DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Title: Potential-of-Mean-Force Description of Ionic Interactions and Structural Hydration in Biomolecular Systems

Author(s): Gerhard Hummer, Dikeos Mario Soumpasis and Angel E. Garcia

Submitted to: "Non-linear Excitations in Biomolecules"
Editions de Physique, Michel Peyrard, Editor
Proceedings of the conference held at Centre de Physique des Houches, France

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Potential-of-Mean-Force Description of Ionic Interactions and Structural Hydration in Biomolecular Systems

Gerhard Hummer*, Dikeos Mario Soumpasis[†], and Angel E. García*,

*Theoretical Biology and Biophysics Group T-10, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545, U.S.A.

[†]Biocomputation Group, Max-Planck-Institute for Biophysical Chemistry, P.O. Box 2841, D-37018 Göttingen, F.R.G.

Internet: hummer@t10.lanl.gov

Phone: +1-505-665-1923

FAX: +1-505-665-3493

1. INTRODUCTION

To understand the functioning of living organisms on a molecular level, it is crucial to dissect the intricate interplay of the immense number of biological molecules. However, large biological macromolecules are not the only players in the field. Most of the biochemical processes in cells occur in a liquid environment formed mainly by water and ions. This solvent environment plays an important role in biological systems [1]. It mediates biochemical reactions and has a strong influence on the structural equilibrium of certain molecules. Therefore, the development and application of theoretical descriptions of solute-solvent interactions provides relevant insight into biological processes.

The potential-of-mean-force (PMF) formalism attempts to describe quantitatively the interactions of the solvent with biological macromolecules on the basis of an approximate statistical-mechanical representation. At its current status of development, it deals with ionic effects on the biomolecular structure and with the structural hydration of biomolecules. The underlying idea of the PMF formalism is to identify the dominant sources of interactions and incorporate these interactions into the theoretical formalism using PMF's (or particle correlation functions) extracted from bulk-liquid systems.

In the following, we shall briefly outline the statistical-mechanical foundation of the PMF formalism and introduce the PMF expansion formalism, which is intimately linked to superposition approximations for higher-order particle correlation functions. We shall then sketch applications, which describe the effects of the ionic environment on nucleic-acid structure. Finally, we shall present the more recent extension of the PMF idea to describe quantitatively the structural hydration of biomolecules. Results for the interface of ice and water and for the hydration of deoxyribonucleic acid (DNA) will be discussed.

2. THE STATISTICAL-MECHANICAL FOUNDATION OF THE POTENTIAL-OF-MEAN-FORCE FORMALISM

The structural equilibrium properties of fluid systems are best described in terms of a hierarchy of particle correlation functions. Within the framework of statistical mechanics applied on classical-mechanical systems, the particle correlation functions can be expressed as integrals over the configuration space of the system. In the following, we consider a monoatomic liquid in a canonical (N, V, T) ensemble. The N particles are subject to interactions described by a potential U depending on the positions $\mathbf{r}_1, \dots, \mathbf{r}_N$ of the N particles. Then, the n -particle correlation functions are defined as

$$g^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n) = \rho_0^{-n} \frac{N!}{(N-n)!} \frac{\int d\mathbf{r}_{n+1} \cdots d\mathbf{r}_N \exp[-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)]}{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N \exp[-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)]}. \quad (1)$$

Here, $\beta = 1/k_B T$ is the inverse temperature; and $\rho_0 = N/V$ is the bulk density of the solvent. No external field is present; and the 1-particle correlation function is independent of the position: $g^{(1)}(\mathbf{r}) \equiv 1$.

To introduce the concept of potentials of mean force, we fix n of the N particles at positions $\mathbf{r}_1, \dots, \mathbf{r}_n$. In the canonical ensemble, the mean force acting on particle i ($1 \leq i \leq n$) is calculated as

$$\mathbf{F}_i(\mathbf{r}_1, \dots, \mathbf{r}_n) = \frac{\int d\mathbf{r}_{n+1} \cdots d\mathbf{r}_N (-\partial U / \partial \mathbf{r}_i) \exp[-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)]}{\int d\mathbf{r}_1 \cdots d\mathbf{r}_N \exp[-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)]}. \quad (2)$$

Changing the order of differentiation and integration yields a potential $W^{(n)}$ for the mean force,

$$\mathbf{F}_i(\mathbf{r}_1, \dots, \mathbf{r}_n) = -\frac{\partial}{\partial \mathbf{r}_i} W^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n), \quad (3)$$

where the n -particle potential of mean force (PMF) is obtained as

$$W^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n) = -k_B T \ln g^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n). \quad (4)$$

This provides an important link between the microscopic structure of the fluid and the thermodynamic properties of inhomogeneous systems. The mean force $\mathbf{F}_i$ integrated over a path $\mathbf{r}_i(\tau)$ connecting $\mathbf{r}_i^1$ and $\mathbf{r}_i^2$ gives the reversible work of bringing particle i from $\mathbf{r}_i^1$ to $\mathbf{r}_i^2$ in the presence of $n-1$ particles constrained to $\mathbf{r}_1, \dots, \mathbf{r}_{i-1}$ and $\mathbf{r}_{i+1}, \dots, \mathbf{r}_n$. Correspondingly, if all particles j are moved from $\mathbf{r}_j^1$ to $\mathbf{r}_j^2$, the free energy (reversible work!) required for this operation is given by $W^{(n)}(\mathbf{r}_1^2, \dots, \mathbf{r}_n^2) - W^{(n)}(\mathbf{r}_1^1, \dots, \mathbf{r}_n^1)$.

