Rice University

Protein Wrapping and Protein Hydration

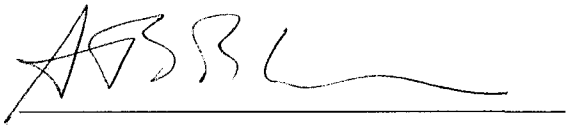
by

Jianping Chen

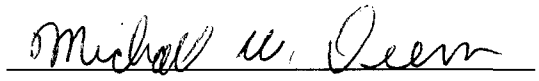
A THESIS SUBMITTED
IN PARTIAL FULFILMENT OF THE
REQUIREMENT FOR THE DEGREE

Master of Science

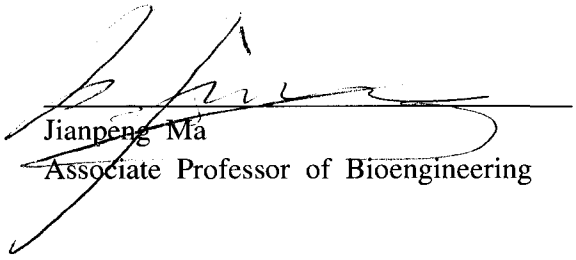
APPROVED, THESIS COMMITTEE:



Ariel Fernández
Karl F. Hasselmann Professor of
Bioengineering



Michael W. Deem
John W. Cox Professor of Biochemical
and Genetic Engineering and
Professor of Physics & Astronomy



Jianpeng Ma
Associate Professor of Bioengineering

Houston, TX
October 2007

UMI Number: 1455225

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform 1455225

Copyright 2008 by ProQuest LLC.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 E. Eisenhower Parkway
PO Box 1346
Ann Arbor, MI 48106-1346

Abstract

Protein Wrapping and Protein Hydration

by

Jianping Chen

Hydrogen bond plays an important role in stability, dynamics and function of protein. Most of backbone hydrogen bonds are well wrapped by nonpolar groups of side chains. However, there are a small portion of hydrogen bonds vulnerable to water attack. Those under-wrapped hydrogen bonds, termed as "dehydron", are sensitive to the change in the local electrostatic environment. Dehydrons constitute a hot spot for protein interactions. They have been identified as structural marker for protein association, proteamic connectivity and protein-ligand binding. The effects of dehydrons on protein hydration shell are assessed by studying the mobility of hydration water. Calculation of water residence times of all residues reveals that dehydrons enhance the water mobility and promote the most intense loosening of hydration shell. Targeting loose hydration shells induced by dehydrons provides a powerful strategy in rational drug design.

Acknowledgement

I am very glad to acknowledge those who contribute to this work. First, I would like to thank my advisor Professor Ariel Fernández for his enthusiastic encouragement and insightful guidance. I would also like to thank Dr. Crespo for great help in performing molecular dynamics and invaluable discussions. I would like to express my gratitude to my fellow graduate students Xi Zhang and Natalia Pietrosemoli for their collaboration and friendship.

I would like to thank all my friends at Rice University for their friendship and their lasting support from the beginning of graduate study.

Finally, I would like to thank all of my family for their undying love and support, that has been unconditionally given to me through all the years.

Table of Contents

Abstract ii
Acknowledgement iii
List of Figures vi
List of Tables vii

1 **Introduction** **1**
1.1 Dehydrons 2
1.2 Hydration water 3

2 **Dehydrons as structural marker for protein interactions** **6**
2.1 Identification of dehydrons 6
2.1.1 The statistical approach 6
2.1.2 Theoretical method 11
2.2 Role of dehydrons in protein interactions 17
2.2.1 Dehydrons as determinants of protein association 17
2.2.2 Dehydron as an indicator of protein interactivity 17
2.2.3 Dehydron patterns of different species 19
2.2.4 Dehydrons as specific targets in drug design 21

3 **Residence times of water molecules in hydration shell** **23**
3.1 Structural and dynamic properties of the hydration water 23
3.2 Simulation study of water residence time in azurin 25
3.2.1 Simulation details 25
3.2.2 Calculation of water mean residence time 26
3.2.3 Results 27
3.3 Water residence time of Myoglobin 30
3.3.1 Molecular dynamics simulation 30
3.3.2 Calculation of water residence times 31
3.3.3 Results 32

4 **Dehydron loosens the hydration shell** **36**
4.1 Methods 36
4.1.1 Molecular simulation 36
4.1.2 Calculation of local residence time 36
4.1.3 Identification of dehydrons 37
4.1.4 Calculation of the dewetting field 38
4.2 Results 39
4.2.1 Water residence times of the SH3 domain 39
4.2.2 Dehydrons and dewetting fields of the SH3 domain 41
4.2.3 Water residence times of ubiquitin 42

5 **Conclusion and implication** **44**

References 46

List of Figures

1.1 Illustration of targetting the nonconserved feature in c-Kit kinase 5

2.1 Desolvation domain of a hydrogen bond 7

2.2 Illustration of the most effective protective motif for an hydrogen bond 9

2.3 Wrapping pattern of the hemoglobin protein and cellular prion protein. 11

2.4 Sensitivity of hydrogen bonds in myoglobin and prion to water exclusion 16

2.5 Relationship between protein interaction and the number of dehydrons. 18

2.6 A comparison of dehydron patterns between different species 20

2.7 HIV-1 Protease bound by the Merck inhibitor Indinavir 22

3.1 Distribution of water residence times around azuri 27

3.2 Residence times of secondary structure elements 28

3.3 Distribution of water residence times of seven atom types as a function of the square root of their average solvent accessible surface area 30

3.4 Distribution of hydration sites around myoglobin. 33

3.5 Distribution of four parameters in the bi-exponential model. 34

3.6 Distribution of hydration sites as a function of their residence times, occupancy and w parameters. 35

4.1 Mean residence times of all residues in SH3 domain 40

4.2 Virtual bond representation of SH3 domain 41

4.3 Dewetting fields of residues paired by hydrogen bonds in SH3 domain 42

4.4 Residence time and dehydron pattern of uniquitin 43

List of Tables

2.1 Wrapping parameters for a list of proteins identified by their PDB entries..... 8

2.2 The extent of hydrogen-bond wrapping of some proteins in PDB. 10

2.3 Density of dehydrons at the protein interfaces. 18

3.1 Residence times of seven atom groups. 29

Chapter 1

Introduction

As a polypeptide folds into a three-dimensional structure, backbone polar groups form a large number of hydrogen bonds. During folding, backbone hydrogen bonds are properly placed in the center of hydrophobic cluster. The low solvent accessibility of hydrogen bonds results in a high bonding energy which makes a significant contribution to protein stability. However, protein does not provide great protection for all hydrogen bonds. The "under-wrapped" hydrogen bonds are exposed to the solvent, and are unstable. Those packing defects are also important for protein structure and function. They serve as a determinant for protein interactions.

Proteins fold and function in the aqueous environment. Hydration plays an important role in protein structure. On the other hand, protein also exerts its influences on the solvent especially on the water molecules on its surface. The geometric constraint imposed on the hydration shell results in a higher density than bulk water. The influence of the protein on the water molecules on the hydration shell is not even, but varies from site to site. Hydration shells containing water molecules with high mobility and short residence are very important region for protein-ligand binding. Upon approach these hydration shells by the hydrophobic group from ligand, water molecules are excluded, which contributes to the binding energy of protein-ligand complex. Since those loose hydration shells are not conserved across

proteins, they can be used as a highly specified target for drug design.

1.1 Dehydrons

Hydrogen bond is the dipole-dipole bond formed between a hydrogen atom chemically bonded to an electronegative atom and another electronegative atom, as exemplified by the O-H:::O, N-H:::N bonds. The hydrogen atom is called the donor of the hydrogen bond, and the electronegative atom O or N is called the acceptor of the hydrogen bond. Compared to other chemical bonds such as covalent and ionic bonds, the typical hydrogen bond is much weaker. For example, the energy of hydrogen bonds between water molecules is about 23.3 kJ/mol [1] about twenty times smaller than the energy of the covalent bond in a water molecule (~492.2 kJ/mol [2]).

Despite its weak strength, hydrogen bond plays an important role in biological system. Water is essential for life. The role of water in life is attributed to its anomalous properties [3], many of which are due to the hydrogen bonding between water molecules. Besides, the hydrogen bonds are very important for the structure and function of biomolecules such as proteins. In 1950s Pauling proposed the models of alpha helix and beta sheet for proteins [4, 5], in which hydrogen bonding was attributed to the critical role. Those models were soon confirmed by the X-ray studies of proteins [6], bolstering the assumption that hydrogen bonds are crucial in protein folding and conformation stability. The detailed illustration of the role of hydrogen bonding in biological system can be seen in Pimentel & Saenger's book[7].

In proteins, most of the backbone hydrogen bonds are well protected from water attack by hydrophobic groups of side chains. These kinds of hydrogen bonds have a high bonding energy and hence are essential for stabilizing the protein conformation. Both statistical and theoretical approaches have been employed to quantify the extent to which the hydrogen bonds are protected by proteins [8 - 10]. The results showed that about 92% of backbone hydrogen bonds were well protected and about 8% of hydrogen bonds are vulnerable to water attack [10]. Those under-wrapped hydrogen bonds are proposed to be central to protein-protein interaction [8-10]. They are termed as "dehydron." When a hydrophobic group approaches to a dehydron, water molecules around the dehydron are excluded and the dehydron turn into a well wrapped hydrogen bond which has a high bonding energy. The net energy gain in this process has been experimentally determined to be close to 4kJ/mol [11], which is significant and comparable to the strength of hydrophobic interaction [12].

1.2 Hydration water

Water molecules interacting with protein may be classified as internal water and water molecules on the protein surface. Internal waters occupy cavities within the protein. They are strongly bound to amino acids and can be regarded as an integral part of the protein. Surface water also called hydration water is loosely "bound" to solvent exposed atoms of the protein and exchanges frequently with bulk water. The residence time of water molecules in hydration shell varies from several picoseconds

to hundreds of picoseconds. Water exclusion from loose hydration shell constitutes an important component in protein interactions.

Protein folds are conserved across distant homologous sequences. Proteins in the kinase family good targets for cancer therapy, for example, possess highly conserved intracellular catalytic domain [23]. The highly conservation of protein folds brings great difficulty to drug design. Side effects always arise when a drug designed to hit an intended target binds to the other proteins. The structure and dynamics of hydration water depend on the protein surface topology as well as the amino acids with which water molecules interact. Protein hydration varies for proteins in the same family. Therefore, targeting loose hydration shell, a non-conserved binding site, provides an effect way of obtaining drug selectivity. Imatinib (Gleevec, STI571) is a very successful drug to treat gastrointestinal stromal tumors (GIST) by targeting c-Kit kinase [24]. However, it still binds to other kinases such as Bcr-Abl [25]. By selectively targeting the nonconserved loose hydration shell in c-Kit (see Fig 1.1), the specificity of Gleevec have been greatly improved [26].

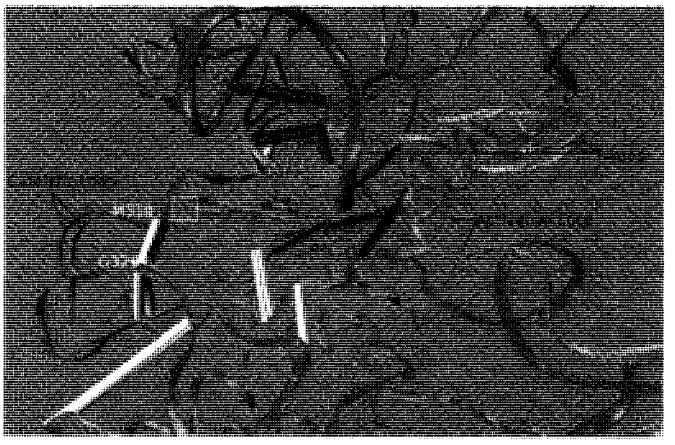


Fig. 1.1 Illustration of targetting the nonconserved feature in c-Kit kinase. A methyl group was added on the yellow rectangle to exclude the water molecules in the loose hydration shell in the vicinity of hydrogen bond C673-G676 (shown as green segment) in c-Kit (shown as cyan ribbon). The corresponding region in Abl (shown as magenta ribbon), in the vicinity of hydrogen bond M318-G321 (shown as light grey segment), does not exhibit the same binding feature [26].

