Biochimica et Biophysica Acta 1850 (2015) 923–931

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Solvation and cavity occupation in biomolecules ☆

Gillian C. Lynch *, John S. Perkyns, Bao Linh Nguyen, B. Montgomery Pettitt *

Sealy Center for Structural Biology and Molecular Biophysics, Departments of Biochemistry and Molecular Biology and Pharmacology and Toxicology, The University of Texas Medical Branch at Galveston, 301 University Blvd, Galveston, TX 77555-0304, USA

ARTICLE INFO

Article history:
Received 25 July 2014
Received in revised form 15 September 2014
Accepted 17 September 2014
Available online 28 September 2014

Keywords:
Integral equations
Proximal radial distribution functions
Molecular dynamics
Solvation

ABSTRACT

Background: Solvation density locations are important for protein dynamics and structure. Knowledge of the preferred hydration sites at biomolecular interfaces and those in the interior of cavities can enhance understanding of structure and function. While advanced X-ray diffraction methods can provide accurate atomic structures for proteins, that technique is challenged when it comes to providing accurate hydration structures, especially for interfacial and cavity bound solvent molecules.
Methods: Advances in integral equation theories which include more accurate methods for calculating the long-ranged Coulomb interaction contributions to the three-dimensional distribution functions make it possible to calculate angle dependent average solvent structure, accurately, around and inside irregular molecular conformations. The proximal radial distribution method provides another approximate method to determine average solvent structures for biomolecular systems based on a proximal or near neighbor solvent distribution that can be constructed from previously collected solvent distributions. These two approximate methods, along with all-atom molecular dynamics simulations are used to determine the solvent density inside the myoglobin heme cavity.
Discussion and results: Myoglobin is a good test system for these methods because the cavities are many and one is large, tens of $Å^3$, but is shown to have only four hydration sites. These sites are not near neighbors which implies that the large cavity must have more than one way in and out.
Conclusions: Our results show that main solvation sites are well reproduced by all three methods. The techniques also produce a clearly identifiable solvent pathway into the interior of the protein.
General significance: The agreement between molecular dynamics and less computationally demanding approximate methods is encouraging.
This article is part of a Special Issue entitled Recent developments of molecular dynamics.

1. Introduction

The solvation of biomolecules is integral to their function. A number of theoretical and computational methods are widely used to complement and interpret experimental results [1]. From the most conceptually simple and computationally inexpensive continuum approximations to an all-atom simulation, computational methods derive thermodynamic quantities and the average structure of the solvent determined by the average structure of the biomolecule itself. It is necessary, therefore, to study the solution properties in order to understand the basic behavior of the system. While significant insight into the average energetics of a biomolecule can be obtained by the use of continuum methods, it is only with a model involving explicit solvent molecules that we can hope to understand the local inter- and intramolecular interactions. This is particularly true for systems with distinct interior cavities with hydration sites that are only transiently occupied. We need to be able to determine the three-dimensional solvent distribution for various biomolecular species. Molecular recognition occurs at interfacial regions. Mutation of a single residue can be the cause of disease or dysfunction. Changes in molecular structure can have effects on the formation of interfaces even when the mutation is far from the interface by changing the solution thermodynamics.

In this article we compare three different methods of computing the solvation of proteins. The methods can be ranked in the order of the complexity of the correlations considered by each. We consider methods with a range of computational complexity and accuracy. Molecular dynamics simulation is a method which in principle is capable of providing most solution thermodynamics for a given potential or force field model and includes all intermolecular correlations of the sample [2]. It is however computationally costly when considering free energy related properties including protein structure prediction [3,4]. Many-body methods such as integral equations provide solvation thermodynamics and free energetics with several orders of magnitude

☆ This article is part of a Special Issue entitled Recent developments of molecular dynamics.

* Corresponding authors at: The University of Texas Medical Branch at Galveston, 301 University Blvd, Galveston, Texas 77555-0304, USA.
E-mail addresses: gclynch@utmb.edu (G.C. Lynch), mpettitt@utmb.edu (B.M. Pettitt).
URL: http://bmb.utmb.edu/pettitt (B.M. Pettitt).

http://dx.doi.org/10.1016/j.bbagen.2014.09.020
0304-4165/

less computation [5]. However, integral equations are inherently approximate [6] and the validity of the approximations must be tested for each class of system. While continuum solvent methods are very computationally inexpensive, their errors are due to approximations which limit or ignore certain intermolecular correlations [7]. Accurate solvation correlations are required to obtain many thermodynamic properties.

Our goal is to determine the solvent probability distribution which includes the location and occupancy of the hydration sites [8,9]. We will use the results of an all-atom molecular dynamics (MD) simulation to provide the most complete and accurate picture of our model systems. The ability to obtain three-dimensional solvent representation for any biomolecular system with the same details of an MD simulation but without the long time and large computer resources is a desirable goal. Two other methods of differing detail and computational complexity, 3-D integral equations and proximal distribution reconstructions are presented here and compared to simulations. We will consider the case of a protein with internal cavities whose occupancies are in equilibrium with the bulk solvent.

The sperm whale myoglobin structure was the first protein crystallography structure solved and so makes a classic test case [10]. Myoglobin is a small heme protein, composed of 153-residues and functions as an intracellular oxygen storage molecule; it also binds other small ligands like CO and NO. The diffusion pathways and ligand binding/exchange mechanism have been investigated, experimentally and computationally, for many years [11–14]. Studies have shown the existence of a large cavity that includes the heme where the ligands bind. A site commonly referred to as the distal pocket has been shown to be often occupied by water [15,11,12]. There are also four distinct pockets or sites referred to as the Xenon binding sites, Xe_1, Xe_2, Xe_3, and Xe_4 [16]. The Xe binding sites have been used to support the idea of a series of hydrophobic cavities but there is kinetic evidence on internal water hindering ligand accessibility [17]. High resolution X-ray diffraction crystallographic structures of myoglobin [15,18] disagree with each other on the occupancy of the internal water molecules, as well as the number of internal water molecules or water sites. The presence of a water molecule near the distal pocket is the key factor in supporting one of the two widely accepted mechanisms for ligand exchange in myoglobin, the direct His gated with water displacement model. Quantitative determination of this functional hydration site occupancy and consequently its implications on the mechanism remains of interest experimentally [17,11] and computationally [19,12]. A spectro-kinetic assay [17] reported that the occupancy of the distal pocket by water is close to unity, and forms a favorable electrostatic interaction with His64, modulating the rate of ligands entering the protein via the direct gated model.

Recent magnetic relaxation dispersion (MRD) experiments [11] suggest that there are four water molecules with relatively long residence times in the cavity in the same volume as the Xe_1 and Xe_3 binding sites, in the apical location, close to the His64, and in a hydration site, below Xe_1. These internal water molecules were found to exchange on the microsecond time scale and thus speculated to have a functional role and not just be a structural factor. MD simulations [19,12] also found multiple water sites with low occupancy that were not identified by X-ray crystallographic structures. Both studies suggested a nanosecond time scale for the water exchange, which is at least an order of magnitude faster than the MRD experiments. The different observations and implications do not resolve the myoglobin cavity occupancy debate and makes myoglobin an attractive structural biological system to test our methods.

The protein myoglobin has no permanent channel through which a water molecule or ligands can fit; it requires dynamic fluctuations to achieve function. Due to the relatively rare nature of the functional fluctuations MD simulation would likely have to sample many rare fluctuations which would be prohibitively expensive. Methods that can efficiently sample the equilibrium properties like occupancy for these nearly isolated parts of the configuration integral are needed.

A method that we will be comparing here which is well suited to the study of solvent structure of systems with complex structure is the 3D integral equation method [20–23]. These methods have been used on a variety of polar molecules in aqueous solution [5,24]. Kovalenko and Hirata have worked on biochemical systems to examine the structure of solvent around proteins [25], as well as to locate water molecules in internal protein cavities [26]. While 3-D integral equation results have become increasingly useful they are inherently approximate [27]. Additional approximations to the method most notably to the Coulomb contributions have been made [28,29,25,26,23]

Another promising method we will compare with simulation in this work is density reconstructions using proximal distribution functions [8,9,30]. In this technique a proximal or near neighbor solvent distribution is collected for a set of solute atom types in a given chemical situation [31,8,32]. Those distributions, suitably normalized, are used to construct a solvation model that contains considerable correlation structure about the solute. Mechanical averages can be taken over the distributions and with a suitable free energy functional, rapid approximate free energies are obtainable [30].

