Solvation Determination for Nucleic Acids and Proteins using HyPred

Rafael O. Soto[a], Hailiang Zhang[b], Joseph Curtis[b]

[a]*Illinois Institute of Technology, Chicago, IL*
[b]*National Institute of Standards and Technology, Gaithersberg, MD*

Abstract

The solvation layer around nucleic acids and proteins is important in maintainting the structure and dynamics of the molecule. This summer I participated in research involving the determination and reproducibility of solvent layers surrounding proteins and nucleic acids. We used a novel method of determining the solvent density termed HyPred. This method determines solvent density through the use of proximal radial distribution functions as well as simulations of immobile molecules. We explored the efficacy of this method using various nucleic acid and protein simulations.

1. Introduction & Techniques

The solvation layer surrounding nucleic acids is vital to their various processes.[6] This is due in part to the high electronegativity of the molecules. The hydration is split up into various layers that each play their own role in interactions including: stabilizing secondary and tertiary structure. Various attempts have been made to study the hydration patterns around nucleic acids.

It is also of interest when trying to do biophysical measurements as it affects: small-angle x-ray scattering (SAXS), and small angle neutron scattering (SANS) data. In addition to this having a method to accurately pre-determine the solvation layer surrounding a protein or nucleic acid would cut down on the necessary time required to equilibrate a solvated structure before molecular dynamics simulations. The motivation for studying the HyPred model is to add implicit solvent functionality to the software suite SASSIE. [2] SASSIE is a suite capable of quickly generating simulation structures from SANS, SAXS data and vice-versa. Adding in quick implicit solvent calculations would extend SASSIE's functionality.

The solvation layer presented in this paper results from the work of Virtanen et. al. termed HyPred (Hydration Predicton).[7] Using simulation data the average electron density around specific atom types in the structure is calculated with respect to a distance r, resulting in a proximal radial distribution function (pRDF). Using this new function of density with relation to atom type the process can be reversed in order to map a hydration layer around a new structure based on equilibrated data. This process would help offset the cost of running long fully solvated simulations in order to equilibrate the hydration layer.

The results in Virtanen's paper point to the existence of compatible pRDF profiles for various proteins. That is one could use a pRDF profile obtained from one protein and use it to determine the hydration layer of a completely different protein. This postulate is also examined in this paper through the simulation of 8 different proteins.

2. Methods

2.1. DNA and RNA Simulations

Nucleic acid simulations were done using NAMD and the CHARMM36 all-atom forcefield optimized for nucleic acids. The B-DNA structure was created using 3D-DART modeling server.[5, 4] The RNA structure was cleaved off from a well-equilibrated long strand of RNA. Structures were solvated using 15 Åpadding in all directions and neutralized with NaCl ions. The structures went through a two part minimization lasting 2ps, as well as a 2ps simulated annealing from 1277 K to 300 K. The structures were then equilibrated in the NPT ensemble for 1ns, followed by an 1ns NVT equilibration. Production runs were done in the NVT ensemble at 300 K lasting 10ns with the molecule fixed in place.Trajectory snapshots were taken every 1ps and were used for analysis.

2.2. Protein Simulations

8 different protein structures were obtained from well equilibrated snapshots. A 1ns NVT equilibration was performed for each structure followed by 2ns production runs in the NVT ensemble. Protein simulations utilized NAMD and the CHARMM27 force field. Trajectory snapshots were written every 1ps and were used for analysis.

2.3. Proximal Radial Distribution Functions

The pRDFs were calculated based on the methodology of Freed et al. A grid of cubes of 0.5Åsidelength was superimposed on the simulation. The density of solvent was calculated every 1-ps for cubes outside the nucleic acid up to a 10Ådistance from the structure. Each cube is assigned to an atom on the exterior surface of the nucleic acid whose scaled van der Waals surface is closest to the center of the cube. The average density throughout the simulation is constructed based on atom type, heavy atoms (C, N, O) as well as a Hydrogen sublass where hydrogen is related to the heavy atom it's bounded to (CH, NH, and OH). Freed et al. calculate the proximal distribution function for CH as :

$$g_{CH}(r) = \frac{1}{N}\sum_{i=1}^{N}\rho_i \tag{1}$$

We used the same methodology, but different nomenclature i.e. HC instead of CH, hydrogen & heavy atom pairing.

