Published in final edited form as:

J Phys Chem Lett. 2011 June ; 2(13): 1626–1632. doi:10.1021/jz200609v.

Fast Calculations of Electrostatic Solvation Free Energy from Reconstructed Solvent Density using proximal Radial Distribution Functions

Bin Lin¹, Ka-Yiu Wong², Char Hu¹, Hironori Kokubo², and B. Montgomery Pettitt^{1,2}

¹ Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030 USA

² Department of Chemistry, University of Houston, Houston, TX 77204-5003 USA

Abstract

Although detailed atomic models may be applied for a full description of solvation, simpler phenomenological models are particularly useful to interpret the results for scanning many, large, complex systems where a full atomic model is too computationally expensive to use. Among the most costly are solvation free energy evaluations by simulation. Here we develop a fast way to calculate electrostatic solvation free energy while retaining much of the accuracy of explicit solvent free energy simulation. The basis of our method is to treat the solvent not as a structureless dielectric continuum, but as a structured medium by making use of universal proximal radial distribution functions. Using a deca-alanine peptide as a test case, we compare the use of our theory with free energy simulations and traditional continuum estimates of the electrostatic solvation free energy.

Keywords

Solute correlation function; protein solvation; electrostatic chemical potential

Solvation free energy plays a central role in protein folding^{1,2}, protein function³ and molecular recognition^{2,4}. It originates from the intermolecular interactions between the solute and the solvent⁵. The free energy is often considered to consist of short-ranged van der Waals or cavity interactions and long-ranged electrostatic interactions^{6,7}. The electrostatic interactions contribute the most to solvation free energy for polar molecules in water like proteins and DNAs⁸. There are two kinds of methods commonly used to evaluate electrostatic solvation free energy (ESFE)^{5,9}. One is explicit solvent free energy simulations, e.g. free energy perturbation¹⁰ (FEP), thermodynamic integration¹¹ (TI) and similar methods^{12–14}. The other involves implicit continuum solvent models such as Poisson-Boltzmann¹⁵ (PB) or the Generalized Born model¹⁶ (GB). The simulation methods are accurate but since averages and correlations come from simulations it is computationally expensive to evaluate. The latter class of implicit methods is fast to evaluate but suffers inaccuracy in cases due to its underlying approximations^{17–19}. Therefore we seek ways to calculate ESFE rapidly as well as accurately by keeping the merits from the aforementioned methods.

Here, we propose a method to rapidly evaluate ESFE of biomolecules based on solvent density reconstruction using proximal radial distribution functions (pRDF). The pRDF is the distribution calculated from solvent molecules to their nearest solute atom^{20–23}. The solvent space is partitioned into Voronoi volumes and the resulting distribution functions are defined perpendicular to the surface formed by respective solute atoms by construction²².

Therefore, proximal radial distribution functions are also called perpendicular radial distribution functions. From previous simulations of compact globular proteins, we found that the pRDFs classified by the chemical identity of solute atoms are nearly universal functions^{20,21,24}. For peptide and protein systems, the pRDFs are quite similar and this universality can be exploited.²⁴ This property can be exploited to reconstruct solvent density of proteins or different conformations of the same polypeptide from pRDFs pre-calculated from explicit solvent simulations²¹. This way, computationally demanding simulations of for instance, the denatured ensemble, may be reduced or in most cases avoided while still allowing one to study the thermodynamic and structural solvation properties. Here we demonstrate ESFE calculations using linear response theory by reconstructing solvent density for five different conformations of deca-alanine from pRDFs. The results are compared to those obtained by explicit solvent free energy simulations and PB calculations.

Deca-alanine is a polypeptide that prefers the alpha helix conformation in vacuum at room temperature for most force field models. In aqueous solution, coil, 3–10 helices and beta sheets can become more stable depending on force field due to the effects from the solvent²⁵. Experimental^{26–28} and computational evidence^{29–31} for solvent effects has been thoroughly accumulated for these conformations. Thus, this peptide serves as a highly useful model for understanding the solvent effects on the free energy landscapes of biomolecules and has relevance to protein folding³². Its size allows us to use explicit solvent free energy simulations to compute EFSE³³ and compare with the method we developed in this work. Based on the conformational manifold, five representative conformations of deca-alanine peptide are chosen from explicit solvent molecular dynamics simulations³³. They consist of an alpha helix structure (H), an extended structure (E) and three structures of intermediate random coil (I1, I2, and I3). The structures are shown in Fig 1.

