Chemical Physics 419 (2013) 59–64

Contents lists available at SciVerse ScienceDirect

Chemical Physics

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

Application of empirical hydration distribution functions around polar atoms for assessing hydration structures of proteins

Daisuke Matsuoka [a], Masayoshi Nakasako [a,b,*]

[a] Department of Physics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan
[b] RIKEN Harima Institute, 1-1-1 Kouto, Mikaduki, Sayo, Hyogo, Japan

ARTICLE INFO

Article history:
Available online 9 January 2013

Keywords:
Hydration structure
Knowledge-based method
Protein hydration
X-ray crystallography

ABSTRACT

To quantitatively characterize hydrogen-bond geometry in local hydration structures of proteins, we constructed a set of empirical hydration distribution functions (EHDFs) around polar protein atoms in the main and side chains of 11 types of hydrophilic amino acids (D. Matsuoka, M. Nakasako, Journal of Physical Chemistry B 113 (2009) 11274). The functions are the ensemble average of possible hydration patterns around the polar atoms, and describe the anisotropic deviations from ideal hydrogen bond geometry. In addition, we defined probability distribution function of hydration water molecules (PDFH) over the hydrophilic surface of a protein as the sum of EHDFs of solvent accessible polar protein atoms. The functions envelop most of hydration sites identified in crystal structures of proteins (D. Matsuoka, M. Nakasako, Journal of Physical Chemistry B 114 (2010) 4652). Here we propose the application of EHDFs and PDFHs for assessing crystallographically identified hydration structures of proteins. First, hydration water molecules are classified with respect to the geometry in hydrogen bonds in referring EHDFs. Difference Fourier electron density map weighted by PDFH of protein is proposed to identify easily density peaks as candidates of hydration water molecules. A computer program implementing those ideas was developed and used for assessing hydration structures of proteins.

1. Introduction

X-ray crystallography has contributed to revealing the hydration structures of proteins at atomic resolution [1]. The distribution and interaction geometry of hydration water molecules facilitate the understanding of how proteins fold and function in aqueous environment. Based on the structural information related to protein hydration, water molecules are argued to contribute significantly to the stability and dynamics of proteins, and to act as mediators for proton transfer in enzymatic reactions and proton pumping through their network of hydrogen bonds (H-bonds) [2,3].

In X-ray crystal structure analyses of proteins, water molecules with small positional fluctuations appear as isolated electron densities in the vicinity of surface atoms or interior cavities of proteins [1,4–6]. We assign density peaks as hydration water molecules, taking the levels of electron density and the geometry of possible H-bonds with polar protein atoms [7–13]. In the primary stage of structure analysis, many electron density peaks resembling those of hydration water molecules appear probably because of the errors in measured diffraction amplitude, ambiguity in the phase set, limitation in resolution, and anisotropy in resolution. Thus, any standard model or theoretical prediction of local hydration structures would be helpful for assessing modeled hydration structures and density peaks in X-ray crystallography.

Although molecular dynamics simulations of proteins immersed in explicit water system [14,15] is an approach to predict possible hydration sites around crystal structures of proteins, the time-consuming procedure is not practical in the rounds of crystallographic structure refinements. The three-dimensional reference interaction site model [16] theoretically predicts hydration structures of proteins, but is still in a pre-matured stage requiring rigorous experimental examination.

To date, a number of crystal structures of proteins with hydration water molecules have been solved at high resolution. As an alternative approach to those theoretical calculations, statistical and systematic analyses on the H-bond geometry of the hydration water molecules in those crystal structures provide a database for assessing hydration structures of proteins. In this regard, we constructed empirical hydration distribution functions (EHDFs) for polar atoms (oxygen and nitrogen) in the main- and the side-chains of 11 types of hydrophilic amino acids through systematic and statistical analyses for 17,984 protein structures determined at high resolution from the Protein Data Bank [17] (Table 1) [18]. EHDFs suggest possible hydration patterns of polar atoms with quantitative descriptions of the anisotropic spreads from the ideal

* Corresponding author at: Department of Physics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan. Tel.: +81 45 566 1679; fax: +81 45 566 1672.
E-mail address: nakasako@phys.keio.ac.jp (M. Nakasako).

0301-0104/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemphys.2012.12.040

Table 1
Polar atoms and EHDFs referenced in the program CK_hydra.

Amino acid	Atom group used as a standard structure	Polar atom	Hybridization/geometry	Lone pairs/bonds with H	Number of possible H-bonds
Peptide bond	C, O, N	O	sp^2/trigonal planar	2/0	2
		N	sp^2/trigonal planar	0/1	1
Glu	CD, OE1, OE2	OE1	sp^2/trigonal planar	2/0	2
		OE2	sp^2/trigonal planar	2/0	2
Asp	CG, OD1, OD2	OD1	sp^2/trigonal planar	2/0	2
		OD2	sp^2/trigonal planar	2/0	2
Arg	NE, CZ, NH2	NE	sp^2/trigonal planar	0/1	1
		NH1	sp^2/trigonal planar	0/2	2
		NH2	sp^2/trigonal planar	0/2	2
Lys	CD, CE, NZ	NZ	sp^3/tetrahedral	0/3	3
His	ND1, CE1, NE2	ND1	sp^2/trigonal planar	0/1	1
		NE2	sp^2/trigonal planar	0/1	1
Trp	ND1, NE1, CE2	NE1	sp^2/trigonal planar	0/1	1
Ser	CA, CB, OG	OG	sp^3/tetrahedral	2/1	3
Thr	CA, CB, OG1	OG1	sp^3/tetrahedral	2/1	3
Tyr	CE2, CZ, OH	OH	sp^2/trigonal planar	1/1	2
Gln	CD, OE1, NE2	OE1	sp^2/trigonal planar	2/0	2
		NE2	sp^2/trigonal planar	0/2	2
Asn	CG, OD1, ND2	OD1	sp^2/trigonal planar	2/0	2
		ND2	sp^2/trigonal planar	0/2	2

