

LA-UR-

96-0240

Title:

Theoretical Studies of the Interaction of Water
with DNA Oligomers and Proteins

RECEIVED

APR 01 1996

OSTI

Author(s):

Angel E. Garcia, T-10
Gerhard Hummer, T-10

Submitted to:

DOE Office of Scientific and Technical
Information (OSTI)

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.

Form No. 836 R5
ST 2629 10/91

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DLG
MASTER

DISCLAIMER

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

Theoretical Studies of the Interaction of Water with DNA Oligomers and Proteins

Angel E. Garcia* and Gerhard Hummer

Abstract

This is the final report of a three-year, Laboratory-Directed Research and Development (LDRD) project at the Los Alamos National Laboratory (LANL). The effects of hydration on the structure and dynamics of DNA oligomers and small proteins have been studied. The hydration of small biomolecules was studied by means of molecular dynamics and Monte Carlo simulations that explicitly include the water-electrolyte solvent. From these simulations and from simulations of pure water and simple solutes in water, a potential of mean force (PMF) was constructed. This PMF approach was used to describe equilibrium hydration patterns around any biomolecule without the need to do extensive simulations. The effect of water on the dynamics of biological macromolecules is understood by combined theoretical and experimental studies. Methodologies that involve the use of molecular dynamics simulations and theoretical approaches to interpret multi-dimensional nuclear magnetic resonance (NMR) and x-ray diffraction data have been developed.

1. Background and Research Objectives

The structure, dynamics and stability of biological macromolecules are influenced by the strong interaction of the biomolecule with the solvent. These interactions are mediated locally by hydrogen bonding, the finite size of the water molecules, and by long-range electrostatic interactions. Structural and bulk water molecules play important roles in determining the structure-function relationships of biomolecules. The solvent bulk properties, like viscosity, determine the ability of a biomolecule to undergo a transition from an equilibrium state to an active metastable state. Also, changes in conformation of biomolecules are limited by the ability of water molecules bound around mobile sites to change conformations and either adopt a new structural pattern or be released and form part of the bulk solvent. In order to understand the function of biomolecules we have to understand each of these fundamental interactions.

* Principal investigator, e-mail: axg@lanl.gov

Given the size of biomolecular systems and the enormous computer time and storage requirements for simulations, it is necessary to extract a set of general principles describing the equilibrium properties of water molecules around biomolecules. This can be accomplished by constructing a potential of mean force (PMF) that describes the equilibrium-averaged interactions between the atoms of the biomolecule and water molecules. The PMF approach has been used successfully to describe the interactions of the charged phosphate groups in DNA, both among themselves and with the excess ions of the solution. The PMF approach describes all known structural transitions of DNA oligomers that are driven by monovalent salts. More complicated extensions of this approach have been used to determine the local ionic concentration around DNA oligomers and to describe the interactions of DNA with divalent cations. The formalism for the generalization of the PMF approach to describe hydration is simple. An exact description of the conditional probability of finding a water molecule at any specific site around a molecule is given by an $N+1$ point correlation function, where N refers to the number of sites in the molecule and $N+1$ is the label of a water molecule. However, the calculation of such a function is not practical, and an approximate description of this function must be obtained. For example, the zero-order approximation describes the probability of finding a water molecule at a specific site as a function of the bulk water density. This zero-order approximation does not contain any structural information. A first-order approximation assumes that any multiple-point-correlation function is given by the Kirkwood Superposition Approximation (KSA). The first-order approximation is accurate when all distances $|x_{ij}|$ are larger than the effective diameter of the interacting particles. However, this approximation fails to give an accurate description of the angular arrangements of more than two particles at close contact. There are other possible approximations that involve the expansion of the multiple-point correlation function in terms of three-point (e.g., Fisher-Kopeliovich approximation) and four-point correlation functions. The algorithms to calculate any three- and four-point correlation functions have been developed recently and implemented for molecular systems.

The main objective of this work was to understand the dynamic and equilibrium properties of water molecules around biomolecular systems and to study the role of water in the dynamics and stability of biomolecules. The structural and dynamical properties of water molecules around DNA oligomers and small proteins were studied by means of molecular dynamics and Monte Carlo simulations. The simulated systems were those for which nuclear magnetic resonance (NMR) spectroscopic and/or diffraction data were available. These simulations were used to interpret the experimental data by taking into account local motions and solvent effects and to test for the accuracy of the models.

2. Importance to LANL's Science and Technology Base and National R&D Needs

Simulations of biological macromolecules using molecular dynamics and Monte Carlo methods provide an important tool for biotechnology and basic research. Accurate and realistic simulations of proteins and other biomolecules in solution allow one to analyze the structure, dynamics, and function of a targeted biomolecule and predict such things as docking sites, interaction energies, and the effects of site-specific mutagenesis. Such simulations would have wide application in protein engineering and drug design. This work complements ongoing LANL programs in protein and nucleic acid dynamics and has applications in high performance computing, genomics, structural biology, and computational biology.

