RESEARCH PAPER

PCCP
www.rsc.org/pccp

Internal hydration of protein cavities: studies on BPTI†

Andrei I. Borodich[a] and G. Matthias Ullmann*[ab]

[a] *IWR-Biocomputing, University of Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany. E-mail: andrei.borodich@iwr.uni-heidelberg.de; matthias.ullmann@iwr.uni-heidelberg.de*

[b] *Structural Biology/Bioinformatics, University of Bayreuth, Germany. E-mail: matthias.ullmann@uni-bayreuth.de*

Received 20th October 2003, Accepted 16th February 2004
First published as an Advance Article on the web 15th March 2004

In this paper, we present a theoretical method to calculate the hydration free energies of the protein cavities and clefts. This method is also used to compute the binding probabilities of buried water molecules. Our approach considers an ensemble of potential binding sites located within the protein and water molecules as ligand particles. The free energy to transfer a water molecule from the gas phase to the protein is calculated from the thermodynamic cycle. The protein is viewed as a dielectric continuum and a water molecule is embedded in it. Electrostatic contributions to the energies of bound water molecules and their interaction energies are computed from the solution of the Poisson equation. The nonpolar contribution and the binding entropy change are assumed to be the same for every bound water molecule. The obtained energies are used in the binding polynomial to calculate the constants for the hydration reactions and the hydration free energies. Calculated values are compared with available measurements for the initial steps of the hydration of BPTI (bovine pancreatic trypsin inhibitor) in the gas phase. The obtained energies are also used to calculate the population probabilities of the hydration sites and compute their titration curves. The results of the studies on BPTI demonstrate that an isolated buried water in the small protein cavity binds to the protein with a lower affinity than a cluster of three water molecules in a large protein cleft.

1 Introduction

Nearly all globular proteins have internal cavities with volumes large enough to accommodate at least one water molecule.[1,2] Some cavities appear to be empty, others contain water molecules as observed in different experiments. X-ray and neutron diffraction studies[3,4] applied to high-resolution crystal structures give valuable information about positions of well-ordered structural waters of average and high occupancy. However, cavities may appear empty if the water molecules bind only temporarily or if they are highly disordered within the cavity. Nuclear magnetic resonance (NMR) techniques, such as measurements of intermolecular nuclear Overhauser effects[5–8] and nuclear magnetic relaxation dispersion studies,[9,10] are used to obtain information about proteins in aqueous solution. NMR methods can detect buried water molecules, even if they are highly disordered, and measure their residence times, orientation order parameters and exchange rates with the bulk solvent. Experimental data on hydration reaction free energies are available from the gas-phase mass-spectrometry.[11,12] This technique provides information about the sequential hydration of an initially dry protein.

It has been accepted that buried waters play both structural and functional roles. They contribute to the conformational stability of proteins and can also be directly involved in enzyme catalysis, facilitate the ligand binding, mediate electron and proton transfer and promote complex formation.[1,2] Therefore, understanding the energetics of water molecule binding is of great biochemical interest.

Molecular dynamics (MD) simulations implemented into the thermodynamic integration and perturbation methods were used successfully to compute free energies of individual water molecules transferred from the bulk solvent into protein cavities.[13–15] Normal-mode analysis applied for an isolated buried water in BPTI (bovine pancreatic trypsin inhibitor) gave the estimates how the rotational and translational degrees of freedom of a single water molecule were transformed into vibrational modes in the protein/water complex.[16,17] Applied to the stepwise protein hydration, general Molecular mechanics (MM) and quantum mechanics (QM) computational schemes helped in analyzing and interpreting the measured data.[12,18]

However, today our knowledge of internal hydration of proteins is far from being complete. Only a few theoretical approaches are used to calculate characteristics of buried waters. Therefore, new insights into this phenomenon are desirable to gain a deeper understanding of the role of buried water molecules in protein structures.

In the present paper, we consider the internal hydration of a protein in terms of ligand binding to a biopolymer. Detected buried waters in the protein interior are viewed as bound to the hydration sites located within the cavities and clefts. The energy of an individual water molecule bound to the protein is calculated using a dielectric continuum approach. The obtained energies are used in the binding polynomial to calculate the constants for the hydration reactions and hydration free energies. The obtained energies are also used to calculate the population probabilities of the hydration sites and compute their titration curves. For testing the method, we study the sequential water binding to the BPTI protein in the gas-phase and compare our results with available measured data.

† Presented at the 81st International Bunsen Discussion Meeting on "Interfacial Water in Chemistry and Biology", Velen, Germany, September 19–23, 2003.

