Hydration Sphere Structure of Proteins: a Theoretical Study

Anikó Lábas[1], Imre Bakó[2*] and Julianna Oláh[1*]

[1] Department of Inorganic and Analytical Chemistry, Budapest University of Technology and Economics, Szent Gellért tér 4, H-1111 Budapest, Hungary

[2] Institute of Organic Chemistry Research Centre for Natural Sciences, Hungarian Academy of Sciences Magyar tudósok körútja 2. H-1519 Budapest, P.O. Box 286, Hungary

[*]corresponding author, to whom corresponding should be addressed

e-mail: julianna.olah@mail.bme.hu; telephone number: +36-1-463-1286

e-mail: bako.imre@ttk.mta.hu; telephone number: +36-1-382-69-81

Abstract

Hydration is essential for the proper biological activity of biomolecules. We studied the water network around insulin (as a model protein) in aqueous NaCl solutions using molecular dynamics simulations and statistical analysis of the topological properties (hydrogen bond neighbor number and the interaction energy between hydrogen-bonded water molecules) of the water network. We propose a simple method to define the hydration layers around proteins. Water molecules in the first and second layers form significantly less, but stronger hydrogen bonds with each other than in the bulk phase. Furthermore, water molecules over the hydrophilic and hydrophobic surface of the protein possess slightly different H-bonding properties, supporting the hypothesis of structural and dynamical heterogeneity of the water molecules over protein surface. The protein molecule perturbs the solvent structure at least up to the fourth-fifth hydration layer. Our data suggest the peculiar role of the second hydration shell.

Keywords: hydrogen-bond network, hydration sphere, MD simulations

Introduction

The appropriate spatial structure is essential for the activity of proteins. It is affected by both intramolecular interactions in the proteins and the intermolecular interactions formed with the solvent molecules, which is water in living cells. It has been well-established that the dominant conformational motions of proteins are profoundly affected by their hydration shell.[1,2] As a consequence, structural changes of the solvent should inevitably affect protein structure and function as well. Indeed, addition of compounds such as inorganic salts, organic molecules, acids or bases to the solution can perturb the structure of liquid water leading to the denaturation of the biomolecule. Among these, the denaturing effect of salts has been the most extensively studied, and more than a century ago Franz Hofmeister ordered the ions according to their ability to precipitate egg-white proteins.[3,4] Kosmotropic ions (e.g. sulfate ion) or water structure makers strengthen the hydrogen-bonding network of bulk water and at the same time decrease the solubility of biomolecules. In contrast chaotropic ions (e.g. nitrate ion) supposedly break the hydrogen-bonding network of bulk water and increase the solubility of biomolecules. Recently, the interfacial tension at the protein−water interface was shown to play a central role in the Hofmeister phenomena.[5] Not only salts, but other chemical agents can also denature proteins. Bennion and Daggett simulated the urea-induced denaturation of chymotrypsin and suggested that the solvent plays various roles in the process. Most importantly, the structure and dynamics of the solvent changed in the solution, and intrusion of the solvent molecules into the hydrophobic core of chymotrypsin was responsible for diminishing the hydrophobic effect and encouraging solvation of the core and thereby changing the intramolecular hydrogen bond network in the protein.[6]

Generally, it is accepted that different levels of hydration occur at a biomolecule. In the first hydration layer water interacts with the external surface of the protein through directional hydrogen-bonding (H-bonding) interaction especially on the hydrophilic surface of the protein, while on the hydrophobic surface of the protein the topology, roughness and spatial constrains of the surface orient the water molecules. As a consequence, the hydrogen-bonded properties of water is influenced significantly by the surface properties of macromolecules resulting in increased mean residence time[7–12] and 10-20% increase of the density of water molecules[13,14] compared to the bulk phase. However, a molecular dynamics study on myoglobin, also showed that only those water molecules have very long residence times that are found in cavities and clefts of the protein; other hydration sites of the protein are characterized by residence times similar to the bulk phase.[15] The water-water hydrogen bonding in the first layer forms a spanning, peptide homogeneously enveloped, percolated network, while lack of biological functions is always connected to the broken

(not percolated) H-bonded network in the first layer.[16] Recent terahertz spectroscopic measurements, a method sensitive to the collective motion of water molecules, indicate that protein disturb the water structure beyond the 1-2 water layers as previously thought.[17] The radius of the dynamic hydration shell was greater than 10 Å for the studied proteins and correlated well with the dipole moment of the protein.[18]

It is obvious from the above overview that gaining a better understanding of the hydration layer structure around proteins could contribute to our understanding of various processes involving protein-solvent interactions such as protein folding and unfolding. A possible way to study the structure of water is graph theory, which has recently been applied to the hydrogen bond network in various solutions and mixtures, e.g. of water, methanol and ethanol solutions adsorbed in microporous silicalite-1[19], of ion aggregates in different high salt solutions[20], and of highly concentrated renal osmolyte solutions.[21] Recently, we studied the mixtures of water and formamide, the simplest model of the peptide bond, and showed that these two compounds form microhomogeneous mixtures, in which the number of hydrogen bonds formed by water and formamide are very similar.[22]

When graph theory is used to study the structure (i.e. hydrogen bond network) of water, the network of interactions is mapped into a graph.[23] The vertices of the graph correspond to the water molecules and the edges to the hydrogen bonds formed by the water molecules. Once the hydrogen bond network is mapped into a graph, a thorough statistical analysis can be carried out in order to get insight into the water structure.

In the present work we extend this theory to explore the hydrogen bond network around a protein to obtain a better understanding of its structure and how it changes from the surface of the protein towards the bulk phase, and investigate how molecular dynamics simulations can give more insight into recent findings by terahertz spectroscopic measurements that proteins disturb the water structure beyond 1-2 water layers. Furthermore, we explore the effect of salt concentration on the properties of the hydrogen bond network around a protein. We chose insulin as a model protein because of its (1) small size (2) importance in human health (3) and as it has a balanced distribution of hydrophobic and hydrophilic patches on its surface. NaCl was selected as a co-solute to the protein, because Na^+ has a relatively high concentration in the cytosole compared to divalent cations and it has a weak hydrate sphere ordering capacity and Cl^- is by far the most common anion in living organisms.

First we carried out a series of molecular dynamics simulations at different salt concentrations on solvated systems (with and without the protein), then for each snapshot taken from the trajectory of

the simulations, we determined the network of hydrogen bonds and transformed it to a graph. Finally, a thorough statistical analysis of the properties of the obtained graphs (hydrogen bond neighbor number and the interaction energy between hydrogen-bonded water molecules) were carried out. Importantly, we present here a simpler approach to define solvent layers around the protein compared to those that have been described in the literature,[24–26] and investigate their hydrogen bond properties layer by layer, which enables us to compare the structure and hydrogen bond properties of these layers to those of reference solutions, which do not include the protein molecule. This methodology enables us to characterize the effect of the protein molecule on the hydrogen bond network and to study the structure of its hydration sphere in a statistical way.

