crystallization of water". In: *Nature materials* 13.7 (2014), pp. 733–739.

[19] CA Tulk et al. "Structural studies of several distinct metastable forms of amorphous ice". In: *Science* 297.5585 (2002), pp. 1320–1323.

[20] Thomas Loerting et al. "A second distinct structural "state" of high-density amorphous ice at 77 K and 1 bar". In: *Physical Chemistry Chemical Physics* 3.24 (2001), pp. 5355–5357.

[21] Philip Ball. "Water—an enduring mystery". In: *Nature* 452.7185 (2008), pp. 291–292.

[22] PG Debenedetti and Metastable Liquids. *Concepts and principles.* 1996.

[23] Pablo G Debenedetti. "Supercooled and glassy water". In: *Journal of Physics: Condensed Matter* 15.45 (2003), R1669.

[24] RJ Speedy and CA Angell. "Isothermal compressibility of supercooled water and evidence for a thermodynamic singularity at- 45 C". In: *The Journal of Chemical Physics* 65.3 (1976), pp. 851–858.

[25] CA Angell, J Shuppert and JC Tucker. "Anomalous properties of supercooled water. Heat capacity, expansivity, and proton magnetic resonance chemical shift from 0 to-38%". In: *The Journal of Physical Chemistry* 77.26 (1973), pp. 3092–3099.

[26] CA Angell and H Kanno. "Density maxima in high-pressure supercooled water and liquid silicon dioxide". In: *Science* 193.4258 (1976), pp. 1121–1122.

[27] Francesco Mallamace et al. "The anomalous behavior of the density of water in the range 30 K¡ T¡ 373 K". In: *Proceedings of the National Academy of Sciences* 104.47 (2007), pp. 18387–18391.

[28] Dazhi Liu et al. "Observation of the density minimum in deeply supercooled confined water". In: *Proceedings of the National Academy of Sciences* 104.23 (2007), pp. 9570–9574.

[29] Jeffrey R Errington and Pablo G Debenedetti. "Relationship between structural order and the anomalies of liquid water". In: *Nature* 409.6818 (2001), pp. 318–321.

[30] Zhenyu Yan et al. "Structure of the first-and second-neighbor shells of simulated water: Quantitative relation to translational and orientational order". In: *Physical Review E* 76.5 (2007), p. 051201.

[31] Anders Nilsson and Lars GM Pettersson. "The structural origin of anomalous properties of liquid water". In: *Nature communications* 6.1 (2015), pp. 1–11.

[32] Andrea Cavagna. "Supercooled liquids for pedestrians". In: *Physics Reports* 476.4-6 (2009), pp. 51–124.

[33] GP Johari, Andreas Hallbrucker and Erwin Mayer. "The glass–liquid transition of hyperquenched water". In: *Nature* 330.6148 (1987), pp. 552–553.

[34] Andreas Hallbrucker, Erwin Mayer and GP Johari. "Glass-liquid transition and the enthalpy of devitrification of annealed vapor-deposited amorphous solid water: a comparison with hyperquenched glassy water". In: *The Journal of Physical Chemistry* 93.12 (1989), pp. 4986–4990.

[35] V Velikov, S Borick and CA Angell. "The glass transition of water, based on hyperquenching experiments". In: *Science* 294.5550 (2001), pp. 2335–2338.

[36] CA Angell. "Liquid fragility and the glass transition in water and aqueous solutions". In: *Chemical reviews* 102.8 (2002), pp. 2627–2650.

[37] O Mishima, LD Calvert and E Whalley. "An apparently first-order transition between two amorphous phases of ice induced by pressure". In: *Nature* 314.6006 (1985), pp. 76–78.

[38] Paul F McMillan et al. "Polyamorphism and liquid–liquid phase transitions: challenges for experiment and theory". In: *Journal of Physics: Condensed Matter* 19.41 (2007), p. 415101.

[39] Osamu Mishima and Yoshiharu Suzuki. "Propagation of the polyamorphic transition of ice and the liquid–liquid critical point". In: *Nature* 419.6907 (2002), pp. 599–603.