DOI: 10.1039/b313238m

2 Theory and methods

2.1 Binding polynomial

The internal hydration of a protein can be regarded as binding n water molecules to the protein interior, which is represented by the reaction

$$\text{Protein} + n(H_2O) \leftrightarrow \text{Protein}(H_2O)_n \qquad (1)$$

We consider internal hydration sites as host and water molecules as ligands. This two component system can be described by a binding polynomial.[19] For a fixed number of sites, a binding polynomial corresponds to a semigrand canonical partition sum representing the thermodynamic system that is open with respect to the ligand particles and closed with respect to the binding sites.[19]

The binding polynomial Ξ is defined as the sum over all possible states of the system with M binding sites hydrated by up to N water molecules,

$$\Xi = 1 + \sum_{n=1}^{N} \lambda^n Q_n. \qquad (2)$$

In eqn. (2), each $Q_n = Q(n,M,T)$ is a canonical partition sum corresponding to the macrostate with n water molecules bound. λ denotes the activity of water molecules. Thermodynamical parameters of the macrostate depend on temperature T.

A partition sum Q_n is calculated according to

$$Q_n = \sum_{i_1} \exp\left(-\frac{G_{i_1}}{RT}\right) \sum_{i_2} \exp\left(-\frac{G_{i_2}}{RT}\right) \exp\left(-\frac{W_{i_1 i_2}}{RT}\right) \cdots \sum_{i_n} \exp\left(-\frac{G_{i_n}}{RT}\right) \exp\left(-\frac{W_{i_1 i_n} + \ldots + W_{i_{n-1} i_n}}{RT}\right) \qquad (3)$$

Here G_i is the energy related to the site i occupied by a water molecule, and W_{ij} is the energy related to the interaction of two occupied sites i and j. R is the molar gas constant.

Two equivalent expressions can be applied for the activity of water molecules, using either the chemical potential μ or the concentration related quantity pW,

$$\lambda = \exp(\mu/RT) = 10^{-\text{pW}} \qquad (4)$$

Eqns. (2)–(4) are commonly used in studies of noncovalent association of ions or small molecules with biopolymers,[20] assuming that each binding contact can be represented by an individual binding site. As soon as the energies of the binding sites in all microstates and the ligand activity are known, one can calculate the binding reaction free energies and the binding probabilities using the polynomial Ξ. For example in case of proton binding, the knowledge of energies for protonating residues (proton binding sites) in all microstates together with the specified proton activity pH allows us to calculate the macroscopic pK-values and titration curves.[21]

2.2 Energies of bound water molecules

The energy of a water molecule bound at a particular hydration site within the protein is the free energy to transfer it from the gas phase to the protein plus the interactions with other incorporated waters. The protein is considered as a dielectric continuum with embedded partial charges. The transfer of a water ligand molecule of a given orientation to the protein is divided into a series of separate steps, as shown in Fig. 1. The molecule is discharged in the gas phase, transferred from the gas phase into the protein and recharged in the protein.[22,23]

The electrostatic contribution to the energy of a bound water molecule is expressed as the difference between the work of discharging and the work of recharging, like in calculations of the solvation free energies.[24,25] It is given by

$$\Delta\Delta G^{\text{ele}}_{\text{g,p}} = \Delta G^{\text{recharging}}_{\text{p,p}} - \Delta G^{\text{discharging}}_{\text{g,g}} \qquad (5)$$

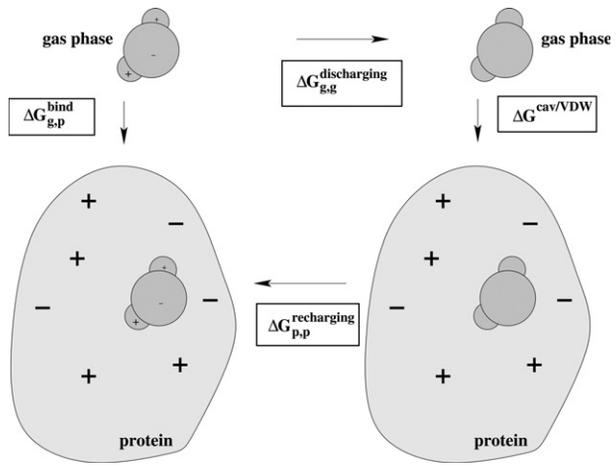

Fig. 1 Thermodynamic cycle showing the dissected contributions to the free energy to transfer a water ligand molecule from the gas phase to the protein interior. In the first step, a ligand molecule is discharged in the gas phase to form a hypothetical nonpolar molecule; the corresponding free energy change is $\Delta G^{\text{discharging}}_{\text{g,g}}$. In the second step, a hypothetical molecule is transferred from the gas phase to the protein; this step includes switching off the ligand-gas phase VDW interactions, formation of a ligand-sized cavity in the protein, and switching on the ligand-protein VDW interactions; the corresponding free energy change is $G^{\text{cav/VDW}}$. In the third step, a ligand molecule is re-charged in protein; the corresponding free energy change is $\Delta G^{\text{recharging}}_{\text{p,p}}$.