We present the theoretical frameworks required next starting with integral equations and ending with simulation. We follow that with a comparison of the results. We conclude with a discussion on the approximations used and the computational convenience of the methods.

2. Integral equation theory

2.1. The integral equation system in 3 dimensions

The model system for which we wish to calculate three dimensional distribution functions and system thermodynamic quantities is a multisite solute protein dissolved at infinite dilution in a multisite solvent, like water. In order to calculate the solvent distribution around the protein solute we will need to know the bulk solvent structure. For this we use the solution of the dielectrically consistent reference interaction site model theory equatons in the hypernetted chain approximation (DRISM/HNC) [33,34], solved in the usual manner for the aqueous solvent alone, giving as its result pair correlation functions $h_{ij}(r_{ij})$, where i and j are any of the n_v sites on the solvent and r_{ij} is the intersite separation distance. Throughout this paper we will denote all solvent intersite functions with small letters.

The solute–solvent structure is written in terms of the direct correlation function $C_{ui}(x, y, z)$ and the total correlation function $H_{ui}(x, y, z)$ which are single valued over the 3-dimensional coordinate system. The subscript u represents the solute molecule and the subscript i represents one atomic site on a solvent molecule. Because the solute is treated as a single species there will be one function pair for each atomic site i present at finite concentration in the solvent, unless solvent molecule symmetry provides some degeneracy. Capital letters will be used to represent the solute–solvent functions.

It is convenient to write the closure equation in terms of the difference or the indirect correlation function

$$T_{ui}(x,y,z) = H_{ui}(x,y,z) - C_{ui}(x,y,z). \tag{1}$$

We write the solute–solvent intermolecular potentials as a sum of pair potential functions depending on interatomic distances, $r_{\alpha i}$,

$$U_{ui}(x,y,z) = \sum_{\alpha}^{n_u} u_{\alpha i}(r_{\alpha i}) \tag{2}$$

where the subscript α can represent any of the n_u atoms on the solute molecule, and the subscript i represents a site on a solvent molecule. The site–site intermolecular pair potentials are of the Lennard–Jones plus Coulomb form,

$$\beta u_{\alpha i}(r_{\alpha i}) = \frac{\beta q_\alpha q_i}{r_{\alpha i}} + 4\beta\epsilon_{\alpha i}\left[\left(\frac{\sigma_{\alpha i}}{r_{\alpha i}}\right)^{12} - \left(\frac{\sigma_{\alpha i}}{r_{\alpha i}}\right)^{6}\right], \tag{3}$$

with q_i being charge on the atom, $\sigma_{\alpha i}$ and $\epsilon_{\alpha i}$ the usual Lennard–Jones parameters and β the energy factor $1/k_BT$, with k_B being the Boltzmann constant and T the absolute temperature.

The HNC equation is considered over the entire 3-dimensional (x,y, z) space and can be written as,

$$C_{ui}(x,y,z) = exp\left\{-\sum_{\alpha}^{n_u}\beta u_{\alpha i}(r_{\alpha i}) + T_{ui}(x,y,z)\right\} - T_{ui}(x,y,z) - 1, \tag{4}$$

and we note that the sum over potential terms is for a single site i on the solvent with the all the sites on the solute.

The form of the OZ equation used is

$$\widetilde{H}\left(k_x,k_y,k_z\right) = \widetilde{C}\left(k_x,k_y,k_z\right)\left(\widetilde{w}(k) + \rho\widetilde{h}(k)\right), \tag{5}$$

where $\widetilde{H}_{ui}\left(k_x,k_y,k_z\right)$ is the i-th element of the $1 \times n_v$ dimensional matrix $\widetilde{H}\left(k_x,k_y,k_z\right)$. The matrix $\widetilde{w}$ has elements, $\widetilde{w}_{ij}(k) = sin\left(kd_{ij}\right)/kd_{ij}$ where k is the distance from the origin in Fourier space, and d_{ij} is the intramolecular distance between site i and site j on a solvent molecule. The value of each $w_{ii}(k)$, on the diagonal of the matrix, is 1. The matrix $\widetilde{h}(k)$ has elements $\widetilde{h}_{ij}(k)$ and the matrix ρ is a diagonal matrix of elements ρ_v, the solvent number density. The matrices ρ, $\widetilde{w}(k)$ and $\widetilde{h}(k)$ are all of order $n_v \times n_v$. The tilde notation denotes Fourier transforms, where

$$\begin{aligned}\widetilde{H}_{ui}\left(k_x,k_y,k_z\right) = \iiint exp\left(i\left(k_x x + k_y y + k_z z\right)\right)\\ \times H_{ui}(x,y,z)dxdydz.\end{aligned} \tag{6}$$

The functions $\widetilde{h}_{ij}(k)$ are also 3-dimensional Fourier transforms, but because they depend only on the intersite distance, their transforms simplify to zeroth order Hankel (spherical Bessel) transforms,

$$\widetilde{h}_{ij}(k) = 4\pi\int_0^\infty r_{ij}^2\frac{sin\left(kr_{ij}\right)}{kr_{ij}}h_{ij}\left(r_{ij}\right)dr_{ij}, \tag{7}$$

where r_{ij} is the distance between atom i on one solute molecule and atom j on another solute molecule.

2.2. Reduction of the OZ equation

In order to reduce computer memory requirements we write Eq. (5) in terms of unique solvent sites. Rewriting the OZ equation in terms of $\widetilde{T}(k_x,k_y,k_z)$ and $\widetilde{C}(k_x,k_y,k_z)$ we have

$$\widetilde{T}\left(k_x,k_y,k_z\right) = \widetilde{C}\left(k_x,k_y,k_z\right)\left\{\left(\widetilde{w}(k)\text{-}I + \rho\widetilde{h}(k)\right)\right\}, \tag{8}$$

where I is the identity matrix. Expanding the matrices and taking advantage of the water model symmetry we write,

$$\begin{aligned}\widetilde{T}_{uO}\left(k_x,k_y,k_z\right) = \widetilde{C}_{uO}\left(k_x,k_y,k_z\right)\rho_v\widetilde{h}_{OO}(r)\\ +2\widetilde{C}_{uH}\left(k_x,k_y,k_z\right)\left(\widetilde{w}_{HO}(k) + \rho_v\widetilde{h}_{HO}(k)\right)\end{aligned} \tag{9}$$

and

$$\begin{aligned}\widetilde{T}_{uH}\left(k_x,k_y,k_z\right) = \widetilde{C}_{uO}\left(k_x,k_y,k_z\right)\left(\widetilde{w}_{OH}(k) + \rho_v\widetilde{h}_{OH}(k)\right)\\ +\widetilde{C}_{uH}\left(k_x,k_y,k_z\right)\left(\widetilde{w}_{HH}(k) + 2\rho_v\widetilde{h}_{HH}(k)\right)\end{aligned} \tag{10}$$

where the subscripts O and H represent the oxygen and hydrogen solvent sites, respectively, with ρ_v being the solvent density.

2.3. Long-ranged function resummation

Because Eq. (3) is a relation between long-ranged (Coulomb dependent) functions we re-sum the equations exactly. We combine all long-ranged pieces in the function,

$$\Phi_{ui} = \sum_{\alpha}^{n_u}\phi_{\alpha i}(r_{\alpha i}), \tag{11}$$

where each component $\phi_{\alpha i}$ in the sum has the asymptotic property

$$\lim_{r_{\alpha i}\to\infty}\phi_{\alpha i}(r_{\alpha i}) = \frac{\beta q_\alpha q_i}{r_{\alpha i}}, \tag{12}$$

and is finite valued at small r_{ui}, so that we can write

$$\beta u^s_{\alpha i}(r_{\alpha i}) = \beta u_{\alpha i}(r_{\alpha i}) - \phi_{\alpha i}(r_{\alpha i}) \tag{13}$$

which we use to separate the total potential into long and short-ranged parts,

$$\beta U^s_{ui}(x,y,z) = \beta U_{ui}(x,y,z) - \Phi_{ui}, \tag{14}$$

where the superscript s indicates a short-ranged function. We also can use Φ_{ui} to separate the long and short parts of the other functions using