[40] Peter H Poole et al. "Phase behaviour of metastable water". In: *Nature* 360.6402 (1992), pp. 324–328.

[41] Srikanth Sastry et al. "Singularity-free interpretation of the thermodynamics of supercooled water". In: *Physical Review E* 53.6 (1996), p. 6144.

[42] Kyung Hwan Kim et al. "Maxima in the thermodynamic response and correlation functions of deeply supercooled water". In: *Science* 358.6370 (2017), pp. 1589–1593.

[43] Sander Woutersen et al. "A liquid-liquid transition in supercooled aqueous solution related to the HDA-LDA transition". In: *Science* 359.6380 (2018), pp. 1127–1131.

[44] Osamu Mishima and H Eugene Stanley. "The relationship between liquid, supercooled and glassy water". In: *Nature* 396.6709 (1998), pp. 329–335.

[45] Paola Gallo and H Eugene Stanley. "Supercooled water reveals its secrets". In: *Science* 358.6370 (2017), pp. 1543–1544.

[46] Paola Gallo et al. "Advances in the study of supercooled water". In: *The European Physical Journal E* 44.11 (2021), pp. 1–36.

[47] Paola Gallo, Thomas Loerting and Francesco Sciortino. *Supercooled water: A polymorphic liquid with a cornucopia of behaviors.* 2019.

[48] A Scala et al. "Free energy surface of supercooled water". In: *Physical Review E* 62.6 (2000), p. 8016.

[49] Jeremy C Palmer et al. "Advances in computational studies of the liquid–liquid transition in water and water-like models". In: *Chemical reviews* 118.18 (2018), pp. 9129–9151.

[50] Jeremy C Palmer et al. "Metastable liquid–liquid transition in a molecular model of water". In: *Nature* 510.7505 (2014), pp. 385–388.

[51] Peter H Poole et al. "Free energy surface of ST2 water near the liquid-liquid phase transition". In: *The Journal of chemical physics* 138.3 (2013), p. 034505.

[52] Yang Liu, Athanassios Z Panagiotopoulos and Pablo G Debenedetti. "Low-temperature fluid-phase behavior of ST2 water". In: *The Journal of Chemical Physics* 131.10 (2009), p. 104508.

[53] Felix Franks. *Water: a matrix of life.* Vol. 21. Royal Society of Chemistry, 2000.

[54] Fausto Martelli et al. "Structural properties of water confined by phospholipid membranes". In: *Frontiers of Physics* 13.1 (2018), pp. 1–8.

[55] MAF Dos Santos et al. "Dynamical aspects of supercooled TIP3P–water in the grooves of DNA". In: *The Journal of chemical physics* 150.23 (2019), p. 235101.

[56] Sotiris Samatas et al. "Water between membranes: Structure and Dynamics". In: *Biomembrane Simulations.* CRC Press, 2019, pp. 69–88.

[57] Fausto Martelli, Carles Calero and Giancarlo Franzese. "Redefining the concept of hydration water near soft interfaces". In: *Biointerphases* 16.2 (2021), p. 020801.

[58] M Paolantoni et al. "Light scattering spectra of water in trehalose aqueous solutions: Evidence for two different solvent relaxation processes". In: *The Journal of Physical Chemistry B* 113.22 (2009), pp. 7874–7878.

[59] Lucia Comez et al. "More is different: experimental results on the effect of biomolecules on the dynamics of hydration water". In: *The journal of physical chemistry letters* 4.7 (2013), pp. 1188–1192.

[60] A Magno and P Gallo. "Understanding the mechanisms of bioprotection: a comparative study of aqueous solutions of trehalose and maltose upon supercooling". In: *The Journal of Physical Chemistry Letters* 2.9 (2011), pp. 977–982.

[61] G Camisasca et al. "Two structural relaxations in protein hydration water and their dynamic crossovers". In: *The Journal of chemical physics* 145.4 (2016), p. 044503.

[62] Dagmar Ringe and Gregory A Petsko. "The 'glass transition'in protein dynamics: what it is, why it occurs, and how to exploit it". In: *Biophysical chemistry* 105.2-3 (2003), pp. 667–680.