The nonpolar contribution $G^{\text{cav/VDW}}$ to the energy of a bound water molecule includes the van der Waals (VDW) dispersion term as well as the free energy to form a ligand-sized cavity at a given solvent density.[26,27] We will take the nonpolar energy into account as a constant value α. Then, the total energy of a bound water molecule is given by

$$\Delta G^{\text{bind}}_{\text{g,p}} = \Delta\Delta G^{\text{ele}}_{\text{g,p}} + \alpha. \qquad (6)$$

Thermodynamic cycle shown in Fig. 1 does not include the entropy change caused by an association of a ligand and the protein. When a single water molecule binds, its nine degrees of freedom (three per atom) are transformed from six translational and rotational degrees of freedom and three internal vibrational modes into nine vibrational modes in the complex.[17] Another source of the binding entropy change is mixing the distinct components.[26] In the binding polynomial approach, it is sufficient to account for the binding entropy change as a constant value Δs referred to an occupied binding site (the small Δs means the entropy change per bound molecule).

Then, the energy of the occupied binding site i, G_i in eqn. (3), takes the form

$$G_i = (\Delta\Delta G^{\text{ele}}_{\text{g,p}} + \alpha - T\Delta s)_i. \qquad (7)$$

2.3 Hydration free energies

Hydration free energies and constants for the hydration reactions measured experimentally are usually related to the sequential hydration, when events of binding a water molecule are detected one by one. Sequential hydration of the protein is represented by a set of reactions

$$\text{Protein}(H_2O)_{n-1} + H_2O \xleftrightarrow{K_n} \text{Protein}(H_2O)_n. \qquad (8)$$

The relationship between constants for the hydration reactions K_n in eqn. (8) and the partition sums Q_n in eqn. (1) are given by the following equations

$$K_n = \frac{Q_n}{Q_{n-1}}; \quad Q_n = \prod_{m=1}^{n} K_m. \qquad (9)$$

The binding polynomial expressed over K_n, which depends on the number of binding sites M and temperature T, takes the form[20,28]

$$\Xi = 1 + \lambda K_1 + (\lambda K_1)(\lambda K_2) + \cdots + \prod_{n=1}^{N} (\lambda K_n). \quad (10)$$

The hydration free energies for the reactions in eqn. (8), are calculated with the equation

$$\Delta G_n = -RT \ln K_n. \quad (11)$$

Here, ΔG_n is the standard reaction free energy of binding one water molecule to the protein that has already $(n-1)$ waters bound. It is independent of the water molecule activity.

The standard condition for the hydration reactions in eqn. (8) is the solution standard state with $C^0 = 1$ M for the ligand particle concentration.[29,30] In this state, $\mu = 0$ and pW $= 0$ in eqn. (4) yield $\lambda = 1$.

2.4 Population probabilities of the hydration sites

We next consider the population probabilities of the hydration sites. To calculate them as the functions of the water activity, we use a statistical equivalence of the system of M hydration sites in a protein surrounded by free water molecules and an ensemble of M spins of a ferromagnet under the influence of an external field. Here, 'the lattice-gas interaction rule'[19] is imposed: any pair of sites interacts as soon as they are both occupied by water molecules.

The current hydration state h of a protein with M distinguishable hydration sites is represented by a vector of M components, $\boldsymbol{x}^h = (x_1^h, x_2^h, \ldots x_M^h)$, with the two possibilities for each x_i^h to be equal to either 1 or 0 depending on whether the site i is occupied by water or empty. Hence, the population probability $\langle x_i \rangle$ of the site i is calculated as its average occupancy[31]

$$\langle x_i \rangle = \frac{\sum_{h=1}^{2^M} x_i^h \exp\left(-\tilde{G}_h(x_i^h)/RT\right)}{\sum_{h=1}^{2^M} \exp\left(-\tilde{G}_h(x_i^h)/RT\right)} \quad (12)$$

The introduced energies $\tilde{G}_h(x_i^h)$ depend on the ligand activity and can be expressed as follows[31,32]

$$\tilde{G}_{\mathrm{h}} = \sum_{i=1}^{M} x_i^h (G_i - \mu) + \sum_{j>i}^{M} \sum_{i=1}^{M-1} x_i^h x_j^h W_{ij} \quad (13)$$

2.5 Computational details

The water ligand binding approach outlined above is applied to the internal hydration of the BPTI. Experimental data obtained with the drift cell technique combined with the mass-spectrometric measurements are available for the first three hydration steps of BPTI in the gas phase.[11] The logarithm of the equilibrium constant, $\ln K_n^{\mathrm{exp}}$, has been plotted at several temperatures and a given pressure of 4.8 mTorr. The measurements are related to the reactions of a single water binding shown in eqn. (8).

For simulations, we use a BPTI crystal structure[3] (PDB code: 5PTI). The total charge of the protein is +6 and corresponds to the charge state of BPTI used in the experimental studies.[11] All waters, except W111, W112, W113 and W122 (buried waters conserved in all high-resolution crystal structures) were removed from the PDB file.