To achieve practical relevance for these formulae, one has to devise ways to calculate $W^{(n)}$ (or, correspondingly, $g^{(n)}$) or find at least approximate representations. However, already the accurate calculation of $g^{(2)}$ for simple (pair additive, spherically symmetric) models for the interaction potential U proves to be a nontrivial task [2]. This is even more so for higher-order correlations ($n \geq 3$). Nevertheless, by means of computer simulations or integral-equation techniques, two- and three-particle correlations can be calculated with reasonable accuracy. This guides the development toward expansion of the higher-order PMF's in terms of expressions of lower-order. This goal is accomplished by the

so-called PMF expansion [3]. We attempt to rewrite our function $W^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n)$ (which is symmetric in its n arguments) as a sum of functions $w^{(m)}$ of m arguments ($1 \leq m \leq n$),

$$W^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n) = \sum_i w^{(1)}(\mathbf{r}_i) + \sum_{\substack{i,j \\ n \geq i > j \geq 1}} w^{(2)}(\mathbf{r}_i, \mathbf{r}_j) + \sum_{\substack{i,j,k \\ n \geq i > j > k \geq 1}} w^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots \quad (5)$$

Inversion of Eq. 5 yields

$$\begin{aligned} w^{(1)}(\mathbf{r}) &= W^{(1)}(\mathbf{r}) \\ w^{(2)}(\mathbf{r}_1, \mathbf{r}_2) &= W^{(2)}(\mathbf{r}_1, \mathbf{r}_2) - W^{(1)}(\mathbf{r}_1) - W^{(1)}(\mathbf{r}_2) \\ w^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) &= W^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) - W^{(2)}(\mathbf{r}_1, \mathbf{r}_2) - W^{(2)}(\mathbf{r}_2, \mathbf{r}_3) - W^{(2)}(\mathbf{r}_3, \mathbf{r}_1) \\ &\quad + W^{(1)}(\mathbf{r}_1) + W^{(1)}(\mathbf{r}_2) + W^{(1)}(\mathbf{r}_3) . \end{aligned} \quad (6)$$

For a homogeneous system, $g^{(1)} \equiv 1$ and $W^{(1)} \equiv 0$. From Eqs. 4, 5, and 6 we obtain an expansion for $W^{(n)}$ in terms of particle correlation functions,

$$\begin{aligned} -\beta W^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n) &= \sum_{\substack{i,j \\ n \geq i > j \geq 1}} \ln g^{(2)}(\mathbf{r}_i, \mathbf{r}_j) \\ &\quad + \sum_{\substack{i,j,k \\ n \geq i > j > k \geq 1}} \ln \frac{g^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k)}{g^{(2)}(\mathbf{r}_i, \mathbf{r}_j) g^{(2)}(\mathbf{r}_j, \mathbf{r}_k) g^{(2)}(\mathbf{r}_k, \mathbf{r}_i)} + \dots \end{aligned} \quad (7)$$

Eqs. 4 and 7 combined yield a product expansion for n -particle correlation functions,

$$\begin{aligned} g^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n) &= \left[\prod_{\substack{i,j \\ n \geq i > j \geq 1}} g^{(2)}(\mathbf{r}_i, \mathbf{r}_j) \right] \\ &\quad \times \left[\prod_{\substack{i,j,k \\ n \geq i > j > k \geq 1}} \frac{g^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k)}{g^{(2)}(\mathbf{r}_i, \mathbf{r}_j) g^{(2)}(\mathbf{r}_j, \mathbf{r}_k) g^{(2)}(\mathbf{r}_k, \mathbf{r}_i)} \right] \dots \end{aligned} \quad (8)$$

Generalized superposition approximations for the n -particle correlation functions are obtained by setting $w^{(n)} \equiv 0$ for $n \geq n_0$. In particular, the Kirkwood superposition approximation (KSA) [4] is obtained for $n_0 = 3$,

$$g_{\text{KSA}}^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \prod_{\substack{i,j \\ 3 \geq i > j \geq 1}} g^{(2)}(\mathbf{r}_i, \mathbf{r}_j) . \quad (9)$$

The Fisher-Kopeliovich superposition approximation [5] corresponds to $n_0 = 4$,

$$g_{\text{FKSA}}^{(4)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \frac{\prod_{\substack{i,j,k \\ 4 \geq i > j > k \geq 1}} g^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k)}{\prod_{\substack{i,j \\ 4 \geq i > j \geq 1}} g^{(2)}(\mathbf{r}_i, \mathbf{r}_j)} . \quad (10)$$

These two approximations form a cornerstone of the PMF formalism to describe biomolecular systems in solution.

3. INTERACTIONS OF NUCLEIC ACIDS WITH IONIC SOLUTIONS

DNA forms an important component of biological systems as the carrier of genetic information. DNA molecules show a remarkable variability in their structure depending on base composition and solution environment [6]. With respect to salt effects on the structure of DNA, the transition of some base sequences from a right-handed B form (low salt) to a left-handed Z form has been particularly well characterized [7]. This structural transition offers an important test ground for any theory attempting to describe ionic interactions with biomolecules.

The PMF formalism can indeed successfully account for the experimentally observed behavior, i.e., the free-energy difference between the two DNA structures depending on ion types and concentrations [8–10]. In the following, we shall briefly outline the general ideas underlying the PMF description of DNA interacting with ionic solutions. More general presentations of the PMF method and its applications on DNA-ion interactions can be found in Refs. [11, 12].

Nucleic acid molecules form negatively-charged polyelectrolytes. The sugar-phosphate backbone connecting the bases carries one negative charge per phosphate group (PO_4^-). As mentioned in the introduction, the key idea of the PMF formalism is to identify the dominant interactions in the system. With respect to the structure dependence on the ionic environment, the charge-charge interaction is expected to be of foremost importance. In low-concentration electrolytes, the negatively-charged phosphate groups are exerting a strong charge repulsion on each other. However, at higher ionic concentrations this picture changes. It is well known from the theory of liquids that at high ionic concentrations the effective interactions, i.e., the PMF's, are of oscillatory behavior reflecting attractive interactions over some distance regions. This qualitative change in the ion-ion PMF's results in a shifted structural equilibrium for certain DNA sequences.