PB calculations were done by solving the linearized Poisson-Boltzmann equation with APBS³⁴. A 0.1 Å was chosen as grid size to balance both speed and accuracy after ensuring convergence³⁵. The solute dielectric constant was set to either 1 or 2 and the solvent dielectric constant is 78.5 as is appropriate for water at ambient conditions. Partial charges were those of the AMBER force field³⁶. Radii were derived from the minimum in the Lennard-Jones potential, given as $2^{-5/6}\sigma$. Explicit solvent free energy simulations were performed previously³³ therefore the method is only briefly outlined here. A linear coupling parameter, λ , was used to couple the solute from its idea gas state to its fully-charged state using λ -replica exchange molecular dynamics. This method has been shown to be able to improve sampling efficiency^{37–39}. The value of λ is dynamically exchanged during the simulation in lieu of a temperature exchange as in a replica exchange simulation⁴⁰. It has been employed previously in this lab to study solvent structure and thermodynamics of amino acid side chain analogs as well as the same five conformations of deca-alanine and has shown advantages over conventional methods^{37,39}. A cubic box of side length 40 Å was considered, the temperature was set at 300 K, and periodic boundary conditions were applied. The total simulation times to obtain the pRDFs were 2 ns for each conformation with a time step of 1 fs. TIP3P water model is used⁴¹. Solute conformations were fixed throughout the simulations.

We first examine the solvent structure of five conformations of deca-alanine by calculating their respective pRDFs. The water distributions are shown in Fig 2 and 3. We find that both solute-water oxygen and solute-water hydrogen pRDFs with respect to oxygen, nitrogen, and carbon atoms on solutes appear nearly identical across all five conformations. They have almost the same shape in terms of peak positions and heights. This is consistent with the confirmed hypothesis that pRDFs are universal functions for proteins or different conformations of similar polypeptides²⁴. The shapes of the peptide pRDFs are nearly identical to those possessed by proteins²¹ and reminiscent of similar atom types in DNA⁴².

We next use average pRDFs to reconstruct solvent densities for all five conformations and compare them to those from the MD simulations on a 3D grid^{20,21}. The grid size is 0.1 Å, consistent with that used in PB calculations. By summing up the water oxygen density and hydrogen density scaled by their respective atomic charges, we obtain a 3D charge density that is related to the ESFE. The spatial distribution comparisons of charge density for the helix and extended conformations are presented in Fig 4. The other conformations have similar characteristics. The charge density computed from MD trajectories are on the left and the charge density reconstructed from pRDFs are on the right. The orientations are the same. Charge densities generally tend to be more concentrated on the oxygen and nitrogen atoms on the solutes while less concentrated on those of carbon and hydrogen atoms. This is a well-known feature of hydrophobic hydration^{43–45}. These densities correspond reasonably well. There are small sporadic spots of charge density on the MD map that arise from statistical noise while the charge density from pRDF reconstruction is more continuous with the noise averaged out in the evaluation of pRDFs. The local charge distribution in the solvation shell due to the existence of the solute is the origin of ESFE and will next be evaluated from the reconstructed solvent density⁴⁶.