[63] Bjarne F Rasmussen et al. "Crystalline ribonuclease A loses function below the dynamical transition at 220 K". In: *nature* 357.6377 (1992), pp. 423–424.

[64] Andrea G Markelz et al. "Protein dynamical transition in terahertz dielectric response". In: *Chemical Physics Letters* 442.4-6 (2007), pp. 413–417.

[65] S-H Chen et al. "Observation of fragile-to-strong dynamic crossover in protein hydration water". In: *Proceedings of the National Academy of Sciences* 103.24 (2006), pp. 9012–9016.

[66] Sow-Hsin Chen et al. "Dynamics of a globular protein and its hydration water studied by neutron scattering and MD simulations". In: *Spectroscopy* 24.1-2 (2010), pp. 1–24.

[67] Giorgio Schirò, Margarita Fomina and Antonio Cupane. "Communication: Protein dynamical transition vs. liquid-liquid phase transition in protein hydration water". In: *The Journal of chemical physics* 139.12 (2013), p. 121102.

[68] Kathleen Wood et al. "Coincidence of dynamical transitions in a soluble protein and its hydration water: direct measurements by neutron scattering and MD simulations". In: *Journal of the American Chemical Society* 130.14 (2008), pp. 4586–4587.

[69] Giorgio Schirò et al. "Translational diffusion of hydration water correlates with functional motions in folded and intrinsically disordered proteins". In: *Nature communications* 6.1 (2015), pp. 1–8.

[70] Marco Lagi et al. "The low-temperature dynamic crossover phenomenon in protein hydration water: simulations vs experiments". In: *The Journal of Physical Chemistry B* 112.6 (2008), pp. 1571–1575.

[71] Pradeep Kumar et al. "Glass transition in biomolecules and the liquid-liquid critical point of water". In: *Physical review letters* 97.17 (2006), p. 177802.

[72] Gaia Camisasca, Margherita De Marzio and Paola Gallo. "Effect of trehalose on protein cryoprotection: Insights into the mechanism of slowing down of hydration water". In: *The Journal of Chemical Physics* 153.22 (2020), p. 224503.

[73] Silvia Corezzi et al. "Protein Hydration in a Bioprotecting Mixture". In: *Life* 11.10 (2021), p. 995.

[74] WJJ Van den Tweel, A Harder and RM Buitelaar. *Stability and stabilization of enzymes: proceedings of an international symposium held in Maastricht, The Netherlands, 22-25 November 1992*. Elsevier, 2013.

[75] Jeffrey L Cleland and Robert Langer. "Formulation and delivery of proteins and peptides: design and development strategies". In: ACS Publications, 1994.

[76] Lorenzo Cordone et al. "Proteins in saccharides matrices and the trehalose peculiarity: biochemical and biophysical properties". In: *Current Organic Chemistry* 19.17 (2015), pp. 1684–1706.

[77] Sergio Giuffrida, Lorenzo Cordone and Grazia Cottone. "Bioprotection can be tuned with a proper protein/saccharide ratio: the case of solid amorphous matrices". In: *The Journal of Physical Chemistry B* 122.37 (2018), pp. 8642–8653.

[78] Lorenzo Cordone et al. "Internal dynamics and protein–matrix coupling in trehalose-coated proteins". In: *Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics* 1749.2 (2005), pp. 252–281.

[79] A Iorio et al. “Slow dynamics of supercooled hydration water in contact with lysozyme: examining the cage effect at different length scales”. In: *Philosophical Magazine* 100.20 (2020), pp. 2582–2595.

[80] Antonio Iorio, Gaia Camisasca and Paola Gallo. “Slow dynamics of hydration water and the trehalose dynamical transition”. In: *Journal of Molecular Liquids* 282 (2019), pp. 617–625.

[81] Kazuhito Kajiwara and Felix Franks. “Crystalline and amorphous phases in the binary systemwater–raffinose”. In: *Journal of the Chemical Society, Faraday Transactions* 93.9 (1997), pp. 1779–1783.

[82] S Di Fonzo et al. “Concentration- Temperature Dependencies of Structural Relaxation Time in Trehalose- Water Solutions by Brillouin Inelastic UV Scattering”. In: *The Journal of Physical Chemistry A* 111.49 (2007), pp. 12577–12583.