Methods

Molecular dynamics simulations (MD). The crystal structure of the monomer, which is the active form of human insulin (PDB code 3I40[27]) was used as a starting structure for the MD simulations. Disulfide bonds were created between $Cys_{6,chainA}$-$Cys_{11,chainA}$, $Cys_{7,chain\ A}$-$Cys_{7,chainB}$ and $Cys_{20,chainA}$-$Cys_{19,chain\ B}$. The protonation state of the titratable amino acid residues were determined using the H++ webserver version 3.2.[28–30] Based on the estimated pK_a values the His_5 residue was doubly protonated in chain B and after visual analysis of their surroundings all other histidine residues were protonated on the ε nitrogen atom. The CHARMM-GUI webserver was used for the system setups and generation of the input files.[31], the NAMD software package[32] with the CHARMM27 force field[33] for the minimization of the structure and dynamics simulations. Hydrogen atoms were added using the standard CHARMM protocol.[34] We chose the TIP3P water model, which is a simple 3-point rigid water model, to simulate water as the non-bonded parameters of protein atom types in the CHARMM27 force field were determined to be in line with the TIP3P water model. As a consequence, when the CHARMM program package is used for modelling proteins the TIP3P water model is by far the most frequently used water model, yielding a reliable description of proteins. Therefore, the protein was solvated by TIP3P water molecules arranged in an octahedral shape with 15 Å edge distances. Three differently solvated protein systems were prepared. One contained only one sodium ion in order to generate a neutral system, while two other systems contained sodium and chloride ions in 0.5 and 1.5 molar concentrations. Although the cytosolic concentration of these ions is much lower, we have chosen these relatively high concentrations to obtain improved statistics for the effect of the desalting in the simulation. The ions were placed by a Monte Carlo approach. Reference systems of the salt solutions with 0 M, 0.5 M and 1.5 M concentrations were also prepared; these did not include the protein molecule. Afterwards, each system was minimized for 10000 steps to eliminate bad initial contacts, followed by a 50 ps long NVT equilibration simulations at 303.15 K temperature. Then 5 ns long NPT Langevin MD simulations were carried out applying 2 fs step size with collecting configurations from every ps. All bonds in the molecules involving hydrogen atoms were kept fixed with the SHAKE[35] algorithm. Periodic boundary conditions (PBC) were used to handle boundary effects. The temperature was set to 303.15 °C in all simulations. The equilibration of the systems was reached by means of temperature reassignment. All of the velocities of the atoms in the systems were periodically reassigned in order to set the desired temperature. Therefore, in every 500 steps the temperature was rescaled during equilibration. The Constant Temperature Control making use of Langevin dynamics. was applied together with the Nose-Hoover Langevin piston pressure control with the target pressure

set to 1.01325 bar. As the random initial velocity distribution used in the MD simulation could influence the obtained results, all MD simulations were carried out with three different initial velocity distributions yielding 3-3 parallel trajectories for all systems studied.

Analysis

Hydrogen bonds. Two water molecules were regarded hydrogen bonded if the H···O distance between the two molecules was smaller than 2.5 Å and O-H···O angle was larger than 120º. This criteria are reasonable for protein simulations where correlation between the distance and the angle criteria have already been shown.[36] However, we have checked the dependency of our results on this definition by analyzing the trajectories obtained for the protein solvated in 0.5 M NaCl solution by setting the O-H···O angle criterion to 130º and 145º as well.

The average hydrogen bond number, N_{HB}, was calculated by averaging the number of hydrogen bonds over the trajectory and over all molecules (Eq. 1):

$$N_{HB}=\frac{<\sum_{i=1}^{N} N_{HB,\,i}>}{2N} \qquad (1)$$

where $N_{HB,i}$ is the number of hydrogen bonds around water molecule i, N is the number of water molecules in the simulated box and < > denotes averaging over all snapshots.

We calculated the hydrogen bond energy between water molecules i and j ($E_{HB,ij}$) for each hydrogen bond using the TIP3P force field following the common practice[37] to identify the hydrogen-bond energy with the interaction energy of the H-bonded molecular pair, even if the H-bond energy cannot be separated from the rest of the pair interaction energy. It was averaged over all hydrogen bonded pairs and over the whole length of the trajectory to yield the average hydrogen bond energy (E_{HB}) using the following equation:

$$E_{HB}=\frac{<\sum_{i=1}^{N}\sum_{j=1}^{N_{HB,i}} E_{HB,ij}>}{<\sum_{i=1}^{N} N_{HB,\,i}>} \qquad (2)$$

$E_{HB,ij}$ is the energy of the hydrogen bond between molecules i and j and the rest of the notation corresponds to Eq. 1.

Definition of solvation layers. As experimental data indicate that proteins disturb considerably the water structure around them, we have developed a methodology to define hydration layers around the protein (**Fig. 1**). Molecules on the surface of the protein (first layer) were defined according to two rules: **(1)** water molecules were assigned to interact with the hydrophilic surface of the protein

if any of the following distances (O_{Water}-$H_{Protein}$ or H_{Water}-$O_{Protein}$ or H_{Water}-$N/S_{Protein}$) was smaller than 2.5 Å and **(2)** water molecules in the first layer interacting with the hydrophobic surface of the protein were defined as having a $C_{Protein}$-O_{Water} (abbreviated to C-O distance from now on) distance smaller than 4.5 Å and not satisfying rule **(1)**. Application of these two rules provided us with a continuous first hydration layer over the surface of the protein. The second hydration layer consisted of the water molecules having an O_{Water}-$O_{Water,layer\ (1)}$ distance smaller than 3.5 Å and not belonging to layer **(1)**. From this on water molecules were assigned to belong to layer n+1 having an $O_{Water,layer(n+1)}$-$O_{Water,layer(n)}$ distance smaller than 3.5 Å and not belong to layer n. 4.5 Å was chosen as the critical C-O distance, because it is very close to the minimum of the partial radial distribution function of the C-O distance in liquid methanol or of the C-O_{water} distance in water-methanol mixtures. It is also the typical C-O distance between the carbon atom of CH_3 groups and the oxgen atom of the water molecule closest to them. However, we have tested the dependency of our results on this distance, and the analyses were performed using 4.0, 4.25, 4.5 and 5 Å criteria as well in the case of insulin system solvated in 0.5 M NaCl solution.

Results

Protein behaviour at different salt concentrations. We compared the behavior of the proteins at various salt concentrations in the MD simulations. The results have been averaged over all snapshots of the productive part of the MD simulation and over the three parallel trajectories carried out at identical salt concentrations. In order to check the stability of the protein conformation during the simulations we have calculated the Root Mean Square Deviation (RMSD) of the atomic positions.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(r_i(t_1) - r_i(t_2)\right)^2}{N}} \quad (3)$$

where N is the number of atoms and $r_i(t)$ is the position of atom i at time t. In all cases t_1 was the position of the protein heavy atoms after the heating up of the system. The comparison of the protein RMSD values for all trajectories is shown in **Table 1**, decomposed according to the various structural elements found in insulin. The structural elements are shown in **Fig. 2**. While the obtained RMSD values are similar in the cases of the neutralized and 0.5M NaCl systems, slightly lower RMSD values have been observed in the case of the 1.5M NaCl system. This could originate from the fact that in this reasonably concentrated 1.5M solution ions have a stronger tendency to absorb on the surface of the protein and slightly stabilize its structure as the ion-surface interaction is much stronger than the H-bonds between the surface and the water molecules.

The RMSD values show that insulin remained in its natural conformation along the MD trajectory. We were also interested to check how other parameters, the radius of gyration and the geometric moment tensor changed along the trajectory, as they could also give insight into structural changes of the protein. Nayar et all showed that the gyration radius of a globular protein (of the 16-residue β-hairpin fragment of the 2GB1 protein) shows a remarkable (at least 1.5-2 Å) difference between its folded and unfolded states.[38] The gyration radius of insulin (10.1 Å±0.1 Å) was quite stable during the whole simulation in all three solution. indicating the stability of the protein structure.

