John von Neumann Institute for Computing

The Influence of Molecular Surface Composition on the Outcome of Poisson Boltzmann Calculations Performed to Obtain Solvation Free Energies

Parimal Kar, Yanjie Wei, Ulrich H.E. Hansmann, Siegfried Höfinger

published in

NIC Workshop 2006,
From Computational Biophysics to Systems Biology,
Jan Meinke, Olav Zimmermann,
Sandipan Mohanty, Ulrich H.E. Hansmann (Editors)
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **34**, ISBN-10: 3-9810843-0-6,
ISBN-13: 978-3-9810843-0-6, pp. 161-164 , 2006.

http://www.fz-juelich.de/nic-series/volume34

The Influence of Molecular Surface Composition on the Outcome of Poisson Boltzmann Calculations Performed to Obtain Solvation Free Energies

Parimal Kar[1], Yanjie Wei[1], Ulrich H.E. Hansmann[1,2], and Siegfried Höfinger[1]

[1] Michigan Technological University,
Department of Physics, 1400 Townsend Drive, Houghton, MI 49931, USA
E-mail: {pkar, yawei, hansmann, shoefing}@mtu.edu

[2] John von Neumann Institute for Computing, FZ Jülich
52425 Jülich, Germany
E-mail: u.hansmann@fz-juelich.de

Poisson Boltzmann is a widely used method to describe solvation effects in structural biology. In the present study we want to focus on the Boundary Element Method within the various set of different Poisson Boltzmann Techniques. Here a fundamental dependence arises from the fact that initially the molecular surface must be computed and discretized in terms of small-sized simple geometric elements, i.e. triangles, tesserae, curved triangles etc. The current study investigates the influence of this critical aspect of molecular surface discretization on the final results of PB/BEM-solvation free energy calculations. Different algorithms are compared to each other as well as to high-level reference states obtained from Quantum Chemistry calculations. Surface resolutions are systematically increased until a sufficient level of convergence is reached. The trends are analyzed with respect to different molecular surface algorithms and varying size and shape of proteins. General conclusions are drawn concerning key-requirements for the molecular surface composition in order to achieve a maximum level of technical robustness.

1 Introduction

Many important processes in biochemistry involve proteins in solution. Theoretical approaches aimed at understanding the structural and mechanical basis of these processes will therefore have to also consider the effect coming from the environment, i.e. from water, membranes, ionic solutions etc. One way of describing solvation effects in the context of biomolecular simulation is based on the Poisson Boltzmann treatment (PB) of macromolecules[1,2]. Especially in the latter description — the *Boundary Element Method* (BEM) — a particular sensitivity arises with respect to the underlying molecular surface computation, which is used for defining the boundary between molecular volume and the environment. While in analogous models used in Quantum Chemistry[3] the average area of these boundary elements has been advised to be on the order of 0.4 $Å^2$, similar rules of thumb seem to be missing for classical PB/BEM approaches of protein structures. We therefore want to investigate in this present study how molecular surface decomposition will affect the net result of PB/BEM. Particular emphasis will be placed on the question if with increasing resolution into boundary elements stable converged PB/BEM results are achievable and how commonly employed molecular surface programs compare to each other in this regard.

2 Methods

Molecular surfaces are computed with two independant programs, the Connolly program[4] and the SIMS program[5]. PB/BEM energies are calculated with program POLCH[6]. Public data bases[7] are used for the download of protein structures in PDB format. A diverse set of different proteins is selected reaching from small-sized peptides of 44 residues to up to large-sized globular proteins of about 500 residues. PDB structures are visualized with program MOLDEN, after having removed HETATM lines, CONNECT lines, ANISOU lines and the footer section, and having deleted all but the 'A' chain in multiple chain pdb files so that column number five is exclusively left with 'A' labels. MOLDEN-visualization of the cleaned PDB files involves selecting force field **Tinker Amber** and writing out a PDB file from within MOLDEN with the option 'Write_With_Hydrogens'. Since MOLDEN always uses the default HIP type in AMBER jargon, we need to convert the HIS residues to HIP ones, as well as to change CYS residues engaged in disulfide bonds to CYX-type residues. After that we make sure that the initial residue is not PRO, and delete initial PROs in case. AMBER non-bonded parameters[8], i.e. charges and van der Waals radii are assigned to all the atoms in the protein structures. The vdW radii are increased by a factor of 1.12 and the charges are scaled down by another factor of 0.9 which was found previously to lead to better agreement with high level Quantum Chemistry reference calculations. Before we execute Connolly's MSROLL program, inner/outer dielectric constants at the molecular boundary must be defined. In the case of water we use the standard values of 1.0 and 80.0. We execute the MSROLL program with different values for the f argument, where f defines the fineness, in other words, the resolution of the surface. When f is decreased, the resolution of the surface becomes better but computational cost will increase. For some very small values of f the Connolly program fails. In such cases we change parameter l which defines a minimal size of triangles to be considered distinct. The p parameter of MSROLL, that is the value of the probe sphere radius remains constant 1.5 (in Angstrom) in all the calculations. We record the analytically calculated surface area and volume and the employed f value and transfer the result file tmp.vet into a human readable format, called COMB_CONN.dt. Then we clean the COMB_CONN.dt file from critical entries, which are triangles that almost coincide and the like. In the next step, we compile the Poisson-Boltzmann program and record the number of boundary elements (triangles) and the number of atoms we have employed. Then we actually run the PB calculation, recording the PB energy and the number of iterations needed for the PB/BEM calculation. Now we refine the -f parameter and repeat the whole procedure. This is continued until convergence of the PB/BEM result is reached. The smaller the -f value becomes, the more BE will be produced, hence the more complicated and long-lasting the PB calculation will be.