The ESFE can be readily calculated from the charge densities obtained from the pRDF reconstruction. This is done according to the microscopic formula originally proposed by Levy et al. as an approximation to free energy perturbation⁴⁷ based on linear response theory⁴⁸:

$$\Delta G_{\text{elec}} = \frac{1}{2} \sum_i q_i \left\langle \sum_j \frac{q_j}{4\pi\epsilon_0 r_{ij}} \right\rangle$$

where q_i is the atomic charge at the solute site i , q_j is the solvent charge at the grid point j , r_{ij} is the distance between solute site i and grid point j . ϵ_0 is the electric constant. $\langle \rangle$ denotes ensemble average. This formula provides a microscopic definition of ESFE when used with the spatially varying charge density^{47,49}. Its advantages and limitations were investigated by Åqvist et al. by comparing results from this method to those from thermodynamic perturbation simulations⁵⁰. It was found that the linear response approximation generally holds well for monovalent ionic solutes but is less accurate for dipolar ones⁵⁰. Recently, we revisited this subject by calculating ESFE of 15 amino acid side chain analogs from linear response approximation and explicit solvent free energy simulations³⁹. We found that despite the fact that the linear response approximation consistently overestimates the absolute ESFE when compared to results from explicit solvent free energy simulations, the deviations tend to be systematic and within a reasonable range^{39,50}. The errors largely cancel when relative solvation free energies are desired.

In addition to the traditional way of grouping solute atoms into oxygen, nitrogen, and carbon atoms (pRDF3), we also tried a different grouping method by adding polar hydrogen connected to nitrogen atoms as a separate group. pRDFs are calculated for these four kinds of atoms and ESFEs are calculated for all five conformations. They are denoted as pRDF4. ESFEs for all five conformations of deca-alanine from the linear response approximation as well as those from FEP/TI simulations and PB calculations are listed in Table 1.

We found that ESFEs determined from pRDF reconstruction do not change much from pRDF3 to pRDF4 for conformation E, I2, and I3 but they improve greatly for conformation I1 and H. Visual analysis shows that conformation I1 and H are more twisted than the other conformations therefore nitrogen atoms on the backbones are less accessible to solvent molecules. This is not the case for the polar hydrogen atoms. The improvement in ESFEs indicates that contributions from polar hydrogen atoms for conformation I1 and H are

significant and therefore cannot be neglected. It also places a cautionary note on future applications: even though grouping solute atoms into oxygen, nitrogen, and carbon atoms has been sufficient for solvation structure analysis of most solutes, and for ESFE calculations of some deca-alanine conformations, it is necessary to include more relevant solute atom types when their relative contributions exposures are apparent. This can be accomplished by *a priori* proximity analysis given the percentage each solute atom contributes to the total ESFE.

We use simulated free energies as our control for the pRDF calculations of ESFE for the five conformations of deca-alanine. Based on the same force field parameters as FEP/TI simulations, PB is able to reproduce the ESFE qualitatively but not quantitatively. This is not surprising and is consistent with the literature about continuum solvation^{39,50}. PB models the solute solvent interface with a computed boundary and treats solvents as a structureless medium. On the contrary, pRDFs, derived from structured solvent distributions, capture a more realistic picture of the solvation structure around solutes. This is demonstrated by the charge distribution around deca-alanine as shown in Fig 5. Not only is the overall solvent charge distribution shown, but also those individual contributions from solute oxygen atoms, nitrogen atoms, and carbon atoms, respectively. The first major positively charged shell is primarily due to water hydrogen atoms strongly attracted by solute oxygen atoms. The second such shell is formed with a major contribution of positive charge by solute nitrogen atoms and a minor contribution of negative charge by solute carbon atoms. The net result is still dominantly positively charged. The only negatively charged shell is comprised of contributions from both solute oxygen and nitrogen atoms, with the former contributing a lesser part and the latter contributing a greater part. Overall, the negatively charged solvent shell dominates the charge imbalance over the positively charged shell in terms of local maximum intensity. The sum of integrations for both is zero as required by electroneutrality.

With the more realistic picture of the solvation as represented by the pRDFs, we find that the pRDF reconstruction ESFE approach performs better than PB. In all five conformations, pRDF reconstruction results in an ESFE in much better agreement with simulation electrostatic free energy than that of PB theory. For conformation E, I2, and I3, the relative errors for pRDF reconstruction compared to FEP results are only 2.3%, 0.1%, and 0.5%, respectively, as opposed to 26.0%, 37.6%, and 25.3%, respectively, for PB theory. The improvement is ascribed to the fact that pRDF reconstruction procedure takes into account the structure of the solvation shell as it is in explicit solvent simulations while PB theory treats the solvation shell as a simple boundary between solute and solvent therefore missing important contributions to solvation free energies from the structured medium. In this context the remaining error in I1 may be due to the choice of a somewhat peculiar conformation. This conformer was also problematic for PB as well as other methods tried and not shown here.