The simplest model for the DNA in the PMF formalism is that of a group of negatively-charged anions centered at the positions of the phosphate atoms. A structural transition of the DNA then corresponds to a rearrangement of these anions. The anions representing the PO_4^- groups are described as restricted-primitive-model (RPM) ions. This pair-interaction model contains the essential features that characterize ions in water, i.e., a hard-sphere interaction modeling the short-range repulsion of hydrated ions and a dielectrically screened long-range Coulomb interaction.

The formal development of the previous section provides us with a tool to find approximate expressions for the free-energy difference between two DNA structures owing to ionic contributions. Other contributions, in particular energetic [13] and vibrational components [14], have to be accounted for differently. For a given solution environment (temperature, ion types and concentrations) the difference in free energy ΔF between two conformations 1 and 2 is calculated as

$$\Delta F(1 \rightarrow 2) = W^{(n)}(\mathbf{r}_1^2, \dots, \mathbf{r}_n^2) - W^{(n)}(\mathbf{r}_1^1, \dots, \mathbf{r}_n^1), \quad (11)$$

where n is the number of phosphate groups (modeled as anions) and $\mathbf{r}_i^1, \mathbf{r}_i^2$ are their initial and final positions, respectively. Employing the expansion Eq. 7 of $W^{(n)}$ in terms of particle correlation functions of anions in RPM electrolytes, we obtain an approximate

expression

$$\Delta F(1 \rightarrow 2) = -k_B T \sum_{\substack{i,j \\ n \geq i > j \geq 1}} \ln \frac{g_{---}^{(2)}(r_i^2, r_j^2)}{g_{---}^{(2)}(r_i^1, r_j^1)} - k_B T \sum_{\substack{i,j,k \\ n \geq i > j > k \geq 1}} \ln \frac{\Gamma_{---}^{(3)}(r_i^2, r_j^2, r_k^2)}{\Gamma_{---}^{(3)}(r_i^1, r_j^1, r_k^1)} + \dots, \quad (12)$$

where

$$\Gamma_{---}^{(3)}(\mathbf{r}, \mathbf{s}, \mathbf{t}) = \frac{g_{---}^{(3)}(\mathbf{r}, \mathbf{s}, \mathbf{t})}{g_{---}^{(2)}(\mathbf{r}, \mathbf{s}) g_{---}^{(2)}(\mathbf{s}, \mathbf{t}) g_{---}^{(2)}(\mathbf{t}, \mathbf{r})}. \quad (13)$$

In the case of 1:1 RPM electrolytes in the concentration regime up to about 4 mol/l, substantial contributions to the free energy come mainly from the pair PMF's. This is a consequence of the reasonable quality of the KSA (Eq. 9) for the RPM system [15]. The three-particle PMF's add only minor corrections to ΔF , as shown in the case of the B- to Z-DNA transition in NaCl [16]. The required particle correlation functions can be calculated accurately by means of integral equation methods such as the hypernetted chain (HNC) or exponential mean-spherical approximation (EXP-MSA).

The approximate expression for the free energy difference in terms of pair PMF's of RPM anions has been used to describe the salt dependent transitions of right-handed B-DNA to left-handed Z-DNA in sodium chloride solution [8] and in solutions containing other alkali cations [10]. The equilibrium between different DNA structures (A, B, C, and Z) has been studied in Ref. [9]. The PMF formalism has been compared with other theoretical approaches (Poisson-Boltzmann equation, counterion-condensation theory) in Ref. [17]. The salt-concentration dependent equilibrium between hairpin, mismatched duplex, and single-stranded DNA has been discussed in Ref. [18]. A similar approach has been used to calculate the salt-dependent melting temperature of DNA (helix-coil transition) within the PMF method [11]. The PMF interaction between charged atoms has been incorporated into force-field descriptions of biomolecules [13]. This allows an approximate representation of the effective, water-averaged Coulomb interactions emerging from a McMillan-Mayer picture of ionic systems.

A similar method has been used to study the harmonic vibrations of left- and right-handed DNA depending on the salt concentration [14]. The effective Coulomb interaction between different phosphate groups has been described by PMF's between anion pairs. In this study of the normal modes of B- and Z-DNA it has been observed that the lowest-frequency eigenmode of B-DNA shows a significant frequency decrease when increasing the salt concentration. This indicates that a soft-mode mechanism might play an important role in the B-Z transition.

4. A STATISTICAL-MECHANICAL THEORY OF BIOMOLECULAR HYDRATION

4.1 Introductory remarks

In the original PMF formalism dealing with ionic effects on DNA, water is described as a dielectric continuum in which the (hydrated) ions are immersed. This simplified picture works well as long as *structural* water does not play an important role in the effects studied.

In the case of DNA, the experimental results (reviewed, for instance, in Refs. [6, 19–21]) show a sequence- and conformation-dependent network of water molecules close to the DNA surface caused by short-range hydrophilic interactions with polar DNA atoms (oxygen, nitrogen). The most prominent feature in the case of B-DNA is the so-called *spine of hydration* [22, 23] covering tracts of A-T base pairs (1) in the minor groove, with water molecules bridging adjacent base pairs and themselves being bridged by other water molecules. Such highly localized water molecules at the DNA surface are thought to affect the structural stability of the DNA and to interfere with the binding of drug molecules to DNA [24].

It becomes evident that water must be seen as an integral part of biomolecular systems. Changes in the biomolecular structure or binding reactions (enzyme catalysis, antibody-antigen binding, drug binding, binding to carrier molecules, etc.) are accompanied by a reorganization of water molecules at the molecular surface. Water mediates interactions between different groups on biomolecules and between different molecules. Our understanding of these biologically important processes is therefore intimately coupled to having insight into the role of water in these systems.

In the following, we shall describe a method to computing quantitative measures of the biomolecular hydration based on the ideas of the PMF formalism. A similar approach has been used to compute ion-density distributions near biological macromolecules [25, 26].

4.2 Density expansion in inhomogeneous liquid systems

We shall study a system comprising a biological macromolecule (or a biomolecular assembly) and water molecules. The biomolecule (solute) is treated as a source of inhomogeneity affecting the structural organization of water in its vicinity. The central quantity to describe the structural equilibrium properties of the water phase is the position-dependent one-particle density of water $\rho(\mathbf{r})$. The primary goal of the theoretical development is to devise approximate expressions for $\rho(\mathbf{r})$ (in terms of known quantities) in the presence of a biological macromolecule (or other solid-like sources of inhomogeneity).