Globular proteins in their native conformation, like insulin, have close-packed structures with quite high number densities and a well-defined shape in the solution. Characterization of the shape of proteins has been in the focus of intense scientific interest for many years. One of the applicable method for the description of the shape of a whole protein uses simple ellipsoids. Here, we apply for this purpose the size of the three main axes of the geometry moment tensor, which is calculated with a similar mathematical construction, as the moment of the inertia tensor, with the difference that each point (i.e. the coordinates of all heavy atoms) is assigned the same mass: a mass of unity, instead of the real mass of the atom. The values of the size of the three main geometrical moments as a function of time are presented in **Fig. 3** in the case of the 1.5 M solution, but we obtained very similar graphs in the case of the other two concentrations as well. The figure shows that that the overall shape of the protein does not change significantly during the simulation time. This method also enables us to characterize the shape of the solvation layers. We performed this for this first three hydration layers together, which is also shown in **Fig. 3**. It is obvious that the shape of the solvation layers closely follow that of the protein, with longer axes.

Number of the water molecules in the solvation layers. We determined the number of water molecules in each shell around the protein (see **Table 2**). As expected as a function of layer number the number of water molecules increases significantly at each salt concentration as we go farther away from the protein. It seems that there is no significant difference between the number of water molecules in the first and second shell around the protein in the case of the neutralized system and the 0.5 M NaCl concentration system, while there seem to be less water molecules around hydrophilic surface of the protein in the simulations with the highest salt concentration. Here it is worth noting, that Na^+ and Cl^- are neither chaotropic nor cosmotropic, thus they have no strong tendency either to be attracted nor to be repelled from the surface of proteins. We examined the number of water molecules associated to the hydrophilic and hydrophobic surface of the protein and it is seen that at all concentrations the number of water molecules at the hydrophobic surface remains identical, but decreases slightly at the hydrophilic surface. There are most likely two different reasons

for this: (1) the solvent accessible surface area of the protein slightly decreases and as such fewer water molecules can access it (2) with increasing salt concentration, the "place" of several water molecules is taken over by ions, which can also occur especially on the hydrophilic surface of the protein. This finding is in accordance with the expectation that ions are repelled from hydrophobic/non-polar regions. As we go farther away from the protein surface we find an increasing difference among the number of molecules in a given shell with increasing salt concentration due to the presence of an increasing number of ions.

Radial distribution function. The oxygen-oxygen (O_{water}-O_{water}) partial radial distribution function (RDF) can be used for characterization of the structure of water around the protein surface.[39,40] We calculated two different types of radial distribution functions to evaluate the extent of the water structure around protein surface. In one of them we calculated the RDF for the interactions between the water molecules in the first layer- and all other water molecules in layers 2-4 to obtain the surface distribution function (SDF). In the other case water-water RDF was calculated only for water molecules in the first layer. Unfortunately, from our simulations we do not know a priori the exact density of water in the simulation box and we could not properly calculate the excluded volume effect, so the raw O_{water}-O_{water} density distribution was determined using the spherical average method according to Eq. 4.:

$$\rho(r) = \frac{N(r)}{(4\pi r^2 dr)} \quad (4)$$

where N denotes the number of particles. It can be seen from both **Figs. 4** and **5** that the curvature of the raw density distribution functions is the same for all interaction types, thus we can assume that the necessary corrections to account for the excluded volume effect would be the same in all cases, thus we can obtain correct trends from the raw distribution functions as well.

In **Fig. 4** the surface distribution functions are shown for the interactions between hydrophobic- and hydrophilic-all other water molecules in layers 2-4 cases. We cannot detect any large changes on these SDFs as a function of salt concentration. The first peak on these SDF is around 2.8 Å and it is significantly more pronounced for the hydrophobic surface water-all water case.

The raw density distribution functions corresponding to the first layer-first layer structure are presented in **Fig. 5**. On the insect we presented the long range behavior of these functions, which show that the long range behavior of these function is the same for different interaction types: hydrophilic-hydrophilic (Hy-Hy), hydrophobic-hydrophobic (Hyb-Hyb) and hydrophilic-hydrophobic (Hy-Hyb); thus these functions can be compared. The H-bonded structure, which can

be characterized by the first peak at around 2.8 Å is more pronounced in the Hyb-Hyb case, which is in good agreement with results of statistical analysis of H-bond strength (see below). The shape of the first peak of the hydrophilic-hydrophilic (Hy-Hy) RDF is considerably deviates from the other two (Hyb-Hyb,Hy-Hyb) RDFs, especially in the range of 3.2-4.5 Å. This difference in shape indicates (1) the uniformity of the interaction between water molecules over the hydrophobic surface of the water molecules (2) larger differences in interaction strengths between water molecules over the hydrophilic surface of the water. This implies that water molecules behave differently over the hydrophobic and hydrophilic surface of the protein and supports the hypothesis of structural and dynamical heterogeneity of the surface.

Average hydrogen bond number. Next, we determined the average number of hydrogen bonds formed by water molecules in the reference aqueous solutions not including the protein, then we analyzed the hydrogen bonds formed by water molecules in the protein-containing systems. In the latter case we determined how many hydrogen bonds are formed between water molecules in the same layer, and between two layers as well. In the case of the first hydration layer the data was even further decomposed that we could see the effect of the protein on the hydrogen bond network. The data is collected in **Table 3**. In the reference solutions, the average hydrogen bond number in the pure water and at 0.5 M NaCl solution is 3.4 which agrees very well with the results of earlier studies.[23,41] At the highest salt solution, the average hydrogen bond number is decreased to 3.14, which clearly originates from the fact that at such a high salt concentration, significant number of hydrogen bonding sites of the water molecules are occupied by the solute ions, i.e. the water molecule itself belongs with an increasing probability to the solvent sphere of an ion.

Water molecules in the 1st solvation shell establish significantly lower number of hydrogen bonds with other water molecules than in the bulk phase (e.g. $n_{HB,ref}$=3.40 at 0.5M), but this effect is overcompensated by the hydrogen bonds established with the protein. In the 2nd shell significantly smaller number of hydrogen bonds are found compared to the bulk phase or the other solvation layers. From the 3rd shell the average hydrogen bond number and the structure of the hydrogen bond network begins to resemble to the reference systems, although the total number of hydrogen bonds is slightly larger. As a result of the increase of the salt concentration the number of the hydrogen bonds (N_{HB}) decreases, and water molecules in the 3rd and 4th solvation shells form similar number of hydrogen bonds than found in the reference systems (aqueous solutions with identical salt concentration). (Here we would like to note that in the reference systems the hydrogen bonds among water molecules that are not in the coordination sphere of sodium or chloride ions are stronger by 1.7- 2.5 kJ/mol than the average hydrogen bond number averaged over all water molecules.) Our

data show that the number of hydrogen bonds in the 1st solvation shell are considerably larger than in the 2nd layer. The number of hydrogen bonds (N_{HB}) formed only within the 1st shell is around 1.8 and the occurrence of a few very large clusters in the cluster size distribution indicate the formation of a percolated network (i.e. a continuous network of interconnected clusters) in good agreement with previous data of Brovchenko et al.[42] We performed the same cluster analysis for solvation layers 2, 3 and 4 separately, taking into account only the in-layer H-bonds, but in these layers the water molecules do not form a percolated network in any of these layers. The average number of hydrogen bonds formed among water molecules in the 2nd solvation shell is remarkably small. This is due to the fact that while water molecules in the 1st layer orient themselves to form as many hydrogen bonds with the protein as possible, the 3rd layer orients itself towards bulk water, but water molecules in the 2nd layer cannot easily accommodate themselves to establish an optimal number of hydrogen bonds. This is best regarded as a transient layer whose properties are determined by a combined effect of the adaptation to the shape of the protein molecule and to the network of water molecules in the bulk solvent.