Chapter 2

Dehydrons as structural marker for protein interactions

Dehydrons are identified as packing defects in protein structure. They are highly sensitive to the water removal, and are important in protein association. Dehydrons are shown to be involved in a variety of biological contexts, including protein aggregation and rearrangement, molecular evolution, and protein-ligand interaction [8-10].

2.1 Identification of dehydrons

2.1.1 The statistical approach

The majority of backbone hydrogen bonds are well protected by the surrounding nonpolar groups from water attack. The level of wrapping could be quantified by counting the number of hydrophobic groups in the dehydration domain of the hydrogen bond. The dehydration domain is defined as two spheres of radius R = 7.2Å centered at α carbon atoms of two residues paired by the hydrogen bond (see Fig. 2.1). The value of R is adopted by relating it to the characteristic length Λ of the solvent-structuring effect due to the presence of a vicinal hydrophobe [13]. By

fixing Λ at 1.8 Å (the effective thickness of a single-layer water cavity) and assuming structuring influence decays exponentially, R is set to be 4Λ = 7.2 Å to reduce the structuring influence to 1% of its maximum value.

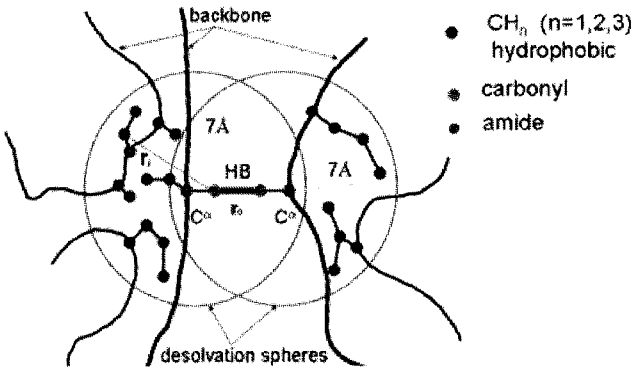


Fig 2.1: Desolvation domain of a hydrogen bond [10].

The extent of wrapping of the hydrogen bonds, ρ, averaged over all backbone hydrogen bonds of a given protein chain is then given as:

$$\rho = C_3/Q \tag{2.1}$$

where C_3 is the total number of three-body correlations (i.e. the number of hydrophobic residues whose α carbon atoms are contained in the dehydration domain) and Q is the number of backbone hydrogen bonds. It is found that 95.66% of the 2918 proteins from RCSB Protein Data Bank (PDB) have almost the same ρ values $\rho = 5.00 \pm 0.38$ (see Table 2.1). The dispersion σ in the extent of protection within a protein is always smaller than 18% of the average ρ.

The nearly constant ρ value reflects a generic feature of protein folds. On the other hand, the dispersion σ arises from the variations in the size of the side chains. Large hydrophobic residues provide better protection for hydrogen bonds than small ones do. The number of protectors for a backbone hydrogen bond ranges from two to

PDB code	C_3	Q	ρ	σ (%)
1aa2	257	52	5.04	10.18
1lou	242	47	5.15	13.05
1ris	230	45	5.11	12.87
1alf	20	4	5.00	10.02
1aue	250	49	5.10	11.80
256b	394	75	5.25	16.05
1aac	170	33	5.15	11.28
1abq	80	15	5.33	14.06
1aoj	39	8	4.87	15.47
1bo3	5	1	5.00	0.00
1ubi	155	31	5.00	10.06
1gb4	80	16	5.00	10.14
1lmb	260	51	5.00	8.67
1srl	40	8	5.00	12.83
2ptl	74	16	4.62	16.33
1crc	136	28	4.85	9.60
1cw6	32	6	5.33	14.02
1vii	30	6	5.00	12.55
1hhh	446	86	5.18	7.68
1mim	318	64	4.96	17.62
1ifb	215	43	5.00	8.83
1hhg	468	95	4.92	11.09
1e4j	225	45	5.00	14.11
1e4k	233	46	5.07	11.15
1gff-1	612	124	4.93	11.58
1csk-A	111	22	5.04	12.01
1c3t	105	21	5.00	10.78
1a6v	172	33	5.21	17.91

Table 2.1: Number of three-body correlations (C3), amide–carbonyl hydrogen bonds (Q), average extent of hydrogen bond protection (ρ) and dispersion in the extent of protection (σ) for a list of proteins identified by their PDB entries [13].

eight (in the case of eight protectors, most of them are alanines). Table 2.1 indicates that the protection of one hydrogen bond provided by five hydrophobic residues exerts a constraint on the packing of stable soluble proteins. The arrangement of hydrophobic residues represents the most effective motif of the protection for hydrogen bonds. As illustrated in Fig. 2.2.a, five hydrophobic residues are placed in a trigonal bipyramid, which shapes the interior for a short and almost collinear amide-carbonyl hydrogen bond. Fig. 2.2.b displays an example of the optimal protective pattern on native hydrogen bonds. The strong Met1-Val17 hydrogen bond of *ubiquitin* (PDB. 1ubi) is protected by five hydrophobic residues: Met1, Ile3, Leu15, Val17, and Pro19, which are arranged in almost the same way as in Fig. 2.2.a.

A more precise approach to quantify the extent of protection of hydrogen bond consists of counting the number of carbonaceous group (CH_i, i = 1, 2, 3) within the desolvation spheres of the hydrogen bonds. It gives $\rho = 15.00 \pm 2.05$ for the 96% of

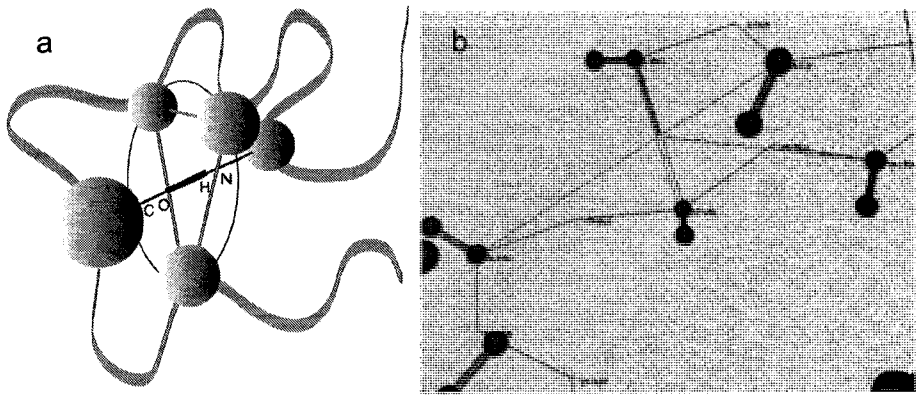


Fig 2.2: (a) Illustration of the most effective protective motif (n=5) for an hydrogen bond. Five balls represent five hydrophobic residues providing protection for the hydrogen bond; (b) Protective clusters for a strong Met1-Val17 hydrogen bond in *ubiquitin* (PDB. 1ubi). Grey balls represent the α and β carbon atom of hydrophobic residues, thick two-color stick connecting two α carbon atoms represents the hydrogen bond (red for carbonyl group, white for amide group), thin blue lines indicate protection provided by hydrophobic residues [reproduced with permission from ref. 8].

the proteins from PDB, with a maximum Gaussian dispersion σ = 3.30 (22%) [9]. Dehydrons are then defined by combining the lowest representative ρ value (12.95) with the maximum dispersion 3.30. Those hydrogen bonds with 9 or less hydrophobic groups in their dehydration domain are categorized as dehydrons. This definition results a low probability of picking dehydrons (probability < 4%) for a protein chosen at random.

Table 2.2 lists a group of well wrapped proteins with $\rho = 15.00 \pm 2.05$ and a group of outliers. Among all the soluble protein examined, only cellular prion proteins and toxins have their ρ smaller than 15.00 – 2.05. In toxins, the conformation stability is determined by disulfide bonds. The most under-wrapped protein (ρ = 7.43) is the antiacetylcho -linesterase toxin from *green mamba* venom (PDB.1fas).

A comparison of the wrapping pattern of human prion protein (PDB. 1qm0) with that of hemoglobin (Hb) β-subunit (PDB. 1bz0, chain B) is displayed in Fig. 2.3. The

PDB ID code	C_3	Q	ρ	σ, %
1aa2	771	52	14.83	10.18
1bz0 (α)	1,450	95	15.26	12.03
1bz0 (β)	1,472	98	14.87	12.00
1lou	726	47	15.44	13.05
1ris	690	45	15.33	12.87
1aue	750	49	15.30	11.80
256b	1,182	75	15.76	16.05
1ubi	465	31	15.00	10.06
1gb4	240	16	15.00	10.14
1srl	120	8	15.00	12.83
2ptl	222	16	13.87	16.33
1crc	408	28	14.57	9.60
1hhh	1,338	86	15.56	12.68
1mlm	954	64	14.90	17.62
1lfb	645	45	14.33	8.83
1hhg	1,404	95	14.77	11.09
1e4j	675	44	15.34	12.11
1e4k	699	46	15.20	11.15
1gff-1	1,836	124	14.81	11.58
1csk-A	333	23	14.47	12.01
1c3t	315	21	15.00	10.78
1fas	171	23	7.43	17.08
1f94	261	25	10.44	22.80
1jwl	300	25	12.00	23.51
1dxo	645	59	10.93	21.8
1dwz	648	60	10.80	24.2
1b10	699	58	12.05	21.3
1qlx	684	58	11.79	19.6
1qm0	648	57	11.37	20.2
1qm1	639	56	11.25	21.4

Table 2.2: The extent of hydrogen-bond wrapping of some proteins in PDB. The upper 21 rows represent good hydrogen-bond wrappers, the middle 3 rows represent toxins, and the last 6 represent cellular prion proteins. The PDB entries of proteins are given in the first column, and the other four columns give the four quantities of the wrapping: the total number of three-body correlations (C_3), the total number of backbone hydrogen bonds (Q), the average wrapping level $\rho = C_3/Q$, and the dispersion σ of the wrapping levels over all the backbone hydrogen bonds in a protein. [9]

hemoglobin protein is a good wrapper, whose β-subunit consists of 96 well wrapped hydrogen bonds (grey bonds in Fig. 2.3.*b*) and 3 dehydrons (green bonds in Fig. 2.3.*b*). To the contrary, the prion protein has 28 dehydrons and 30 well wrapped hydrogen bonds. The high proportion of dehydrons renders the structure vulnerable to water attack and prone to rearrangement. Helix 1 (residues 143-156) and helix 3 (residues 199-228) are two regions rich in dehydrons (see Fig. 2.3.*d*): the former has been singled out as probable site for rearrangement [14], the latter is assumed to define the epitope for protein-X binding [15].