[83] Laura Lupi et al. “Dynamics of biological water: Insights from molecular modeling of light scattering in aqueous trehalose solutions”. In: *The Journal of Physical Chemistry B* 116.25 (2012), pp. 7499–7508.

[84] Sara Emanuela Pagnotta, Angel Alegrìa and Juan Colmenero. “Dynamical behavior of highly concentrated trehalose water solutions: A dielectric spectroscopy study”. In: *Physical Chemistry Chemical Physics* 14.9 (2012), pp. 2991–2996.

[85] C Branca et al. “Anomalous cryoprotective effectiveness of trehalose: Raman scattering evidences”. In: *The Journal of chemical physics* 111.1 (1999), pp. 281–287.

[86] Adrien Lerbret et al. “How homogeneous are the trehalose, maltose, and sucrose water solutions? An insight from molecular dynamics simulations”. In: *The Journal of Physical Chemistry B* 109.21 (2005), pp. 11046–11057.

[87] Sau Lawrence Lee, Pablo G Debenedetti and Jeffrey R Errington. “A computational study of hydration, solution structure, and dynamics in dilute carbohydrate solutions”. In: *The Journal of chemical physics* 122.20 (2005), p. 204511.

[88] Adrien Lerbret et al. “Influence of homologous disaccharides on the hydrogen-bond network of water: complementary Raman scattering experiments and molecular dynamics simulations”. In: *Carbohydrate Research* 340.5 (2005), pp. 881–887.

[89] Jason Malsam and Alptekin Aksan. “Hydrogen bonding and kinetic/thermodynamic transitions of aqueous trehalose solutions at cryogenic temperatures”. In: *The Journal of Physical Chemistry B* 113.19 (2009), pp. 6792–6799.

[90] Silvia Di Fonzo et al. “Water dynamics and structural relaxation in concentrated sugar solutions”. In: *Food Biophysics* 8.3 (2013), pp. 183–191.

[91] Robert Zwanzig. “Memory effects in irreversible thermodynamics”. In: *Physical Review* 124.4 (1961), p. 983.

[92] Hazime Mori. “Transport, collective motion, and Brownian motion”. In: *Progress of theoretical physics* 33.3 (1965), pp. 423–455.

[93] Wolfgang Gotze. *Complex dynamics of glass-forming liquids: A mode-coupling theory.* Vol. 143. OUP Oxford, 2008.

[94] David R Reichman and Patrick Charbonneau. “Mode-coupling theory”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2005.05 (2005), P05013.

[95] Liesbeth Janssen. “Mode-coupling theory of the glass transition: A primer”. In: *Frontiers in Physics* 6 (2018), p. 97.

[96] W Gotze and L Sjogren. “Scaling properties in supercooled liquids near the glass transition”. In: *Journal of Physics C: Solid State Physics* 21.18 (1988), p. 3407.

[97] JP Hansen and IR McDonald. *Theory of simple liquids third edition.* 2006.

[98] Ulf Bengtzelius, W Gotze and A Sjolander. “Dynamics of supercooled liquids and the glass transition”. In: *Journal of Physics C: solid state Physics* 17.33 (1984), p. 5915.

[99] Frank H Stillinger and Pablo G Debenedetti. “Glass transition thermodynamics and kinetics”. In: *Annu. Rev. Condens. Matter Phys.* 4.1 (2013), pp. 263–285.

[100] Walter Kob and Hans C Andersen. “Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture I: The van Hove correlation function”. In: *Physical Review E* 51.5 (1995), p. 4626.

[101] Walter Kob and Hans C Andersen. “Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture. II. Intermediate scattering function and dynamic susceptibility”. In: *Physical Review E* 52.4 (1995), p. 4134.

[102] P Gallo and M Rovere. “Mode coupling and fragile to strong transition in supercooled TIP4P water”. In: *The Journal of chemical physics* 137.16 (2012), p. 164503.

[103] CA Angell. “Relaxation in liquids, polymers and plastic crystals—strong/fragile patterns and problems”. In: *Journal of Non-Crystalline Solids* 131 (1991), pp. 13–31.

