International Journal of Biological Macromolecules 89 (2016) 137–143

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

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

# When proteins are completely hydrated in crystals

Oliviero Carugo*

*Department of Chemistry, University of Pavia, Pavia, Italy and Department of Structural and Computational Biology, Vienna University, Vienna, Austria*

ARTICLE INFO

*Article history:*
Received 22 January 2016
Received in revised form 15 March 2016
Accepted 21 April 2016
Available online 22 April 2016

*Keywords:*
Crsytallographic resolution
Crystallography
Hydration
Protein data bank
Protein structure
Solvent
Water

ABSTRACT

In the crystalline state, protein surface patches that do not form crystal packing contacts are exposed to the solvent and one or more layers of hydration water molecules can be observed. It is well known that these water molecules cannot be observed at very low resolution, when the scarcity of experimental information precludes the observation of several parts of the protein molecule, like for example side-chains at the protein surface. On the contrary, more details are observable at high resolution. Here it is shown that it is necessary to reach a resolution of about 1.5–1.6 Å to observe a continuous hydration layer at the protein surface. This contrasts previous estimations, which were more tolerant and according to which a resolution of 2.5 Å was sufficient to describe at the atomic level the structure of the hydration layer. These results should prove useful in guiding a more rigorous selection of structural data to study protein hydration and in interpreting new crystal structures.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

In physiological conditions, globular proteins are surrounded by water molecules and cannot perform their chemical and biological roles in the absence of at least one hydration layer [29]. Consequently, protein hydration has been extensively studied with both experimental and computational methods [1].

Most of the information on hydration structure has been obtained with crystallographic studies. Several statistical analyses have been published, based on room temperature crystal structures [30–32,22,34,14,19]. More recently, structural analyses are routinely performed at low temperature (typically 100 K), to minimize radiation damage due to bright synchrotron sources [18,17,7]. New statistical surveys of the structures determined with cryo-crystallography have also been published. It arose that more water molecules are detectable at low temperature, though previous observations at room temperature have been generally confirmed [25,27,28,23]. Interestingly, the growth of water molecule numbers at low temperature was not detected in a previous analysis, where few low temperature crystal structures (only 33 compared to 873 room temperature structures) were compared though linear multiple regression [6].

The effects of lowering the temperature of protein crystals on the structures have been examined extensively [16,15,33,21,35]. It emerged that low temperature structures are systematically different from room temperature structures, though many differences are relatively minor. At low temperature, the crystal volume diminishes slightly, inter-atomic contacts between symmetry related molecules are more numerous, and the conformational disorder of some side chains is different. More water molecules are detected in low temperature crystal structures, especially in the second hydration sphere [13].

However, cryo-crystallography became a routine technique and nearly all new protein crystal structures deposited in the Protein Data Bank [2,3] are determined at temperature close to 100 K. Therefore, given that the positions of the hydration water molecules do not change by lowering the temperature, [24] and although these data might provide a partially biased view of protein flexibility, it is obviously necessary to mine this enormous amount of experimental information in order to detect interesting structural trends.

Several issues must be addressed in data mining the Protein Data Bank [11]. It is mandatory to define the criteria for the selection of the protein crystal structures. The latter ones must be reasonably well hydrated if one is looking for realistic statistical trends. For example, it is probable better to exclude structures refined at very low crystallographic resolution, where few water molecules are detectable in the electron density maps.

Crystallographic resolution is by far the most used quality indicator in statistical surveys of protein structures. For example, only

* Correspondence to: Department of Structural and Computational Biology, Vienna University, Campus Vienna Biocenter 5, 1030 Vienna, Austria.
*E-mail address:* oliviero.carugo@univie.ac.at

http://dx.doi.org/10.1016/j.ijbiomac.2016.04.061
0141-8130/© 2016 Elsevier B.V. All rights reserved.

crystal structures refined at resolution better than 1.5 Å have been considered in a detailed analysis of water molecules buried in the protein core [5]. A threshold of 2.0 Å was used in the analysis of buried chlorides [4]. Other threshold values have been used too and the selection of the threshold is a compromise between the need of accuracy, which improves with smaller and more stringent values, and the need of large datasets, which enlarge with higher and less severe threshold values. It must be admitted, however, that threshold values are often quite arbitrary.

When dealing with the analysis of protein hydration, it is necessary to identify the most correct value of resolution, in order to include (nearly) only crystal structures of proteins that have well hydrated surfaces and to exclude proteins that have an incomplete hydration layer. This problem is addressed in the present communication by computing several figures of merit, which monitor the hydration degree, at various resolution levels.

## 2. Methods

All protein crystal structures were downloaded from the Protein Data Bank [2,3] according to the following criteria: (i) entries with nucleic acids were discarded, (ii) only crystal structures determined at low temperature (80–120 K) were considered; (ii) only monomers with one molecule per asymmetric unit were retained. I imposed the restriction on the data collection temperature since the hydration degree may depend on temperature [13]. I analyzed only monomers, since the reduction of sequence redundancy is obviously simpler and more effective, if oligomeric assemblies are discarded. Sequence redundancy was limited to pairwise 30% sequence identity.