The biomolecule is represented by a set of positional coordinates $\{\mathbf{r}_{i\alpha}\}$ of N_α atoms of type α and M different atom types α . We shall assume a rigid equilibrium structure of the molecule, as known from X-ray or neutron diffraction crystallography, NMR, or theoretical modeling. This assumption can be relaxed by performing an additional average over representative ensembles of structures, as, for instance, obtained from NMR spectroscopy.

To keep the notation simple, we shall develop the theory for an atomic solvent. The generalization to molecular solvents (water) is straightforward, either by introducing additional angular coordinates or by using an atom-site description. We consider a canonical ensemble of N solvent particles at positions $\{\mathbf{r}_i\}$ within the framework of statistical mechanics of classical-mechanical systems. The total potential energy of the system com-

(1) The following abbreviations for the DNA bases are used: Adenine (A), thymine (T), guanine (G), and cytosine (C). The deoxyribose in the sugar-phosphate backbone of DNA is denoted by $d(\cdots)$. The strand polarity of the backbone is represented by writing the base sequence in the 5' to 3' direction. Dinucleotide steps XY are written as 5'-d(XpY), where p represents the sugar-phosphate linkage. A dot denotes Watson-Crick base pairing (i.e., A-T and G-C). The DNA molecules studied in this work are all double stranded forming Watson-Crick base pairs.

prising solute and solvent molecules is denoted by U . For pair-additive interactions, it can be split into two parts such that the solute appears as an external field acting on the solvent molecules. The conditional solvent density at a point $\mathbf{r}_1$ can be expressed as an integral over the configuration space of the solvent molecules,

$$\rho(\mathbf{r}_1|\{\mathbf{r}_{i_\alpha}\}) = N \frac{\int d\mathbf{r}_2 \cdots d\mathbf{r}_N \exp[-\beta U(\{\mathbf{r}_i\}, \{\mathbf{r}_{i_\alpha}\})]}{\int d\{\mathbf{r}_i\} \exp[-\beta U(\{\mathbf{r}_i\}, \{\mathbf{r}_{i_\alpha}\})]}. \quad (14)$$

We now introduce higher-order particle correlation functions describing the correlations of sets of atoms $\{N_\alpha\}$ and water, as generalized for multiparticle systems from Eq. 1,

$$\rho(\mathbf{r}|\{\mathbf{r}_{i_\alpha}\}) = \rho_0 \frac{g^{(1;\{N_\alpha\})}(\mathbf{r}, \{\mathbf{r}_{i_\alpha}\})}{g^{(\{N_\alpha\})}(\{\mathbf{r}_{i_\alpha}\})}, \quad (15)$$

where $\rho_0 = N/V$ is the bulk density of the solvent and

$$\begin{aligned} g^{(\{N_\alpha\})}(\{\mathbf{r}_{i_\alpha}\}) &= \left(\prod_{\alpha=1}^M N_\alpha! \rho_\alpha^{-N_\alpha} \right) \frac{\int d\{\mathbf{r}_i\} \exp[-\beta U(\{\mathbf{r}_i\}, \{\mathbf{r}_{i_\alpha}\})]}{\int d\{\mathbf{r}_{i_\alpha}\} d\{\mathbf{r}_i\} \exp[-\beta U(\{\mathbf{r}_i\}, \{\mathbf{r}_{i_\alpha}\})]}, \quad (16) \\ g^{(1;\{N_\alpha\})}(\mathbf{r}_1, \{\mathbf{r}_{i_\alpha}\}) &= N \rho_0^{-1} \left(\prod_{\alpha=1}^M N_\alpha! \rho_\alpha^{-N_\alpha} \right) \frac{\int d\mathbf{r}_2 \cdots d\mathbf{r}_N \exp[-\beta U(\{\mathbf{r}_i\}, \{\mathbf{r}_{i_\alpha}\})]}{\int d\{\mathbf{r}_{i_\alpha}\} d\{\mathbf{r}_i\} \exp[-\beta U(\{\mathbf{r}_i\}, \{\mathbf{r}_{i_\alpha}\})]}, \quad (17) \end{aligned}$$

with $\rho_\alpha = N_\alpha/V$. Eq. 15 is analogous to the well-known formula for conditional probabilities of events A and B ,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (18)$$

The exact calculation of the higher-order correlation functions $g^{(\{N_\alpha\})}$ and $g^{(1;\{N_\alpha\})}$ is not possible with currently available means. However, the PMF expansion Eq. 8 provides the tool to expand them in terms of lower-order correlations. We obtain a product expansion,

$$\begin{aligned} \rho(\mathbf{r}|\{\mathbf{r}_{i_\alpha}\}) &= \rho_0 \left[\prod_{\alpha=1}^M \prod_{i_\alpha=1}^{N_\alpha} g^{(1;\alpha)}(\mathbf{r}, \mathbf{r}_{i_\alpha}) \right] \\ &\times \left[\prod_{\alpha=1}^M \prod_{i_\alpha=1}^{N_\alpha} \prod_{\beta=1}^M \prod_{i_\beta=1+\delta_{\alpha\beta}i_\alpha}^{N_\beta} \frac{g^{(1;\alpha,\beta)}(\mathbf{r}, \mathbf{r}_{i_\alpha}, \mathbf{r}_{i_\beta})}{g^{(1;\alpha)}(\mathbf{r}, \mathbf{r}_{i_\alpha}) g^{(\alpha,\beta)}(\mathbf{r}_{i_\alpha}, \mathbf{r}_{i_\beta}) g^{(\beta;1)}(\mathbf{r}_{i_\beta}, \mathbf{r})} \right] \cdots, \quad (19) \end{aligned}$$

where $\delta_{\alpha\beta}$ is the Kronecker symbol. This expression relates the conditional solvent density to products of correlation functions involving one solvent particle and one, two, etc., solute atoms. The first term in the expansion Eq. 19 is the bulk-density of the solvent, the next order yields a product over pair correlations involving individual solute atoms; and the triplet term involves a product over all distinct pairs of solute atoms.