The grouping strategy of solute atoms in the pRDF reconstruction also allows us to study individual contributions to the ESFE from different solute atoms, as depicted in Fig 5. The breakdown for pRDF3 is listed in Table 2. For all five conformations, the major contribution (about 70%) comes from solute oxygen atoms. The next major contribution (about 25%) comes from solute nitrogen atoms. This is not surprising because oxygen atoms have a higher negative charge than nitrogen atoms. Contributions from solute oxygen atoms vary for different conformations, with conformation I2 and I3 being the largest and conformation H and I1 being the least. This is due to the fact that the more extended conformations of I2 and I3 allow more Vorinói volume for oxygen atoms increasing its contribution while the more twisted conformations of H and I1 allow the least. The contribution for conformation E is in between these two groups of similar conformation. Contributions from nitrogen atoms

are more conformation independent than those from oxygen atoms due to exposure as well. They are all almost about -20 kcal/mol, with conformation H being the least, -16.2 kcal/mol and conformation I2 being the highest, -23.5 kcal/mol. The contributions from solute carbon atoms are nearly the same. The positively charged carbonyl carbon atoms make their contributions to ESFE positive, which means less soluble and more hydrophobic. They represent only a small portion of the total ESFE.

Aside from the accuracy shown above, the speedup of ESFE calculation by pRDF reconstruction procedure compared to other methods is enormous. Roughly 10's of thousands of CPU hours are needed for the explicit solvent free energy simulation (data based on the massively parallel supercomputer Bigben on Pittsburgh Supercomputing Center, a Cray XT3 MPP system). Several minutes are needed for Poisson-Boltzmann calculation on a 2 GHz Intel Core 2 Duo, and only seconds are needed for pRDF reconstruction and free energy evaluation on the same CPU. The characteristic of fast evaluation of electrostatic solvation free energy by pRDF reconstruction has significant implications on drug design/discovery where the balance between accuracy and speed is critical to the success of screening large databases of small-molecule drug candidates⁵¹.

In summary, we have presented a novel method to evaluate the ESFE of different conformations of deca-alanine rapidly. It is based on the observation that pRDFs are universal functions for proteins and for polypeptides. The solvent density reconstructed from the pRDF includes the approximate structure of the solvation shell so that the method contains more information than the continuum model. We have demonstrated that ESFE of deca-alanine can be obtained fairly accurately from a pRDF reconstruction procedure when compared to explicit solvent free energy simulations or PB calculations, but is more rapid to evaluate.

Acknowledgments

This work was supported in part by the National Institutes of Health (GM 037657), and the Robert A. Welch Foundation (E-1028). This research was performed in part using Molecular Science Computing Facility at PNNL and in part by National Science Foundation through TeraGrid resources provided by Pittsburgh Supercomputing Center and Texas Advanced Computing Center. Structure representations are prepared with VMD⁵².

References

1. Hu CY, Kokubo H, Lynch GC, Bolen DW, Pettitt BM. *Protein Sci.* 2010; 19:1011–1022. [PubMed: 20306490]
2. Levy Y, Onuchic JN. *Annu Rev Biophys Biomol Struct.* 2006; 35:389–415. [PubMed: 16689642]
3. Krone MG, Hua L, Soto P, Zhou RH, Berne BJ, Shea JE. *J Am Chem Soc.* 2008; 130:11066–11072. [PubMed: 18661994]
4. Aleksandrov A, Schuldt L, Hinrichs W, Simonson T. *Biophys J.* 2009; 97:2829–2838. [PubMed: 19917238]
5. McQuarrie, DA. *Statistical Mechanics.* University Science Books; Sausalito, Calif: 2000.
6. Deng Y, Roux B. *J Phys Chem B.* 2004; 108:16567–16576.
7. Shirts MR, Pitera JW, Swope WC, Pande VS. *J Chem Phys.* 2003; 119:5740–5761.
8. Honig B, Nicholls A. *Science.* 1995; 268:1144–1149. [PubMed: 7761829]
9. Chipot, C.; Pohorille, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology.* Springer; New York: 2007.
10. Zwanzig RW. *J Chem Phys.* 1954; 22:1420–1426.
11. Straatsma TP, Berendsen HJC, Postma JPM. *J Chem Phys.* 1986; 85:6720–6727.
12. Bennett CH. *J Comput Phys.* 1976; 22:245–268.
13. Shirts MR, Pande VS. *J Chem Phys.* 2005; 122