We have checked the dependence of the results on the applied H-bond criteria. In **Table 4** we have collected the data obtained with various C-O distances used to allocate water molecules over the hydrophobic surface of the protein. It can be observed that the C-O distance primarily influences the average H-bond neighbor number in the first shell, taking longer distance the n_{HB} slightly increases. The same effect is observed in the second shell although to a lesser extent, but no effect can be seen in the case of the farther layers. The reason observed for the first two layers arises from the fact that by changing the C-O distance the number of water molecules slightly changes over the hydrophobic surface of the protein, some molecules may be assigned to layer 1 or layer 2 depending on the exact C-O distance, but at larger distances this effect diminishes. However, it is important to emphasizes that our conclusions do not depend on the exact value of the C-O distance between 4-5 Å.

We have also investigated how the obtained results depend on the angle criteria used to determine hydrogen bonds. The data in **Table 5** show that with using a larger angle criterion, the average H-bond neighbor number slightly decreases as only more “perfect” H-bonds are identified as H-bonds (i.e. a smaller number of them is found), but the observed trends do not change at all.

Changes in hydrogen bond energy as a function of the distance from the surface of the protein. For the characterization of the hydrogen bonded interaction we can use the strength of the hydrogen bond. This strength is identified by the interaction energy of the two hydrogen bonded water molecules. In order to sign the perturbation effect of the protein we have investigated the average hydrogen bond energies in each layer and between the layers. Student's t-test, a rigorous statistical

probe, was used to prove that the observed differences between the obtained values are significant. Data in **Table 6** show that the strongest hydrogen bonds between water molecules are formed in the pure liquid water reference system, and addition of solutes (either protein or salt) decreases the hydrogen bond strength, and this decrease becomes more significant with increasing salt concentration. Furthermore, **Table 7** shows that changing the angle criteria does not influence the observe trends. We can decompose this value to calculate the hydrogen bond energy strength between water molecules associated to the hydrophilic or to the hydrophobic surface of the protein. It is clear from these data that hydrogen bonds among water molecules at the hydrophobic surface are much stronger than at the hydrophilic surface. This is due to the protein-water interactions at the hydrophilic surface, where water molecules orient themselves to interact mainly with protein surface groups. However, over the hydrophobic surface water molecules mainly interact with each other leading to more optimal hydrogen bonded arrangements. Furthermore, the average hydrogen bond strength in the first layer is mainly governed by interactions of the hydrophobic water molecules, and suggests that the interaction between a hydrophobic and hydrophilic water molecule is also very strong. These results are in a good agreement with the differences of the water-water partial radial distribution functions calculated for the various parts (hydrophobic or hydrophilic) of the protein surface (**Fig. 5**).

We also observe considerable difference in the strength of the interactions between and in the layers (e.g. in the case of the 0.5M system: inside the 2^{nd} layer -14.88 kJ/mol and between the 2^{nd} and 3^{rd} layers -15.34 kJ/mol, respectively), which indicates that the perturbation effect caused by the protein shape is more pronounced in interactions in the layers than between the layers. This phenomenon is in the 4^{th} and 5^{th} layers begin to disappear. In these layers the strength of the hydrogen bonds is getting closer and closer to the values observed in the reference systems. (Here we would like to note that in the reference systems the hydrogen bond strength between water molecules that are not in the coordination sphere of sodium or chloride ions is only 0.08-0.13 kJ/mol stronger than the average hydrogen bond strength averaged over all water molecules.) Due to the salt concentration, the average hydrogen bond strength decreases in the same way as in the reference systems.

Relevance of our results to physiological solutions

After having examined numerous properties of the hydrogen bond network around insulin, used as a model protein, it is worth putting our results into a wider context and consider their relevance to physiological and other solutions. The data show that proteins disturb the H-bond network at least up to five water layers, which imply a large volume of water and reasonably large distances between the protein molecules. We calculated the concentration of the protein in our systems to be $8.360*10^{-}$

^{3} M, 8.762*10^{-3} M and 9.595*10^{-3} M in the neutral, 0.5 M and 1.5M NaCl solutions, respectively. As an example of a physiologically relevant solution we could consider blood, whose typical protein concentration can be estimated the following way. The most abundant proteins in blood are albumins (constituting about 55% of blood proteins), which are present in 3-5 g/dL concentration. The molar weight of albumins is around 65 kDalton (or 65000 g/mol). Taking 5 g/dL concentration this would yield a concentration of 7.69*10^{-4} M. Most other proteins present in blood have a much higher molecular weight (e.g. the molar mass of globulins is between 93-1193 kDalton), thus they increase only slightly the molar protein concentration of blood or they are present in much smaller quantities e.g. typical blood level of insulin between meals is 57*10^{-4} M.[43] This implies that blood is a slightly more diluted protein solution than the insulin solution studied by us, thus the 5 water layers is likely to be present around protein in blood and most likely in other physiologically relevant solutions. It is worth keeping in mind, though, that blood and other physiological solutions contain a variety of other co-solutes, e.g. sugars and other small molecules, which also influence the water network in them. Furthermore, from smaller peptides one may prepare much more concentrated solutions, where the individual peptide molecules may be closer to each other and there could be less than 8-10 hydration layers between two peptide molecules (i.e. 5 layers belonging to peptide 1 and five layers belonging to peptide would yield a separation of 10 layers). In this case it is very likely that proteins would seriously influence each other's solvation spheres, and in this case our conclusions would not be valid. A separate study would be needed to study the effect of peptide concentration on the structure of the hydration shells of peptides.

Conclusions

In this work we have studied the topological properties of the water layers around protein molecules as a function of sodium chloride concentration, using insulin as a model system. Our statistical analysis shows a significant difference among the hydrogen-bonded properties of water molecules in the first, second and farther solvation layers. We can also show that water molecules over the hydrophilic and hydrophobic surface of the protein possess slightly different H-bonding properties, supporting the hypothesis of structural and dynamical heterogeneity of the water molecules over the protein surface. The effect of the protein on the hydrogen bonded water network exist at least up to 4 layers, which is in accordance with recently reported sub-terahertz spectroscopic measurements.

Acknowledgements

A.L. thanks the financial support for the Richter Gedeon Talentum Foundation. J.O. was supported by the Bolyai János Research Scholarship and by NKFIH Grant No. 115503. I.B. was supported by OTKA Grant No. 108721.

References

1 Frauenfelder H, Sligar SG & Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* **254**, 1598–1603.

2 Austin RH, Beeson KW, Eisenstein L, Frauenfelder H & Gunsalus IC (1975) Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–5373.

3 Hofmeister F (1888) On the understanding of the effect of salts. Second report. On regularities in the precipitating effect of salts and their relationship to their physiological behavior. *Naunyn-Schmiedebergs Arch. Exp. Pathol. Pharmakol* **24**, 247–260.