3. Results

3.1. pRDF Profiles RNA & DNA

The pRDF profiles of the nucleic acid simulations (Fig. 1 and 2) show distinct similarities with respect to certain atom types, namely: O, N, and HN. The pRDF of RNA also bears a high degree of similarity with respect to the protein pRDFs discussed in the following section. The peaks for O, HC are nearly identical as well as the shapes for C, N and there is a more general similarity for HO, HN. DNA is largely the same except that the magnitude of the main HO peak is much larger than the other simulations examined. The C profile is also not as well pronounced as in other simulations. mv

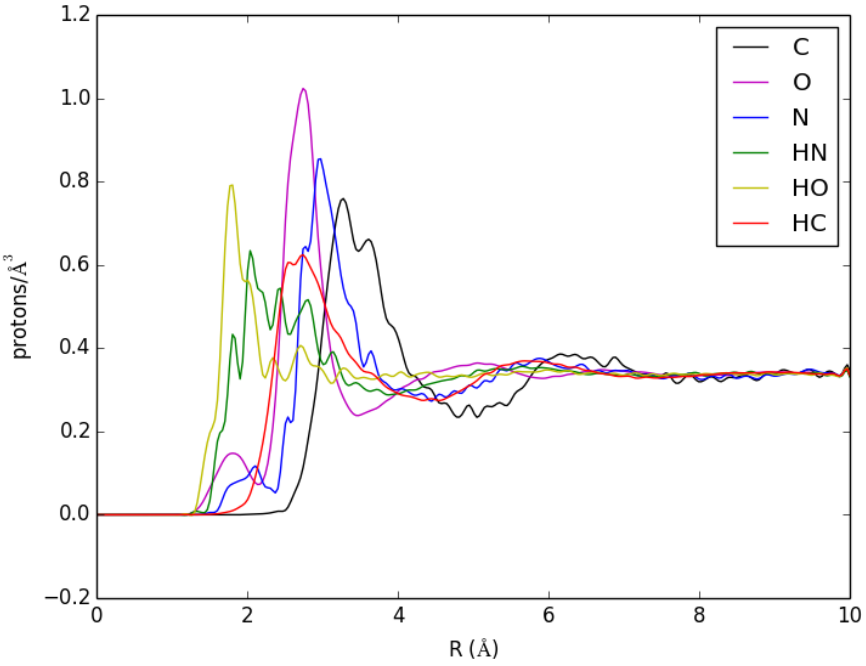


Figure 1: proximal radial distribution function for RNA

3.2. Protein pRDFs

The protein pRDFs (Fig. 3 - 6) show highly similar profiles for O with a small peak at 1.8 Åand a much larger one another at 2.7 Å, these numbers match the results from Virtanen's HyPred paper. We also observe highly similar profiles for HN, HC, and similar shaped profiles for HN, and N for a highly diverse set of protein structures. Virtanen et al. noticed this similaritly and it led them to two ideas; that the pRDF set of one protein can be used to calculate the hydration layer of a different protein, and that there may exist a set of universal pRDFs that will accomodate any protein.

3.3. Scattering Profile for DNA

4. Discussion and Next Steps

This work validates our process of collecitng pRDF data from simulations and will allow us to move forward with SASSIE implementation. The next step is to see whether in fact one protein or nucleic acid's pRDF profile can accurately predict that of another. There's considerable evidence for this being the case as shown in figures BLANK AND BLANK, these are the the pRDF profiles for every simulation from this study in one graph for O and N. It's plain that for some atom types the predictability would be highly accurate. However some pRDFs do vary considerably as is the case for C, and HO. Further study is needed in order to see how accurately these profiles predict the hydration shell. We can then potentially implement these separate pRDFs into SASSIE in order to give it a large range of structures it can work with.