$$C^s_{ui}(x,y,z) = C_{ui}(x,y,z) + \Phi_{ui} \tag{15}$$

which implies

$$T^s_{ui}(x,y,z) = T_{ui}(x,y,z) - \Phi_{ui} \tag{16}$$

which defines $C^s_{ui}(x,y,z)$ and $T^s_{ui}(x,y,z)$. We note that for numerical solutions it must be possible to Fourier transform Φ_{ui}, and we designate the result as $\widetilde{\Phi}_{ui}$. Eqs. (15) and (16) are used in Eq. (9) to give

$$\begin{aligned}\widetilde{T}^s_{uO}\left(k_x,k_y,k_z\right) + \widetilde{\Phi}_{uO} = \left(\widetilde{C}^s_{uO}\left(k_x,k_y,k_z\right) - \widetilde{\Phi}_{uO}\right)\rho_v\widetilde{h}_{OO}(k)\\ +2\left(\widetilde{C}^s_{uH}\left(k_x,k_y,k_z\right) - \widetilde{\Phi}_{uH}\right)\\ \times\left(\widetilde{w}_{HO}(k) + \rho_v\widetilde{h}_{HO}(k)\right),\end{aligned} \tag{17}$$

which can be rearranged into the form

$$\widetilde{T}^s_{uO}\left(k_x,k_y,k_z\right) = \widetilde{T}^{s\dagger}_{uO}\left(k_x,k_y,k_z\right) - \widetilde{\Theta}_{uO}, \tag{18}$$

where

$$\begin{aligned}\widetilde{T}^{s\dagger}_{uO}\left(k_x,k_y,k_z\right) = \widetilde{C}^s_{uO}\left(k_x,k_y,k_z\right)\rho_v\widetilde{h}_{OO}(k)\\ +2\widetilde{C}^s_{uH}\left(k_x,k_y,k_z\right)\\ \times\left(\widetilde{w}_{HO}(k) + \rho_v\widetilde{h}_{HO}(k)\right),\end{aligned} \tag{19}$$

and

$$\widetilde{\Theta}_{uO} = \widetilde{\Phi}_{uO}\left(1 + \rho_v\widetilde{h}_{OO}(k)\right) + 2\widetilde{\Phi}_{uH}\left(\widetilde{w}_{HO}(k) + \rho_v\widetilde{h}_{HO}(k)\right). \tag{20}$$

Similarly, Eq. (10) gives

$$\widetilde{T}^s_{uH}\left(k_x,k_y,k_z\right) = \widetilde{T}^{s\dagger}_{uH}\left(k_x,k_y,k_z\right) - \widetilde{\Theta}_{uH}, \tag{21}$$

where

$$\begin{aligned}\widetilde{T}^{s\dagger}_{uH}\left(k_x,k_y,k_z\right) = \widetilde{C}^s_{uO}\left(k_x,k_y,k_z\right)\left(\widetilde{w}_{OH}(k) + \rho_v\widetilde{h}_{OO}(k)\right)\\ +\widetilde{C}^s_{uH}\left(k_x,k_y,k_z\right)\left(\widetilde{w}_{HH}(k) + 2\rho_v\widetilde{h}_{HH}(k)\right),\end{aligned} \tag{22}$$

and

$$\widetilde{\Theta}_{uH} = \widetilde{\Phi}_{uO}\left(\widetilde{w}_{OH}(k) + \rho_v \widetilde{h}_{OH}(k)\right) + \widetilde{\Phi}_{uH}\left(1 + \widetilde{w}_{HH}(k) + 2\rho_v \widetilde{h}_{HH}(k)\right). \tag{23}$$

We also write (Eq. (4)) in terms of short and long-ranged contributions. Given that Θ_{ui} and $T^{s\dagger}_{ui}$ are the inverse Fourier transforms of $\widetilde{\Theta}_{ui}(k_x, k_y, k_z)$ and $\widetilde{T}^{s\dagger}_{ui}(k_x, k_y, k_z)$ respectively, they can be combined after the transform process to give

$$T^{s}_{ui}(x,y,z) = T^{s\dagger}(x,y,z) - \Theta_{ui}(x,y,z), \tag{24}$$

which can be used with Eqs. (14),(15) and (16) to rewrite Eq. (4) in the more useful form

$$C^{s}_{ui}(x,y,z) = exp\left\{-\beta U^{s}_{ui}(x,y,z) + T^{s}_{ui}(x,y,z)\right\} - T^{s}_{ui}(x,y,z) - 1. \tag{25}$$

2.4. Fourier transformation of the long-ranged contribution

The function $\Phi_{\alpha i}(x, y, z)$ is a sum of component functions $\phi_{\alpha i}(r_{\alpha i}) = \phi_{\alpha i}(|r_i - s_\alpha|)$ where each $\phi_{\alpha i}(|r_i - s_\alpha|)$ is isotropic about a fixed point s_α in (x, y, z) -space. We Fourier transform each of these components separately using a coordinate translation to a new origin s_α using $r'_{\alpha i} = -s_\alpha$. This gives

$$\begin{aligned}\widetilde{\phi}_{\alpha i}\left(k_x, k_y, k_z\right) &= \iiint exp(ik \cdot r_i)\phi_{\alpha i}(|r_i - s_\alpha|)dr_i \\ &= exp(ik \cdot s_\alpha)\iiint exp\left(ik \cdot r'_{\alpha i}\right)\phi_{\alpha i}\left(|r'_{\alpha i}|\right)dr'_{\alpha i} \\ &= exp(ik \cdot s_\alpha)\widetilde{\phi}_{\alpha i}(k).\end{aligned} \tag{26}$$

Using the linearity of Fourier transforms and Eq. (11) and have

$$\widetilde{\Phi}_{ui}\left(k_x, k_y, k_z\right) = \sum_{\alpha}^{n_u} exp(ik \cdot s_\alpha)\widetilde{\phi}_{\alpha i}(k). \tag{27}$$

Next Eq. (27) in Eqs. (20) and (23), gives

$$\begin{aligned}\widetilde{\Theta}_{uO} &= \sum_{\alpha}^{n_u} exp(ik \cdot s_\alpha)\left\{\widetilde{\phi}_{\alpha O}(k)\left(1 + \rho_v \widetilde{h}_{OO}(k)\right) + 2\widetilde{\phi}_{\alpha H}(k)\left(\widetilde{w}_{HO}(k) + \rho_v \widetilde{h}_{HO}(k)\right)\right\} \\ &= \sum_{\alpha}^{n_u} exp(ik \cdot s_\alpha)\widetilde{\theta}_{\alpha O}(k),\end{aligned} \tag{28}$$

and

$$\begin{aligned}\widetilde{\Theta}_{uH} &= \sum_{\alpha}^{n_u} exp(ik \cdot s_\alpha)\left\{\widetilde{\phi}_{\alpha O}(k)\left(\widetilde{w}_{OH} + \rho_v \widetilde{h}_{OH}(k)\right) + \widetilde{\phi}_{uH}(k)\left(1 + \widetilde{w}_{HH}(k) + 2\rho_v \widetilde{h}_{HH}(k)\right)\right\} \\ &= \sum_{\alpha}^{n_u} exp(ik \cdot s_\alpha)\widetilde{\theta}_{\alpha H}(k),\end{aligned} \tag{29}$$

which define the component functions $\widetilde{\theta}_\alpha(k)$ and $\widetilde{\theta}_\alpha(k)$ of $\widetilde{\Theta}_u$ and $\widetilde{\Theta}_u$, respectively. The similarity of Eqs. (28) and (29) to Eq. (27) allows us to write the inverse Fourier transform of $\widetilde{\Theta}_{ui}\left(k_x, k_y, k_z\right)$ as

$$\Theta_{ui}(x,y,z) = \sum_{\alpha}^{n_u} \theta_{\alpha i}(r_{\alpha i}), \tag{30}$$

provided that the Fourier function pairs $\widetilde{\theta}_{\alpha i}(k)$ and $\theta_{\alpha i}(r)$ exist and it is implied that the $\theta_{\alpha i}(r_{\alpha i})$ are functions of $|r_i - s_\alpha|$, each one isotropic around the site position s_α.