4 Baldwin RL (1996) How Hofmeister ion interactions affect protein stability. *Biophys. J.* **71**, 2056–2063.

5 Bogár F, Bartha F, Násztor Z, Fábián L, Leitgeb B & Dér A (2014) On the Hofmeister Effect: Fluctuations at the Protein–Water Interface and the Surface Tension. *J. Phys. Chem. B* **118**, 8496–8504.

6 Bennion BJ & Daggett V (2003) The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5142–5147.

7 Otting G, Liepinsh E & Wüthrich K (1991) Protein hydration in aqueous solution. *Science* **254**, 974–980.

8 Denisov VP & Halle B (1996) Protein hydration dynamics in aqueous solution. *Faraday Discuss.* **103**, 227–244.

9 Wiesner S, Kurian E, Prendergast FG & Halle B (1999) Water molecules in the binding cavity of intestinal fatty acid binding protein: dynamic characterization by Water17O and 2H magnetic relaxation dispersion. *J. Mol. Biol.* **286**, 233–246.

10 Henchman RH & McCammon JA (2002) Structural and dynamic properties of water around acetylcholinesterase. *Protein Sci.* **11**, 2080–2090.

11 Pal SK & Zewail AH (2004) Dynamics of water in biological recognition. *Chem. Rev.* **104**, 2099–2123.

12 Russo D, Hura G & Head-Gordon T (2004) Hydration dynamics near a model protein surface. *Biophys. J.* **86**, 1852–1862.

13 Rossky PJ & Cheng Y-K (1998) Surface topography dependence of biomolecular hydrophobic hydration. *Nature* **392**, 696–699.

14 Makarov V, Pettitt BM & Feig M (2002) Solvation and hydration of proteins and nucleic acids: A theoretical view of simulation and experiment. *Acc. Chem. Res.* **35**, 376–384.

15 Makarov VA, Andrews BK, Smith PE & Pettitt BM (2000) Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys. J.* **79**, 2966–2974.

16 Oleinikova A & Brovchenko I (2011) What Determines the Thermal Stability of the Hydrogen-Bonded Water Network Enveloping Peptides? *J. Phys. Chem. Lett.* **2**, 765–769.

17 Ebbinghaus S, Kim SJ, Heyden M, Yu X, Heugen U, Gruebele M, Leitner DM & Havenith M (2007) An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci.* **104**, 20749–20752.

18 Sushko O, Dubrovka R & Donnan RS (2015) Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected. *J. Chem. Phys.* **142**, 55101.

19 Wang C-H, Bai P, Siepmann JI & Clark AE (2014) Deconstructing Hydrogen-Bond Networks in Confined Nanoporous Materials: Implications for Alcohol–Water Separation. **118**, 19723–19732.

20 Choi J-H & Cho M (2014) Ion aggregation in high salt solutions. II. Spectral graph analysis of water hydrogen-bonding network and ion aggregate structures. *J. Chem. Phys.* **141**, 154502.

21 Lee H, Choi J-H, Verma PK & Cho M (2015) Spectral Graph Analyses of Water Hydrogen-Bonding Network and Osmolyte Aggregate Structures in Osmolyte-Water Solutions. *J. Phys. Chem. B* **119**, 14402–14412.

22 Bakó I, Oláh J, Lábas A, Bálint S, Pusztai L & Bellissent Funel MC (2017) Water-formamide mixtures: Topology of the hydrogen-bonded network. *J. Mol. Liq.* **228**, 25–31.

23 Bakó I, Bencsura Á, Hermannson K, Bálint S, Grósz T, Chihaia V & Oláh J (2013) Hydrogen bond network topology in liquid water and methanol: a graph theory approach. *Phys. Chem. Chem. Phys.* **15**, 15163–15171.

24 Mezei M (2003) A new method for mapping macromolecular topography. *J. Mol. Graph. Model.* **21**, 463–472.

25 Willard AP & Chandler D (2010) Instantaneous Liquid Interfaces. *J. Phys. Chem. B* **114**, 1954–1958.

26 Sega M, Kantorovich SS, Jedlovszky P & Jorge M (2013) The generalized identification of truly interfacial molecules (ITIM) algorithm for nonplanar interfaces. *J. Chem. Phys.* **138**, 44110.

27 Timofeev V., Chuprov-Netochin R., Samigina V., Bezuglov V., Miroshnikov K. & Kuranova IP (2010) X-ray investigation of gene-engineered human insulin crystallized from a solution containing polysialic acid. *Acta Crystallogr.,Sect.F* **66**, 259–263.

28 Anandakrishnan R, Aguilar B & Onufriev A V (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40**, W537-541.

29 Myers J, Grothaus G, Narayanan S & Onufriev A (2006) A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* **63**, 928–938.

30 Gordon JC, Myers JB, Folta T, Shoja V, Heath LS & Onufriev A (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, W368-371.

31 Jo S, Kim T, Iyer VG & Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865.

32 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L & Schulten K (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.

33 MacKerell AD, Banavali N & Foloppe N (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56**, 257–265.

34 Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G,

Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM & Karplus M (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614.

35 Ryckaert J-P, Ciccotti G & Berendsen HJ. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341.

36 De Loof H, Nilsson L & Rigler R (1992) Molecular dynamics simulation of galanin in aqueous and nonaqueous solution. *J. Am. Chem. Soc.* **114**, 4028–4035.

37 Wendler K, Thar J, Zahn S & Kirchner B (2010) Estimating the Hydrogen Bond Energy. *J. Phys. Chem. A* **114**, 9529–9536.

38 Nayar D & Chakravarty C (2014) Sensitivity of local hydration behaviour and conformational preferences of peptides to choice of water model. *Phys. Chem. Chem. Phys.* **16**, 10199–10213.

39 Rani P & Biswas P (2014) Shape dependence of the radial distribution function of hydration water around proteins. *J. Phys. Condens. Matter* **26**, 335102.

40 Rani P & Biswas P (2015) Local Structure and Dynamics of Hydration Water in Intrinsically Disordered Proteins. *J. Phys. Chem. B* **119**, 10858–10867.

41 Abhishek Rastogi AKG and SJS (2011) *Hydrogen bond interactions between water molecules in bulk liquid, near electrode surfaces and around ions* InTech.

42 Oleinikova A, Brovchenko I, Smolin N, Krukau A, Geiger A & Winter R (2005) Percolation Transition of Hydration Water: From Planar Hydrophilic Surfaces to Proteins. *Phys. Rev. Lett.* **95**, 247802.

43 Iwase H, Kobayashi M, Nakajima M, Takatori T, Rindfrey H, Lang H, Leybold K, Rick W & Staudinger HJ (2001) The ratio of insulin to C-peptide can be used to make a forensic diagnosis of exogenous insulin overdosage. *Forensic Sci. Int.* **115**, 123–127.

Table 1. Average protein RMSD values calculated for backbone heavy atoms in Å in the three parallel 5 ns long MD simulations with their standard deviations for all systems.