4.3 The water-density profile at the interface of ice and water

We shall first illustrate the applicability of the expansion formula Eq. 19 in the context of a planar solid-liquid interface formed by ice and water. A comparison with reference data from computer simulations of the same system allows us to assess the quality of the density

Figure 1: A snapshot of an MC simulation of the ice-water interface using 1000 water molecules. The simulation box comprises one ice and one liquid water layer. The second ice layer arises from the use of periodic boundary conditions.

expansion method under well-defined conditions. In particular, we are able to investigate the reliability of the truncation of the expansion after the two- and three-particle level in the case of water. Results for the ice-water interface have been presented in Ref. [27]. Here, we shall review this study to give an instructive example for applications of the density expansion methods.

To model the interface, we fix water oxygens at ideal hexagonal-ice Ih lattice positions $\mathbf{r}_1, \dots, \mathbf{r}_n$, but do not specify molecular orientations. The thermal disorder in the interfacial region of the solid phase is therefore limited to reorientation of the water molecules. This ice layer is embedded into liquid water, giving rise to a structured density distribution in the liquid phase. This distribution has been calculated both using the PMF expansion formalism and extensive Monte-Carlo (MC) computer simulations.

The application of the general expression Eq. 19 is simplified since only water molecules are present. We obtain an expression for the water-oxygen and hydrogen density in the liquid phase in terms of two- and three-particle correlation functions of bulk water,

$$\rho_X(\mathbf{r}|\mathbf{r}_1, \dots, \mathbf{r}_n) \approx \rho_X \prod_{i=1}^n g_{XO}^{(2)}(\mathbf{r}, \mathbf{r}_i) \prod_{j=1}^{n-1} \prod_{k=j+1}^n \frac{g_{XOO}^{(3)}(\mathbf{r}, \mathbf{r}_j, \mathbf{r}_k)}{g_{XO}^{(2)}(\mathbf{r}, \mathbf{r}_j) g_{OO}^{(2)}(\mathbf{r}_j, \mathbf{r}_k) g_{OX}^{(2)}(\mathbf{r}_k, \mathbf{r})}, \quad (20)$$

where $X=O$ or H . Since only the pair but not the triplet correlations of fluid phases are accessible to an experimental determination using currently available techniques, we performed computer simulations of bulk water to obtain the correlation functions. $g^{(2)}$ and $g^{(3)}$ were calculated for the simple-point-charge (SPC) model of water [28], as described in Ref. [27].

To provide reliable density data for the interface of ice Ih and water that allow a careful testing of the PMF expansion method, a series of MC simulations of the interfacial system has been performed. We position 240 water molecules in an ideal ice-Ih lattice. The oxygens are constrained to lattice sites; but the molecules are free to rotate. A rectangular box is used with sides $L_x = 2.245$ nm and $L_y = 2.333$ nm. The two basal planes of the ice layer are oriented perpendicular to the z axis. The volume above the plane is filled with mobile water molecules. The temperature is kept at 298 K. System sizes of 600 and 1000 water molecules in total at an average density of $\rho_0 = 33.33 \text{ nm}^{-3}$ are

Figure 2: Water-oxygen density at the ice-water interface in units of the bulk-water density ρ_0 . z measures the distance from the closest layer of oxygen atoms in the ice phase. (+), PMF expansion including only pair correlations; ($\diamond$), PMF expansion including pair and triplet correlations. MC simulations: (—), GRF, $N = 600$, $r_c = 0.9$ nm; (---), GRF, $N = 600$, $r_c = 0.8$ nm; (- - -), GRF, $N = 1000$, $r_c = 0.9$ nm; ($\cdots$), Ewald, $N = 600$.

studied, resulting in $L_z = 3.436$ and 5.726 nm, respectively. Periodic boundary conditions are applied in all three directions, such that the system effectively consists of an ice layer (1.192 nm thick) in contact with liquid water on both sides. Fig. 1 shows a snapshot of an MC simulation with the liquid phase enclosed by two ice layers. The charge-charge interaction has been treated with Ewald summation [29] and a generalized reaction field (GRF) method [27, 30] using different cutoff radii r_c .

The z -dependent oxygen and hydrogen density in layers parallel to the interface (basal plane) obtained from four simulations with different system sizes and charge interactions is shown in Figs. 2 and 3. Also shown are the corresponding results obtained from PMF calculations, truncating Eq. 20 at the pair and triplet level.

Not surprisingly, if only the pair correlations are included in the PMF expansion, we observe serious disagreement between simulations and PMF calculations. In particular, the oxygen density shows two distinct peaks: The first one is too close and the second one is too weak. The PMF hydrogen density calculated with pairs alone shows somewhat better agreement with the simulations.

However, if the three-particle corrections are applied in the PMF expansion, the agreement drastically improves. The first peak in the oxygen density is quantitatively reproduced (position and height). Also, the strong second peak in the hydrogen density is found in quantitative agreement with the simulations. The first peak of the hydrogen density is somewhat too small. We ascribe this discrepancy to problems with the discretization of the $g_{OOH}^{(3)}$ correlations. These are known with a resolution of 0.02 nm, which results

Figure 3: Water-hydrogen density at the ice-water interface in units of $2 \rho_0$. Details as in Fig. 2.

in large jumps at small OH pair distances only poorly described by linear interpolation. For larger distances ($z > 0.4$ nm), the agreement between the PMF expansion and the Monte-Carlo simulation is somewhat less satisfactory. However, qualitative agreement is observed. In particular, the computer simulations consistently show minima in the oxygen and hydrogen densities at $z = 0.75$ and 0.8 nm, respectively. These minima also appear in the PMF calculations at the right positions but more shallow.