Any form for the functions $\phi_{\alpha i}(r_{\alpha i})$ may be chosen as long as they meet the asymptotic requirements described above and have suitable Fourier transform pairs. For this work we use [35]

$$\phi_{\alpha i}(r_{\alpha i}) = \frac{\beta q_\alpha q_i}{r_{\alpha i}} \mathrm{erf}(\gamma r_{\alpha i}), \tag{31}$$

for which the Fourier transform is

$$\widetilde{\phi}_{\alpha i}(k) = \frac{4\pi\beta q_\alpha q_i}{k^2} exp\left(\frac{-k^2}{4\gamma^2}\right). \tag{32}$$

This simplification allows us to pre-calculate $\beta U^{s}_{ui}(x, y, z)$ and $\Theta_{ui}(x, y, z)$, which in turn allows the solution of the OZ and HNC equations can be solved as in terms of constant and short-ranged functions only, using Eqs. (19), (22), (24) and (25). No further approximations were made in this derivation.

The component functions $\phi_{\alpha i}$ and $\theta_{\alpha i}$ are calculated most easily by using the same r and k-space grids as for the solvent–solvent functions and interpolating the $\theta_{\alpha i}(r_{\alpha i})$ functions onto the 3D (x, y and z) coordinate system.

2.5. A 3D bridge function

The HNC equation in the form we use here (Eq. 25) is an approximation to the exact closure expression for the direct correlation function with the missing term, the bridge function, being set to zero. Bridge diagrams can show improvements over HNC theories [36–39], in particular sometimes even yielding numerical solutions when using the HNC equation proves difficult or impossible. As in previous work [27,40] we use the HNCB closure, which replaces a large proportion of the diagrams of the exact bridge function with a simple approximation which gives solutions for a wide range of phase points and appears to improve the result,

$$\begin{aligned}B_{uk}(r_{uk}) &= -\frac{1}{2}\sum_{ij}\rho_i\rho_j \int H_{ui}(r_{ui})h_{ik}(r_{ik}) \times H_{uj}\left(r_{uj}\right)h_{jk}\left(r_{jk}\right)dr_i dr_j, \\ &= -\frac{1}{2}\left(\sum_i \rho_i \int H_{ui}(r_{ui})h_{ik}(r_{ik})dr_i\right) \times \left(\sum_i \rho_i \int H_{uj}\left(r_{uj}\right)h_{jk}\left(r_{jk}\right)dr_j\right),\end{aligned} \tag{33}$$

which can be calculated by simple Fourier methods and updated as the calculation proceeds. We note that $h_{ij}(r_{ij})$, $H_{ui}(r_{ui})$ and $B_{uk}(r_{uk})$ are short-ranged. This approximation is used in Eq. (4) as

$$C_{ui}(x,y,z) = exp\left\{-\sum_{\alpha}^{n_u}\beta u_{\alpha i}(r_{\alpha i}) + T_{ui}(x,y,z) + B_{ui}(x,y,z)\right\} - T_{ui}(x,y,z) - 1. \tag{34}$$

2.6. Thermodynamic quantities

We will examine excess internal energies and Kirkwood-G integrals calculated with respect to each solvent site which are calculated using [6]

$$\frac{\langle U^{ex}\rangle}{N} = \frac{1}{2}\rho_v \sum_i \int (H_{ui}(x,y,z) + 1)U_{ui}(x,y,z)dxdydz, \tag{35}$$

where i represents each solvent site, and

$$G_{ui} = \int H_{ui}(x,y,z)dxdydz. \tag{36}$$

3. Proximal radial distribution functions

In this section we briefly review the theory underlying the proximal radial distribution functions (pRDF), and the additional criteria in the proximal search algorithm.

As reviewed above the partial occupancy of internal water molecules at the distal pocket of myoglobin is established, yet the quantitative occupancy, the number of hydration sites and their correlation patterns remains a challenge. The proximal distribution function method (like I.E.s and MD) considers perturbation of the solvent distribution by the protein, consequently the effects by the protein conformational charges. The different hydration correlations within a confined structural pocket of a protein are a result of the physical and chemical interactions. Water correlations have shown a rich variety at the various heterogeneous surface sites of proteins [2,41]. The polar groups (charged or partially charged) establish strong electrostatic interactions and layering of the water molecules. In order to explicitly quantify the effects of a confined structural situation within a biological system on the protein–water pair correlation function, we consider the conditional pair correlation function that describes the solvent structure closest to a protein atom at a distance r, or equivalently perpendicular to the protein surface, $g_\perp(r)$ [31]. This is the first member of a physical cluster hierarchy form of the partition function written in terms of near neighbors, next near neighbors etc. Here we only consider the first term as an approximation to the series. The convergence of the series has been considered previously and generally the suitably normalized first member of the hierarchy captures the major features of the distribution of solvent [9]. The distribution of solute around solvent converges much differently [42].

These $g_\perp(r)$, or perpendicular radial distribution functions (pRDFs), can later be used to reconstruct the solvent density distribution around, as well as interior to, proteins or other biological complexes [9,43]. Although the solvent density distribution may be unique to individual proteins, the pRDFs, are approximately transferable [30] across globular proteins. We want to utilize the approximate universality of the pRDFs to predict the hydration structures in a confined region within biological assemblies. In this case we want to map the internal water density distribution near the heme distal pocket. Successfully reconstructing the water density distribution at the myoglobin distal pocket using the pRDFs is a challenge. Understanding the occupancy level of the internal water can provide evidence that further tests proposed ligand entering and escaping mechanisms.

The proximal distributions are a near neighbor case of the quasi-component distributions [44]. Consider a solution of a polyatomic solute molecule and N solvent molecules. The solute–solvent pair correlation function $g(r_{ij})$ describes the relative probability of finding a solvent molecule i at a given distance r away from a specific solute atom j. It is easily computed from a trajectory according to the equation,

$$g\left(r_{ij}\right) = \frac{1}{4\pi r^2 \Delta r N}\sum_{t=0}^{T}\sum_{j=1}^{N}\delta\left[|\vec{r}_i(t)-\vec{r}_j(t)|-r\right], \tag{37}$$

where T is the total simulated time, $\vec{r}_i(t)$ and $\vec{r}_j(t)$ represent the position vectors of solute atom i and solvent atom j at time t, respectively, and $1/4\pi r^2\delta r$ is the normalization volume of a spherical shell of width δr. For large non-spherical proteins, the ease of use and interpretation of $g(r_{ij})$ can suffer complications from the volume element as well as coupled correlations since the $g(r_{ij})$ implicitly depends on the distribution of other atomic sites on the solute. To approximate the local correlations and account for the normalization of the volume for a non-spherical solute, we use a different quantity, a perpendicular or proximal radial distribution (pRDF), $g_\perp(r_{ij})$, which gives the probability of finding a solute atom closest to a solvent atom. This roughly defines a perpendicular to the protein surface, and in this sense the pRDF explicitly takes into account the solute surface feature closest to any given solvent molecule. Moreover, $g_\perp(r_{ij})$ suffers from having the excluded volume of the protein in the normalization of the solvent distribution around the solute surface atoms; one expects the radial distribution to approach unity after a distance corresponding to the radius of gyration. In contrast, the $g_\perp(r_{ij})$ has local characteristics of the proximal protein surface and does not depend on the rest of the protein volume. We define the pRDF as

$$g_\perp\left(r_{ij}\right) = \sum_{t=0}^{T}\sum_{j=1}^{N}\frac{\delta\left(inf\left[|\vec{r}_i(t)-\vec{r}_j(t)|-r\right]_{i=1,N_P}\right)}{\delta\tau\left(\vec{r}_j(t),k\right)}, \tag{38}$$

where N_P is the number of solute atoms, $\delta\tau(\vec{r}_j(t),k)$ is the volume around solvent molecule j at instantaneous time t. $inf\left[\left|\vec{r}_i(t)-\vec{r}_j(t)\right|\right]$ yields the minimum distance vector between any solvent molecule j and solute atoms i at instant t. And k is the solute atom that is closest to the solvent atom j

$$inf\left[\left|\vec{r}_i(t)-\vec{r}_j(t)\right|\right] = \left|\vec{r}_k(t)-\vec{r}_j(t)\right|. \tag{39}$$

It is nontrivial to compute the $g_\perp(r_{ij})$ described in Eq. (38) because the volume element has to be solved for each solvent molecule j at any instant t. A computationally cost effective solution is to calculate the solvent positions on a three-dimensional grid prior to computing an averaged perpendicular distribution function. The pair correlation function computed in this way defines an averaged water distribution at distance r perpendicular to a protein surface, measured with respect to a reference frame attached to the protein. The $g_\perp(r_{ij})$ in this sense is considered to be a conditional pair distribution function between a protein surface (a fixed condition) and the water patterns around it. Not only is the pre-averaged water distribution easier to compute instantaneously on a grid, the averaged water distribution function is also more relevant to X-ray diffraction crystallographic density distributions [9]. The detailed theory and computing procedure of $g_\perp(r_{ij})$ is described in reference [9]. We have found that the hydration around C, N and O atom types on the surface of a globular protein is transferable to other proteins [9,45]. From the pre-computed pRDF for specific protein atom types, one can reconstruct a model for a three-dimensional solvent density distribution $\rho(r_{uvw})$ around other protein solutes using the surface atom specific functions

$$\rho(r_{uvw}) = g_\perp^X(r') \tag{40}$$

where r' takes the minimum value of $|r_k - r_{uvw}|$ for each grid point for all protein atoms i. X is the atom type of protein atom k, for which $r' = |r_k - r_{uvw}|$ ($X = C, N, O, S$, etc) The indices u, v, w denote the grid along the x, y and z directions, respectively. We have tested and find the optimal grid spacing for this purpose to be 0.5 Å([9]).