Chain	Structure	Neutral	0.5 M	1.5 M
A	**Helices**	**1.72** ± 0.97	**1.74** ± 1.20	**1.44** ± 1.07
A	**Turn**	**1.56** ± 0.24	**1.69** ± 0.45	**1.28** ± 0.28
A	**Loop**	**3.67** ± 3.42	**2.98** ± 1.77	**2.32** ± 1.54
B	**Helix**	**1.87** ± 0.87	**1.96** ± 0.89	**1.66** ± 0.79
B	**Loops**	**1.87** ± 0.80	**2.00** ± 0.79	**1.53** ± 0.91

Table 2. Number of water molecules in shells with their standard deviations around the protein averaged over all snapshots and three parallel MD simulations and total number of water molecules/ions in the simulation box as a function of salt concentrations

Water layer	Neutral	0.5 M	1.5 M
1	**269** ±3	**269** ±4	**263** ±4
1 ***hydrophobic***	**149** ±3	**150** ±3	**150** ±2
1 ***hydrophilic***	**120** ±1	**119** ±3	**113** ±2
2	**297**± 4	**296** ±4	**287** ±3
3	**389** ±4	**386** ±5	**374** ±3
4	**493** ±5	**490**± 5	**478** ±4
5	**611** ±5	**607** ±4	**594** ±5
N_{water}	**6445**	**6104**	**5459**
N_{sodium}	**1**	**61**	**182**
$N_{chloride}$	**0**	**60**	**181**
$N_{water,\ reference}$	**6657**	**6397**	**5800**

Table 3. The average number of hydrogen bonds with their standard deviations formed between protein and 1st solvation layer, and between the rest of the solvation shells at different salt concentrations. Data is averaged over the three parallel MD simulations and over all snapshots. "All" denotes the sum of all hydrogen bonds towards any partner. Solvation layers are designated by numbers.

Hydrogen bond partners	Neutral	0.5 M	1.5 M
1^{st} - 1^{st}	**1.84** ± 0.01	**1.81** ± 0.01	**1.76** ± 0.00
1^{st} - protein	**0.56** ± 0.02	**0.55** ± 0.02	**0.55** ± 0.01
1^{st} - $protein^{hydrophilic}$	**1.25** ± 0.04	**1.26** ± 0.06	**1.26** ± 0.03
1^{st} - all water	**3.05** ± 0.02	**2.98** ± 0.01	**2.84** ± 0.03
$1^{st,hydrophilic}$ - all water	**2.54** ± 0.02	**2.49** ± 0.02	**2.36** ± 0.02
$1^{st,hydrophobic}$ - all water	**3.46** ± 0.01	**3.37** ± 0.02	**3.18** ± 0.02
1^{st} – all (protein+ water)	**3.61** ± 0.02	**3.53** ± 0.01	**3.39** ± 0.02
2^{nd} - 2^{nd}	**1.18** ± 0.01	**1.14** ± 0.01	**1.06** ± 0.01
2^{nd} - all water	**2.91** ± 0.01	**2.79** ± 0.02	**2.57** ± 0.02
3^{rd} - 3^{rd}	**1.17** ± 0.01	**1.13** ± 0.01	**1.04** ± 0.01
3^{rd} – all water	**3.69** ± 0.01	**3.56** ± 0.04	**3.28** ± 0.03
4^{th} - 4^{th}	**1.17** ± 0.01	**1.13** ± 0.01	**1.04** ± 0.01
4^{th} – all water	**3.68** ± 0.03	**3.54** ± 0.05	**3.26** ± 0.05
reference systems	**3.40** ± 0.01	**3.40** ± 0.01	**3.14** ± 0.01

Table 4. The average number of hydrogen bonds using various C···O distance criteria for identifying hydrophobic water molecules in the first hydration shell in protein simulations at 0.5 M NaCl concentration. "All" denotes the sum of all hydrogen bonds towards any partner. Solvation layers are designated by numbers.

Hydrogen bond partners	5.00 Å	4.50 Å	4.25 Å	4.00 Å
1st - all water	3.07	2.99	2.93	2.85
2nd - all water	2.82	2.80	2.79	2.77
3rd - all water	3.58	3.58	3.58	3.58
4th - all water	3.57	3.57	3.57	3.57

Table 5. The average number of hydrogen bonds at 0.5 M NaCl concentration calculated using various angle criteria for defining a hydrogen bond. "All" denotes the sum of all hydrogen bonds towards any partner. Solvation layers are designated by numbers.

Hydrogen bond partners	120°	130°	145°
1st - 1st	1.63	1.54	1.27
1st - all water	2.93	2.77	2.27
2^{nd} - 2^{nd}	1.14	1.07	0.85
2^{nd} - all water	2.54	2.38	1.92
3^{rd} - 3^{rd}	1.13	1.06	0.84
3^{rd} - all water	3.58	3.36	2.72
4^{th} - 4^{th}	1.14	1.06	0.85
4^{th} - all water	3.57	3.36	2.72

Table 6. The average energy (kJ/mol) of hydrogen bonds with their standard deviations formed between molecules within the same layer and between molecules in neighboring layers.

	Neutral	0.5 M	1.5 M
1^{st} - 1^{st}	**-15.73** ± 0.02	**-15.68** ± 0.01	**-15.63** ± 0.03
$1^{st,hydrophylic}$ - $1^{st,hydrophylic}$	**-15.07** ± 0.04	**-14.96** ± 0.03	**-14.91** ± 0.04
$1^{st,hydrophobic}$ - $1^{st,hydrophobic}$	**-15.78** ± 0.04	**-15.70** ± 0.04	**-15.53** ± 0.04
1^{st} - 2^{nd}	**-15.14** ± 0.02	**-15.03** ± 0.01	**-14.83** ± 0.03
2^{nd} - 2^{nd}	**-14.95** ± 0.01	**-14.87** ± 0.01	**-14.68** ± 0.02
2^{nd} - 3^{rd}	**-15.49** ± 0.01	**-15.34** ± 0.01	**-15.05** ± 0.03
3^{rd} - 3^{rd}	**-15.01** ± 0.02	**-14.88** ± 0.01	**-14.65** ± 0.01
3^{rd} - 4^{th}	**-15.50** ± 0.01	**-15.34** ± 0.01	**-15.03** ± 0.02
4^{th} - 4^{th}	**-15.20** ± 0.02	**-14.98** ± 0.01	**-14.65** ± 0.01
4^{th} - 5^{th}	**-15.51** ± 0.01	**-15.35** ± 0.01	**-15.01** ± 0.01
Reference systems	**-15.91** ± 0.01	**-15.67** ± 0.01	**-15.55** ± 0.01

Table 7. The average energy (kJ/mol) of hydrogen bonds at 0.5 M NaCl concentration calculated using various angle criteria for defining a hydrogen bond. Solvation layers are designated by numbers.