It becomes evident that it is essential to retain the water-triplet correlations in the expansion Eq. 20. They contain the information about the preferentially tetrahedral, ice-like organization of water molecules caused by hydrogen bonding. Three water molecules in close contact typically form isosceles triangles with edges of 0.275 , 0.275 , and 0.45 nm length rather than the KSA prediction of equilateral triangles with edges of 0.275 nm length. This KSA prediction results in the erroneous peak of the oxygen density at small z values and the underestimation of the contact peak at $z = 0.26$ nm. The inclusion of the water-triplet correlations on the other hand results in agreement of phase and amplitude of the density distribution. The quality of the expansion results is particularly high in the interfacial region, where the strongest variations in the density profile occur. This observation is of special relevance for further studies of biomolecular hydration where the major emphasis lies on the identification of highly localized water.

4.4 Hydration of biological macromolecules: Results for DNA

A prohibitively large number of correlation functions would be required to describe the hydration of solvated macromolecules such as proteins and nucleic acids using Eq. 19 directly. As a consequence, it is important to simplify the theoretical description with regard to the number of different atom types α used to model the macromolecules, recalling

that the main interactions between water and a solvated macromolecule are hydrogen bonds.

In the case of nucleic acids, we classify the atoms as polar (oxygen, nitrogen) and nonpolar (carbon). The hydrogen bonds involving oxygen and nitrogen atoms are within narrow bands concerning energy, bond length, and bond angle; and they are similar to those formed between water molecules. In the following we shall equate all nitrogen and oxygen atoms to water oxygen with respect to their interactions with water. The nonpolar atoms (carbon) are represented as excluded volume (unit step function for $g^{(2)}$, no triplet correction). This treatment of the nonpolar atoms might require some refinement in the case of large hydrophobic surface regions, e.g., by using correlation functions of methane in water. However, in the case of nucleic acids studied here, the molecular surface is essentially hydrophilic justifying the simplification. The negative net charge of the phosphate groups PO_4^- on the DNA is expected to be well represented by its four oxygens, each carrying a negative partial charge. As a consequence, we use Eq. 20 (disregarding excluded volume effects of the nonpolar atoms) to compute the water density, where $r_1, \dots, r_n$ are the positions of the electronegative atoms oxygen and nitrogen on DNA.

Some results for the hydration of DNA oligomers are described in Ref. [31]. Preliminary results for the hydration of proteins have been presented [32]. A comparison with published simulation results for canonical B-DNA oligomers by Forester and McDonald [33] has been submitted for publication.

In this study we use water pair and triplet correlations computed separately in collaboration with P. Procacci and G. Corongiu at IBM, Kingston [34], from an MD simulation of the highly refined *ab initio* Niesar-Corongiu-Clementi model of water, which includes n -body polarization effects and yields very good results for thermodynamic, structural, spectroscopic, and transport properties of the liquid [35]. The extensive pair and triplet correlational database thus obtained at a discretization of 0.02 nm step size is used in Eq. 20 along with multilinear interpolation subroutines to obtain the results described next.

In this work we shall report results for DNA duplexes in the B-family of conformations. Water-oxygen densities are computed at vertices of a three-dimensional cylindrical volume with a height of 4.2 nm and a radius of 1.2 nm. Global helix axes are obtained from diagonalization of the molecules' tensor of inertia as discussed in Ref. [36]. Grid spacings are used of $\Delta r = 0.015$ nm in radial, $\Delta z = 0.03$ nm in axial, and $\Delta\phi = 2\pi/100$ in angular direction. The atomic coordinates have been taken from the Brookhaven database for biomolecular structures. The carbon atoms are surrounded by a sphere of radius 0.25 nm, from which water oxygens are excluded. The triplet correction in Eq. 20 is applied to triangles with all edges between 0.23 nm and 0.79 nm, which is the range considered in the water triplet simulations [34].

In Fig. 4, we present plots of the main features of DNA structural hydration obtained for the Dickerson B-dodecamer $\text{d}(\text{CGCGAATTCGCG})_2$ [22] re-refined by Westhof [37] and two relatives, $\text{d}(\text{CGCATATATGCG})_2$ [38] and $\text{d}(\text{CGCAAAAATGCG})_2$ [39], of which the so-called "up" structure is analyzed. The water-oxygen densities are depicted on cylindrical surfaces with different radii around the DNA molecules using different colors (green for low, blue for bulk, and red for high densities). The densities are averaged over 0.2 nm cylindrical shells.

In all cases, we find a marked *spine of hydration* in the narrow minor groove containing the A-T base pairs, which splits into two less pronounced side-by-side water *ribbons* in

Figure 4: Water-oxygen distributions around B-DNA structures on cylindrical surfaces with radii 6, 8, 10, and 12 Å. The densities are averaged over cylindrical shells with radii 4–6, 6–8, 8–10, and 10–12 Å, respectively. Green stands for low, blue for bulk, and red for high densities, as indicated by the color scale. The DNA phosphate groups are shown in light red, the sugar atoms in white, adenine in red, thymine in green, cytosine in yellow, and guanine in blue. Crystal structures: (A) d(CGCGAATTCGCG)₂, (B) d(CGCATATATGCG)₂, and (C) d(CGCAAAA-ATGCG)₂; Canonical B-DNA: (D) d[(AATT)₃]₂ and (E) d[(CCGG)₃]₂.

the wider minor groove at the two ends of the structures, where the G·C base pairs are situated. By and large, the major grooves do not contain strongly localized water, and localization along the sugar-phosphate backbone is less pronounced except for high probability regions at the phosphates. O-3' and O-5' neighborhoods as well as the sugars are free of strongly localized water. All these findings are in complete agreement with the experimental data.