The $g_\perp^X(r)$ for reconstruction purposes is defined as a set of functions of different protein atomic species. We decompose the protein into various atomic types: non-polar carbon C, backbone nitrogen and oxygen O, negatively charged amino acid side chain oxygens and polar positively charged amino acid side chain nitrogens, and sulfur S. The additional classes of protein atom types over the previous C,H,N and O set improve reconstructions of water density distributions in this near neighbor scheme [44]. We also find that reconstructions of water density from the side chain analogs-water pair radial distribution functions more closely resemble simulated solvent density distribution [46]. This most recent method is used in the current work to reconstruct the solvent density distribution within different confined regions of myoglobin to scan for internal hydration sites.

4. Molecular dynamics simulations

The initial coordinates of the myoglobin molecule are those of sperm whale myoglobin (Mb) with carbon monoxide CO (PDB ID 2MGK) [15]. The crystallographic waters and ions were removed and the system was protonated and solvated using the PSFGEN plugin in VMD [47] using the CHARMM27 force field [48] with TIP3P waters [49]. The 2553 protein atoms, including the HEME and CO, were solvated with 15960 water molecules and one chlorine ion was added to neutralize the system in a rectilinear box of dimensions $72 \times 82.8 \times 89.5$ Å^3. The molecular dynamics simulations were performed with NAMD 2.8 [50] in the NPT ensemble using a Langevin thermostat to control the temperature at 300 K and a Langevin Nosé–Hoover piston for constant pressure ($P = 1.01$ bar). The covalent bonds were kept fixed with the SHAKE algorithm and periodic boundary conditions were enforced. The van der Waals interactions were cut-off at a distance of 12 Å and the long-range electrostatics were determined using the particle-mesh Ewald summation with a grid spacing of 1 Å. Following the procedure used successfully in our previous studies of BPTI [27] two different simulations were performed. In the first molecular dynamics simulation all the atoms were allowed to move and the simulation was run with a 1 fs time step for 3 ns after equilibration. This simulation was run for only a few nanoseconds so as to remove the crystal contacts. An average structure of the final nanosecond was determined from the trajectory snapshots that were saved every 500 steps. This yields a time averaged "solution" structure of this biomolecule with this force field. Because both the IE and the pRDF methods apply to single conformations, the time-averaged "solution" protein structure, without any waters, was used for all subsequent calculations, including the molecular dynamics simulations. A second all-atom molecular dynamics simulation was performed in which the average protein molecule from the first simulation was solvated with 16783 TIP3P waters and 1 chloride ion. For this simulation, the "solution" time-averaged conformation of the macromolecule was centered at the origin and all of the protein atoms were fixed, i.e., they were not allowed to move; the atoms of all the other species were allowed to move, and the simulation was performed as described above. Any solvent later determined in the interior of the protein occurs as a result of diffusion and not from pre-assignment. This second simulation was performed for 50 ns and the last 30 ns was used for the determination of solvent densities presented below. The myoglobin H_*ui*(x,y,z)functions were determined by computing the solvent positions in a plane of 0.5 Å cubes centered about two values. The bulk density was obtained from a 20 Å cubic box in the upper corner of the simulation box; in a bulk solvent region away from the biomolecule.

5. Results and discussion

The heme moiety of myoglobin resides in a cavity that is transiently accessible. Experiments and simulations have shown that ligands, like the O_2, CO, and NO, can migrate into and out of the heme cavity via a variety of connected pathways. Yet, outside of the four hydration sites that are consistently observed, the cavity tends to be low in water occupancy even though the volume of that space is large enough to accommodate many water molecules.

The recent experiments [11] have shown that there are four internal regions where a water molecule can have a relatively long residence time and can exchange with the bulk. These internal water molecules are believed to be located at polar sites which include one of the xenon-binding sites and are several Å apart, implying more than one single pathway into and out of the heme-cavity exists. We will use three different methods to map the solvation of the myoglobin molecule, in particular the internal cavities and some possible paths into and out of the heme pocket.

Molecular dynamics simulations of detailed atomic models of biomolecular systems may be used to determine a description of the molecular solvation. Such calculations are computationally expensive whereas simple phenomenological models, such as our pRDF method, are especially useful for fast calculation of the solvent density and the electrostatic solvation free energy while retaining most of the accuracy of the all-atom simulations. We wish to compare the solvation densities obtained from three different methods we studied here: IEs, pRDFs, and molecular dynamics simulations. Because the IE and pRDF methods are applied to rigid molecules, the molecular dynamics simulation used to determine the solvent properties were performed for an all-atom simulation where the protein, including the heme, were held fixed and the water molecules and ions were allowed to move. The conformation of the protein is the average structure of myoglobin from the all-atom, flexible, simulation described above; this conformation is used by all three methods for the determination of the protein–solvent density.

The IE results for our average myoglobin structure were evaluated at infinite dilution in TIP3P DRISM/HNC water using the method described above, including our improved treatment of the long-ranged Coulombic tails. The calculations were initially converged with a grid spacing and range using 128^3 points, spaced evenly in all dimensions with a spacing of 0.5 Å. The solvent densities were determined with data structures of 256^3 points using the HNCB closure. The IE method samples the ensemble of solvent configurations in all regions of space and will show finite probability for the population of internal cavities whenever the equation of state gives favorable free energetics in the volume. However, there are known problems with the IE method. In an enclosed cavity, the IE method may incorrectly determine population in that volume. Many approximate theories tend to over- or under-estimate the cavity population [24,26,27]. The essential approximation in the method involves missing terms in the density expansions which result in incorrectly weighted correlations that are typically most noticed in regions completely surrounded by solute sites at close range. There are missing terms in all orders of the density for RISM-like theories, including the low order terms that contribute directly to the correlation between the solvent and the solute. Despite this, the IE theories are extremely beneficial in locating solvent densities in interior regions of biomolecular systems that would be computationally expensive using more detailed models like MD simulations.

Average pRDFs for C, N, O, and our newly developed representations for the specific atomic types: non-polar carbon C, backbone nitrogen and oxygen O, negatively charged amino acid side chain oxygens and polar positively charged amino acid side chain nitrogen atoms, and sulfur S representations are used to reconstruct the solvent densities for our myoglobin conformation. The density functions for the MD simulations were obtained by tabulating positions in a plane of 0.5 Å cubes, the same as used for the pRDF and the IE calculations, centered about the plane. The bulk density was obtained from a 20.0 Å cubic box cut from the upper corner of the simulation box in a volume away from the protein. The solvent distribution for two different planes, cut through the solute molecule are presented in Figs. 1 and 2. The planes chosen were the $z = -4.0$ and $y = 13.5$ which highlight the solvation sites at the Xe_3 and the apical sites, respectively. As a structural comparison of these three methods plots of the results centered on the planes cut through the data are plotted in Figs. 1(a)–(c) for the $z = -4.0$ plane and 2(a)–(c) for the $y = 13.5$ planes for the IE, pRDF, and MD densities, respectively.