[104] C Austen Angell. “Formation of glasses from liquids and biopolymers”. In: *Science* 267.5206 (1995), pp. 1924–1935.

[105] Francis W Starr, Johannes K Nielsen and H Eugene Stanley. “Hydrogen-bond dynamics for the extended simple point-charge model of water”. In: *Physical Review E* 62.1 (2000), p. 579.

[106] Francis W Starr, Francesco Sciortino and H Eugene Stanley. “Dynamics of simulated water under pressure”. In: *Physical Review E* 60.6 (1999), p. 6757.

[107] P Gallo, M Rovere and S-H Chen. “Dynamic crossover in supercooled confined water: Understanding bulk properties through confinement”. In: *The Journal of Physical Chemistry Letters* 1.4 (2010), pp. 729–733.

[108] P Gallo, D Corradini and M Rovere. “Fragile to strong crossover at the Widom line in supercooled aqueous solutions of NaCl”. In: *The Journal of chemical physics* 139.20 (2013), p. 204503.

[109] M De Marzio et al. “Mode coupling theory and fragile to strong transition in supercooled TIP4P/2005 water”. In: *The Journal of chemical physics* 144.7 (2016), p. 074503.

[110] Li Liu et al. "Pressure dependence of fragile-to-strong transition and a possible second critical point in supercooled confined water". In: *Physical review letters* 95.11 (2005), p. 117802.

[111] Berni J Alder and Thomas Everett Wainwright. "Studies in molecular dynamics. I. General method". In: *The Journal of Chemical Physics* 31.2 (1959), pp. 459–466.

[112] Aneesur Rahman. "Correlations in the motion of atoms in liquid argon". In: *Physical review* 136.2A (1964), A405.

[113] MP Allen and DJ Tildesley. "Computer simulation of liquids: Oxford university press.[Google Scholar]". In: (1989).

[114] Herman JC Berendsen et al. "Molecular dynamics with coupling to an external bath". In: *The Journal of chemical physics* 81.8 (1984), pp. 3684–3690.

[115] HJC Berendsen, JR Grigera and TP Straatsma. "The missing term in effective pair potentials". In: *Journal of Physical Chemistry* 91.24 (1987), pp. 6269–6271.

[116] Alex D MacKerell Jr et al. "All-atom empirical potential for molecular modeling and dynamics studies of proteins". In: *The journal of physical chemistry B* 102.18 (1998), pp. 3586–3616.

[117] Alexander D Mackerell Jr, Michael Feig and Charles L Brooks III. "Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations". In: *Journal of computational chemistry* 25.11 (2004), pp. 1400–1415.

[118] Nicolas Foloppe and Alexander D MacKerell Jr. "All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data". In: *Journal of computational chemistry* 21.2 (2000), pp. 86–104.

[119] Scott E Feller and Alexander D MacKerell. "An improved empirical potential energy function for molecular simulations of phospholipids". In: *The Journal of Physical Chemistry B* 104.31 (2000), pp. 7510–7515.

[120] Olgun Guvench et al. "Additive empirical force field for hexopyranose monosaccharides". In: *Journal of computational chemistry* 29.15 (2008), pp. 2543–2564.

[121] Olgun Guvench et al. "CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses". In: *Journal of chemical theory and computation* 5.9 (2009), pp. 2353–2370.

[122] Joan M Montes de Oca, Francesco Sciortino and Gustavo A Appignanesi. "A structural indicator for water built upon potential energy considerations". In: *The Journal of Chemical Physics* 152.24 (2020), p. 244503.

[123] Gaia Camisasca et al. "Structure and slow dynamics of protein hydration water". In: *Journal of Molecular Liquids* 268 (2018), pp. 903–910.

[124] Kurt Binder, Jörg Baschnagel and Wolfgang Paul. "Glass transition of polymer melts: test of theoretical concepts by computer simulation". In: *Progress in Polymer Science* 28.1 (2003), pp. 115–172.

[125] Ralf Metzler and Joseph Klafter. “The random walk’s guide to anomalous diffusion: a fractional dynamics approach”. In: *Physics reports* 339.1 (2000), pp. 1–77.