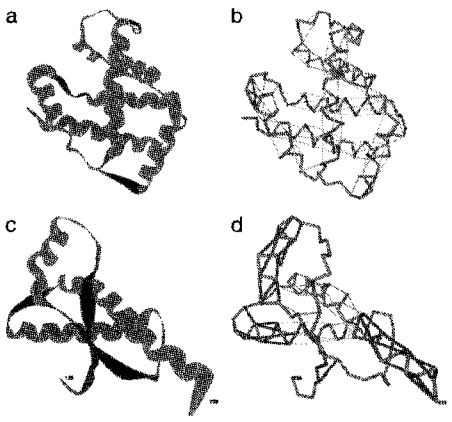


Fig 2.3: Wrapping pattern of the hemoglobin protein (PDB. 1qm0) and cellular prion protein (PDB. 1bz0). a and c display the ribbon structure of two proteins. In b and d, the backbone is represented as virtual bonds joining consecutive α carbon atoms (shown as grey thick lines). The light-grey thin lines represent the well wrapped hydrogen bonds, and the green lines represent the under-wrapped hydrogen bonds [9].

2.1.2 Theoretical method

The more accurate way of defining dehydrons is based on the extent of change in Coulomb energy of the hydrogen bonds with the presence of a new hydrophobe from a binding partner. To begin, a Cartesian coordinate system was built to calculate the effective permittivity ε for the Coulombic interaction between the donor and acceptor of hydrogen bond: the carbonyl oxygen atom with effective charge q is placed at the center of coordinates; the carbonyl-amide hydrogen bond is along the x-axis and the amide hydrogen atom is placed at the position $\mathbf{r}$ (r = 1.4-2.1 Å) away from the oxygen atom; a discrete number of identical spherical hydrophobic units of radius $d/2$ (d is discussed below) are arranged around the hydrogen bond with their centers at fixed positions $\mathbf{r}_j$, j = 1, 2, …, n.

The evaluation the electric field $E(\mathbf{r})$ at position $\mathbf{r}$ involves determination of the solvent structuring effect induced by the solvent-hydrophobe interface in $\mathbf{k}$-space and an inverse Fourier transformation. To begin, the Poisson equation $\mathrm{div}[\varepsilon(\mathbf{r})E(\mathbf{r})] = 4\pi q\delta(\mathbf{r})$ is replaced by the equation:

$$\mathrm{div}[\int K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\})E(\mathbf{r}')d\mathbf{r}'] = 4\pi q\delta(\mathbf{r}) \tag{2.1}$$

where the kernel $K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\})$ represents the correlations of the electric field $E(\mathbf{r})$ at position $\mathbf{r}$ with the field $E(\mathbf{r}')$ at position $\mathbf{r}'$ and depends parametrically on the fixed hydrophobe positions. In the absence of hydrophobic units, $K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\}) = K(\mathbf{r}, \mathbf{r}')$ which decays as $\exp(-\|\mathbf{r}-\mathbf{r}'\|/\xi)$ (ξ is the characteristic correlation length). In the limit $\xi \rightarrow 0$, $K(\mathbf{r},\mathbf{r}') \sim \delta(\mathbf{r}'-\mathbf{r})$, and equation (1) becomes the standard Poisson equation.

By taking into account the relationship between diffraction and dielectric, $K(\mathbf{r},\mathbf{r}') = K(\mathbf{r}-\mathbf{r}')$ is identified as:

$$K(\mathbf{r}-\mathbf{r}') = \int \exp[i\mathbf{k}\bullet(\mathbf{r}-\mathbf{r}')]L(\mathbf{k})d\mathbf{k} \tag{2.2}$$

where the distribution $L(\mathbf{k}) = [(\varepsilon_w - \varepsilon_0)/(1 + \varepsilon_w\|\mathbf{k}\|^2\xi^2/\varepsilon_0) + \varepsilon_0]$ reflects that low frequency radiation has a dominant influence in water polarizability. Here ξ is fixed at 5 Å, and ε_w, ε_0 are the permittivities of water and vacuum, respectively.

By introducing a characteristic length Λ to assess the water-structuring effect, we obtain the correlation kernel with n hydrophobic units at fixed position:

$$K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\}) = [\int \exp[i\mathbf{k}\bullet(\mathbf{r}-\mathbf{r}')]L(\mathbf{k})d\mathbf{k}] * [1+\sum_{j=1,2,\ldots,n}\Gamma_j(\mathbf{r}, \mathbf{r}')] \tag{2.3}$$

where $\Gamma_j(\mathbf{r}, \mathbf{r}') \sim \exp[-(\|\mathbf{r} - \mathbf{r}_j\|+\|\mathbf{r}' - \mathbf{r}_j\|)/\Lambda]$ for $\|\mathbf{r} - \mathbf{r}_j\|$ and $\|\mathbf{r}' - \mathbf{r}_j\| > d/2$. Λ is fixed at 1.8 Å, the effective thickness of a single-layer cavity.

Then Fourier transformation of $K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\})$ results in $K(\mathbf{k}, \{\mathbf{k}_j\})$:

$$K(\mathbf{k}, \{\mathbf{k}_j\}) = \int \exp(-i\mathbf{k}\bullet\mathbf{r}) \int \exp(-i\mathbf{k}\bullet\mathbf{r}') K(\mathbf{r}, \mathbf{r}', \{\mathbf{r}_j\}) d\mathbf{r}' d\mathbf{r}$$

$$= L(\mathbf{k}) + L(\mathbf{k}) \otimes \{ \sum_{j=1,\ldots,n} \exp(-i\mathbf{k}\bullet\mathbf{r}_j)[1/(1+\|\mathbf{k}-\mathbf{k}_j\|^2\Lambda^2)] \} \qquad (2.4)$$

where the symbol $\otimes$ stands for convolution.

Now the electric field can be obtained by performing an inverse Fourier transformation on $[K(\mathbf{k}, \{\mathbf{k}_j\})]^{-1}$:

$$\int E(\mathbf{r}) d\mathbf{r} = -4\pi q \int \exp(-i\mathbf{k}\bullet\mathbf{r}) \|\mathbf{k}\|^{-2} [K(\mathbf{k}, \{\mathbf{k}_j\})]^{-1} d\mathbf{k} \qquad (2.5)$$

After retaining the real part in the residue calculation at the first-order poles $\mathbf{k} = \pm i(\varepsilon_0/\varepsilon_w)^{1/2}\xi^{-1}$ ($k=\|\mathbf{k}\|$) and $\mathbf{k}=\mathbf{k}_j \pm i\Lambda^{-1}$, $E(\mathbf{r})$ is given by:

$$E(\mathbf{r}) = (q/r^2)[(\varepsilon_0^{-1} - \varepsilon_w^{-1})\Omega(\{\mathbf{r}_j\})(1+r/\xi)\exp(-r/\xi) + \varepsilon_w^{-1}] \qquad (2.6)$$

where

$$\Omega(\{\mathbf{r}_j\}) = \Pi_{j=1,\ldots,n}[1+\exp(-\|\mathbf{r}_j\|/\Lambda)][1+\exp(-\|\mathbf{r}-\mathbf{r}_j\|/\Lambda)] \qquad (2.7)$$

and $r = \|\mathbf{r}\|$. An effective permittivity can be defined as:

$$\varepsilon = [(\varepsilon_0^{-1} - \varepsilon_w^{-1})\Omega(\{\mathbf{r}_j\})(1+r/\xi)\exp(-r/\xi) + \varepsilon_w^{-1}]^{-1} \qquad (2.8)$$

For a fixed $\mathbf{r}$, the effective permittivity ε depends only on the function $\Omega(\{\mathbf{r}_j\})$ contributed by the hydrophobic residues. The maximization of $\Omega(\{\mathbf{r}_j\})$ leads to the lowest ε. To calculate the maximum value Ω^* for $\Omega(\{\mathbf{r}_j\})$, the maximum $\Omega^*(\{\boldsymbol{n}\})$ of $\Omega(\{\mathbf{r}_j\})$ for each fixed n is computed. The minimum distance d between any two hydrophobes is preserved. The value of d is taken to be 5 or 6 Å according to the minimal distance between α carbon atoms of two nonadjacent residues in standard secondary or tertiary structure motifs. Since no hydrogen bond is protected by less than 3 hydrophobes, the computation is carried out for $n \geq 3$. The result gives the maximum value Ω^* at $n = 5$:

$$\Omega^*(3) < \Omega^*(4) < \Omega^*(5) > \Omega^*(6) > \Omega^*(7) > \bullet\bullet\bullet \qquad (2.9)$$

This theoretical result confirms the above statistical finding which shows that 5 hydrophobes can provide the most effective protection for a hydrogen bond. The maximization of $\Omega^*(\{n\})$ at $n = 5$ represents the best compromise between the number of protectors within the desolvation domain and the distance between protectors and hydrogen bond.

With the knowledge of the effective permittivity ε, the Coulombic energy change for a hydrogen bond due to the presence of a new hydrophobe can be assessed as follows:

a), define a desolvation domain D_k with the boarder ∂D_k around the kth hydrogen bond in the same way as in the statistical approach.

b), introduce an extra "test" hydrophobe at position **R**. Then the set of vector positions of n_k hydrophobic groups surrounding the hydrogen bond is extended from $\{r_j, j = 1, 2, \ldots, n_k\}$ to $\{ r_j, j = 1, 2, \ldots, n_k; \mathbf{R}\}$, which changes the effective permittivity from $\varepsilon = \varepsilon(\{\mathbf{R}_j\}_{j=1,2,n_k})$ to $\varepsilon = \varepsilon(\{\mathbf{R}_j, \mathbf{R}\}_{j=1,2,n_k})$.

c), calculate the gradient $\nabla_{\mathbf{R}}(1/\varepsilon)|_{\mathbf{R}=\mathbf{R}_0}$ taken with respect to a perpendicular approach by the test hydrophobe at position $\mathbf{R} = \mathbf{R}_0$ to the center of hydrogen bond, where position $\mathbf{R}_0$ is located at the intersection C of the plane perpendicular to the hydrogen bond with the desolvation domain border ∂D_k.

d), a quantity M_k, which describes the sensitivity of the Coulombic energy for the kth hydrogen bond to the perturbation of the environment, is given by the maximum

gradient:

$$M_k = \text{Maximum}_{R_o \in C} \| \nabla_{\mathbf{R}} (1/\varepsilon)|_{\mathbf{R}=\mathbf{R}_o} \| \tag{2.10}$$

The Coulombic energy is given by

$$E_{coul}(k, \mathbf{R}) = - (4\pi\varepsilon(\mathbf{R}))^{-1} qq'/r_0(k) \tag{2.11}$$

where q, q' and $r_0(k)$ are fixed for the kth hydrogen bond.

Then we get

$$\nabla_{\mathbf{R}}(1/\varepsilon) = [4\pi r_0(k)/qq'] \bullet [-\nabla_{\mathbf{R}}(E_{coul}(k, \mathbf{R}))] = [4\pi r_0(k)/qq'] \bullet F_k(\mathbf{R}) \tag{2.12}$$

Thus the quantity M_k measures the maximum force exerted on the test hydrophobe by the kth hydrogen bond.

The calculation of the quantity M_k for the hydrogen bonds from a sample of 702 proteins of moderate sizes ($52 < N < 110$, N is the number of amino acids) is then performed. 8% of the hydrogen bonds are found to have

$$M_k \geq \lambda /10 \tag{2.13}$$

where $\lambda = (\varepsilon_0^{-1} - \varepsilon_w^{-1})/2$ Å defines an extreme sensitivity in the ideal case in which the completely solvated hydrogen bond ($\varepsilon = \varepsilon_w$) would be fully dehydrated ($\varepsilon = \varepsilon_0$) by a 2-Å displacement of a hydrophobe.