We use the following computational scheme to obtain the hydration free energies. First, we calculate for every water molecule $\Delta\Delta G_{\mathrm{g,p}}^{\mathrm{ele}}$ and W_{ij} (see subsection 2.2) in a continuum approach using the Poisson equation. Explicit equations for the computations are analogous to those for proton binding.[32] The water molecules are represented by the TIP3P model.[33] Second, we use the obtained energies to calculate the partition sums up to Q_n according to eqn. (3). Here, only electrostatic components of G_i and W_{ij} are taken into account. Third, we use the obtained partition sums to calculate constants for the hydration reactions K_n^{ele} according to eqn. (9) and record a set of the hydration free energies $\Delta G_n^{\mathrm{ele}}$ in eqn. (11). Finally, we derive the nonpolar energy α and the binding entropy change Δs by fitting the measurements and re-calculate K_n and ΔG_n.

The nonpolar energy is significantly smaller than the electrostatic contributions (see Section (3)). Therefore, α can be taken the same for every bound water molecule. This reasoning suggests also the interaction energies, W_{ij}, of water molecules bound to sites i and j assumed to be purely electrostatic. The binding entropy change is not small, but its values in the sequential hydration do not vary drastically. Therefore, Δs also can be taken the same for each bound water molecule.

The computational scheme to obtain the population probabilities of the hydration sites again uses the energies of water molecules bound to the protein together with the chemical potential of water vapor at constant pressure. Calculations rely on eqns. (12) and (13).

Constants for the hydration reactions in eqn. (9), hydration free energies in eqn. (11), and the population probabilities of the internal hydration sites in eqn. (12) are computed at nine different temperatures.

The numerical solution of the Poisson equation is calculated by a finite-difference method using the MEAD program suite.[34] Atomic charges and radii of the protein atoms are taken from the CHARMM force field[35] and from the Bondi[36] parameter set, respectively. In calculations, we use the dielectric constant $\varepsilon = 1$ for the gas phase and $\varepsilon = 4$ for the protein. The last value has been used extensively in the continuum electrostatic calculations for protonated residues and provided a good agreement with measurements.[32,34]

3 Results and discussion

BPTI is a small protein composed of 58 amino acids that form a single polypeptide chain. Because of the relatively small number of atoms and many high resolution structures available, it is among the most popular systems used both in experimental[6,10] and computational[37–39] studies of various aspects of protein dynamics and function.

There are four internal water molecules conserved in all high-resolution crystal structures of BPTI.[3,4] These water molecules are also detected by NMR in solution.[5,6] The geometry of the hydrated interior of BPTI is shown in Fig. 2. The three water molecules, W111, W112 and W113, are located in a large cleft and form a hydrogen bonded cluster. They are hydrogen-bonded to PRO8, TYR10, ASN43, LYS41 and ASN44.[4] One isolated structural water molecule, W122, is located in a small protein cavity. It is hydrogen bonded to CYS38, CYS14, THR11.[4]

Constants for the hydration reactions and hydration free energies

Following the experimentalists,[11,18] we assume that in the sequential hydration setup, the four buried water molecules conserved in the BPTI crystal structures are the most likely participants of the first hydration steps, rather than those located at the exposed protein surface. Arguments supporting this statement rely on the hydration energy values for buried and exposed water binding sites.[11,40]

Experimental constants for the hydration reactions have been determined for the first three hydration steps. In our

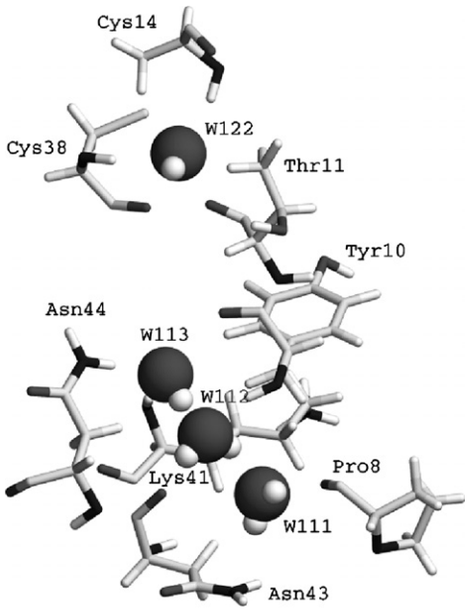

Fig. 2 Geometry of the hydrated interior of the BPTI with four conserved water molecules and surrounding residues.

calculations, we consider the binding of four water molecules. Thus, the first three calculated binding constants correspond to the three experimental constants for the hydration reactions. The binding polynomial with $M = N = 4$ has the form

$$\Xi = 1 + K_1^{\mathrm{ele}}\frac{\lambda}{\kappa} + K_1^{\mathrm{ele}}K_2^{\mathrm{ele}}\frac{\lambda^2}{\kappa^2} + K_1^{\mathrm{ele}}K_2^{\mathrm{ele}}K_3^{\mathrm{ele}}\frac{\lambda^3}{\kappa^3} + K_1^{\mathrm{ele}}K_2^{\mathrm{ele}}K_3^{\mathrm{ele}}K_4^{\mathrm{ele}}\frac{\lambda^4}{\kappa^4} \qquad (14)$$

In eqn. (14), $\kappa = \exp((\alpha - T\Delta s)/RT)$. Assuming the same κ for every bound water molecule, we have a common multiplier $1/\kappa$ in eqn. (3) powered in n in the partition with n water molecules bound.