geometry in H-bonds. In addition, the probability distribution functions of hydration water molecules (PDFHs) over hydrophilic surfaces of a protein molecule was computed by simply summing the EHDFs around solvent accessible polar atoms. PDFHs enveloped more than 90% of the crystal water sites identified in protein crystal structures [19].

This promising result of PDFHs encouraged us to apply EHDFs for assessing hydration structures of proteins visualized in crystallographic studies. One of the application is to classify hydration water molecules regarding the deviation of their H-bond geometry from the ideal. The other provides potential monolayer hydration volume over hydrophilic surfaces of protein, which assists the search for electron density peaks assignable as hydration water molecules. These knowledge-based method would be a helpful tool to analyze and understand the geometries in protein hydration.

2. Implementation

To utilize EHDFs in crystallographic structural studies, we developed the program 'CK_hydra', which was written using the FORTRAN 77 language, through the incorporation of several subroutines for hydration structure analyses of proteins coded in our previous studies [18,19]. The CK_hydra program uses the reference structure models of atom groups in the main- and side-chains of 11 hydrophilic amino acids (Table 1), and the EHDFs fixed to the reference models (Fig. 1). The program only requires the coordinate file of a target protein molecule in the PDB format.

The EHDF around each polar atom of the reference model in Table 1 is described in the polar coordinate system (r, φ, θ) centered at the atom (Fig. 1) [18], and is illustrated as contour maps projected to the r–θ, r–φ and φ–θ planes. The contour value is the ratio of the numbers of water molecules enveloped by the contour and used in the EHDF construction. For instance, the EHDF contour at level 0.6 means that the contour envelopes 420,000 hydration water molecules out of the 700,000 used in the EHDF construction (Fig. 1A).

2.1. Application of EHDF for assessing H-bond geometry

When a water molecule is within H-bond distance (2.4–3.4 Å) [18] from a polar atom in any group listed in Table 1, the molecule is transferred to the coordinate system of the reference model of the group by computing a translation vector and a rotation matrix to optimally overlap the group with the reference. Then, the position of the transferred hydration water molecule in the coordinate system is calculated for assessing the hydration geometry using EHDF. As EHDFs have large values near the ideal H-bond geometry of polar atoms and small apart from the ideal, they are useful for assessing how the H-bond geometry of a hydration water molecule is close to the ideal.

2.2. PDFH-weighted difference Fourier electron density map

In X-ray crystallography, hydration water molecules are picked up using difference Fourier F_o–F_c electron density maps ($\Delta\rho_{F_o-F_c}$) [20]. The PDFH of a protein molecule, hereafter designated f_{PDFH}, suggests volumes potentially involved in monolayer hydration in the vicinity of the hydrophilic surfaces and cavities of the protein [19]. Thus, conventional $\Delta\rho_{F_o-F_c}$ map weighted by PDFHs would contribute to distinguish whether each electron density peak is assignable as a water molecule of monolayer hydration or not. Here, we define $\Delta\rho_{F_o-F_c}$ map weighted by f_{PDFH} as

$$\Delta\rho_{\text{weighted}}(\mathbf{r}_i) = [\Delta\rho_{F_o-F_c}(\mathbf{r}_i) - \Delta\rho_{\min}] \times \mathbf{H}(\mathbf{f}_{\mathbf{PDFH}}(\mathbf{r}_i) - \mathbf{f}_{\min}) \quad (1)$$

where $\vec{r}_i$ was the position vector for the ith voxel in electron density map. $H(x)$ is the Heaviside's step function taking 1 when $x > 1$ or otherwise taking 0, and the cut-off levels of the density maps $\Delta\rho_{F_o-F_c}$ and f_{PDFH} are designated as $\Delta\rho_{\min}$ and $f_{\min}$, respectively. The voxel size of f_{PDFH} was set to that of $\Delta\rho_{F_o-F_c}$ map. The weighted density map is written in the map.dn6 format [21] to be visualized easily. The basic idea of this approach was proposed about 20 years ago in the program AQUARIUS2 with the small number of crystal structures of proteins [22].

2.3. Computation

We tested the CK_hydra program through applying to crystal structures determined at different resolution: the light-oxygen-voltage sensing domain 1 (LOV1) of phototropin2 refined at a resolution of 2.00 Å (PDB accession code 2Z6D) [23], glutamate dehydrogenase (GDH) at 1.80 Å (1EUZ) [19,24], and the Gly74Cys/Cys188Ser mutant of arylmalonate decarboxylase (AMDase) at 1.45 Å (PDB accession code 3IXL) [25] (Table 2). They were, of course, excluded from the construction of EHDFs.

Computations were carried out using a system composed of a Xeon 5160 (3.0 GHz) CPU (Intel). For each crystal structure, the