On the other hand, 91.60% of hydrogen bonds have their M_k lie in the range [0, λ/100). These hydrogen bonds are relatively less sensitive to the water removal compared to those 8% of hydrogen bonds. Therefore, the dehydrons are defined as the hydrogen bonds which satisfy Eq. (2.13). The dehydron pattern of the hemoglobin and the cellular prion proteins derived by assessing the sensitivity of the hydrogen bonds to water removal is shown in Fig. 2.4. These results coincide very well with the

above ones derived by the statistical approach. The examination of an exhaustive nonredundant structural sub-database of 2808 entries shows that two approaches give exactly the same results in ~91% of the complexes and monomeric proteins. However, discrepancies arise in the remaining 9% of the cases. For example, unlike the theoretical method, the statistical approach does not identify dehydrons at the adhesive site responsible for the single contact between the HIV-1 capsid protein P24 and the antibody light chain FAB25.3. The statistical approach fails, because it does not take into account the geometrical distribution of protectors. Hydrogen bonds surrounded by a group of uneven distributed protecotors can be dehydrons, even the number of protectors exceeds average ρ. These possible dehydrons can not be identified by the statistical approach.

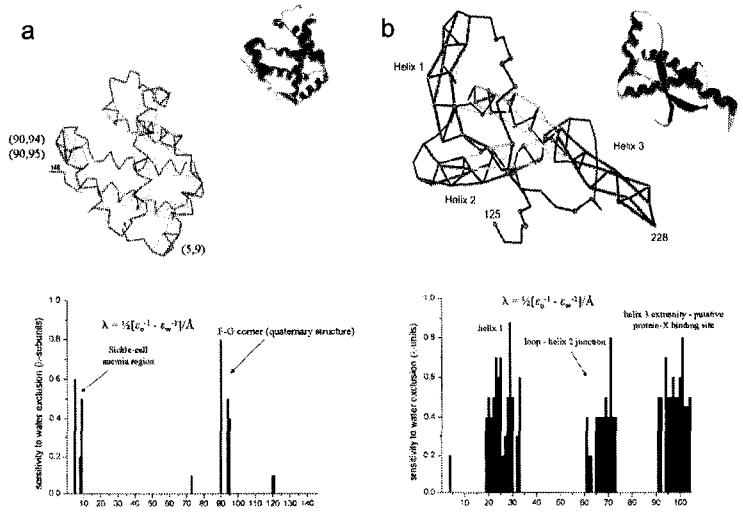


Fig. 2.4: Sensitivity of hydrogen bonds in myoglobin and prion to water exclusion. Ribbon and virtual bond representations are given for each protein. In the virtual bond representations, hydrogen bonds with high sensitivity are shown as green segments, while those insensitive hydrogen bonds are represented by thin light-blue segments [13].

2.2 Role of dehydrons in protein interactions

2.2.1 Dehydrons as determinants of protein association

As defined in the second method, dehydrons are those bonds whose Coulomb energy could change dramatically with the presence of a new hydrophobe from a binding partner. Therefore, dehydrons could serve as good indicator for binding sites. By examining the protein-protein interface of 212 complexes from an exhaustive database, it is found that the density of dehydrons in the interface is higher than the overall density of dehydrons (Table 2.3): 77 complexes have $\delta_{int}/\delta > 1.5$, some even has 7 times higher density of dehydrons in the interface. 92.9% of the PDB complexes have higher density of dehydrons at the protein-protein interface than the average density for individual monomeric partners [9]. The dehydrons in the interface of monomeric proteins become well wrapped in the complexes. The high density of dehydrons in the interface indicate that the exclusion of water from the structurally defective region plays an important role in protein-protein association.

2.2.2 Dehydron as an indicator of protein interactivity

As a strong indicator of protein interactivity, dehydrons could further provide insight into the pattern of proteomic connectivity. A systematic investigation of the dehydron and interaction patterns of all monomeric PDB domains from the yeast

Complex name (PDB ID code)	Y_{int}	Y	$C_{3,int}$	δ, 10^{-3} (Å^{-2})	δ_{int}, 10^{-3} (Å^{-2})
HLA antigen A-2 + β_2-microglobulin (1i4f)	6	36	66	1.58	3.21
Ig light-chain dimer (1jvk)	8	26	72	1.78	3.54
Transthyretin dimer (1bm7)	5	14	57	1.01	3.55
Insulin dimer (6ins)	4	12	51	2.80	4.61
HIV-1 protease dimer + inhibitor (1a30)	7	11	78	1.87	4.91
SIV protease dimer (1siv)	4	14	45	1.06	2.65
CheY complex (1fqw)	4	10	42	1.02	6.07
Defensin dimer (1dfn)	8	14	63	2.72	10.01
Antitrypsin polymers (1d5s)	14	22	165	1.01	2.76
Bombyxin (1bon)	4	5	45	0.60	3.02
FcγRIII receptor + Ig (1e4k) B–C	7	22	57	0.97	7.08
Colicin + ligand (1bxl)	6	12	57	0.92	3.97
Colicin + ligand (1emv)	5	11	60	0.86	3.60
Serpin + ligand (1as4)	14	31	183	1.40	2.02
Troponin heterodimer (1pon)	6	10	69	1.34	4.54
MHC, antigen + receptor (1lm9), A–D	3	22	42	0.84	2.21
MHC, antigen + ligand (1lm9), A–C	3	19	27	2.00	6.12
De novo protein of α-2D (1qp6)	8	12	96	1.65	3.67

Table 2.3: Density of dehydrons at the protein interfaces. Protein complexes are identified by their PDB ID codes. Y_{int} is the number of dehydrons at the interface that become well wrapped in complexes; Y is the total number of dehydrons in the separated components; $C_{3,\,int}$ is the number of three-body correlations at the interface which involves dehydrons and the overexposed hydrophobic residues in the interaction partners; δ is the density of dehydrons given as the number of dehydrons averaged over the total solvent-exposed surface area of all components, δ_{int} is density of dehydrons at the interface [9].

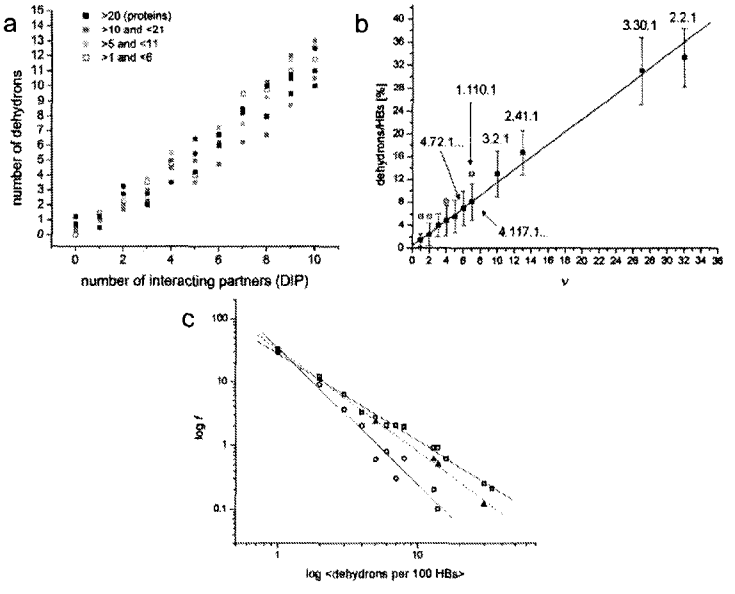


Fig. 2.5: Relationship between protein interaction and the number of dehydrons. (a) Correlation between the average number of dehydrons of a domain fold and the number of its interacting partners reported in Database of Interacting Proteins (DIP). (b) Correlation between the average percentage of dehydrons and the connectivity v of SCOP families represented in the PDB. The numbers in the graph are the SCOP IDs. (c) Distribution of SCOP families according to their average number $\langle r \rangle$ of dehydrons per 100 hydrogen bonds. f = f($\langle r \rangle$) is the fraction of families with $\langle r \rangle$ dehydrons. □ - *H.sapiens*; ▲- *M. musculus*; ○ - *E. coli* [16].

proteome found that domain connectivity is proportional to the average number of dehydrons in the family [16]. Fig 2.5.a shows a correlation of the number of dehydrons in all monomeric PDB domains from the yeast proteome and the number of their interacting partners obtained from the Database of Interacting Proteins (DIP). The correlation coefficient is 0.88 and the dispersion is 0.29.

The almost linear correlation between the the number of dehydrons and the number of interaction also holds for structural families (or Structural Classification of Proteins (SCOP) superfamilies [17, 18]) (see Fig.2.5.b) [16]. The numbers in the figure are the SCOP IDs for some protein domains. These two results suggest that the domain connectivity is measured by the average number of dehydrons $\langle r \rangle$ in a given family. Fig. 2.5.c displays the distribution of protein families according to their $\langle r \rangle$. The fraction $f = f(\langle r \rangle)$ of protein families follows the same power law $f(\langle r \rangle) \propto \langle r \rangle^{-\gamma}$ as the distribution of protein connectivity [19], which further confirm the relationship between $\langle r \rangle$ and proteomic interactivity. The index γ for *H. sapiens*, *M.musculus* and *E. coli* are respectively 1.44, 1.49 and 2.10.

2.2.3 Dehydron patterns of different species

Protein folds are conserved across species [20, 21]. The wrapping level of protein folding domains, however, shows significant differences among species [16]. The number of dehydrons for a specific domain is greater in high-level species. This trend is shown clearly in Fig. 2.6. The differences of dehydron patterns are measured by the

ratio r of dehydrons to backbone hydrogen bonds in the following comparison. The Src homology 3 (SH3) domain in the nematode *Caenorhabditis elegans* (PDB. 3sem) has a ratio r = 3.2%, which is much smaller than the ratio r = 16.6% in the human SH3 domain (PDB. 5hck) (Fig. 2.6 a and b). Fig. 2.6 c and d show the case for the ubiquitin protein of two species: human (PDB. 1ubi, r = 18.7) and *E. coli.* (PDB. 1foz, r = 10.2). The wrapping patterns of hemoglobin protein from human (PDB. 1bz0, r = 2.9) and *Paramecium* (PDB. 1dlw, r = 0) are shown in Fig. 2.6 e and f. The greater number of dehydrons in high-level species indicates more complexity of their interaction network, since dehydrons determine the proteomic connectivity as shown above.

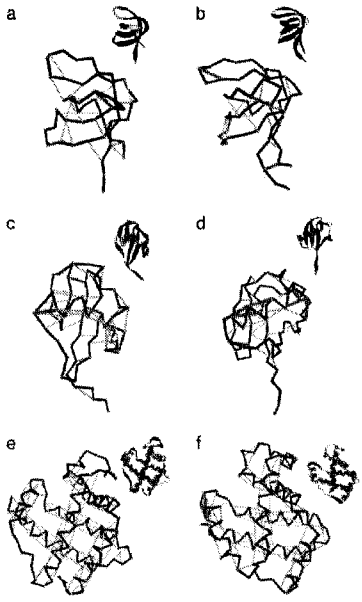


Fig. 2.6: A comparison of dehydron patterns between different species. (a) , (b). Dehydrons in SH3 domains from nematode *C. elegans* (PDB. 3sem) and from *H. sapiens* (PDB. 5hck); (c) , (d). Dehydrons in ubiquitin from *E. coli* (PDB. 1foz) and from *H. sapiens* (PDB. 1ubi); (e) , (f). Dehydrons in hemoglobin from *Paramecium* (PDB. 1dlw) and from *H. sapiens* (PDB. 1bz0, chain B) [16].

2.2.4 Dehydrons as specific targets in drug design

Due to the conservation of folds across paralog proteins, one ligand always binds to the target protein as well as other proteins. The indiscriminative binding of the ligand to different proteins leads to side effects in drug-based inhibition. This kind of side effect can be reduced by selectively targeting nonconserved structure features. Dehydrons as a determinant of protein interactivity are often not conserved across paralog proteins. Therefore wrapping of those nonconserved packing defects by the ligand provides a powerful strategy for drug design [20].