The cross-sectional plane for the $z = -4.0$ slice in Fig. 1 clearly displays the solvation on the side opposite of the Heme, in the vicinity of the Xe_3 occupation site. For all three methods there is a solvent density cluster in the vicinity of the Xenon site; the density appears as "islands" with different ridges at different heights. This cluster contains the ridges with the highest peaks on the interior of the protein. As discussed above, the IE results, Fig. 1(a), clearly displays the highest and widest peaks compared to both the pRDF and the MD methods with a single high peak closest to the exterior region and a wider cluster with different ridge heights as we move further into the cavity. The peaks on the exterior of the protein are also larger for the IE method compared to any of

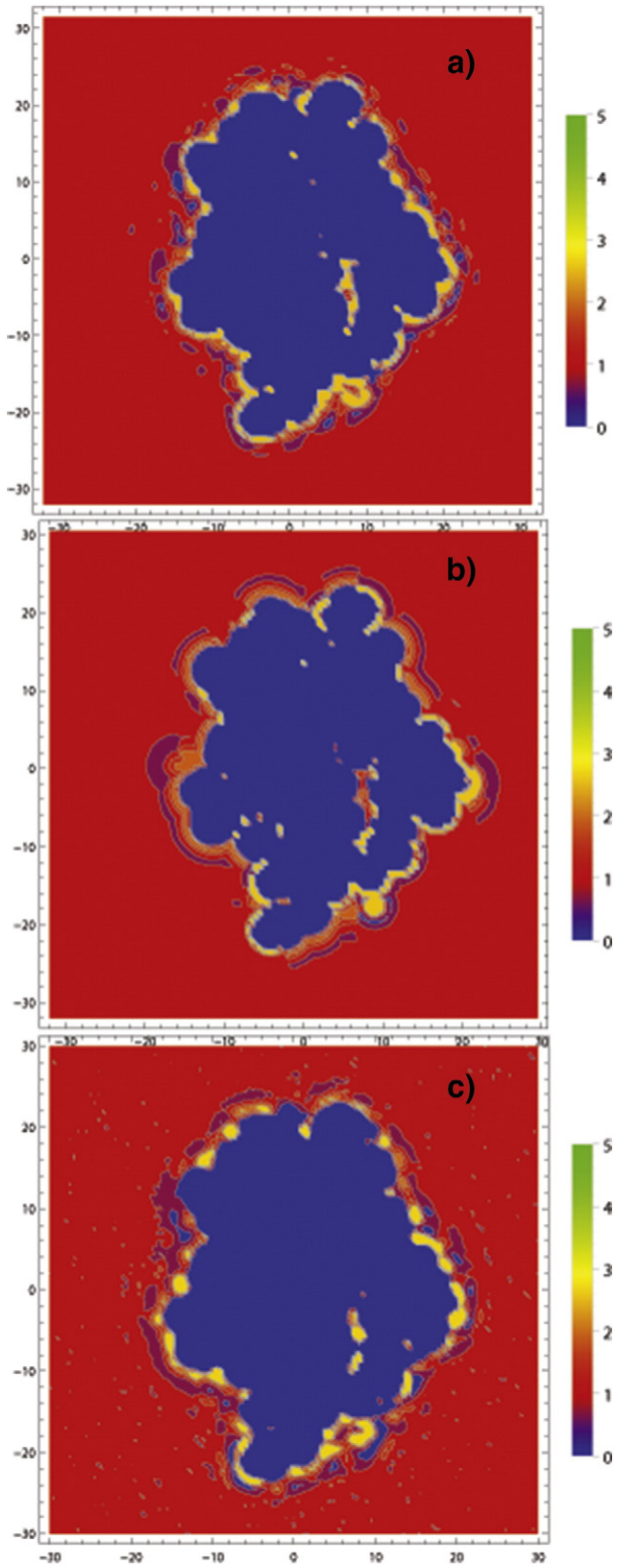

Fig. 1. Solvent density plots for slice z = −4.0 for IE method a), for pRDF method b), for MD method c).

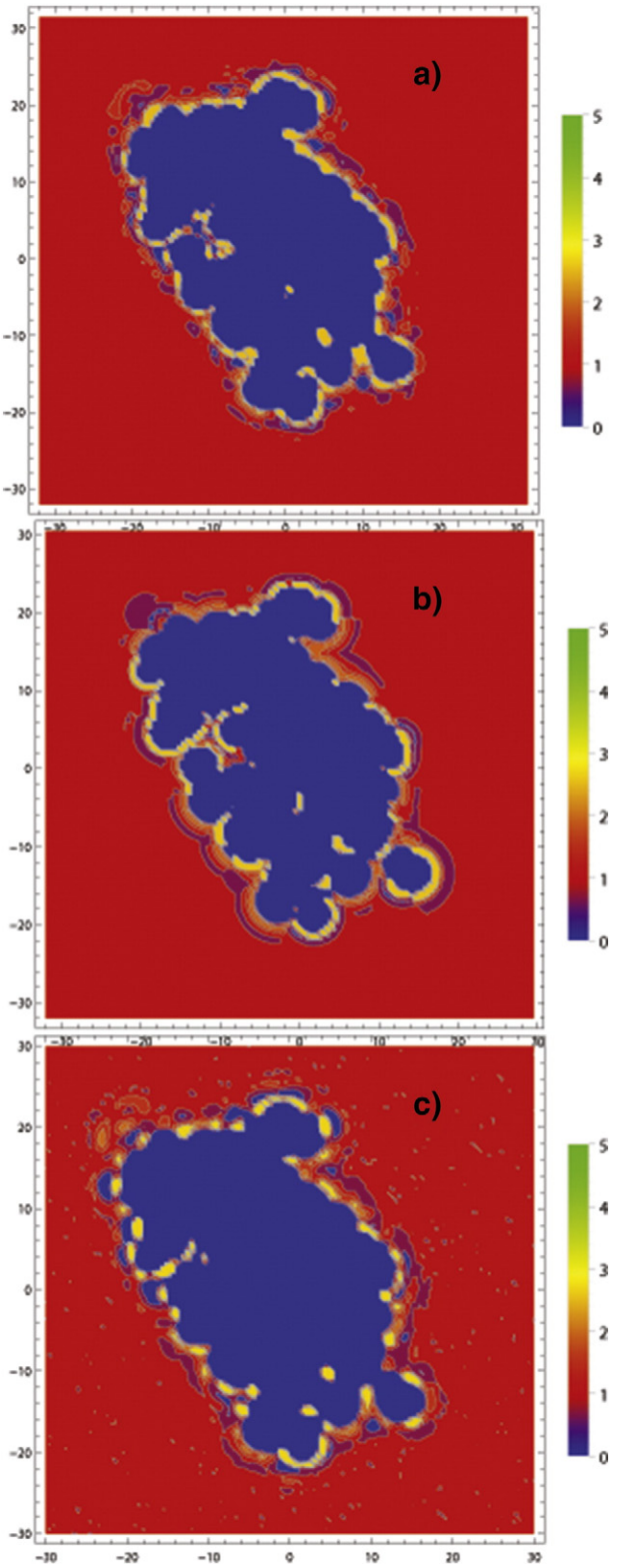

Fig. 2. Solvent density plots for slice y = 13.5 for IE method a), for pRDF method b), for MD method c).

the other techniques. For the pRDF results, Fig. 1(*b*), there are two "island" clusters in this location with ridges of comparable heights. The MD results, Fig. 1(*c*), also has two density clusters in this volume but the cluster with the higher peak is the one at the location deeper into the cavity and not closer to the bulk as observed for the IE results. The IE and pRDF results also show a few small peaks deeper into the pocket but these peak heights are significantly smaller than those observed in the "island" clusters. The third, most interior cluster, is not seen in the IE results but the IE interior density is larger and wider than both the pRDF and the MD results and encompasses these minor peaks.

The Xe_3 is the location of the long-lived water molecules and this solvent occupancy is clearly supported by the densities observed by

our three methods. Although the Xe_3 is mostly buried in the protein with very limited access to the bulk solvent there is an opening from the exterior into the cavity where this Xenon atom would be located. As shown in Fig. 1 the solvent can, and does, enter the interior and the hydration sites via this entrance on the back side of the heme. For these simulations there was no pre-solvating of the cavities; the solvation of the interior cavities is observed by either pre-solvating particular sites or by diffusion into the cavity. As we have observed in earlier calculations [27], the IE and the MD methods tend to corroborate cavity solvation densities if there is a pathway; in the case of these results the MD occupancies result from water diffusing into those hydration sites during the simulation for the conformation of the protein used. The experimental data observes solvent molecules in this area and all three methods clearly observe solvent density in this interior region of the cavity which is close to the Xe_3 site location. Interestingly, the IE and pRDF calculations also determine a second high density peak in the interior. This is the second highest peak for both methods and the IE peak is only slightly lower in height than the highest peak observed in the Xe_3 region. The MD results do not display this peak. The pRDF results also show other density probabilities that are not observed in either the IE or the MD results. The pRDF density calculations are based on atom types, C, N, and O, and spatial availability. These regions of density are not accessible to the MD results for the length of the simulation, 50 ns, presented here.