	120°	130°	145°
1^{st} - 1^{st}	-15.66	-16.79	-18.21
1^{st} – 2^{nd}	-15.11	-16.04	-17.42
2^{nd} – 2^{nd}	-14.86	-16.08	-17.58
2^{nd} – 3^{rd}	-15.32	-16.20	-17.54
3^{rd} – 3^{rd}	-14.86	-16.12	-17.58
3^{rd} – 4^{th}	-15.32	-16.20	-17.50
4^{th} – 4^{th}	-15.16	-16.08	-17.33
4^{th} – 5^{th}	-15.37	-16.08	-17.25
5^{th} – 5^{th}	-14.91	-15.66	-16.87

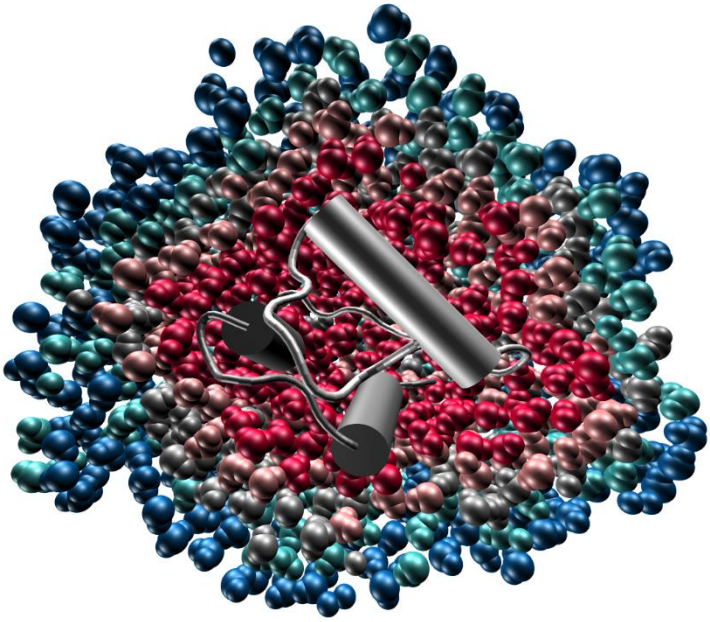


Figure 1. Schematic representation of the hydration sphere layers around the protein.

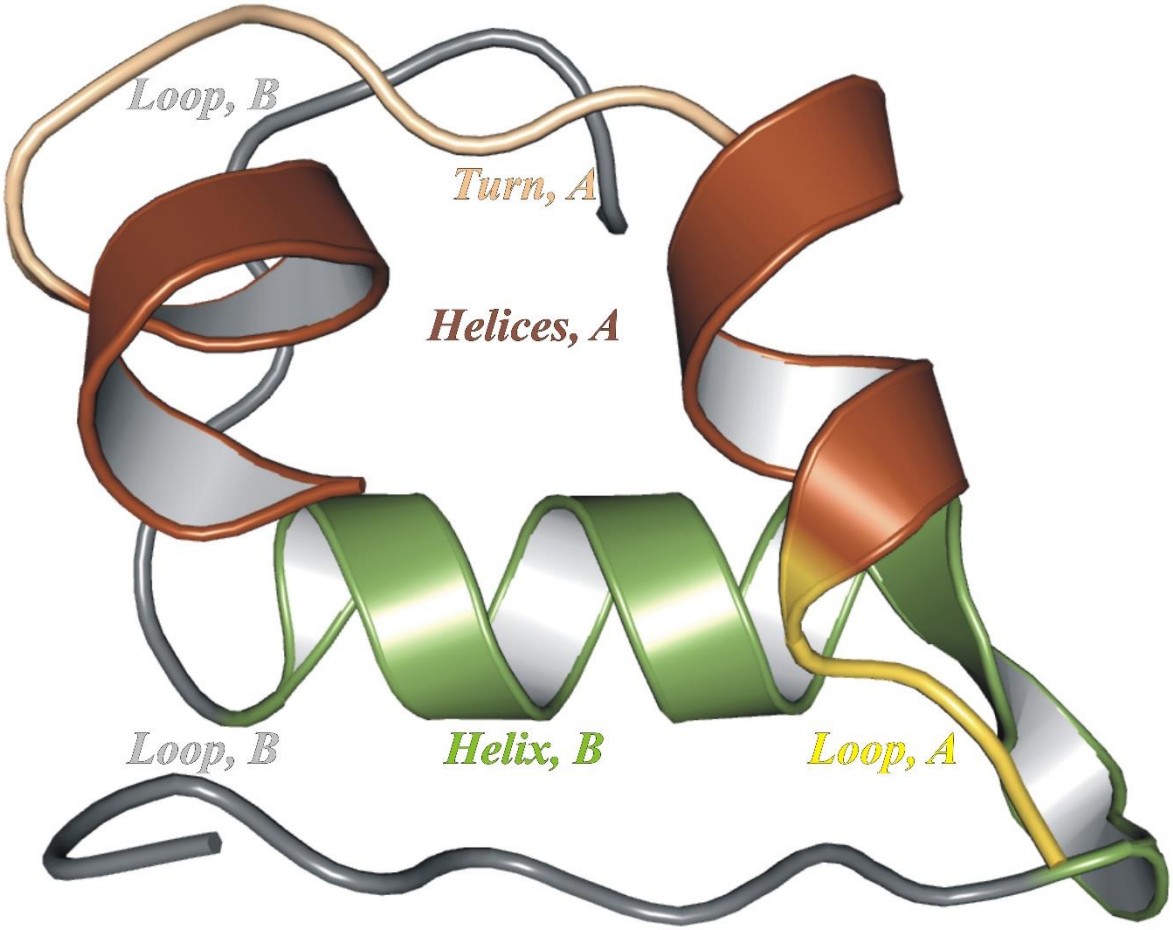


Figure 2. Structural elements in the insulin monomer. Colors used to depict the various structural elements in chain A : brown – helices; wheat – turn; yellow – loop; while in chain B: green – helix; grey – loops.