The inhibitor always serves as a dehydron wrapper when they bind to proteins. In 631 of the 814 protein-inhibitor complexes examined, nonpolar groups of inhibitors are present in the desolvation domain of dehydrons. The hydrogen bonds in the binding cavities of the 488 complexes lack protection from the side chains of protein. Those packing defects are compensated by the hydrophobes from inhibitors. Fig. 2.7 gives an illustration of this situation for the HIV-1 protease bound by the Merck inhibitor Indinavir (PDB. 2bpx) [21]. The inhibitor wraps three dehydrons in the binding cavity of each monomer: Ala28-Arg87 and Asp29-Asn88 in the neighborhood of the catalytic site (Asp25) and Gly49-Gly52 in the flap. It contributes 12 nonpolar groups to the dehydron Gly49-Gly52, 10 to the dehydron Ala28-Arg87, and 8 to the dehydron Asp29-Asn88. These three underwrapped hydrogen bonds are strengthened by the water removal induced by the inhibitor. The inhibitor improves the packing of protein, although it is beyond the designer's expectation.

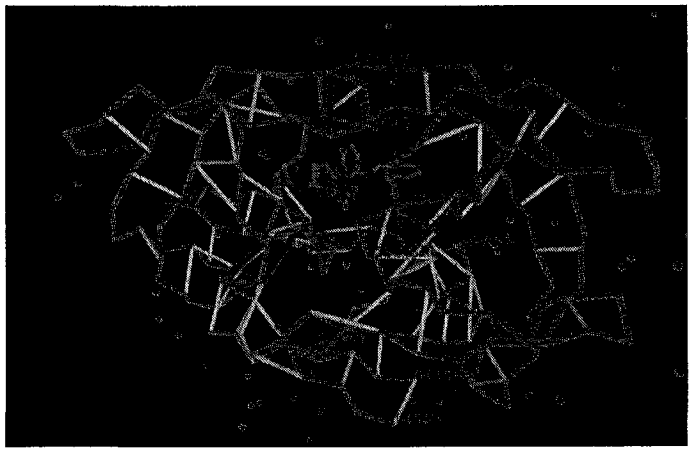


Fig. 2.7: HIV-1 Protease bound by the Merck inhibitor Indinavir (PDB. 2bpx) [20].

The PDB contains 440 non-redundant pairs of human paralogs identified by employing a 30% minimal sequence similarity filter. 269 pairs are different from each other in the location or presence of at least one dehydron at the binding cavity, and the difference expands to 2 or more dehydrons for 203 of them. The high level of nonconservation of packing defects leads them to be a good candidate for highly specific targets.

Chapter 3

Residence times of water molecules in hydration shell

3.1 Structural and dynamic properties of the hydration water

Hydration plays a very important role in the structure and function of proteins [27, 28]. Water is essential for the folding, structure, stability, dynamics and activity of proteins. Thus, the study of structural and dynamic properties of the hydration water is necessary for the comprehension of the protein functionality. Experiments using a variety of techniques have provided many important results: Svergun *et al.* observed a 10%-20% higher density for the water on the protein surface by X-ray and neutron-scattering [29]; nuclear magnetic relaxation data suggest a long residence time (10–100 ps) for hydration water [30, 31].

However, there are some shortcomings inherent in experimental techniques. Diffraction experiments measure the time-averaged behavior of water, and can only detect those water molecules that have resided for a period of time on the order of the time resolution of equipments. In addition, the interpretation of experiment results usually depends on models. For example, it is difficult to assess the contribution of the disordered water molecules to the crystallographic electron density maps [32].

Molecular dynamics simulation as a complementary tool to experimental techniques can provide an atomically detailed description of water structure and dynamics. A simulation study of hydration structure found that water on the protein surface exhibits coherent hydrogen-bond patterns [33]. The structural properties of water up to 6 Å away from the protein are shown to be distinguished from those of bulk water [34, 35].

Many computational efforts to study the dynamics of protein hydration have been focused on calculating the residence time of water molecules in the hydration shell [36 - 42]. Most of hydration sites are occupied by water molecules with residence times 5-50 ps [36, 37, 40, 41]. However, there are also a few hydration sites where water molecules with a residence time in the range of 100-500 ps are found [40, 41]. Controversy arises when researchers try to figure out what factors determine the residence time. Some authors showed that residence time depends on the physicochemical properties of the residues around the water molecules. In crambin [36] and plastocyanin [37], the average residence time is found to follow the sequence: $\tau_{charged} \geq \tau_{polar} > \tau_{nonpolar} \approx \tau_{bulk}$. However, this relationship was not confirmed by the simulation results of the lung surfactant polypeptide SP-C [38]. Many researchers relate the residence time with the geometry or the topology of the protein surface. By analyzing the sites of water in 56 crystallographic protein structures, Kuhn *et al.* found that the ordered water molecules are more likely to be in "groove" positions [39]. "Clefted" and "buried" water molecules with low solvent accessibility have a longer residence time than "exposed" water molecules with high solvent accessibility

[40, 41]. In the next two sections, we will present two examples showing residence times mainly determined by protein surface topology.

3.2 Simulation study of water residence time in azurin

3.2.1 Simulation details

The MD simulation of solvated azurin was performed with the DLPROTEIN simulation package by using the GROMOS force field with the set of parameters denoted "37c"[43]. The starting coordinates for one monomer of Pseudomonas aeruginosa oxidized azurin were obtained from the Protein Data Bank (PDB. 4azu). The protein was then solvated using 3, 658 SPC/E water molecules in a truncated octahedral boundary box [44]. After equilibration in the NPT ensemble, the system was relaxed to 0 K by a steepest descent quenching procedure, and slowly increased to 300 K. The SHAKE algorithm was used to constrain bond lengths [45]. Equations of motion were integrated using time step of 1.0 fs. The temperature was maintained at 300 K by using the Nosé-Hoover thermostat[46], and the pressure was kept at 1.0 atm. The short-range van der Waals interaction was calculated by choosing a cutoff distance of 10.0 Å, while the Ewald sum method was used to handle the electrostatics with an accuracy of 10^{-5}[47]. After 100 ps of thermalization, 500 ps molecular dynamic simulation was performed. Configurations were saved every 0.1 ps.

3.2.2 Calculation of water mean residence time

Water residence time is represented by the mean residence time of water in the first coordination shell of protein atoms. The first coordination shell around a protein atom is defined as a sphere whose radius $\mathbf{r}_{sh}$ is given by:

$$\mathbf{r}_{sh} = \mathbf{r}_{ex} + \mathbf{r}_{OH} + \Delta\,\mathbf{r}_{RT} \tag{3.1}$$

where $\mathbf{r}_{ex}$ is the minimum exclusion distance between the solute and the water oxygen atoms, which is determined from the radial distribution function, $\mathbf{r}_{OH}$ = 1.0 Å is the length of O-H bond in a water molecule, and $\Delta\,\mathbf{r}_{RT}$ = 0.5 Å is a correction due to positional fluctuation.

To calculate the water mean residence time, the "*survival probability function*" $P_\alpha(t)$ is introduced:

$$P_\alpha(t) = \sum_{j=1}^{N_W} \frac{1}{N-m+1} \sum_{n=1}^{m} P_{\alpha,j}(t_0, t_0 + t', \Delta t) \tag{3.2}$$

where $t = m\Delta t$ and $t' = n\Delta t$, Δt = 0.1 ps is the time interval for which configurations are saved; the binary function $P_\alpha(t_0, t_0 + t', \Delta t)$ is equal to 1 when the water molecule j stay in the shell α from time t_0, $t_0 + t'$, and 0 otherwise; t_0 can take the values from 0 to T, which is the total time length of trajectory; $N = T/\Delta t$ is the total number of configurations; N_W is the total number of water molecules.

$P_\alpha(t)$ represents the average number of water molecules that stay in the shell for a time t or longer. $P_0 = P_\alpha(t = 0)$ is the average number of water molecules in the shell α.

The water mean residence time τ is then obtained by fitting $P_\alpha(t)$ with a single exponential function:

$$P_\alpha(t) = P_0 \, e^{-t/\tau} \tag{3.3}$$

3.2.3 Results

Fig. 3.1 shows the distribution of water residence time of azurin. Most of atomic sites are characterized by water residence time lower than 20 ps, while few atoms (less than 20) are characterized by a residence time longer than 100 ps or longer. Those long-life water molecules are trapped into holes or deep crevices of the protein surface.

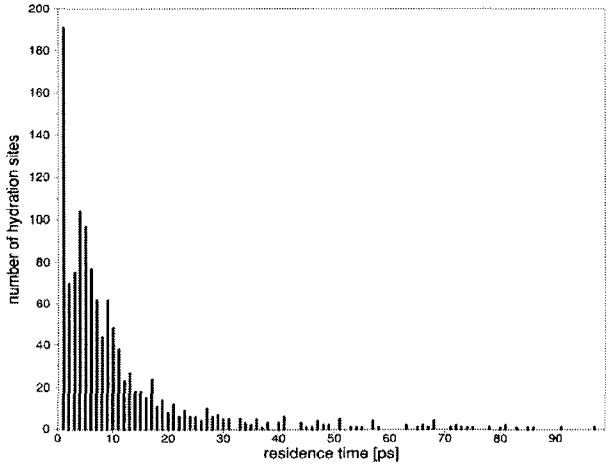


Fig. 3.1: Distribution of water residence times around azuri [40].

The mean residence times of side chain atoms (grey bars) and backbone atoms (black bars) averaged over the secondary structure elements were displayed in Fig. 3.2. The results for side chains and backbone are similar and show the main peak in the

same position. The large values of residence time are observed in the region between Val49 and Asp55, that includes β strand 4 (S4) and the loop 4 (L4). The high peak of water residence time of S4 is largely due to the internal water molecule that permanently hydrogen bonded to the hydroxyl group of Ser51. On the other hand, water molecules trapped in the cavities near L4 also have long residence times. The similar residence time values for different secondary elements suggest there is no strong correlation between water residence time and secondary structure.

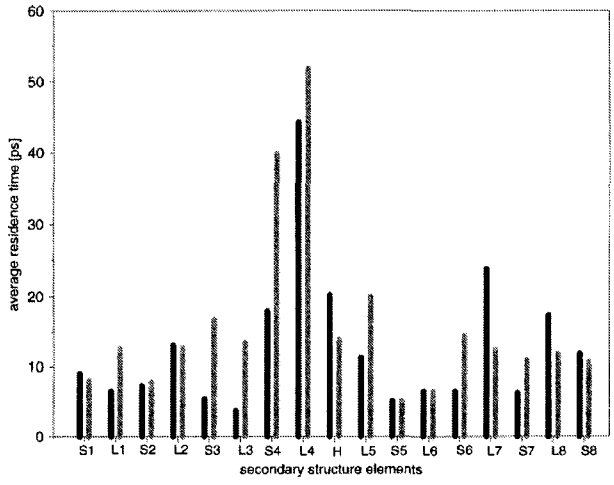


Fig. 3.2 Residence times of secondary structure elements [40].

The water residence time for different atom types is analyzed by calculating the average values of the water mean residence time of seven groups from the side chains and the backbone. The five groups from side chains are polar oxygen (OH), polar nitrogen (NH_2), charged oxygen (COO^-), charged nitrogen (NH_3^+) and carbon atoms (CH_2 and CH_3), and the two groups from the backbone are polar oxygen (CO) and polar nitrogen (NH). The results are reported in Table 3.1. The ranking relation follows:

$$\tau_{OH} > \tau_{O^-} > \tau_{N^+} = \tau_{NH} > \tau_{CH2/CH3} > \tau_{CO} > \tau_{NH2}$$

This result agrees well with those for crambin [36]:

$$T_{charged} > \tau_{polar} > \tau_{non\text{-}polar}$$

but different from those for BPTI [42]:

$$\tau_{polar} > \tau_{non\text{-}polar} > \tau_{charged}$$

The large variances for the average residence time for different atom types (see Table 5) suggest the water residence time is determined by other factors besides atomic characteristics.