14. Widom B. *J Chem Phys.* 1963; 39:2808–2812.
15. Born M. *Z Phys.* 1920; 1:45.
16. Still WC, Tempczyk A, Hawley RC, Hendrickson T. *J Am Chem Soc.* 1990; 112:6127–6129.
17. Hunenberger PH, McCammon JA. *J Chem Phys.* 1999; 110:1856–1872.
18. Baker NA. *Curr Opin Struct Biol.* 2005; 15:137–143. [PubMed: 15837170]
19. Feig M, Brooks CL. *Curr Opin Struct Biol.* 2004; 14:217–224.
20. Lounnas V, Pettitt BM, Phillips GN. *Biophys J.* 1994; 66:601–614. [PubMed: 8011893]
21. Makarov VA, Andrews BK, Pettitt BM. *Biopolymers.* 1998; 45:469–478. [PubMed: 9577228]
22. Mehrotra PK, Beveridge DL. *J Am Chem Soc.* 1980; 102:4287–4294.
23. Virtanen JJ, Makowski L, Sosnick TR, Freed KF. *Biophys J.* 2010; 99:1611–1619. [PubMed: 20816074]
24. Lin B, Pettitt BM. *J Chem Phys.* 2011:134.
25. Levy Y, Jortner J, Becker OM. *Proceedings of the National Academy of Sciences of the United States of America.* 2001; 98:2188–2193. [PubMed: 11226214]
26. Ingwall RT, Scheraga HA, Lotan N, Berger A, Katchals E. *Biopolymers.* 1968; 6:331–368. [PubMed: 5641934]
27. Platzer KEB, Scheraga HA, Andreatt Rh, Ananthan Vs. *Macromolecules.* 1972; 5:177–187.
28. Blondelle SE, Forood B, Houghten RA, PerezPaya E. *Biochemistry.* 1997; 36:8393–8400. [PubMed: 9204887]
29. Soman KV, Karimi A, Case DA. *Biopolymers.* 1991; 31:1351–1361. [PubMed: 1816873]
30. Daggett V, Levitt M. *J Mol Biol.* 1992; 223:1121–1138. [PubMed: 1538392]
31. Doruker P, Bahar I. *Biophys J.* 1997; 72:2445–2456. [PubMed: 9168021]
32. Chakraborty A, Kortemme T, Baldwin RL. *Protein Sci.* 1994; 3:843–852. [PubMed: 8061613]
33. Kokubo H, Hu CY, Pettitt BM. *J Am Chem Soc.* 2011; 133:1849–1858.
34. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. *Proc Natl Acad Sci U S A.* 2001; 98:10037–10041. [PubMed: 11517324]
35. Mohan V, Davis ME, Mccammon JA, Pettitt BM. *J Phys Chem.* 1992; 96:6428–6431.
36. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. *J Am Chem Soc.* 1995; 117:5179–5197.
37. Hu CY, Lynch GC, Kokubo H, Pettitt BM. *Proteins.* 2010; 78:695–704. [PubMed: 19790265]
38. Jiang W, Hodoscek M, Roux B. *Journal of Chemical Theory and Computation.* 2009; 5:2583–2588.
39. Lin B, Pettitt BM. *J Comput Chem.* 2011; 10.1002/jcc.21668
40. Mitsutake A, Sugita Y, Okamoto Y. *Biopolymers.* 2001; 60:96–123. [PubMed: 11455545]
41. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J Chem Phys.* 1983; 79:926–935.
42. Feig M, Pettitt BM. *J Mol Biol.* 1999; 286:1075–1095. [PubMed: 10047483]
43. Pratt LR, Chandler D. *J Chem Phys.* 1977; 67:3683–3704.
44. Choudhury N, Pettitt BM. *J Am Chem Soc.* 2005; 127:3556–3567. [PubMed: 15755177]
45. Choudhury N, Pettitt BM. *J Am Chem Soc.* 2007; 129:4847–4852. [PubMed: 17385863]
46. Roux B, Yu HA, Karplus M. *J Phys Chem.* 1990; 94:4683–4688.
47. Levy RM, Belhadj M, Kitchen DB. *J Chem Phys.* 1991; 95:3627–3633.
48. Hansen, JP.; McDonald, IR. *Theory of Simple Liquids.* 3. Academic Press; Amsterdam, the Netherlands: 2006.
49. Smith PE, Pettitt BM. *J Phys Chem.* 1994; 98:9700–9711.
50. Åqvist J, Hansson T. *J Phys Chem.* 1996; 100:9512–9521.
51. Shoichet BK. *Nature.* 2004; 432:862–865. [PubMed: 15602552]
52. Humphrey W, Dalke A, Schulten K. *J Mol Graph.* 1996; 14:33–38. [PubMed: 8744570]