The electrostatic components of the binding free energies of the four conserved buried water molecules and the interaction energies are listed in Table 1. Their absolute values are large and exceed significantly the typical values of a nonpolar contribution which is about 2–3 kcal mol^{-1} as recently obtained from the solubility of water in liquid hydrocarbons.[41]

The small nonpolar energies allow us to use the same value of α in eqn. (6) for every bound water molecule, which makes calculations of the partition sums and the binding constants more feasible. Moreover, it is of advantage to use averaged α, if one takes it from fitting the experimental data, because several conformations of the protein have contributed to mass-spectra.[11] Indeed, the dominating electrostatic interactions are long-ranged and expected to be less sensitive to conformational fluctuations, while the short-range interactions contributing to the VDW and cavity formation energies require averaging.

The calculated $\ln K_n^{\mathrm{ele}}$ and derived $\Delta G_n^{\mathrm{ele}}$ for the four hydration steps are given in Table 2 at various temperatures. It contains also the experimental $\ln K_n^{\mathrm{exp}}$, extracted from mass-spectra, together with derived $\Delta G_n^{\mathrm{exp}}$.

Table 1 Electrostatic contributions to the energies of four bound water molecules in BPTI. All values are given in kcal mol^{-1}

Site	G_i^{ele}	$W_{ij}, j > i$			
W111	−10.21	0	−1.54	−0.31	−0.04
W112	−8.07		0	−5.06	−0.06
W113	−9.42			0	−0.09
W122	−7.83				0

Table 2 Calculated $\ln K_n^{\mathrm{ele}}$ and measured $\ln K_n^{\mathrm{exp}}$ (from ref. 11) for the sequential hydration of BPTI in the gas phase at different temperatures, together with the corresponding binding free energies, $\Delta G_n^{\mathrm{ele}}$ and $\Delta G_n^{\mathrm{exp}}$, respectively

Step	0 → 1	1 → 2	2 → 3	3 → 4
T/K	$\ln K_n^{\mathrm{ele}}$			
228.3	24.02	22.7	22.01	17.49
232.6	23.56	22.29	21.61	17.16
238.1	22.99	21.79	21.12	16.77
243.4	22.48	21.33	20.68	16.41
248.8	21.96	20.87	20.23	16.05
253.2	21.57	20.52	19.89	15.76
257.7	21.17	20.17	19.55	15.49
263.2	20.71	19.76	19.15	15.17
270.3	20.14	19.26	18.65	14.77
T/K	$\ln K_n^{\mathrm{exp}}$			
228.3	15.8	13.5	12.9	—
232.6	14.7	13.0	12.3	—
238.1	13.8	12.2	11.8	—
243.4	12.9	11.7	11.1	—
248.8	12.0	11.1	10.2	—
253.2	11.2	10.1	10.0	—
257.7	10.3	9.9	9.5	—
263.2	9.2	9.1	8.8	—
270.3	8.5	8.3	8.3	—
T/K	$\Delta G_n^{\mathrm{ele}}$/kcal mol^{-1}			
228.3	−10.90	−10.30	−9.986	−7.933
232.6	−10.89	−10.30	−9.986	−7.930
238.1	−10.88	−10.31	−9.993	−7.932
243.4	−10.87	−10.32	−10.00	−7.936
248.8	−10.86	−10.32	−10.00	−7.931
253.2	−10.85	−10.32	−10.00	−7.930
257.7	−10.84	−10.33	−10.01	−7.933
263.2	−10.83	−10.33	−10.01	−7.930
270.3	−10.82	−10.34	−10.02	−7.930
T/K	$\Delta G_n^{\mathrm{exp}}$/kcal mol^{-1}			
228.3	−7.168	−6.124	−5.852	—
232.6	−6.793	−6.007	−5.684	—
238.1	−6.529	−5.772	−5.583	—
243.4	−6.240	−5.659	−5.369	—
248.8	−5.931	−5.486	−5.042	—
253.2	−5.634	−5.081	−5.030	—
257.7	−5.275	−5.070	−4.865	—
263.2	−4.811	−4.758	−4.601	—
270.3	−4.565	−4.457	−4.457	—

Three major contributions cause the difference between $\Delta G_n^{\mathrm{exp}}$ and $\Delta G_n^{\mathrm{ele}}$: a correction term to adjust the difference in standard states used in measurements and calculations,[42,43] the nonpolar energy of a bound water molecule, and the binding entropy change. The relationship has the form

$$\Delta G_n^{\mathrm{exp}} = \Delta G_n^{\mathrm{ele}} + RT\ln\frac{p_w}{p^0} + \alpha - T\Delta s \qquad (15)$$

The correction term, $RT\ln(p_w/p^0)$, is the free energy of an ideal gas going from a pressure p^0 corresponding to a density of 1 M to a given pressure $p_W = 6.3 \times 10^{-6}$ atm used in the experiment. This term adjusts the difference between the gaseous standard state of $p_0 = 1$ atm and the solution standard state of $C^0 = 1$ M.[42] The nonpolar energy α and the binding entropy change Δs in eqn. (15) has been discussed above. They

come from the $1/\kappa$ multiplier in eqn. (14), $\alpha - T\Delta s = -RT \ln(1/\kappa)$.