	τ [ps]	(σ_τ)	N_{co}	(σ_{Nco})
Side chain polar O (OH)	33	(28)	3.6	(0.9)
Side chain charged O (COO^-)	22	(21)	4.4	(1.0)
Side chain charged N (NH_3^+)	18	(21)	6.1	(0.9)
Backbone polar N (NH)	18	(16)	2.4	(0.6)
Side chain apolar C (CH_2, CH_3)	14	(26)	5.4	(2.0)
Backbone polar O (CO)	12	(18)	3.1	(1.1)
Side chain polar N (NH_2)	9	(12)	4.4	(1.3)

Table 3.1: Residence times of seven atom groups. τ, the mean value of residence time; σ_τ, the standard deviation of residence time; N_{co}, the average coordination number given as the number of water molecules averaged over all configurations; σ_{Nco} the standard deviation of coordination number [40].

The solvent-accessible surfaces for the atoms in the seven groups defined in the previous section are calculated. Their relationship with the water residence time is displayed in Fig. 3.3. A general feature from Fig. 3.3 a – d is: atoms with solvent-accessible surface larger than 16 Å are characterized by a residence time smaller than 20 ps. For the atomic sites with solvent-accessible surface smaller than 16 Å, the residence time varies from several ps to 100 ps. The water molecules around the low accessible atoms can be well protected so that the rate of exchange with the bulk water is reduced and the residence time is prolonged. On the other hand, the

variance for the residence time of low accessible atoms means other factors make an important contribution to water residence time of atoms with small solvent-accessible surface.

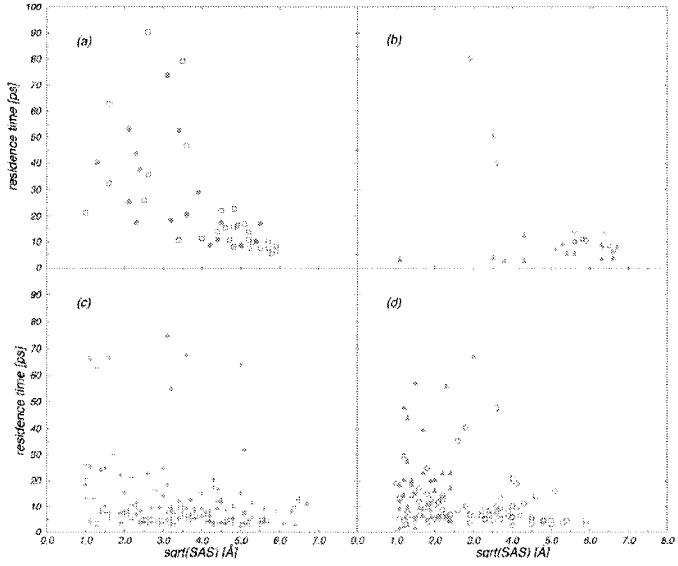


Fig. 3.3: Distribution of water residence times of seven atom types according to their square root of average solvent accessible surface area. (a), • represents charged oxygen atom in COO^-; □ represents oxygen atom in OH. (b)▲, ◊ represent nitrogen atoms in NH_2 and NH_3^+ respectively. (c) + represents the carbon atom in CH_2 and CH_3. (d) ▲, ○ represent the nitrogen atom in NH and the oxygen atom in CO respectively [40].

3.3 Water residence time of Myoglobin

3.3.1 Molecular dynamics simulation

The simulations were performed with CHARMM-23 using the CHARMM-23 all-hydrogen force field [48]. A single molecule of carbomonoxy myoglobin (PDB.

2mgk) was solvated by 3717 explicit TIP3P water molecules in a periodical box of dimensions 60.4 Å × 54.7 Å × 40.7 Å [49]. The system went through a series of energy minimization and thermalization, and was then equilibrated at 294 K for 200 ps. No constraints were applied to bond lengths. Equations of motion were integrated with a 0.5-fs time step. The electrostatic interaction was calculated using a 13-Å cutoff and a potential-based switching function beginning at 10 Å. The following simulation last for 1.1 ns. Coordinates and velocities were saved to the trajectory every 0.1 ps.

3.3.2 Calculation of water residence times

The MD trajectory was reimaged into the rectangular box in the way that the protein was fitted to a consistent frame reference. A three-dimensional number density distribution is obtained by mapping the coordinates of the water oxygen atoms to the three-dimensional rectangular grid with a 0.5 Å grid step. Every grid cell in the bulk water region has at least 50 counts over the entire trajectory.

Proteins hydration sites are defined as the grid cells where the solvent density has a local maximum and satisfy some conditions. The nearest distances between the hydration sites to any protein atoms should be smaller than 5 Å; the distances between two hydration sites should be larger than 1 Å (two grid steps); the maximum density of the hydration sites should be equal to or larger than 150% of the bulk water density.

A time correlation function is introduced to calculate the water residence time:

$$P(\tau) = \int_0^\infty \delta(N(t) - N(t+\tau))dt \tag{3.4}$$

where $N(t)$ is the index of the water molecule found in the hydration site at time t, and $\delta(N(t) - N(t+\tau))$ takes the value of 1 when $N(t)$ and $N(t+\tau)$ refer to same water molecule and 0 otherwise. The correlation function was fitted by a double exponential function

$$P(\tau) = W_0[(1-w)\exp(-k_1\tau) + w\exp(-k_2\tau)] \tag{3.5}$$

In this model, water molecules are classified as those stay in the hydration shell for a long time and those enter and leave the shell immediately. The short and long residence time are given by $\tau_1 = 1/k_1$ and $\tau_2 = 1/k_2$. W_0 represents the average time site is occupied, while $1-w$ and w represent the weights for the fast and slow components.

3.3.3 Results

A total of 294 hydration sites around myoglobin were identified (see Fig. 3.4). The hydration sites are distributed according to the shape of the protein, except in some regions. As Fig. 13 shows, there is a high density of hydrations sites on top of the protein, and only few hydration sites are on the right region. Residues in the low site-density areas are highlighted in yellow. In contrast, crystal hydration water molecules are distributed much more uniformly. The differences between the simulated and crystal hydration sites may arise from the absence of the crystal packing effects in the simulation.

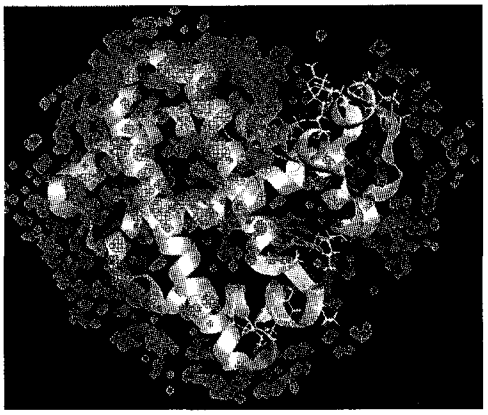


Fig. 3.4: Distribution of hydration sites around myoglobin. Exposed residues without hydration sites are colored as yellow [41].

The hydration sites are not occupied all the time, and no more than 164 of 294 sites are occupied at any moment. Fig. 3.5 shows distributions of the parameters in the biexponential model for 294 hydrations sites: average site occupancy (W_0), short and long residence times (t_1 and t_2), and the weight of two components w. As a comparison, those parameters for the bulk water sites are: W_0 is 49%, t_1 = 0.34 ps, t_2 = 4.1 ps, and w = 0.39. The peaks in the distributions of W_0, t_1 and t_2 are very close to the values of the bulk water sites, which indicates that many hydration sites are bulk-like. The residence times (t_2) for most hydration sites are in the range [5ps, 35ps], and comparably larger t_2 are found for the sites occupied by the bound waters. The typical average occupancy of the hydration sites varies from 15% to 59%. The distribution of w peaks at a value significantly larger than w for bulk water sites, which suggests that most hydration sites are strongly influenced by the protein and are occupied mainly by waters with prolonged residence times (t_2).

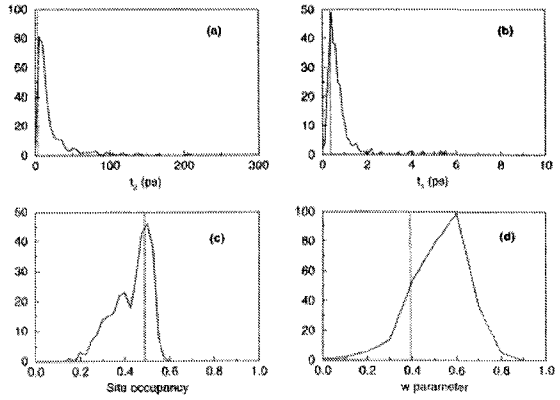


Fig. 3.5: Distribution of four parameters in the bi-exponential model. (a) long residence time (t_2). (b) short residence time (t_1). (c) site occupancy. (d) w parameter. Vertical lines represent the values of four parameters for bulk water [41].

There is no simple correlation between residence time and site density. High density sites are not always characterized by high residence time [41]. Sites with different residence times were then mapped onto the protein surface. As Fig. 3.6 shows, sites with very long residence times ($t_2 \geq 80$ ps) are located in the cavities or in the grooves and concave regions, while the low residence time sites ($t_2 \leq 10$ ps) are found in convex areas. There are similar relations between solvent accessibility and other parameters: sites with low occupancy ($W_0 \leq 0.3$) or large weight ($w \geq 0.6$) are buried, while sites with high occupancy ($W_0 \geq 0.5$) or small weight ($w \leq 0.4$) are exposed. The three groups of buried sites (high t_2, low W_0 and large w) are not identical, although they have some sites in common. The water residence times of hydration sites are mainly determined by the locations of the hydration sites: sites in buried regions have longer residence times than those exposed to the solvent.

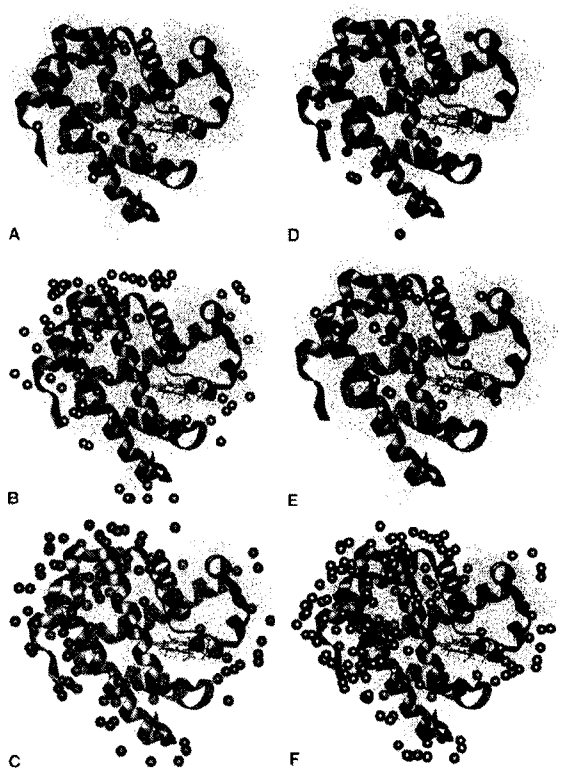


Fig. 3.6: Distribution of hydration sites according to their residence times (t_2), occupancy and w parameters. (A) sites with high t_2. (B) sites with low t_2. (C) sites with high occupancy. (D) sites with low occupancy. (E) sites with high w parameter (F) sites with low w parameter [41].