Another interesting result of the computation concerns the sequence dependence of the *spine of hydration* in the A+T-rich central regions of the molecules. Reflecting the different geometrical arrangement of the electronegative atoms, the 5'-d(TpA) dinucleotide organizes the water in its minor groove less than the 5'-d(ApT) dinucleotide, with the latter being a strong promoter of water localization. As can be seen particularly for d(CG-CATATATGCG)₂, the continuity of the high probability (*spine*) region is interrupted at the 5'-d(TpA) steps. Disruption of the *spine of hydration* at the 5'-d(TpA) step of another oligomer, namely d(CGATTAATCG)₂, has experimentally been found [40] and the weak tendency of 5'-d(TpA) to support a *spine* formation has also been observed in simple energy calculations [41]. However, it must be stressed that energetic considerations alone do not suffice in dense many-body systems where the central quantities are free energies. (PMF's are free energies *by definition* and often dominated by entropy contributions).

In Fig. 4 also, data are shown from computations with canonical (Arnott-Hukins) B-form structures [42] for d[(AATT)₃]₂ and d[(CCGG)₃]₂. Unlike crystallographic B-DNA, canonical B-DNA does not exhibit variations in the local helical parameters. In particular, the minor groove width is constant while crystallographic structures show a distinct sequence dependence with the minor groove in general being narrow at A·T and wide at G·C base pairs. Many features of DNA structural hydration described above are also found with the canonical structures. A·T sequences again exhibit the most pronounced localization (well-defined minor-groove *spines*); whereas, the minor groove of C·G base pairs is covered with a *double ribbon* of water molecules.

What we consider to be a particularly interesting result of our computations is the consistently observed unfavorable effect of the 5'-d(TpA) step on the hydration *spine* formation. It is present both in the case of crystal and canonical B structures, although less distinct for the latter. As has already been mentioned, it agrees both with X-ray studies [40] and simple (*in vacuo*) energy calculations [41]; and it could play a key role in many puzzling DNA structural phenomena. We mention here two examples.

Leroy *et al.* [43] have found that base-pair lifetimes in the T₃A₃ center of a DNA oligomer are much smaller than for the inverted sequence A₃T₃. Burkhoff and Tullius [44] report that A₄T₄ is better protected from cleavage by hydroxyl radicals than T₄A₄. We believe that this behavior is related to the absence of strongly localized water molecules, which as a byproduct results in poor stacking. If stacking were the principal cause, the 5'-d(TpA) *dinucleotide* in solution should stack as well as 5'-d(ApT) since these mini helices are much more free to optimize their geometry; but as has been found by NMR [45] 5'-d(TpA) shows less geometric overlap than 5'-d(ApT). However, the presence of a strongly localized water molecule in the minor groove at 5'-d(ApT) steps could explain the longer base-pair lifetimes. For the bases to open, this localized water molecule must break its *hydrogen bonds* to the bases, resulting in a higher energetic barrier to be overcome.

5. CONCLUSIONS

We have reviewed the formal basis of the PMF method and described applications on ionic effects in biomolecular systems. Expressing the Coulomb interactions between the negatively-charged phosphate groups on DNA as solvent-averaged PMF's, we achieve a conceptually clear and technically simple description of complex systems involving DNA molecules in electrolytic solution. This description captures the essential features, as established by comparison of experimental results and theoretical calculations for various salt-induced structural transitions of DNA [11]. Moreover, incorporation of the PMF's into a force field description of biomolecules allows efficient ways to study salt concentration dependent harmonic and anharmonic motions [14].

However, many important properties of biomolecular systems are strongly related to hydration phenomena. We have also presented a method to study the structural hydration of biomolecules. It approximates the water-density distribution near a macromolecule in solution (or, generally, at a solid-liquid interface) as a product expansion in terms of particle correlation functions. A calculation of the water-density profile at the interface of ice and water illustrates that by inclusion of three-particle corrections quantitative agreement with results of extensive computer simulations is achieved. In particular, position and magnitude of the high-density region are predicted correctly by the expansion formula.

This observation is of particular relevance for the study of biomolecular hydration, where a major interest lies in the identification of regions of strong water localization at the biomolecular surface. In this work, we have described some results for the hydration of B-DNA. We have made observations in agreement with the experimental studies. The narrow minor groove of B-DNA exhibits the most pronounced localization of water. In the case of A-T base-pair regions it shows a peaked, *spine*-like hydration, which is replaced by two double *ribbons* of high water density in C-G regions. Interestingly, 5'-d(TpA) steps are found to organize water less than 5'-d(ApT) steps.

We would like to point out two important aspects of the PMF hydration method. First, the required particle correlation functions need to be calculated only once (for given ρ and T). Then they can be used for various geometries and systems, minimizing the computational effort necessary. Second, the expansion method is local in space, i.e., to compute the density at a given position no explicit information about the density at any other point is required. Put together, this results in a computationally efficient method which allows to study large numbers of molecules (e.g., different mutants of a protein), large molecules or molecular assemblies (tRNA, complexes between proteins and nucleic acids), and localized regions in space (e.g., reaction centers). Various studies of this kind are planned for the near future.

Acknowledgments

We wish to thank Drs. T. M. Jovin, M. Neumann, R. Klement, P. Procacci, G. Corongiu, and E. Clementi for collaborations and discussions that greatly contributed to the realization of the project. This work has been funded by the Department of Energy (U.S.), the Bundesministerium für Forschung und Technologie, and the Max-Planck-Society (both F.R.G.).