3. Scientific Approach and Results

The goal of this project was to develop fast and efficient software that can be used to perform simulations and study DNA oligomers and small proteins in solution. First, the hydration structure of solutions of simple solutes are studied and analyzed by multiple-point correlation functions. The hydration patterns of large biomolecular systems are then compared to the hydration pattern of these simple solutes. The simplest way to transfer the hydration pattern of large subgroups in molecules is by using multiple-point correlation functions among polar and non-polar groups in the biomolecule and water.

From the results obtained by the analysis and interpretation of simulations on large biomolecules and simple solutes, together with the analysis of experimental data on biomolecular systems, a potential of mean force (PMF) is constructed that gives an approximate description of the average hydration structure of the biomolecule. This PMF allows the determination of the hydration structure of any molecular surface without the need of simulations. Such a PMF is particularly useful in determining the hydration structure of large proteins, DNA or biomolecular assemblies for which simulations are not possible. The effect of water on the dynamics of biomolecules is understood by combined theoretical and experimental studies.

Methods have been developed to interpret two-dimensional nuclear-Overhauser-enhancement (2DNOE) data from NMR spectroscopy experiments. This interpretation does not make any assumptions concerning the dynamics of the molecule, other than those used in the potential energy function that describes the system in solution.

A major effort of this project was to study details of the hydration patterns of a biological macromolecule's surface groups. These hydration patterns were studied by doing

extensive molecular dynamics simulations of biomolecules in solution. First, the hydration of charged, polar and non-polar groups on the biomolecule's surface were studied. An approximate, but accurate description of the hydration patterns around atomic centers on the surface of a biomolecule were described by means of pair- and triplet-correlation functions between biomolecule atoms and water atoms. The arrangements of water molecules around specific groups (such as bases in DNA or the amino acid backbone of a protein) were studied in terms of probability distributions defined on a set of spherical coordinates centered and oriented around three atoms in the specific group.

Second, triplet- and quadruplet-correlation functions involving simple solutes found on the surface of biomolecules were calculated. These m-point correlation functions were used to approximate the local water density around a macromolecule containing surface groups similar to the simple solutes studied. Statistically significant descriptions of the hydration spine around the minor groove of DNA molecules have been obtained. These correlation functions may include one-, two- or three-donor atom(s) in the DNA minor groove, and three-, two- or one-water molecule(s) belonging to the first or second hydration shells of those groups. Both PMF and molecular dynamics simulation studies of peptides in alpha helical and beta hairpins conformations were used to test the accuracy of the PMF results. The PMF results are in excellent agreement with simulation data.

Third, hydration profiles calculated from simulations of small biomolecular systems like DNA hairpins, short DNA oligomer duplexes (8 to 12 base pairs) and small globular proteins (such as crambin, T4 lysozyme and myoglobin) were obtained. The PMF results were also tested against simulation data and crystallographic data. A simulation of a parallel-stranded DNA oligomer in solution has been completed. The system consisted of an 8-base-pair DNA oligomer, 14 sodium ions and 2484 water molecules. In total, this system has 7976 atoms. This simulation has been extended for 180 picoseconds. The configurational data was analyzed in terms of atom-water, radial-distribution functions around every atom in the DNA oligomer, and in terms of the local water density around the oligomer.

The triplet-correlation functions of simple aqueous solutions have been further refined. These functions have been used in the PMF approach to describe the hydration of large biomolecules. Simulations on proteins and DNA oligomers in solution have been conducted to compare the water density profiles obtained from our approximate theoretical expansion, from the PMF approach, and from simulation results obtained for model systems.

Another effort on this project has been the use of effective potentials to study hydrated biomolecules without explicitly including water. An extension of the PMF approach to study the effective conformational energy surface of small peptides has been completed. The free energy surface obtained by the effective potential including hydration effects via a mean

potential of force (MPF) is in excellent agreement with similar calculations that explicitly include water. The execution time of the calculations that make use of the MPF are two orders-of-magnitude faster than the calculations that explicitly include water.

We also studied hydrophobic hydration. The structure of water molecules in the vicinity of nonpolar solutes gives rise to the characteristic thermodynamics of hydrophobic hydration and plays a governing role in a host of biological processes, such as protein folding and micellar self-assembly. Current approaches, such as integral-equation or density-functional theories, are limited because of the complexity involved in treatment of molecular fluids. The inhomogeneous nature of their structure presents a further challenge for application of such theories to associating fluids.

We utilize a PMF expansion to predict the structural aspects of water organization in the vicinity of nonlinear, nonpolar solutes. We find that for a weakly interacting inhomogeneity, decomposition of n-particle PMF into pair-, triplet-, etc., contributions closed at the three-particle level fails to capture the underlying physics of hydration of complex solutes. An alternative proximity-type approximation, which was based on the physical arguments about the sensitivity of water structure to the shape and excluded volume of the solute, was devised. These approximations not only capture the underlying physics of hydrophobic hydration, but also have the potential to be applied as closure techniques for n-particle correlation functions in the realm of integral-equation theory of classical fluids.

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.