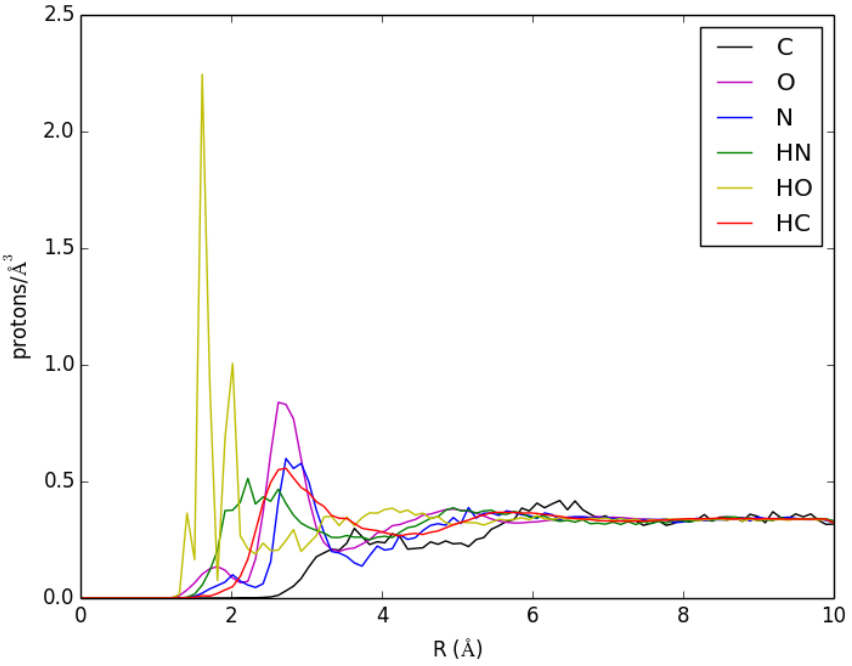


Figure 2: proximal radial distribution function for DNA

5. References

[1] Chen, J., Im, W., & Brooks, C. L. (2005). Application of torsion angle molecular dynamics for efficient sampling of protein conformations. Journal of Computational Chemistry, 26(15), 156578. doi:10.1002/jcc.20293

[2] Curtis, J. E., Raghunandan, S., Nanda, H., & Krueger, S. (2012). SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. Computer Physics Communications, 183(2), 382389. doi:10.1016/j.cpc.2011.09.010

[3] Hart, K., Foloppe, N., & Baker, C. (2011). Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. Journal of Chemical , 8(1), 348362. doi:10.1021/ct200723y.Optimization

[4] Maffeo, C., & Aksimentiev, A. (n.d.). Introduction to MD simulation of DNA protein systems.

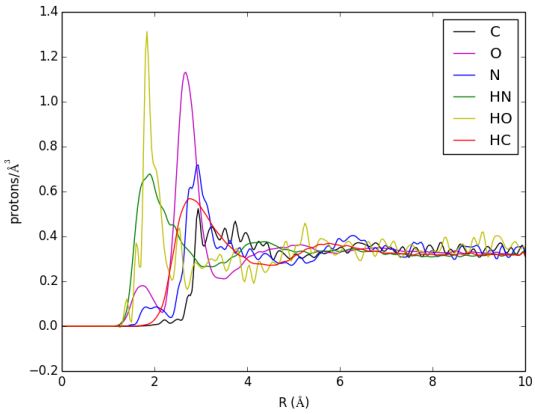
[5] M. van Dijk and A.M.J.J. Bonvin (2009) "3D-DART: a DNA structure modelling server", Nucl. Acids Res., 37 (Web Server Issue):W235-W239 doi:10.1093/nar/gkp287

[6] Schlick, T. (2003). Molecular Modeling and Simulation: An Interdisciplinary Guide (2nd ed.). New York: Springer.