The results of water residence times of azurin and myoglobin may be not comparable: they are based on the study of different proteins; the definition of residence times are different; the simulations were performed using different packages. However, these two analyses reveal some common features: the water residence time is determined by several factors such as atomic characteristics and protein surface topology, the major factor influencing the residence time is the topology properties of the protein surface.

Chapter 4

Dehydron loosens the hydration shell

4.1 Methods

4.1.1 Molecular simulation

The GROMACS program with optimized potential for liquid simulations (OPLS) force field was used to perform the computations [50, 51]. The starting coordinates were prepared by placing the protein in a preequilibrated cell of explicit presented water molecules and counterions [52]. The SHAKE algorithm was used to constrain bond lengths [45]. Computations were then performed by integrating the equations of motion with a 2-fs time step in an NPT ensemble with box size of $8\times8\times8$ nm^3 under periodical boundary conditions. The temperature was maintained at 300 K using the Nosé-Hoover thermostat [46], and the pressure was kept at 1atm using a weak-coupling algorithm. TIP3P model was adopted to describe water molecules [49]. The long-range electrostatics were treated using the particle mesh Ewald summation method [53]. A total of 15 ns of molecular dynamic simulation was performed.

4.1.2 Calculation of local residence time

The hydration shell for residue i was defined as a spherical domain $D(i)$ of 6- Å radius (~width of three water layers) centered at the α-carbon atom of residue i. The computations for a range of radii were also performed. The mean residence time $\langle\tau\rangle_i$ of water molecules in shell i is obtained as follows:

$$\langle\tau\rangle_I = \int \tau f_i(\tau)d\tau \,/ \int f_i(\tau)d\tau\,,$$

$$\int_0^\tau f_i(\tau')d\tau' = P_i(0) - P_i(\tau), \tag{4.1}$$

$$P_i(\tau) = \Theta^{-1}\int_0^\Theta \left[\sum_{\substack{v(t)\in U(i,t)\\ w(t+\tau)\in U(i,t+\tau)}} \delta(v(t), w(t+\tau))\right] dt\,,$$

where $f_i(\tau)$, is the number of water molecules with a life time τ; $P_i(\tau)$ is the number of water molecules with a life time $\geq \tau$ and $P_i(0)$ is the total number of water molecules that have visited the domain $D(i)$; $v(t)$ and $w(t+\tau)$ are the indices of water molecules in $D(i)$ at times t and $t+\tau$ respectively; $U(i, t)$ and $U(i, t+\tau)$ denote the collection of indices of water molecules in $D(i)$ at times t and $t+\tau$ respectively; δ is the Kronecker symbol ($\delta(v(t),w(t+\tau)) = 1$ if $v(t),w(t+\tau)$ denote the same water molecule and 0, otherwise); Θ = 10 ns is the total time length of the trajectory used to calculate residence times.

4.1.3 Identification of dehydrons

The statistical approach in section 2.1.1 was employed to identify dehydrons. In this work, the radius of the desolvation domain R was set to be 6 Å. Two thirds of the

backbone hydrogen bonds in soluble proteins have $\rho = 26.6 \pm 7.5$ nonpolar groups in their desolvation domains. Dehydrons were then defined as hydrogen bonds with $\rho \leq 19$.

4.1.4 Calculation of the dewetting field

The dehydration propensity of the backbone hydrogen bonds can be assessed by calculating the change in Coulomb energy of the hydrogen bonds upon removal of surrounding waters. For a hydrogen bond wrapped by K nonpolar groups in postions $\{\mathbf{R}_j\}_{j=1,\ldots,K}$, a new test hydrophobe was introduced to quantify the Coulomb energy change as we did in the theoretical method of identifying dehydrons. When the test hydrophobe placed at postion $\mathbf{R}$ approaches the hydrogen bond, the Coulomb energy of the hydrogen bond increases due to the decrease in permittivity. The dewetting field $\Phi(\mathbf{R})$, referred to the force exerted on the hydrophobe by the hydrogen bond, is given by

$$\Phi(\mathbf{R}) = (qq'/4\pi r)\nabla_{\mathbf{R}}(\varepsilon^{-1}(\{\mathbf{R}_j, \mathbf{R}\}_{j=1,\ldots,K})) \tag{4.2}$$

where q and q' are the effective charges of the donor and acceptor of the hydrogen bond, r is the bond length, and the permittivity between two atoms ε, given by formula (2.8)

$$\varepsilon(\{\mathbf{R}_n\}) = [(\varepsilon_0^{-1} - \varepsilon_w^{-1})\Omega(\{\mathbf{R}_n\})(1+r/\xi)\exp(-r/\xi) + \varepsilon_w^{-1}]^{-1} \tag{4.3}$$

where ε_0 and ε_w are the permittivities of vacuum and bulk water, ξ is the characteristic length for water-dipole reorientation influence, here fixed at 5 Å, $\Omega(\{\mathbf{R}_n\}) = \Pi_{n=1,\ldots,}$

$[1+\exp(-||\mathbf{R}_n||/\Lambda)]$ *$[1+\exp(-||\mathbf{r}-\mathbf{R}_n||/\Lambda)]$, where Λ = 1.8Å is the characteristic length for water structuring around a nonpolar group.

4.2 Results

4.2.1 Water residence times of the SH3 domain

The water residence time of the SH3 domain (PDB. 1srl) was analyzed. The results are shown in Fig. 4.1. The average water residence time for all residues is 44 ps. The majority of residues have their residence times lying between 30 and 50 ps. The distribution is relatively narrow, compared to the previous results of azurin and myoglobin. The largest residence time is ~60 ps, much smaller than that of myoglobin (456.5 ps). Difference between our results and those of azurin and myoglobin are largely due to the definition of water residence time. In the analysis of water residence times of azurin and myoglobin, their relatively small hydration shell containing only one or two bound water molecules can have very long residence time. On the contrary, many waters included in the hydration shell, as a result of large radius we used, tend to reduce the possible maximum residence time. The longest residence times are found for two exposed charged residues (Lys27 and Lys28) that are tightly hydrated. Some polar residues in cavities (like Ser58) also have long residence time, since their hydrogen-bonding to water molecules in the shell is well shielded. In contrast, exposed residues with delocalized charges such as Arg19 are surrounded by loosely

bound waters. On the other hand, hydrophobic residues at cavities with large curvature radius tend to show high dewetting propensity [54], and hence are characterized by short residence time. This situation is exemplified by the low residence time of Trp43 in the hydrophobic pocket. Three buried hydrophobic residues Ala12, Ile56 and Val61 are devoid of waters, and hence their residence times are zero.

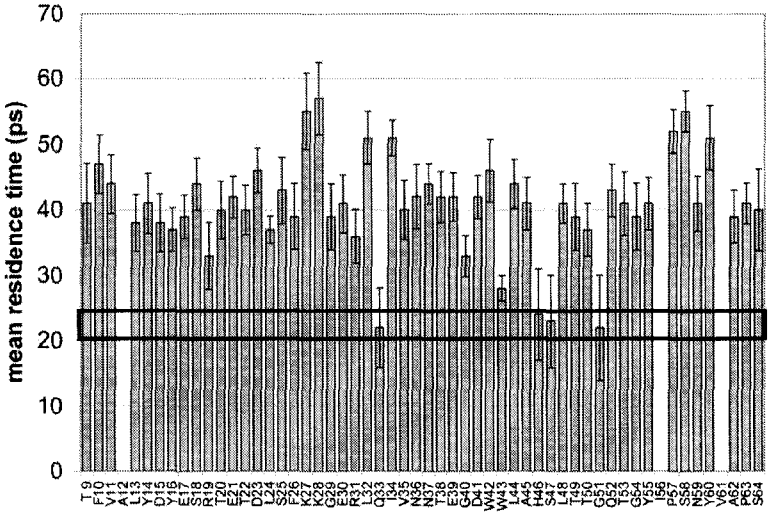


Fig. 4.1 Mean residence times of all residues in SH3 domain (PDB. 1srl) [55].

Four residues: Gln33, His46, Ser47 and Gly51, whose residence times lie in the marked region, have the shortest residence times ($<\tau>$ < 24ps) and the largest variances ($<(\tau - <\tau>)^2>^{1/2}$ > 6 ps), which indicates a highly fluctuating environment and a high dewetting propensity. This observation agrees well with the following results which show that the two backbone hydrogen bonds formed between these four residues are lack of the protection from side chains and become favorably dehydrated.

Computations with hydration-shell radii in the range of 5.6–6.3 Å result in the same four hot spots (Gln33, His46, Ser47, and Gly51).

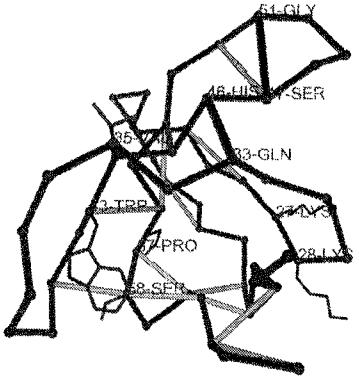


Fig. 4.2: Virtual bond representation of SH3 domain (PDB. 1srl). Two dehydrons Gln33-His46 and Ser47-Gly51 are shown as thick green lines [55].

4.2.2 Dehydrons and dewetting fields of the SH3 domain

Two dehydrons Gln33-His46 and Ser47-Gly51 were identified in SH3 domain (shown as green bonds in Fig. 4.2). Calculation of the dewetting field reveal that these two dehydrons generate the most pronounced dewetting fields (see Fig. 4.3). The high dewetting fields of the two dehydrons represent a high dehydration propensity. The incomplete wrapping of the dehydrons renders them vulnerable to water attack. Dehydration decreases the overall environmental polarizability, and strengthens the Coulomb energy of the dehydrons.

The above analyses lead us to the following conclusions: (a) the mobility of hydration water is not uniform around the protein surface, but depends on the surface topology and physiochemical properties of residues; (b) dehydrons are surrounded by

hydration waters with high mobility; (c) dehydrons are characteristic of high dewetting propensity.

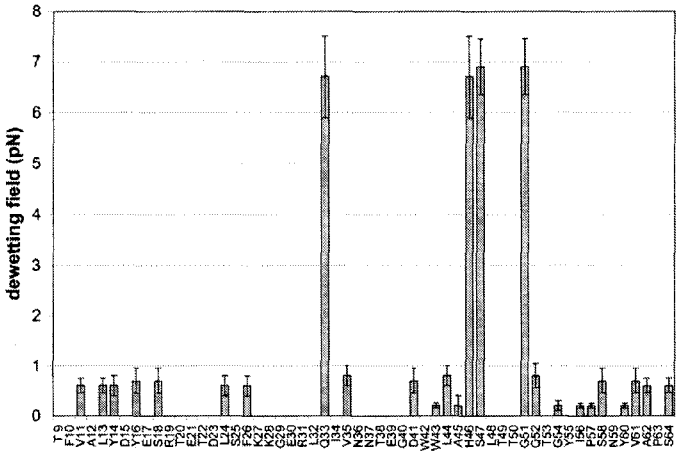


Fig. 4.3: Dewetting fields of residues paired by hydrogen bonds in SH3 domain [55].

4.2.3 Water residence times of ubiquitin

A similar analysis was performed for the ubiquitin protein (PDB. 1ubi). The residence time results and dehydron pattern are displayed in Fig. 4.4. 9 of total 11 dehydrons are singled out as dewetting hot spots. The remaining 2 dehydrons Glu24-Asp52 and Pro19-Ser57 have slightly larger residence time. The low water mobility at the four hydration shells may be rationalized in this way: the dehydration demands of the two charged side chains prolong the lifetime of water molecules in the case of Glu24-Asp52, while the protection from the side chain of Pro19 contributes to the increase in residence time of water in the other case.