We use eqn. (15) together with data in Table 2 to get two unknowns, α and Δs, from the least squares solution of 27 linear equations of the form

$$\alpha - T\Delta s = \Delta G_n^{\text{exp}} - \Delta G_n^{\text{ele}} - RT \ln \frac{p_w}{p^0}, \qquad (16)$$

written at 9 temperatures for each of 3 hydration steps. The fitting procedure yields $\alpha = 0.31$ kcal mol^{-1} and $\Delta s = -45.7$ cal mol^{-1} K^{-1}. The minimized least squares deviations norm is 0.5. It is an overall error of fitting for both parameters.

Looking at eqn. (15), we can say that the calculated relative free energies, $\Delta G_n^{\text{ele}} - \Delta G_{n-1}^{\text{ele}}$, reproduce well the sequential hydration and temperature trends observed experimentally. Indeed, being referred to the standard state for measurements, they are equal to the shifted and titled experimental relative free energies, $\Delta G_n^{\text{exp}} - \Delta G_{n-1}^{\text{exp}}$. Deviation is within 0.5 kcal mol^{-1}. This finding gives credence to the approach we have implemented.

Population probabilities of the four hydration sites in BPTI

We next discuss the titration of four water binding sites in BPTI and their population probabilities. To set the activity range for the titration, we calculate the values of the chemical potential μ of water vapor at given pressure of 4.8 mTorr and T ranging from 228.3 K to 270.3 K. The data for the excess chemical potential of liquid water $\mu_{\text{ex}}(H_2O,l)$ at $p = 1$ atm and various T are taken from ref. 43 (authors have used the expanded ensemble method in MC and MD simulations). Starting from the liquid state at $p = 1$ atm, we consider a process where water is evaporated at a given temperature T followed by a vapor isothermal expansion from 1 atm to 6.3×10^{-6} atm. The chemical potential of water vapor $\mu(H_2O,g)$ at constant pressure is given by

$$\mu = \mu_{\text{ex}} + \mu_{\text{evap}} + \mu_{\text{expan}} \qquad (17)$$

The equation of state for the perfect gas is used to obtain μ_{expan}. For μ_{evap} the value is taken from the literature.[44] Table 3 contains data for the calculated μ together with pW, which is related to the concentration of water and is given by

$$\text{pW} = -\frac{1}{\ln(10)} \frac{\mu}{RT}. \qquad (18)$$

The population probabilities of four hydration sites in BPTI are computed according to eqn. (12). At different temperatures, the obtained $\langle x_i \rangle$ as the functions of pW show a similar behavior, therefore the titration curves only at two temperatures are given in Fig. 3.

The most striking result is that at any temperature three water molecules bind to the protein nearly simultaneously followed by the binding of a fourth water. The titration curves of strongly interacting water molecules W111, W112 and W113 in Fig. 3 reach their half-hydration values at pW $\approx$ 9.3 (given $T = 243.4$ K). The titration curve of the isolated water molecule W122 takes a half-hydration value at pW = 7.1 (given $T = 243.4$ K).

For every hydration site the population probabilities $\langle x_i \rangle$ show the temperature dependence as a function of pW. This observation is in agreement with the temperature dependence of the binding constants for four hydration steps (see $\ln K_n^{\text{ele}}$ in Table 1). Also, the small changes in $\langle x_i \rangle$ going from $T = 232.6$ K to $T = 243.6$ K in Fig. 3 are in accordance with the linear temperature dependence of pW in Table 3.

Now we compare the activity values in the calculations with those used in the measurements. Again, we need to account for the difference in standard states used in the binding polynomial approach and in the experimental setup. There is a shift in the chemical potential in eqn. (17) with respect to the gaseous standard state, $p_0 = 1$ atm. The relationship between μ used in calculations (see Table 3) and μ', referred to measurements, is given by

$$\mu' = \mu + RT \ln \frac{p_w}{p^0}, \qquad (19)$$

The positive shift in μ, eqn. (19), causes the negative shift in pW according to eqn. (18). Therefore, the pW′ values, referred to measurements, are $10.14 - 6.49 = 3.65$ (at $T = 232.6$ K) and $9.78 - 6.51 = 3.27$ (at $T = 243.4$ K). In both cases, all four sites are hydrated with the probability 1, as shown in Fig. 3.

As discussed above, binding a water W122 to the protein cavity takes place only after a cluster of other three waters has formed. Therefore, the measurements of the first three constants for the hydration reactions in ref. 11 are very likely attributed to the cluster of the water molecules W111, W112 and W113. However, finding the sequence of these reactions requires additional studies.