[126] Biswajit Santra et al. “Local structure analysis in ab initio liquid water”. In: *Molecular Physics* 113.17-18 (2015), pp. 2829–2841.

[127] Berk Hess et al. “GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation”. In: *Journal of chemical theory and computation* 4.3 (2008), pp. 435–447.

[128] Claudia Rocchi, Anna Rita Bizzarri and Salvatore Cannistraro. “Water dynamical anomalies evidenced by molecular-dynamics simulations at the solvent-protein interface”. In: *Physical Review E* 57.3 (1998), p. 3315.

[129] Anna Rita Bizzarri and Salvatore Cannistraro. “Molecular dynamics of water at the protein- solvent interface”. In: *The Journal of Physical Chemistry B* 106.26 (2002), pp. 6617–6633.

[130] Dario Corradini et al. “Microscopic mechanism of protein cryopreservation in an aqueous solution with trehalose”. In: *Scientific reports* 3.1 (2013), pp. 1–10.

[131] S Perticaroli et al. “Hydration and aggregation of lysozyme by extended frequency range depolarized light scattering”. In: *Journal of Non-Crystalline Solids* 407 (2015), pp. 472–477.

[132] L Comez et al. “Molecular properties of aqueous solutions: a focus on the collective dynamics of hydration water”. In: *Soft Matter* 12.25 (2016), pp. 5501–5514.

[133] Silvina Cerveny and Jan Swenson. “Water dynamics in the hydration shells of biological and non-biological polymers”. In: *The Journal of chemical physics* 150.23 (2019), p. 234904.

[134] Samir Kumar Pal, Jorge Peon and Ahmed H Zewail. “Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution”. In: *Proceedings of the National Academy of Sciences* 99.4 (2002), pp. 1763–1768.

[135] Luyuan Zhang et al. “Mapping hydration dynamics around a protein surface”. In: *Proceedings of the National Academy of Sciences* 104.47 (2007), pp. 18461–18466.

[136] Rene Costard, Ismael A Heisler and Thomas Elsaesser. “Structural dynamics of hydrated phospholipid surfaces probed by ultrafast 2D spectroscopy of phosphate vibrations”. In: *The journal of physical chemistry letters* 5.3 (2014), pp. 506–511.

[137] Torsten Siebert et al. “Range, magnitude, and ultrafast dynamics of electric fields at the hydrated DNA surface”. In: *The journal of physical chemistry letters* 7.16 (2016), pp. 3131–3136.

[138] Gottfried Otting, Edvards Liepinsh and Kurt Wuthrich. “Protein hydration in aqueous solution”. In: *Science* 254.5034 (1991), pp. 974–980.

[139] Sheh-Yi Sheu et al. “Surface topography effects of globular biomolecules on hydration water”. In: *The Journal of Physical Chemistry B* 123.32 (2019), pp. 6917–6932.

[140] W Doster, S Cusack and W Petry. “Dynamic instability of liquidlike motions in a globular protein observed by inelastic neutron scattering”. In: *Physical review letters* 65.8 (1990), p. 1080.

[141] Francesco Mallamace et al. “Role of the solvent in the dynamical transitions of proteins: the case of the lysozyme-water system”. In: *The Journal of chemical physics* 127.4 (2007), 07B617.

[142] Gaia Camisasca et al. “Slow dynamics and structure of supercooled water in confinement”. In: *Entropy* 19.4 (2017), p. 185.

[143] Wolfgang Gotze. *Complex dynamics of glass-forming liquids: A mode-coupling theory.* Vol. 143. Oxford University Press on Demand, 2009.

[144] Dietmar Paschek and Alfons Geiger. “Simulation study on the diffusive motion in deeply supercooled water”. In: *The Journal of Physical Chemistry B* 103.20 (1999), pp. 4139–4146.

[145] M De Marzio et al. “Microscopic origin of the fragile to strong crossover in supercooled water: The role of activated processes”. In: *The Journal of chemical physics* 146.8 (2017), p. 084502.