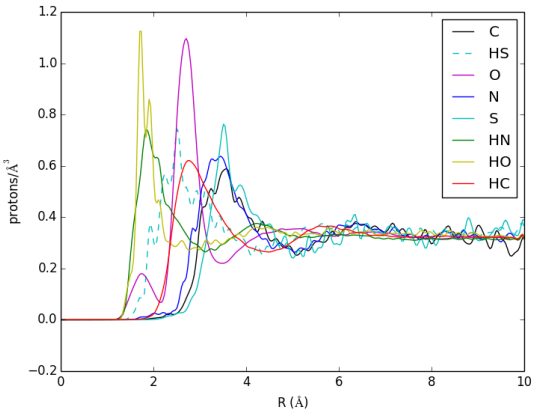
[7] Virtanen, J. J. (2012). Modeling The Hydration Layer Around Proteins and Nucleic Acids. Disseration Publishing, (March).

[8] Virtanen, J. J., Makowski, L., Sosnick, T. R., & Freed, K. F. (2010). Modeling the hydration layer around proteins: HyPred. Biophysical Journal, 99(5), 16119. doi:10.1016/j.bpj.2010.06.027

[9] Watson, M. C., & Curtis, J. E. (2013). Rapid and accurate calculation of small-angle scattering profiles using the golden ratio. Journal of Applied Crystallography, 46(4), 11711177. doi:10.1107/S002188981301666X