The cross-sectional plane for $y = 13.5$ was chosen to elucidate the solvation of the apical site. The results for this slice are displayed in Fig. 2 using the same order: Fig. 2(a) is from the IE calculations, Fig. 2(b) is from the pRDF calculations, and Fig. 2(c) is from the MD simulations. All three methods clearly display solvent density in the vicinity of the apical site. But, more interesting is the solvent densities clearly map a pathway into the heme pocket. The IE results show the path with ridges of similar height until the density is located entirely inside the cavity. Inside the cavity the results then display a second density "island". Fig. 2(b) represents the results for the pRDF method which also finds solvent density in the apical region. The density also maps a very clearly defined pathway produced during the 3ns equilibration. The MD results, Fig. 2(c), show a similar pathway into the cavity but, not surprisingly, the interior "island" peak heights are lower than those closer to the exterior and lower than the IE values. The pRDF and MD results also determined two density islands in this region of the protein with peak heights second only the highest peak displayed in the apical region. The myoglobin molecule, like all proteins, has interior cavities that are accessible to the solvent and can be populated if the solvent diffuses into those interior regions. This is clearly displayed by the additional solvent density peaks seen for this slice and may represent one of those peripherally located hydration sites suggested by Kaieda and Halle [11] that may be necessary for the transport of ligands and solvent across a cavity as large as the myoglobin cavity for which the hydration sites are tens of Å apart.

Myoglobin is an example of a biological molecule with an internal prosthetic cavity with a volume large enough to accommodate many solvent molecules but which has been found to be underoccupied, with only four hydration sites regardless of the technique used to determine the solvent occupancy. For myoglobin, the other factor of interest, is that the heme region is the main site of biological function. In Fig. 3 the location of the three interior solvent density sites determined by our methods is represented as red spheres and the Xenon binding sites are represented as green spheres. Our results indicate that for the two slices chosen, there are buried solvent binding sites for myoglobin and they span the large cavity. There have been numerous hypotheses about the hydrophobicity of the heme cavity being the reason for the low occupancy of that volume coupled to the gating of that region for entrance and exit of ligands. The pRDF and the IE methods are two computational methods that determine a solvent path into the heme pocket by identifying the high-probability solvent density locations for a given conformation of the protein.

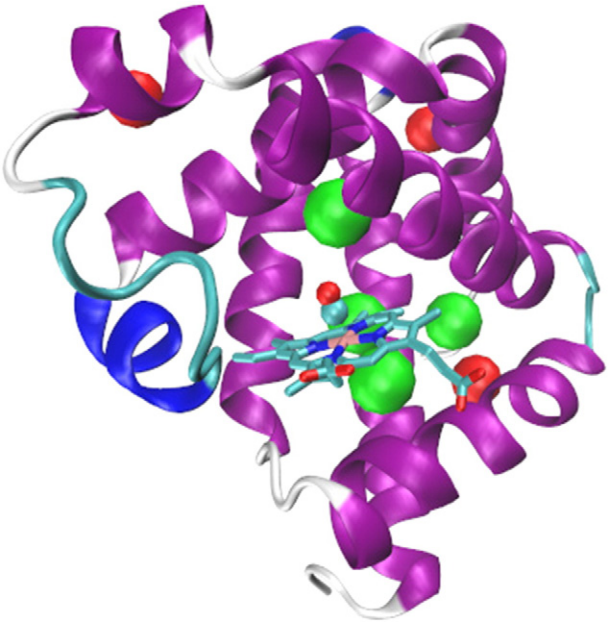

Fig. 3. Cartoon representation of CO bound protein with Xe atoms as green balls and solvent density sites as large red balls.

In this study we have been most interested in the heme cavity occupation versus the solvent density of the protein's exterior. With simulations the interior pockets may not be populated during short time scales. If by chance there is diffusion into the pocket, those events tend to be rare. The X-ray crystallography results do not resolve significant solvent population in the heme pocket. From our earlier work on BPTI [27], we found that if internal cavities are populated, as determined experimentally, and that population is preset at the beginning of the simulation, there is a high probability that the cavity remains populated throughout the simulation and that population density is confirmed by the IE method. For the myoglobin structure used there were no such experimental solvent densities in the molecule's interior and, as such, none were placed there at the start of the simulations. Our results indicate that there are multiple paths into the cavity which is in agreement with the work of Elber on possible entry and exit paths for ligands [51]. If the mechanism of exchange of the diatomic ligands depends on migration to and from the heme-bound iron our results indicate that there may be multiple pathways into and out of the pocket for one single protein conformation.

6. Conclusions

This work is the first comparison of these three different theoretical methods for calculating the solvent distribution around an atomic site model of a biomolecule. Our 3D IE method with exact Coulomb interactions [27] and proximal radial distribution function reconstruction [46] are compared to all-atom molecular dynamics simulations. The solutions for both the IE and pRDF techniques are easily and efficiently determined for small and large molecules and comparison to all-atom MD simulations provides a determination of how well they predict the solvent distribution around and inside myoglobin. While simulations may take CPU days, the integral equations are typically finished in 10s of minutes and the pRDF reconstructions in seconds or less.

Comparison of the different methods showed that the computed solvent densities for both the Xe_3 and apical regions, two of the interior solvent hydration sites, are reproduced by all of the methods. It is well known that the IE method overestimates some peak heights and that is observed in our calculations here. But, the hydration locations, matching the cavity solvent locations, are reproduced by all three

methods. The IE and pRDF methods, as expected, did determine solvent densities that were not quantitatively in perfect agreement with the simulation results. However, as we have observed in other studies, [27] these may be either an overestimation of the approximate method or hydration site locations that the MD simulations can either not access or do not observe because they are low probability events and the simulations can never be run long enough to properly observe statistically. The results also clearly determine a solvent pathway via the apical site from the exterior into the cavity. We have also determined a secondary interior hydration site, not in the cavity, that may correspond to one of the auxiliary pathways used by the myoglobin molecule to shuttle molecules into and out of its large cavity [11]. While interior sites have been studied by integral equations [24] these results provide the first systematic study of the solvent density probabilities among the IE, pRDF, and MD methods. The corroboration of the results indicate that these approximate methods that can be successfully applied to large biomolecular systems with multiple hydration sites at a computational cost that is much less than direct simulation.

Acknowledgment

The Robert A. Welch Foundation (H-0037), the National Science Foundation (CHE-1152876) and the National Institutes of Health (GM-037657) are thanked for partial support of this work. A portion of this work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. In particular, the calculations were performed on the machines at the Texas Advanced Computing Center and the National Institute for Computational Sciences.