Using the SIMS program is similar except that the molecular surface and the solvent excluded volume are computed numerically and that the probe radius, the smooth radius and the dot-density are defined within a small argument file to the program. Analogous to modifying the -f parameter in MSROLL is changing the dot-density in SIMS. Higher values of the dot-density will provoke higher surface resolutions. The remainder of the procedure described above is identical when using SIMS. We record dot-density, number of BE, PB energy, number of iterations, SASA and volume for comparison.

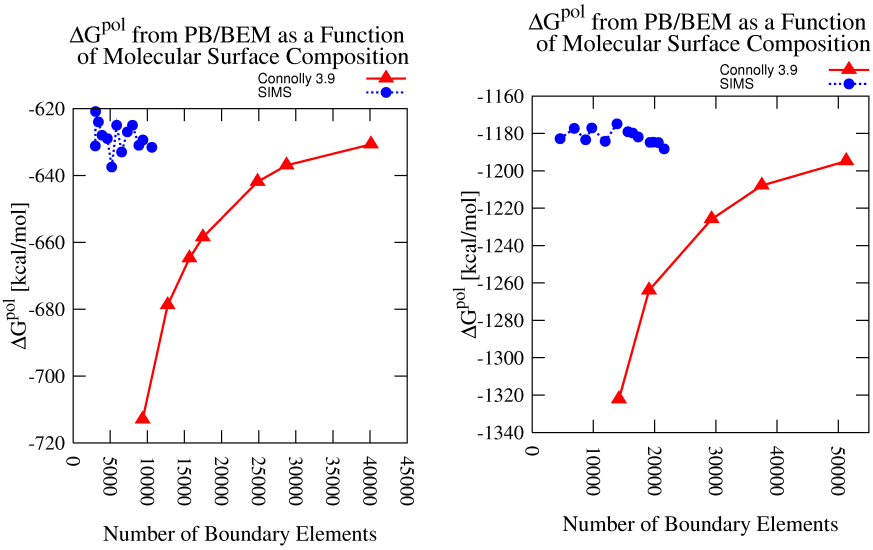

Figure 1. Variation of PB/BEM results as a function of molecular surface composition. Left panel: pdb code 1OTF, 62 residues; Right panel: pdb code 1WN2, 121 residues; Connolly's MSROLL program (red triangles) as well as the SIMS program (blue spheres) lead to the same PB/BEM result upon convergence. The SIMS program, however, does so with smaller numbers of BEs to consider.

3 Results and Conclusions

A diverse ensemble of different pdb structures is examined for sensitivity of PB/BEM calculations with respect to molecular surface composition. As an example two typical cases are shown in Figure 1 (pdb codes 1OTF and 1WN2 with 62 and 121 residues respectively). The general trend is similar in either case. Connolly's MSROLL program (red triangles) reaches a plateau value in a continuous mode, while the SIMS program (blue spheres) approaches converged PB/BEM results in an alternating fashion when the surface resolution is steadily increased. Both programs lead to the same net PB/BEM result, hence the description should be of equal quality. Considerable dependence on this critical aspect is demonstrated clearly ($\pm$ 100 kcal/mol uncertainty when working off the convergence domain). The SIMS program appears to reach convergence much faster than the Connolly program, which would have important implications on computational overall performance. The required average size of boundary elements for achieving sufficient levels of numerical accuracy lies in the range of 0.4 $Å^2$ or below, hence is comparable to recommended values from Quantum Chemistry.

Acknowledgments

This work was supported in part by the National Institutes of Health Grant GM62838.

References

1. J. Warwicker, H. C. Watson, *Calculation of the electric potential in the active site cleft due to α-helix dipoles*, J. Mol. Biol. **157**, 671–679 (1982).
2. R. J. Zauhar, R. S. Morgan, *A new method for computing the macromolecular electric potential*, J. Mol. Biol. **186**, 815–820 (1985).
3. *The PCM model*, European Summerschool in Quantum Chemistry, Book II, Björn O. Roos and Per-Olof Widmark, Eds **561–597**, 2000 (.)
4. M. L. Connolly, *Computation of molecular volume*, J. Am. Chem. Soc. **107**, 1118- (1985).
5. Y. N. Vorobjev, J. Hermans, *SIMS: computation of a smooth invariant molecular surface*, Biophys. J. **73**, 722-732 (1997).
6. S. Höfinger, *Solving the Poisson Boltzmann Equation with the Specialized Computer Chip MD-GRAPE-2*, J. Comp. Chem. **26**, 1148-1154 (2005).
7. *http://mbs.cbrc.jp/pdbreprdb-cgi/reprdb_menu.pl* all default criterions except **Number of residues less than 40 excluded – NO**, **Include MUTANT – NO**, **Exclude COMPLEX**, **Exclude FRAGMENT**, **Include NMR – NO**, **Include Membrane Proteins – NO**
8. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules*, J. Am. Chem. Soc. **117**, 5179–5197 (1995).