International Journal of Biological Macromolecules 99 (2017) 160–165

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Protein hydration: Investigation of globular protein crystal structures

Oliviero Carugo[a,b,*]

[a] Department of Chemistry, University of Pavia, viale Taramelli 12, I-27100 Pavia, Italy
[b] Department of Structural and Computational Biology, Vienna University, Campus Vienna Biocenter, A-1030 Vienna, Austria

ARTICLE INFO

Article history:
Received 10 January 2017
Received in revised form 10 February 2017
Available online 23 February 2017

Keywords:
Structural biology
Water chemistry
Protein hydration

ABSTRACT

The positions of water molecules have been analyzed in high quality protein X-ray crystal structures. About 70% of these water molecules are in contact with protein atoms at the protein surface and constitute the first hydration layer. About 20% of them are close to the first hydration layer but are not in contact with protein atoms and constitute the second hydration layer. The rest of the water molecules are either buried in the protein core or close to hetero-atoms (inorganic ions and small organic molecules). Upper layers (third, fourth, etc.) are not observed in the dataset of protein crystal structures examined here. Water molecules of both layers are not, in general, surrounded by a tetrahedral arrangement of atoms, as it should be expected on the basis of the electronic structure of water. Usually there are less than four atoms around water molecules and even when there are four atoms, the stereochemistry is often distorted. Water molecules are more mobile than protein atoms, more in the second hydration layer than in the first.

1. Introduction

Since globular proteins exert their functions in extremely concentrated aqueous solutions, their evolution has clearly been constrained by their necessity to be soluble and to avoid aggregation and crystallization *in vivo* [1]. It has been estimated that only 60–70% of the volume of a cell is occupied by water molecules [2,3], with the consequence that collisions and interactions (physiological and not) between macromolecules are extremely frequent and numerous. Interestingly, several years ago, McConkey introduced the expression "quinary" structure to indicate protein-matrix contacts, which severely influence protein stability, flexibility, and function, despite the thermodynamics of each, single intermolecular interaction is modest [4,5]. It is perhaps surprising that this macromolecular density is comparable with that of protein crystals, where water molecules can occupy 30–70% of the volume [6–8]. Obviously, this does not mean that protein crystals are realistic models of physiological macromolecular crowding, since protein molecules are not free to move in the crystal. However, the crowding level is similar and drastically different from the dilute solutions that are commonly analyzed in most of the biochemical spectroscopic studies.

The interaction between globular proteins and water is crucial to ensure protein solubility and water, in physiological conditions, may be rather different from bulk, liquid water. It is interesting to remember, at this regard, studies performed nearly fifty years ago that showed how a considerable fraction of water (0.3–0.4 g of water per gram of protein) was unfreezable in protein solutions [9]. This implies that water is not just a solvent for globular proteins but it is also a partner, which determines their structures and functions. A good example of this starring role of water is the case of the water molecules buried in the protein core [10,11], which are so common that buried water deserved the name of "21th type of residue" [12].

Understandably, several studies have been devoted also to the analysis of protein surface hydration, which resulted in contradictory results. On the one hand, there are studies that suggest that the hydration layer is very thin, and on the other hand, there are studies that suggest that the hydration layer is extremely thick.

At least in part, these contradictions may depend on the definition of hydration layer. In particular, spectroscopic techniques can determine the fraction of water molecules, the properties of which are affected by the presence of a protein in solution, and different spectroscopic methods, by measuring different properties, might present different results.

According to Laage, the perturbation induced by a protein is short-ranged and involves only the first layer of water molecules, within 3.5–4.0 Å [13]. Moreover, water molecules are quite mobile. If the rotational dynamics in ice is one million slower than in cold water, both rotational and translational dynamics are only 2–3

* Correspondence address: Department of Chemistry, University of Pavia, viale Taramelli 12, I-27100 Pavia, Italy.
E-mail address: oliviero.carugo@univie.ac.at

http://dx.doi.org/10.1016/j.ijbiomac.2017.02.073
0141-8130/

times slower in the hydration layer [13]. The hydration dynamics is however quite heterogeneous. Water molecules located in deep pockets of the protein surface are much less mobile than water molecules close to planar or convex patches of surfaces [13].

This description of protein hydration seems to be confirmed by experiments on protein solutions at high concentration. Confinements slow down systematically this fast water dynamics. Two dimensional spectroscopy was used to study water hydration on a protein in water-glycerol solutions [14]. Increasing the grycerol fraction increases the confinement of water near the protein surface. A threefold dynamics showdown was observed in going from pure water to 80% glycerol volume fraction. Ultrafast infrared spectroscopy showed an increasing retardation in reorientation dynamics of pure water confined in reverse micelles of decreasing size [15]. Hydration dynamics slows down to timescales ranging from ps to ns in protein encapsulated in reverse micelles examined by NO-NMR, suggesting the formations of water clusters [16].

Few studies on protein hydration in intact cells show similar trends. An NMR study of water in *E. coli* and a halophilic organism showed that hydration dynamics is retarded 15 folds relative to diluted protein solutions [3]. A neutron scattering study showed somehow different results: water translational dynamics is almost bulk-like in *E. coli* and retarded 250 folds on the halophilic organism [17].