Our calculations demonstrate that a spatially isolated water molecule W122 is represented by a sole titration curve corresponding to a hydration site that weakly interacts with others. At the same time, a spatial cluster of hydrogen bonded water molecules W111, W112 and W113 is represented by the titration curves of strongly interacting hydration sites.

As any thermodynamic process, the internal hydration of BPTI is driven by the free energy changes coming from the initial state of the fully dehydrated protein to the final state with four water molecules bound. The locations of the hydration sites are imposed by the geometry of the protein, so that four water molecules occupy a cleft and a cavity. Thus, internal hydration of the protein can be viewed as a phenomenon caused by the energetics and geometry as well. The cluster of the bound water molecules in the cleft and the atoms of nearby protein residues form a hydrogen-bonded network. It is a consequence of the three steps of hydration of the dry BPTI. The hydrogen-bonded network involving the cluster of three water molecules is the structural basis for the high affinity of this hydration site cluster for water.

4 Conclusions

In this paper, a ligand binding approach is outlined to study the hydration of protein cavities and clefts. For testing the method, we consider the binding of water molecules to the BPTI interior. Calculations give correct values of the constants for the hydration reactions and hydration free energies for the water binding to the protein in the gas phase. The difference between the calculated hydration free energies and those extracted from mass-spectra is about 0.5 kcal mol^{-1}.

The first three steps of BPTI hydration in the gas phase correspond to the binding of three strongly interacting water

Table 3 Calculated water vapor chemical potential μ and pW values at a given pressure $p_W = 4.8$ mTorr and various temperatures. Data for μ_{ex} at the standard pressure $p_0 = 1$ atm are taken from ref. 43

T/K	228.3	232.6	238.1	243.4	248.8	253.2	257.7	263.2	270.3
μ/kcal mol^{-1}	−10.76	−10.79	−10.84	−10.89	−10.94	−10.97	−11.01	−11.06	−11.12
pW	10.30	10.14	9.95	9.79	9.61	9.47	9.34	9.19	8.99
μ_{ex}/kcal mol^{-1}	−7.38	−7.31	−7.23	−7.14	−7.07	−7.00	−6.93	−6.85	−6.75

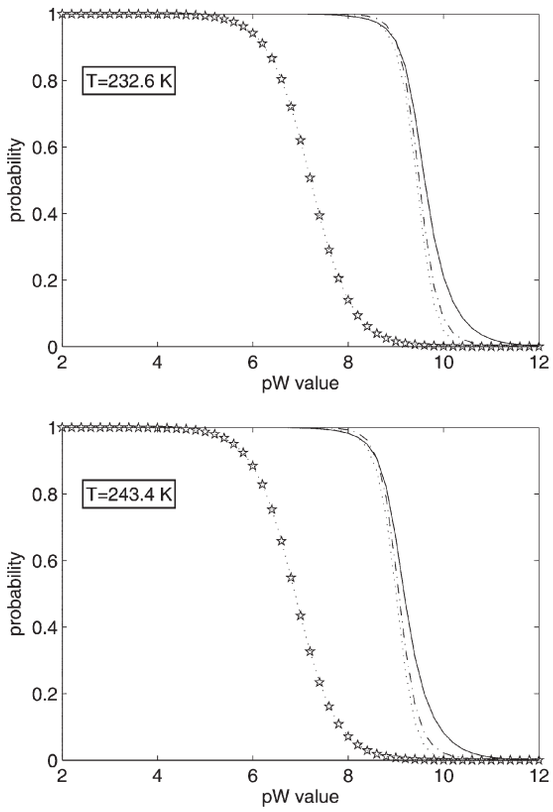

Fig. 3 Population probabilities $\langle x_i \rangle$ of four internal hydration sites of BPTI as the functions of pW at two different temperatures. Data for W111, W112, and W113 are depicted with the solid, dashed and dash–dot lines respectively. Data for W122 are represented by the dotted lines marked with pentagrams.

molecules located in a protein cleft. An isolated buried water molecule occupying a small hydrophobic protein cavity is bound with a lower affinity than the cluster of water molecules in cleft. Electrostatic interaction of water molecules with the protein and with each other is the driving force of the internal hydration of BPTI.

The method that we have presented here gives a good agreement with measurements, reproducing well the sequential hydration and temperature trends observed experimentally. We think that this approach provides a general way of computing affinities of water binding sites and can also be applied to other proteins.

Acknowledgements

This work was supported by the DFG (UL174/2 and UL174/4). Authors thank Prof. Donald Bashford for providing his program MEAD and Nicolas Calimet for discussions. A.I. Borodich is grateful to Prof. Alexander Lyubartsev for sending his articles and helpful discussions. The computation was done on the HELICS supercomputer (IWR, Heidelberg University).