References

[1] M. Feig (Ed.), Modeling Solvent Environments: Applications to Simulations of Biomolecules, Wiley-VCH Verlag GmbH & Co. KGaA, 2010.
[2] C.L. Brooks, M. Karplus, B.M. Pettitt, Proteins: a theoretical perspective of dynamics, structure and thermodynamics, Vol. 71 of Advances in Chemical Physics, Wiley, 1988.
[3] A. Pohorille, L.R. Pratt, Cavities in molecular liquids and the theory of hydrophobic solubilities, J. Am. Chem. Soc. 112 (13) (1990) 5066–5074.
[4] K. Lindorff-Larsen, S. Piana, R.O. Dror, D.E. Shaw, How fast-folding proteins fold, Science 334 (6055) (2011) 517–520.
[5] J. Howard, B. Pettitt, Integral equations in the study of polar and ionic interaction site fluids, J. Stat. Phys. 145 (2) (2011) 441–466.
[6] J.P. Hansen, I.R. McDonald, Theory of Simple Liquids, Academic Press, 2006.
[7] J.J. Howard, J.S. Perkyns, B.M. Pettitt, The behavior of ions near a charged walls-dependence on ion size, concentration, and surface charge, J. Phys. Chem. B 114 (18) (2010) 6074–6083.
[8] V. Lounnas, B.M. Pettitt, Distribution function implied dynamics versus residence times and correlations—solvation shells of myoglobin, Proteins Struct. Funct. Genet. 18 (2) (1994) 148–160.
[9] V.A. Makarov, B.K. Andrews, B.M. Pettitt, Reconstructing the protein–water interface, Biopolymers 45 (7) (1998) 469–478.
[10] J. Kendrew, G. Bodo, H. Dintzis, R. Parrish, H. Wyckoff, D. Phillips, A three-dimensional model of the myoglobin molecule obtained by X-ray analysis, Nature 181 (4610) (1958) 662–666.
[11] S. Kaieda, B. Halle, Internal water and microsecond dynamics in myoglobin, J. Phys. Chem. B 117 (47) (2013) 14676–14687.
[12] M. Lapelosa, C.F. Abrams, A computational study of water and co migration sites and channels inside myoglobin, J. Chem. Theory Comput. 9 (2) (2013) 1265–1271.
[13] Q.H. Gibson, R. Regan, R. Elber, J.S. Olson, T.E. Carver, Distal pocket residues affect picosecond ligand recombination in myoglobin—an experimental and molecular-dynamics study of position 29 mutants, J. Biol. Chem. 267 (31) (1992) 22022–22034.
[14] R. Elber, M. Karplus, Enhanced sampling in molecular-dynamics—use of the time-dependent hartree approximation for a simulation of carbon-monoxide diffusion through myoglobin, J. Am. Chem. Soc. 112 (25) (1990) 9161–9175.
[15] M.L. Quillin, R.M. Arduini, J.S. Olson, G.N. Phillips Jr., High-resolution crystal structures of distal histidine mutants of sperm whale myoglobin, J. Mol. Biol. 234 (1) (1993) 140–155.
[16] R.F. Tilton, I.D. Kuntz, G.A. Petsko, Cavities in proteins—structure of a metmyoglobin–xenon complex solved to 1.9-a, Biochemistry 23 (13) (1984) 2849–2857.
[17] R.A. Goldbeck, S. Bhaskaran, C. Ortega, J.L. Mendoza, J.S. Olson, J. Soman, D.S. Kliger, R.M. Esquerra, Water and ligand entry in myoglobin: assessing the speed and extent of heme pocket hydration after co photodissociation, Proc. Natl. Acad. Sci. U. S. A. 103 (5) (2006) 1254–1259.
[18] S.E.V. Phillips, B.P. Schoenborn, Neutron diffraction reveals oxygen–histidine hydrogen bond in oxymyoglobin, Nature 292 (2) (1981) 81–82.
[19] M.A. Scorciapino, A. Robertazzi, M. Casu, P. Ruggerone, M. Ceccarelli, Heme proteins: the role of solvent in the dynamics of gates and portals, J. Am. Chem. Soc. 132 (14) (2010) 5156–5163.
[20] D. Beglov, B. Roux, Numerical solution of the hypernetted chain equation for a solute of arbitrary geometry in three dimensions, J. Chem. Phys. 103 (1995) 360.
[21] M. Ikeguchi, J. Doi, Direct numerical solution of the Ornstein–Zernike integral equation and spatial distribution of water around hydrophobic molecules, J. Chem. Phys. 103 (1995) 5011.
[22] D. Beglov, B. Roux, Solvation of complex molecules in a polar liquid: an integral equation theory, J. Chem. Phys. 104 (1996) 8678.
[23] C.M. Cortis, P.J. Rossky, R.A. Friesner, A three-dimensional reduction of the ornstein–zernike equation for molecular liquids, J. Chem. Phys. 107 (1997) 6400.
[24] F. Hirata, Molecular Theory of Solvation, Springer, 2003.
[25] T. Imai, A. Kovalenko, F. Hirata, Solvation thermodynamics of protein studied by the 3D-RISM theory, Chem. Phys. Lett. 395 (2004) 1.
[26] T. Imai, R. Hiraoka, A. Kovalenko, F. Hirata, Water molecules in a protein cavity detected by a statistical–mechanical theory, J. Am. Chem. Soc. 127 (2005) 15334.
[27] J.S. Perkyns, G.C. Lynch, J.J. Howard, B.M. Pettitt, Protein solvation from theory and simulation: exact treatment of coulomb interactions in three-dimensional theories, J. Chem. Phys. 132 (6) (2010) 064106.
[28] A. Kovalenko, F. Hirata, Potentials of mean force of simple ions in ambient aqueous solution. i. Three-dimensional reference interaction site model approach, J. Chem. Phys. 112 (2000) 10391.
[29] A. Kovalenko, F. Hirata, Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional interaction site model, J. Chem. Phys. 110 (1999) 10095.
[30] B. Lin, K.-Y. Wong, C. Hu, H. Kokubo, B.M. Pettitt, Fast calculations of electrostatic solvation free energy from reconstructed solvent density using proximal radial distribution functions, J. Phys. Chem. Lett. 2 (13) (2011) 1626–1632.
[31] S. Swaminathan, D.L. Beveridge, A theoretical study of the structure of liquid water based on quasi-component distribution functions, J. Am. Chem. Soc. 99 (26) (1977) 8392–8398.
[32] W.R. Rudnicki, B.M. Pettitt, Modeling the DNA–solvent interface, Biopolymers 41 (1) (1997) 107–119.
[33] J.S. Perkyns, B.M. Pettitt, A dielectrically consistent interaction site theory for solvent–electrolyte mixtures, Chem. Phys. Lett. 190 (6) (1992) 626.
[34] J.S. Perkyns, B.M. Pettitt, A site–site theory for finite concentration saline solutions, J. Chem. Phys. 97 (10) (1992) 7656.
[35] K.-C. Ng, Hypernetted chain solutions for the classical one-component plasma up to $\Gamma = 7000$, J. Chem. Phys. 61 (1974) 2680.
[36] J. Perkyns, B.M. Pettitt, Computationally useful bridge diagram series for the structure and thermodynamics of Lennard–Jones fluids, Theor. Chem. Accounts 96 (1997) 61.
[37] J.S. Perkyns, K.M. Dyer, B.M. Pettitt, Computationally useful bridge diagram series ii. Diagrams in H-bonds, J. Chem. Phys. 116 (2002) 9404.
[38] K.M. Dyer, J.S. Perkyns, B.M. Pettitt, Computationall useful bridge diagram series. iii. Lennard–Jones mixtures, J. Chem. Phys. 116 (2002) 9413.
[39] J. Perkyns, B.M. Pettitt, Erratum: computationally useful bridge diagram series for the structure and thermodynamics of Lennard–Jones fluids, Theor. Chem. Accounts 99 (1998) 207.
[40] J.J. Howard, G.C. Lynch, B.M. Pettitt, Ion and solvent density distributions around canonical B-DNA from integral equations, J. Phys. Chem. B 115 (3) (2011) 547–556.
[41] T.M. Raschke, Water structure and interactions with protein surfaces, Curr. Opin. Struct. Biol. 16 (2) (2006) 152–159.
[42] K.M. Dyer, B.M. Pettitt, Proximal distributions from angular correlations: a measure of the onset of coarse-graining, J. Chem. Phys. 139 (21) (2013).
[43] V. Lounnas, B.M. Pettitt, G.N. Phillips, A global-model of the protein–solvent interface, Biophys. J. 66 (3) (1994) 601–614.
[44] T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallogr. A: Found. Crystallogr. 47 (Pt 2) (1991) 110–119 (journal Code: 8305825; Chemical Name: 0 (Proteins)).
[45] Bin Lin, B.M. Pettitt, On the Universality of proximal Radial Distribution Functions of Proteins, J. Chem. Phys. 134 (2011) 106101/1–106101/2.
[46] B.-L. Nguyen, B.M. Pettitt, Effects of acids, bases, and heteroatoms on proximal radial distribution functions of proteins, J. Phys. Chem. B (2014) (in preparation).
[47] W. Humphrey, A. Dalke, K. Schulten, VMD—visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[48] N. Foloppe, J.A.D. MacKerell, All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data, J. Comput. Chem. 21 (2) (2000) 86–104.
[49] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[50] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. KalÃ©, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (16) (2005) 1781–1802.
[51] R. Elber, Ligand diffusion in globins: simulations versus experiment, Curr. Opin. Struct. Biol. 20 (2) (2010) 162–167.