On the contrary, quite a different picture of protein hydration emerged on the basis of terahetrz spectroscopy and related computational studies. Sushko and colleagues observed 2–3 hydration layers in lysozyme, myoglobin, and bovine serum albumin [18]. Molecular dynamics simulations of the protein hydration shell coupled with THz data showed an hydration layers about 10 Å [19,20]. Ding and colleagues determined that the hydration-shell thickness is 11–17 Å in alanine-rich peptides [21]. A thickness up to 20 Å was observed also by Ebbinghaus and colleagues [22]. Even more distant interactions (20–40 Å) between proteins and water molecules was predicted in a computational study of the cross-correlation between solute and water dipoles [23].

Here I adopt a completely different approach. I describe a statistical survey of a carefully selected ensemble of protein crystal structures deposited in the Protein Data Bank [24,25]. The hypothesis underlining the present study is that water molecules that are observed in the electron density maps cannot be considered liquid water, since the positions of the water molecules in the liquid state cannot be conserved in time and/or space. Waters detected in electron density maps are therefore part of the hydration layer that covers the protein surface.

It is interesting to remember here that crystallographic experiments allow the determination of atomic positions that are sufficiently occupied on the time scale of the experiment (the time necessary to perform the experiments) and on the spatial scale of the crystal (the entire crystal and not only a small fraction of unit cells). The electron density that can be interpreted as a water molecule indicates a position that is occupied by a water molecule, not necessary the same during the experiment, since the water might rapidly exchange with other water molecules. However, this position is occupied by a water molecule most of the time and in most of the unit cells; this position is not occupied by water in the liquid state. Liquid water, which is abundant in protein crystals, occupies large channels that cross the entire crystal, since the protein crystal packing is sub-optimal. In other words, in protein crystals, there are two types of water molecules, those that occupy well defined positions and are not in the liquid state and those that are really liquid and do not occupy well defined positions.

During the last few years, crystallographic data have been losing some of their importance in the biophysical and biochemical community, since it was increasingly thought that they are just snapshots of a flexible structure, which evolves in time and space. If the flexibility of a protein cannot be denied or neglected – just think at the access of water molecules to internal cavities of globular proteins – it must also be considered that a snapshot is not an unreal structure. On the contrary, it is a real structure, an experimental results, which, like any other experimental result has limitations and approximations (and perhaps also systematic errors), but it is a real picture of a real phenomenon, the full description of which needs a series of snapshots. As a consequence, a collection of snapshots is an effective picture of the reality, like it has been pointed out long ago, at the dawn of structure correlation studies [26,27].

One might criticize this approach since the snapshots collected in structural databases may be uncorrelated to each other. This is like to say that a collection of random frames from several movies cannot reproduce a new, realistic (and interesting) movie. But reality is not a movie, with well-defined plot and characters, it is just a random assembly of snapshots and the random collection of structural data from databases is thus an effective strategy to sample the reality. Perhaps it is can also be argued that, whenever possible, it is preferable to use experimental data deposited in databases than to create virtual snapshots with computational techniques that may oversimplify the molecules and their interactions.

Based on the data deposited in the Protein Data Bank, this article shows that two hydration layers are commonly observed. The first hydration layer of water molecules, which interact directly with protein atoms, is larger than the second hydration layer, the water molecules of which interact with water molecules of the first hydration layers and not with protein atoms. Upper layers (third, fourth, etc.) are not observed in the dataset of protein crystal structures examined here.

Additionally, I observed that water molecules of both hydration layers are not surrounded by a tetrahedral arrangement of atoms, as it can be expected on the basis of the electronic structure of water. On the contrary, usually there are less than four atoms around water molecules and even when there are four atoms, the stereochemistry is, on average, severely distorted.

2. Results and discussion

2.1. Numbers of observations

In the 262 structures examined here, there are 88,450 water molecules. 69.1% of them are in the first hydration layer, 21,7% in the second hydration layer, 4.2% were buried in the protein core, and 5% were not classified (they were not in contact with the protein or with a first layer water; a partial, visual inspection of a subset of these data suggested that these waters are in contact with a hetero-group, like a cofactor or an inorganic cation/anion, that was in contact with the protein). Only one water was observed in the third hydration layer, which was, as a consequence, disregarded in the present communication.

More than 90% of the waters detected in protein X-ray crystallography are therefore at the surface of the proteins and a large fraction is in direct contact with protein atoms. Fig. 3 shows that the number of waters in both the first and the second hydration layer increases if the number of residues present in the asymmetric unit increases. This relationship is nearly linear for the first hydration layer (Pearson correlation coefficient close to one), while the linearity is weaker for the second hydration layer (Pearson correlation coefficient close to 0.5). The gradient is much higher for the waters of the first hydration layer. Therefore, their number increases faster when the residue number increases. In other words, there is a systematic prevalence of the first hydration layer on the second.

Both the first and the second hydration layers are not continuous on the protein surface. This obviously depend on the crystal packing interactions. When two symmetry related protein molecules are