FIG 1.
Five representative conformations of deca-alanine. Peptide backbone is highlighted as a red tube; Alpha helix is highlighted as a red ribbon.

FIG 2. Water oxygen pRDFs for five different conformations of deca-alanine with respect to (a) oxygen atoms, (b) nitrogen, and (c) carbon atoms on deca-alanine.

FIG 3. Water hydrogen pRDFs for five different conformations of deca-alanine with respect to (a) oxygen atoms, (b) nitrogen, and (c) carbon atoms on deca-alanine.

FIG 4. Charge density around the helix and extended conformation of deca-alanine at the same level of charge density. (a) and (c): charge density calculated from MD simulation; (b) and (d): charge density computed from pRDF reconstruction procedure. Similar characteristics are observed for other conformations.

FIG 5. Radial charge distributions around deca-alanine. The red line represents the contribution by solute oxygen atoms, the blue line represents the contribution by solute nitrogen atoms, the cyan line represents the contribution by solute carbon atoms. The black line is the total contributions from all three components.

Table 1

Electrostatic solvation free energies (ESFE) of deca-alanine. FEP stands for Free Energy Perturbation, TI for Thermodynamic Integration, PB for Poisson Boltzmann with solute dielectrics of 2 or 1, pRDF3 for the pRDF reconstruction method with solute atoms grouped by oxygen, nitrogen, and carbon atoms. pRDF4 is pRDF3 with an additional solute group of polar amide hydrogen atoms. The units are kcal/mol.

Method\Structure	E	I1	I2	I3	H
FEP	-69.6±0.3	-65.9±0.3	-88.2±0.4	-77.8±0.3	-49.7±0.2
TI	-69.8±0.4	-66.3±0.3	-88.5±0.5	-78.2±0.3	-49.7±0.2
pRDF4	-67.6	-58.7	-87.8	-80.4	-50.9
pRDF3	-68.3	-54.2	-88.7	-79.2	-55.1
PB($\epsilon=2$)	-51.7	-42.6	-55.0	-58.1	-33.6
PB($\epsilon=1$)	-72.16	-59.85	-81.9	-82.08	-44.0

Table 2

Individual contribution to total ESFE by solute oxygen, nitrogen, and carbon atoms, respectively. The unit is kcal/mol.

Contribution\Structure	E	I1	I2	I3	H
O	-50.8	-37.9	-68.6	-59.8	-42.1
N	-20.9	-20.2	-23.5	-22.9	-16.2
C	3.3	3.9	3.4	3.5	3.2
Total	-68.4	-54.2	-88.7	-79.2	-55.1