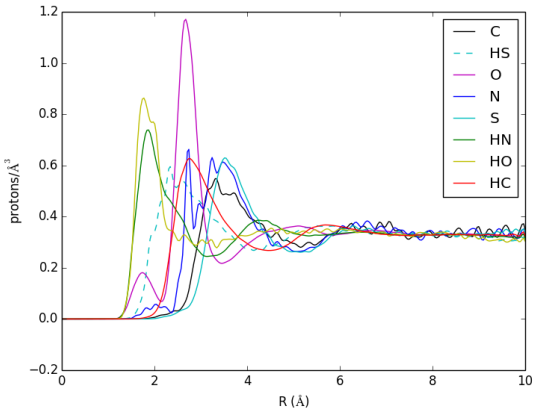


(a) pRDF for

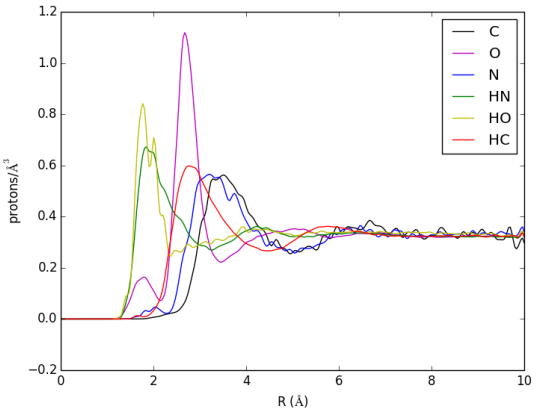


(b) pRDF for

Figure 3: pRDFs

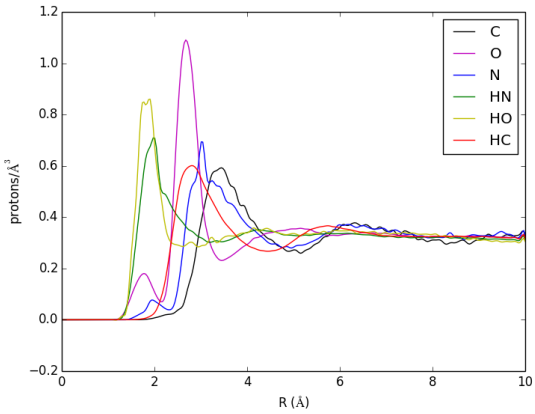


(a) pRDF for

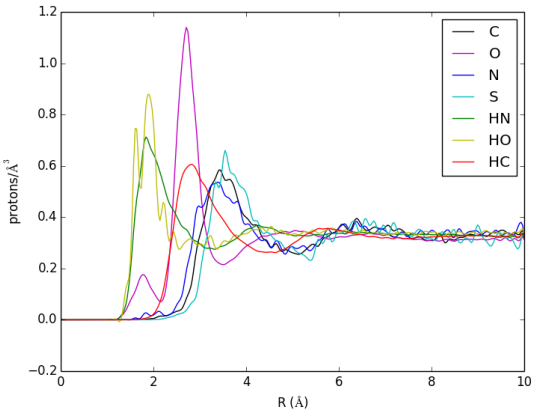


(b) pRDF for

Figure 4: pRDFs

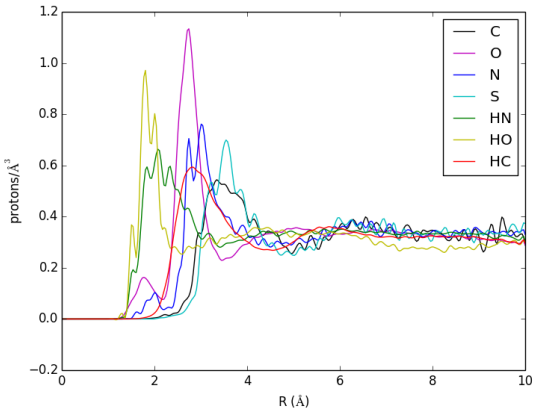


(a) pRDF for

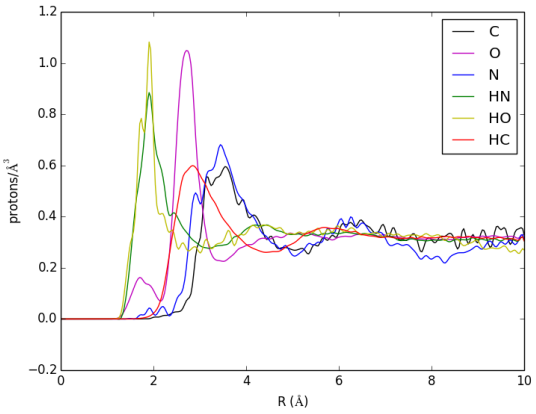


(b) pRDF for

Figure 5: pRDFs

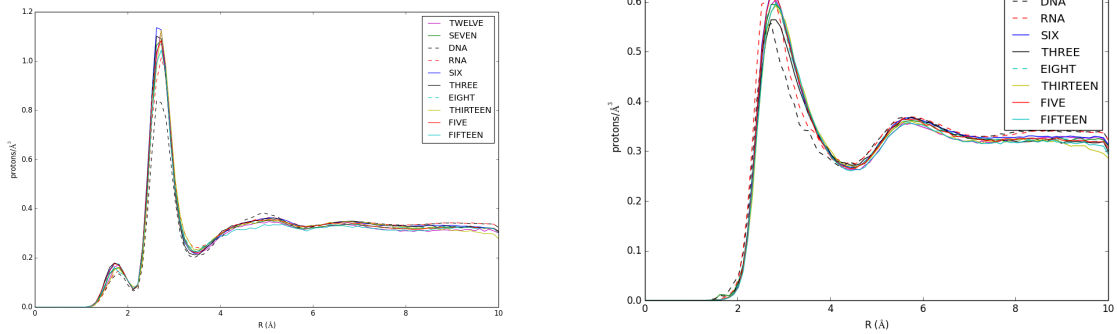


(a) pRDF for



(b) pRDF for

Figure 6: pRDFs



(a) pRDF for

(b) pRDF for

Figure 7: pRDFs calculated for oxygen(a) and hydrogen-carbon(b) from every simulation (protein and nucleic acid) in this study.