I removed incomplete structures, lacking atoms and residues or with zero occupancy atoms, since these residues and atoms are very often at the surface of the protein and are therefore crucial for the detection of hydration water molecules. For similar reasons, I discarded structures where protein atoms had B-factors larger than 70 Å or larger than five times the average B-factor of the protein atoms. These thresholds on the B-factors are necessary, since the hydration of structures with large flexible moieties is unreliable. In particular, an atom with a B-factor of 70 Å$^2$ has a mean displacement amplitude from its average position larger than 0.94 Å, which is closer to the shortest covalent bond distance (C—H). I also rejected proteins shorter than 50 amino acids.

Structures were classified into 22 groups according to their resolution. The first group included only structures refined at a resolution equal to or better than 1.0 Å The last group was formed only by structures refined at a resolution lower than 3.0 Å. The other groups comprised structures in 0.1 large resolution ranges ranging from 1.0 to 3.0 Å. The largest group, 129 structures, was in the 1.70–1.80 Å resolution range. The smallest group, 3 structures, included structures refined at a resolution lower than 3.0 Å. The average dimension of the groups was 51 (standard error = 9).

Residues involved in crystal packing contacts were identified with the program CPC [10] and solvent accessible surface areas were computed with NACCESS [20]. A protein atom located at the protein surface and not involved in crystal packing contacts was considered to be close to a hydration water molecule if their interatomic distance was minor than 4.5 Å. A residue was considered to be hydrated if at least one of its atoms was close to at least one water molecule.

I computed the propensities of the solvent accessible residues to be close to water molecules according to the following equation:

$$propensity = \frac{n_{r,w}/n_w}{n_r/n} \quad (1)$$

where $n_{r,w}$ is the number of residues of type "r" that are close to water molecules, $n_w$ is the total number of residues that are close to water molecules, $n_r$ is the total number of residues of type "r", and $n$ is the total number of residues. Consequently, an amino acid with a propensity values larger than 1 tend to be close to a water molecules, while an amino acid with a propensity lower than 1 does not tend to be close to a water molecule.

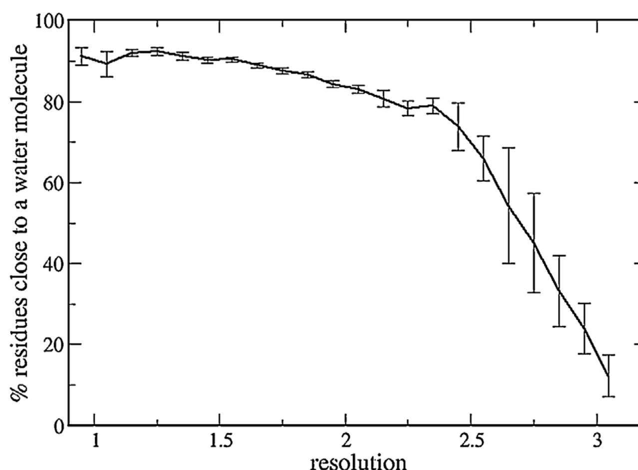


**Fig. 1.** Relationship between resolution and percentage of residues that are solvent accessible and close to a water molecule. Vertical bars indicate estimated standard errors.

## 3. Results

First, it is mandatory to identify which are the residues that are really exposed to the solvent in the crystal structure. Some of the residues that are at the protein surface are not exposed to the solvent, since they are at the interface between two proteins related by a symmetry operation. In other words, they are buried in crystal packing contacts. I used the program CPC, which has been used in other studies [8,9,36], to identify residues that are at the surface of the proteins and that are buried in crystal packing contacts and that are, therefore, not accessible to the solvent. I considered accessible to the solvent all the other residues that are at the surface of the protein and are not involved in crystal packing interactions.

Then, I computed for each protein crystal structure the percentage of amino acids that are really solvent accessible and close to a water molecule. The average percentages for all residues are shown in Table 1 and plotted in Fig. 1, for various resolution ranges. Fig. 2 shows the average percentages for each amino acid type.

Table 1 and Fig. 1 show that at very high resolution, most of the residues are close to water molecules. If resolution decreases, the percentage of residues close to water molecules is nearly constant, slightly above 90%, until resolution of 1.5–1.6 Å. Then it decreases slowly up to 2.3–2.4 Å resolution and it fall steeply at lower resolution.

The points of Fig. 1 can be fitted very well (Pearson correlation coefficient = 0.996) by a sigmoid function:

$$y = -281.0 + \frac{90.4 + 281.0}{1 + \left(\frac{x}{3.6}\right)^{7.15}}. \quad (2)$$

This indicates that the maximum percentage of amino acids that are really solvent accessible and are close to a water molecule is equal to 90.4 and that this percentage decreases very little when resolution diminishes to 1.5 Å (89.7). The same percentage diminishes considerably when resolution decreases to 2.4 Å (71.1), falls steeply at worse resolution and reaches a value of 0 at resolution larger than 3 Å. It must be observed that this function cannot be used to predict the percentage of amino acids that are really solvent accessible and are close to a water molecule outside the resolution

Download English Version:

https://daneshyari.com/en/article/1985714

Download Persian Version:

https://daneshyari.com/article/1985714

Daneshyari.com