References

1 M. A. Williams, J. M. Goodfellow and J. M. Thornton, *Protein Sci.*, 1994, **3**, 1224.
2 A. A. Rashin, B. H. Rashin, A. Rashin and R. Abagyan, *Protein Sci.*, 1997, **6**, 2143.
3 A. Wlodawer, J. Walter, R. Huber and L. Sjölin, *J. Mol. Biol.*, 1984, **180**, 301.
4 A. Wlodawer, J. Deisenhofer and R. Huber, *J. Mol. Biol.*, 1987, **193**, 145.
5 G. Otting and K. Wuthrich, *J. Am. Chem. Soc.*, 1989, **111**, 1871.
6 G. Otting, E. Liepinsh and K. Wuthrich, *J. Am. Chem. Soc.*, 1991, **113**, 4363.
7 J. A. Ernst, R. T. Club, H.-X. Zhou, A. M. Gronenborn and G. M. Clore, *Science*, 1995, **267**, 1813.
8 J. A. Ernst, R. T. Club, H.-X. Zhou, A. M. Gronenborn and G. M. Clore, *Science*, 1995, **270**, 1848.
9 B. Halle and V. P. Denisov, *Biophys. J.*, 1995, **69**, 242.
10 V. P. Denisov, K. Venu, J. Peters, H. D. Horlein and B. Halle, *J. Phys. Chem. B*, 1997, **101**, 9380.
11 J. Woenckhaus, R. R. Hudgins and M. F. Jarrold, *J. Am. Chem. Soc.*, 1997, **119**, 9586.
12 D. Liu, T. Wyttenbach, P. E. Barran and M. T. Bowers, *J. Am. Chem. Soc.*, 2003, **125**, 8458.
13 L. Zhang and J. Hermans, *Proteins: Struct. Funct. Gen.*, 1996, **24**, 433.
14 V. Helms and R. C. Wade, *Proteins: Struct. Funct. Gen.*, 1998, **32**, 381.
15 B. Roux, M. Nina, R. Pomes and J. C. Smith, *Biophys. J.*, 1996, **71**, 670.
16 S. Fischer and C. S. Verma, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 9613.
17 S. Fischer, J. C. Smith and C. S. Verma, *J. Phys. Chem. B*, 2001, **105**, 8050.
18 Y. Mao, M. A. Ratner and M. F. Jarrold, *J. Am. Chem. Soc.*, 2000, **122**, 2950.
19 T. L. Hill, *An Introduction to Statistical Thermodynamics,*, Dover Publ. Inc., New York, 1986.
20 D. Poland, *Biopolymers*, 2001, **58**, 477.
21 G. M. Ullmann, *J. Phys. Chem. B*, 2003, **107**, 1263.
22 W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127.
23 D. Sitkoff, K. A. Sharp and B. Honig, *J. Phys. Chem.*, 1994, **98**, 1978.
24 G. Hummer, L. R. Pratt and A. E. Garcia, *J. Phys. Chem.*, 1995, **99**, 14188.
25 M. Nina, D. Beglov and B. Roux, *J. Phys. Chem. B*, 1997, **101**, 5239.
26 R. A. Pierotti, *Chem. Rev.*, 1976, **76**, 717.
27 M. Ikeguchi, S. Shimizu, S. Nakamura and K. Shimizu, *J. Phys. Chem. B*, 1998, **102**, 5891.
28 D. Poland, *J. Chem. Phys.*, 2000, **113**, 4774.
29 M. K. Gilson, J. A. Given, B. L. Bush and J. A. McCammon, *Biophys. J.*, 1997, **72**, 1047.
30 P. Bongrand, *Rep. Prog. Phys.*, 1999, **62**, 921.
31 P. Beroza, D. R. Fredkin, M. Y. Okamura and G. Feher, *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 5804.
32 G. M. Ullmann and E. W. Knapp, *Eur. Biophys. J.*, 1999, **28**, 533.
33 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
34 D. Bashford and K. Gerwert, *J. Mol. Biol.*, 1992, **224**, 473.
35 A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586.
36 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441.
37 B. Brooks and M. Karplus, *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 6571.
38 R. M. Brunne, E. Liepinsh, G. Otting, K. Wuthrich and W. F. van Gunsteren, *J. Mol. Biol.*, 1993, **231**, 1040.
39 C. A. Schiffer and W. F. van Gunsteren, *Proteins: Struct. Funct. Gen.*, 1999, **36**, 501.
40 J. Woenckhaus, Y. Mao and M. F. Jarrold, *J. Phys. Chem. B*, 1997, **101**, 847.
41 C. Amovilli and F. M. Floris, *Phys. Chem. Chem. Phys.*, 2003, **5**, 363.
42 G. Hummer, L. R. Pratt and A. E. Garcia, *J. Phys. Chem.*, 1996, **100**, 1206.
43 A. P. Lyubartsev, A. Laaksonen and P. N. Vorontsov-Velyaminov, *Mol. Phys.*, 1994, **82**, 455.
44 P. W. Atkins, *Physical Chemistry*, Oxford University Press, Oxford, 5th edn., 1994.