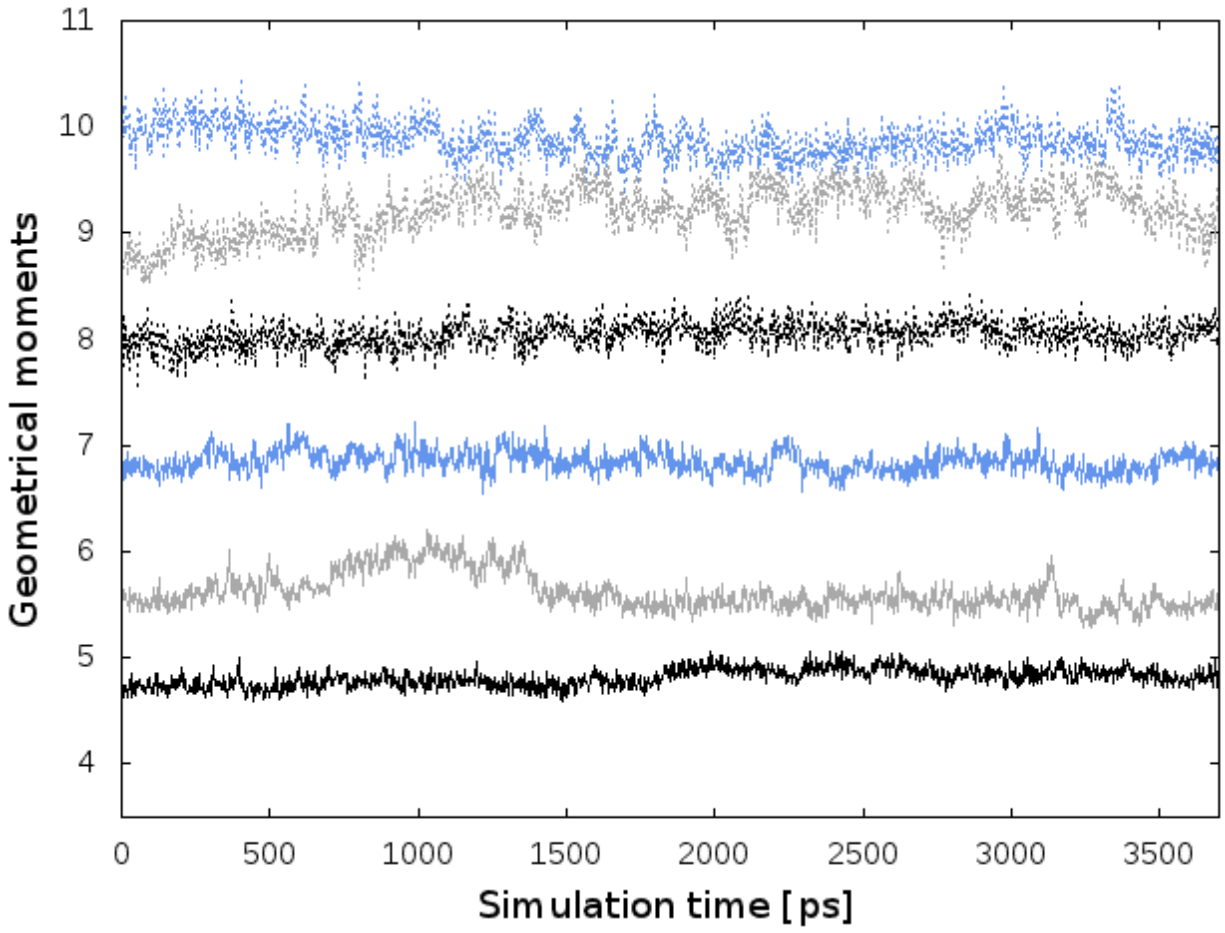
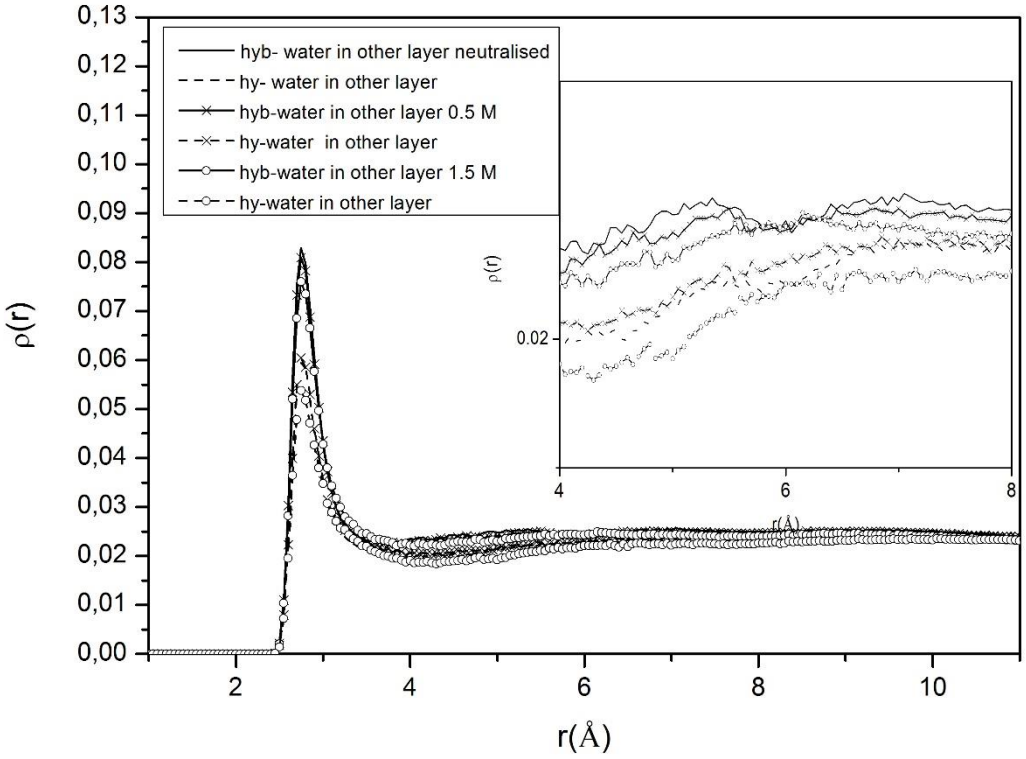


Figure 3. The inertia moments during the simulations. Inercia momements belong to protein shown with lines. Data include the protein and the first three water layers are represented with dotted lines. The three moment components are shown in black, gray and ice blue.



549

550 **Figure 4. Decomposition of the raw density distribution function (of-the O_{water}-O_{water}**
551 **distance) of water molecules in the first hydration layer interacting with water molecules in**
552 **outer layers according to the hydrophilic (it forms H-bonds with the protein) or hydrophobic**
553 **(it forms no H-bonds with the protein) nature of the water molecules in the first solvation**
554 **layer at various salt concentrations.**

555

556

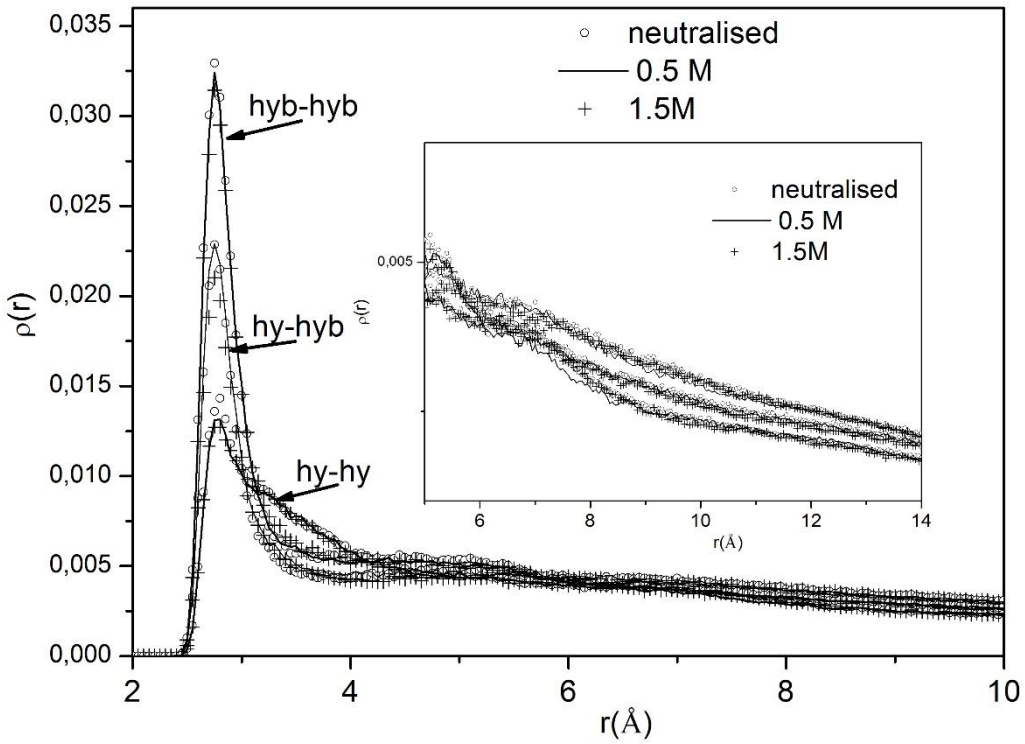


Figure 5 Decomposition of the raw density distribution function (of-the O_{water}-O_{water} distance) among water molecule in the first hydration shell according to the nature of interacting water molecules: Hydrophobic water: it forms no H-bonds with the protein, hydrophylic water: it forms H-bonds with the protein at various salt concentrations.