References

- [1] Rupley J. A. and Careri G., *Adv. Protein Chem.* **41** (1991) 37–172.
- [2] Hansen J.-P. and McDonald I. R., *Theory of Simple Liquids* (Academic Press, London, UK, 1986).
- [3] Münster A., *Statistical Thermodynamics* (Springer, Berlin, 1969), Vol. 1, p. 338.
- [4] Kirkwood J. G., *J. Chem. Phys.* **3** (1935) 300–313.
- [5] Fisher I. Z. and Kopeliovich B. L., *Dokl. Akad. Nauk. SSSR* **133** (1960) 81–83 [*Sov. Phys.-Dokl.* **5** (1960) 761–763].
- [6] Saenger W., *Principles of Nucleic Acid Structure* (Springer, Berlin, 1984).
- [7] Jovin T. M., Soumpasis D. M., and McIntosh L. P., *Ann. Rev. Phys. Chem.* **38** (1987) 521–560.
- [8] Soumpasis D. M., *Proc. Natl. Acad. Sci. USA* **81** (1984) 5116–5120.
- [9] Soumpasis D. M., Wiechen J., and Jovin T. M., *J. Biomol. Struct. Dyn.* **4** (1987) 535–552.
- [10] Soumpasis D. M., Robert-Nicoud M., and Jovin T. M., *FEBS Lett.* **213** (1987) 341–344.
- [11] Soumpasis D. M., Garcia A., Klement R., and Jovin T., “The potentials of mean force (PMF) approach for treating ionic effects on biomolecular structures in solution”, *Theoretical Biochemistry and Molecular Biophysics*, Beveridge D. L. and Lavery R., Eds. (Adenine Press, Schenectady, NY, 1990), pp. 343–360.
- [12] Soumpasis D. M., “Formal aspects of the potentials of mean force approach”, *Computation of Biomolecular Structures*, Soumpasis D. M. and Jovin T. M., Eds. (Springer, Berlin, 1993), pp. 223–239.
- [13] Klement R., Soumpasis D. M., von Kitzing E., and Jovin T. M., *Biopol.* **29** (1990) 1089–1103.
- [14] García A. E. and Soumpasis D. M., *Proc. Natl. Acad. Sci. USA* **86** (1989) 3160–3164.
- [15] Hummer G. and Soumpasis D. M., *J. Chem. Phys.* **98** (1993) 581–591.
- [16] Hummer G., Ph.D. thesis (University of Vienna, Austria, 1992).
- [17] Soumpasis D. M., *J. Biomol. Struct. Dyn.* **6** (1988) 563–574.
- [18] Garcia A. E., Gupta G., Soumpasis D. M., and Tung C. S., *J. Biomol. Struct. Dyn.* **8** (1990) 173–186.
- [19] Texter J., *Prog. Biophys. Mol. Biol.* **33** (1978) 83–97.
- [20] Westhof E. and Beveridge D. L., *Water Science Reviews*, Franks F., Ed. (Cambridge Univ. Press, Cambridge, UK, 1990), Vol. 5, pp. 24–136.

- [21] Berman H. M., *Current Opinion Struct. Biol.* **4** (1994) 345–350.
- [22] Drew H. R. and Dickerson R. E., *J. Mol. Biol.* **151** (1981) 535–556.
- [23] Kopka M. L., Fratini A. V., Drew H. R., and Dickerson R. E., *J. Mol. Biol.* **163** (1983) 129–146.
- [24] Marky L. A. and Kupke D. W., *Biochem.* **28** (1989) 9982–9988.
- [25] Klement R., Soumpasis D. M., and Jovin T. M., *Proc. Natl. Acad. Sci. USA* **88** (1991) 4631–4635.
- [26] Klement R., “Computation of ionic distributions around charged biomolecular structures using the PMF approach”, *Computation of Biomolecular Structures*, Soumpasis D. M. and Jovin T. M., Eds. (Springer, Berlin, 1993), pp. 207–222.
- [27] Hummer G. and Soumpasis D. M., *Phys. Rev. E* **49** (1994) 591–596.
- [28] Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., and Hermans J., “Interaction models for water in relation to protein hydration”, *Intermolecular Forces: Proceedings of the 14th Jerusalem Symposium on Quantum Chemistry and Biochemistry*, Pullman B., Ed. (Reidel, Dordrecht, Holland, 1981), pp. 331–342.
- [29] Ewald P., *Ann. Phys.* **64** (1921) 253–287.
- [30] Hummer G., Soumpasis D. M., and Neumann M., *J. Phys.: Condens. Matt.* **23A** (1994) A141–A144.
- [31] Hummer G. and Soumpasis D. M., “A new approach to calculate the hydration of DNA molecules”, *Structural Biology: The State of the Art; Proceedings of the Eighth Conversations in the Discipline Biomolecular Stereodynamics*, Sarma R. H. and Sarma M. H., Eds. (Adenine Press, Schenectady, NY, 1994), Vol. 2, pp. 273–278.
- [32] García A. E., Hummer G., and Soumpasis D. M., *Biophys. J.* **66** (1994) A130.
- [33] Forester T. R. and McDonald I. R., *Mol. Phys.* **72** (1991) 643–660.
- [34] Soumpasis D. M., Procacci P., and Corongiu G., IBM DSD report, October 1991, IBM.
- [35] Niesar U., Corongiu G., Clementi E., Kneller G. R., and Bhattacharya D. K., *J. Phys. Chem.* **94** (1990) 7949–7956.
- [36] Soumpasis D. M., Tung C.-S., and García A. E., *J. Biomol. Struct. Dyn.* **8** (1991) 867–888.
- [37] Westhof E., *J. Biomol. Struct. Dyn.* **5** (1987) 581–600.
- [38] Yoon C., Privé G. G., Goodsell D. S., and Dickerson R. E., *Proc. Natl. Acad. USA* **85** (1988) 6332–6336.
- [39] DiGabriele A. D., Sanderson M. R., and Steitz T. A., *Proc. Natl. Acad. Sci. USA* **86** (1989) 1816–1820.

- [40] Quintana J. R., Grzeskowiak K., Yanagi K., and Dickerson R. E., *J. Mol. Biol.* **225** (1992) 379–395.
- [41] Chuprina V. P., *Nucleic Acids Res.* **15** (1987) 293–311.
- [42] Arnott S. and Hukins D. W. L., *Biochem. Biophys. Res. Commun.* **47** (1972) 1504–1509.
- [43] Leroy J.-L., Charretier E., Kochoyan M., and Guéron M., *Biochem.* **27** (1988) 8894–8898.
- [44] Burkhoff A. M. and Tullius T. D., *Nature* **331** (1988) 455–457.
- [45] Hosur R. V., Govil G., Hosur M. V., and Viswamitra M. A., *J. Mol. Struct.* **72** (1981) 261–267.