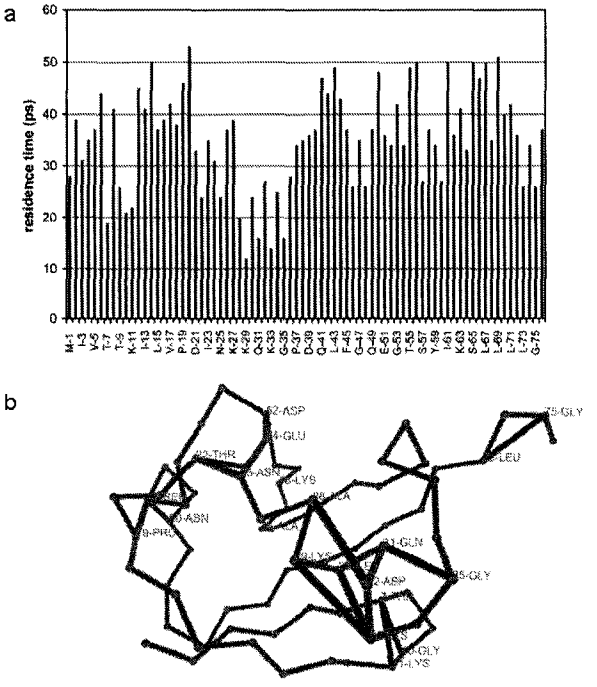


Fig. 4.4: Residence time (a) and dehydron pattern (b) of uniquitin (PDB. 1ubi) [55].

Chapter 5

Conclusion and Implication

Hydration is essential for the structure and function of proteins. The study of properties of hydration water helps to clarify many issues related to protein interactions. The mobility of hydration water depends on both surface topology and physiochemical properties of residues. The longest residence times are found for exposed charged residues that are tightly hydrated. Some polar residues in cavities also have long residence time, since their hydrogen-bonding to water molecules in the shell is well shielded. On the other hand, hydrophobic residues at cavities with large curvature radius tend to show high dewetting propensity, and hence are characterized by short residence time. The shortest residence times are found for the residues paired by dehydrons. The water residence time analysis reveals that dehydrons enhance the mobility of water molecules and generate the most pronounced loosening of hydration shell.

Hydration shell loosening promoted by nonconserved dehydrons provides a new strategy for rational drug design. Wrapping the packing defects by the hydrophobic group of ligand enhances the protein-ligand binding affinity. The nonconservation of packing effects make them specific target for drug discovery.

Protein interactivity is measured by the extent of wrapping. Highly under-wrapped proteins have many interacting partners, because they need partners to

maintain their structure integrity. The over-expression of highly under-wrapped genes will render a higher probability of aberrant misfolding and aggregation. Wrapping provides a new sight to understand dosage imbalance effects.

References

[1] S. J. Suresh and V. M. Naik, Hydrogen bond thermodynamic properties of water from dielectric constant data. J. Chem. Phys. 113, 9727 (2000).

[2] P. Maksyutenko, T. R. Rizzo, and O. V. Boyarkin, A direct measurement of the dissociation energy of water. J. Chem. Phys. 125, 181101 (2006).

[3] Luck, W. A. P. (1985). In Water and Ions in Biological Systems, ed. A. Pullman, V. Vasileui and L.Packer, New York: Plenum, p. 95.

[5] L. Pauling and R. B. Corey, The pleated sheet, a new layer configuration of polypeptide chains. Proc. Natl. Acad. Sci. USA 37, 251 (1951).

[4] L. Pauling, R. B. Corey and H. R. Branson, The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc. Natl. Acad. Sci. USA 37, 205 (1951).

[6] M.F. Perutz, New X-Ray Evidence on the Configuration of Polypeptide Chains: Polypeptide Chains in Poly-γ-benzyl-L-glutamate, Keratin and Hemoglobin. Nature 167, 1053 (1951).

[7] Jeffrey, G. A., Saenger, W. (1991) Hydrogen bonding in biological structures. New York: Springer-verlag

[8] Ariel Fernandez and R. Stephen Berry, Extent of Hydrogen-bond protection in folded proteins: a constraint on packing architectures. Biophys. J. 83, 2475 (2002).

[9] Ariel Fernandez and Harold A. Scheraga, Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. Proc. Natl. Acad. Sci. USA 100, 113

(2003).

[10] Ariel Fernandez and Ridgway Scott, Dehydron: a structurally encoded signal for protein interaction. Biophys. J. 85, 1914 (2003).

[11] Ariel Fernandez and L. Ridgway Scott, Adherence of packing defects in soluble proteins. Phys. Rev. Lett. 91, 018102 (2003).

[12] James T. Kellis Jr, Kerstin Nyberg, DaSa Sali and Alan R. Fersht Contribution of hydrophobic interactions to protein stability. Nature 333, 784 (1988).

[13] Ariel Fernandez, Tobin R. Sosnick and Andres Colubri, Dynamics of Hydrogen Bond Desolvation in Protein Folding. J. Mol. Biol. 321, 659 (2002).

[14] T. Clackson and J. A. Wells, A hot spot of binding energy in a hormone-receptor interface. Science 267, 383-386 (1995).

[15] C. M. Dobson, Protein misfolding, evolution and disease. Trends Biochem. Sci. 24, 329 (1999).

[16] Ariel Fernandez, Ridgway Scott, and R. Stephen Berry, The nonconserved wrapping of conserved protein folds reveal a trend toward increasing connectivity in proteomic networks. Proc Natl Acad Sci U S A 101, 2823 (2004).

[17] A. Murzin, S. E. Brenner, T. J. P. Hubbard and C. Chothia, SCOP: A structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247, 536 (1995).

[18] L. LoConte, S. E. Brenner, T. J. P. Hubbard, C. Chothia and A. G. Murzin, SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res. 30, 264 (2002).

[19] S.N. Dorogovtsev and J.F.F. Mendes, Evolution of Networks: from biological networks to the Internet and WWW, Oxford University Press, 2003, ISBN 0-19-851590-1

[20] Ariel Fernández, Incomplete protein packing as a selectivity filter in drug design. Structure, 13, 1829 (2005).

[21] A. Wlodawer and J. Vondrasek, Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu. Rev. Biophys. Biomol. Struct. 27, 249 (1998).

[22] Ariel Fernández, Angela Sanguino, Zhenghong Peng, Alejandro Crespo, Eylem Ozturk, Xi Zhang, Shimei Wang, William Bornmann and Gabriel Lopez-Berestein, Rational Drug Redesign to Overcome Drug Resistance in Cancer Therapy: Imatinib Moving Target, Cancer Research 67, 4028 (2007).

[23] A. Bennasroune et al., Tyrosine kinase receptors as attractive targets of cancer therapy, Crit. Rev. Oncol. Hematol. 50, 23 (2004).

[24] Samir Attoub et al. The c-kit Tyrosine Kinase Inhibitor STI571 for Colorectal Cancer Therapy. Cancer Res. 62, 4879 (2002).

[25] Thomas Schindler, William Bornmann, Patricia Pellicena, W. Todd Miller, Bayard Clarkson, and John Kuriyan Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. Science 289, 1938 (2000).

[26] Ariel Fernández et al., An anticancer C-kit kinase inhibitor is re-engineered to make it more active and less cardiotoxic. J. Clin. Invest., in press (2007)

[27] F. Franks, Protein stability: the value of 'old literature.' Biophys. Chemist. 96, 117 (2002).

[28] M. F. Chaplin, Opinion: Do we underestimate the importance of water in cell biology? Nature Rev. Mol. Cell Biol. 7, 861 (2006).

[29] D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai, Protein hydration in solution: Experimental observation by x-ray and neutron scattering Proc. Natl. Acad. Sci. U.S.A. 95, 2267 (1998).

[30] G. M. Clore and G. M. Gronenborn, Localization of bound water in the solution structure of the immunoglobulin binding domain of streptococcal protein G: Evidence for solvent-induced helical distortion in solution. J. Mol. Biol. 223, 853 (1998).

[31] K. Wuthrich, G. Otting and E. Liepinsh, Protein hydration in aqueous solution Faraday Discuss. 93, 35 (1992).

[32] G. N. Phillips, Jr. and B. M. Pettitt,Structure and dynamics of the water around myoglobin Protein Sci. 4, 149 (1995).

[33] T. Yokomizo, M. Nakasako, T. Yamazaki, H. Shindo and J. Higo, Hydrogen-bond patterns in the hydration structure of a protein. Chem. Phys. Lett. 401, 332 (2005).

[34] V Lounnas, B M Pettitt, and G N Phillips, Jr, A global model of the protein-solvent interface. Biophys. J. 66, 601 (1994).

[35] C. Schro..der, T. Rudas, S. Boresch, and O. Steinhauser, Simulation studies of the protein-water interface. I. Properties at the molecular resolution. J. Chem. Phys. 124, 234907 (2006).

[36] Garcia AE, Stiller L. Computation of the mean residence time of water in the hydration shells of biomolecules. J Comput Chem 14, 1396 (1993).

[37] C. Rocchi, A. R. Bizzarri, and S. Cannistraro, Water residence times around

copper plastocyanin: a molecular dynamics simulation approach Chem. Phys. 214, 261 (1997).

[38] H. Kovacs, A. E. Mark, and W. F. van Gunsteren,Solvent structure at a hydrophobic protein surface. Proteins: Struct., Funct., Genet. 27, 395 (1997).

[39] L. A. Kuhn, M. A. Siani, M. E. Pique, C. L. Fisher, E. D. Getzoff, and J. A. Tainer, The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. J. Mol. Biol. 228, 13 (1992).

[40] A. Luise, M. Falconi, and A. Desideri, Molecular dynamics simulation of solvated azurin: Correlation between surface solvent accessibility and water residence times. Proteins: Struct., Funct., Genet. 39, 56 (2000).

[41] V. A. Makarov, B. K. Andrews, P. E. Smith, and B. M. Pettitt, Residence Times of Water Molecules in the Hydration Sites of Myoglobin. Biophys. J. 79, 2966 (2000).

[42] Brunne RM, Liepinsh E, Otting G, Wutrich K, van Gunsteren WF. Hydration of proteins. A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. J Mol Biol 231, 1040 (1993).

[43] van Gunsteren WF, Berendsen HJC. Groningen Molecular Simulation (GROMOS) Library Manual. Groningen, Biomos, (1987).

[44] Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. J Phys. Chem. 91, 6269 (1987).

[45] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian

equations of motions of a system with constraints: molecular dynamics of N-alkanes. J Comp. Phys. 23, 327 (1977).

[46] W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A 31, 1695 (1985).

[47] Frenkel D, Smit B. Understanding molecular simulation. From algorithms to applications. San Diego: Academic Press; 1996. 347 p.

[48] A. D. MacKerell, et al., Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. FASEB J. 6A, 143 (1992).

[49] Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926 (1983).

[50] E. Lindahl, B. Hess, and D. van der Spoel, Gromacs 3.0: A package for molecular simulation and trajectory analysis. J. Mol. Model. 7, 302 (2001).

[51] W. L. Jorgensen and J. Tirado-Rives. The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. J. Am. Chem. Soc. 110, 1657 (1988).

[52] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926 (1983).

[53] T. Darden, D. York, and L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089 (1993)

[54] K. Lum, D. Chandler, and J. D. Weeks, Hydrophobicity at Small and Large

Length Scales. J. Phys. Chem. B 103, 4570 (1999)

[55] Ariel Fernández, Jianping Chen, Alejandro Crespo, Solvent-exposed backbone loosens the hydration shell of soluble folded proteins J. Chem. Phys. 126, 245103 (2007)