[146] Gaia Camisasca et al. “Translational and rotational dynamics of high and low density TIP4P/2005 water”. In: *The Journal of chemical physics* 150.22 (2019), p. 224507.

[147] Yuya Shinohara et al. “Viscosity and real-space molecular motion of water: Observation with inelastic x-ray scattering”. In: *Physical Review E* 98.2 (2018), p. 022604.

[148] Srikanth Sastry, Pablo G Debenedetti and Frank H Stillinger. “Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid”. In: *Nature* 393.6685 (1998), pp. 554–557.

[149] P Gallo, R Pellarin and M Rovere. “Slow dynamics of a confined supercooled binary mixture: Direct space analysis”. In: *Physical Review E* 67.4 (2003), p. 041202.

[150] Paola Gallo, Andrea Attili and Mauro Rovere. “Mode-coupling behavior of a Lennard-Jones binary mixture upon increasing confinement”. In: *Physical Review E* 80.6 (2009), p. 061502.

[151] CJ Roberts and Pablo G Debenedetti. “Structure and dynamics in concentrated, amorphous carbohydrate- water systems by molecular dynamics simulation”. In: *The Journal of Physical Chemistry B* 103.34 (1999), pp. 7308–7318.

[152] KT Wikfeldt, Anders Nilsson and Lars GM Pettersson. “Spatially inhomogeneous bimodal inherent structure of simulated liquid water”. In: *Physical Chemistry Chemical Physics* 13.44 (2011), pp. 19918–19924.

[153] Takashi Tokushima et al. “High resolution X-ray emission spectroscopy of liquid water: The observation of two structural motifs”. In: *Chemical Physics Letters* 460.4-6 (2008), pp. 387–400.

[154] Anders Nilsson and Lars GM Pettersson. “Perspective on the structure of liquid water”. In: *Chemical Physics* 389.1-3 (2011), pp. 1–34.

[155] A Taschin et al. “Evidence of two distinct local structures of water from ambient to supercooled conditions”. In: *Nature communications* 4.1 (2013), pp. 1–8.

[156] GE Walrafen. “Raman spectral studies of the effects of temperature on water structure”. In: *The Journal of Chemical Physics* 47.1 (1967), pp. 114–126.

[157] Josef N Stern, Markus Seidl-Nigsch and Thomas Loerting. “Evidence for high-density liquid water between 0.1 and 0.3 GPa near 150 K”. In: *Proceedings of the National Academy of Sciences* 116.19 (2019), pp. 9191–9196.

[158] L. Tenuzzo, G. Camisasca and P. Gallo. “Protein-Water and Water-Water Long-Time Relaxations in Protein Hydration Water upon Cooling—A Close Look through Density Correlation Functions”. In: *Molecules* 25.19 (2020), p. 4570.

[159] Sciortino Francesco, Saika-Voivod Ivan and Poole Peter. “Study of the ST2 model of water close to the liquid-liquid critical point”. In: *arXiv preprint arXiv:1108.5961* (2011).

[160] Fausto Martelli. “Unravelling the contribution of local structures to the anomalies of water: The synergistic action of several factors”. In: *The Journal of chemical physics* 150.9 (2019), p. 094506.

[161] Lorenzo Tenuzzo, Gaia Camisasca and Paola Gallo. “Protein hydration water - focus on low density and high density local structures upon cooling”. In: *In preparation* ().

[162] Eli Shiratani and Masaki Sasai. “Growth and collapse of structural patterns in the hydrogen bond network in liquid water”. In: *The Journal of chemical physics* 104.19 (1996), pp. 7671–7680.

[163] Tigran V Chalikian. “Structural thermodynamics of hydration”. In: *The Journal of Physical Chemistry B* 105.50 (2001), pp. 12566–12578.

[164] Fernando Luis Barroso da Silva et al. “Understanding and controlling food protein structure and function in foods: perspectives from experiments and computer simulations”. In: *Annual review of food science and technology* 11 (2020), pp. 365–387.

[165] Pooja Rani and Parbati Biswas. “Local structure and dynamics of hydration water in intrinsically disordered proteins”. In: *The Journal of Physical Chemistry B* 119.34 (2015), pp. 10858–10867.