A FAST AND GENERAL CONTINUUM APPROACH FOR DESCRIBING ELECTROSTATIC EFFECTS IN MOLECULAR DYNAMICS SIMULATIONS OF BIOMOLECULES

Sergio A. Hassan, Daqun Zhang, Ernest L. Mehler and Harel Weinstein, Department of Physiology and Biophysics,
Mount Sinai School of Medicine, New York, New York 10029

The Screened Coulomb Potential Implicit Solvent Model (SCP-ISM) was recently proposed as the basis for the rigorous derivation of a continuum treatment of electrostatic effects in solvated macromolecules. The SCP-ISM avoids the fundamental difficulty of requiring a boundary between the solvated macromolecule and the solvent. The SCP-ISM was originally developed for Monte Carlo simulations, and is extended here to carry out Molecular Dynamics (MD) simulations. In the initial algorithm the effective Born radii, required for calculating the self-energy terms, was based on the degree of exposure of the atoms to the solvent calculated from the solvent accessible surface areas of atoms. To reduce CPU requirements and simplify the calculation, an alternative approach is proposed that is based on a solvent contact model. This approach allows the electrostatic energy to be completely expressed as a pair-wise function, without compromising the quality of the results. This new description makes possible the rapid and easy calculation of the energy and its derivatives allowing MD simulations of macromolecules in biologically realistic time frames. The complete SCP-ISM and the adaptation for MD simulations will be presented. Preliminary results of long simulations of a 17-amino acid peptide Dynorphin and a 56-amino acid Protein G will illustrate the utility of this approach in comparison to simulations with explicit water.

Introduction

By providing the dynamic evolution of a system, Molecular Dynamics (MD) simulations allow the evaluation of time dependent as well as thermodynamic properties and provide a natural link to experiment. The solvent around a macromolecule modulates its dynamics and also its stabilization. Because of the highly demanding computational requirements needed to simulate a macromolecule in an explicit representation of the solvent, implicit solvent models (ISM) provide an attractive alternative that should reduce substantially CPU time and allow more realistic trajectories to be calculated. The SCP-ISM is a general continuum approach that was developed with the aim of being of general applicability [1-3]. **The performance of the model has been assessed in a number of tests carried out on different size scales (single amino acids, peptides and proteins) and in a number of comparison with explicit water calculations, CD, X-ray and NMR experiments and also with other computational approaches as, e.g., Poisson-Boltzmann calculations [1,4,5]. The model has also being used to rationalize pharmacological and biological data [6,7].**

Because of the promising results in all these tests, using mainly Monte Carlo simulations, the SCP-ISM is extended here for use with MD simulations. A preliminary validation is reported by comparing the results from simulations of a peptide and a protein using the SCP-ISM and explicit solvent.

Continuum Electrostatics of a Macromolecule in a Polar Solvent

Because of the derivation from the microscopic to the macroscopic realm, the SCP-ISM is described in terms of effective screening functions and does not require either a solute/solvent boundary or a definition of the so-called internal dielectric constant. The resulting description consists of the macromolecule embedded in a dielectric that permeates all of space and is completely characterized by a dielectric function $\varepsilon(r)$.

The total electrostatic energy is given by:

$$E_T = \frac{1}{2}\sum_{i\neq j}^{N} \frac{q_i q_j}{D(r_{ij})r_{ij}} + \frac{1}{2}\sum_{i=1}^{N} \frac{q_i^2}{R_{i,B}}\left[\frac{1}{D(R_{i,B})} - 1\right] = E^{\text{int}} + E^{\text{self}}$$

Where $D(r)$ is the screening function and R is the Born radius of atom i

Calculation of forces in the framework of the SCP-ISM requires the calculation of gradient of E_T in the conformational space of the macromolecule.

Molecular Dynamics Simulation of a Peptide and a Proteins using the SCP-ISM

The calculation of forces involves the calculation of the partial derivatives of E_T with respect to the interparticle distances. The approach was implemented in the CHARMM force field for use with the param22 all-atom representation. As a preliminary assessment of the performance of the SCP-ISM in MD simulations two systems were studied, and the results compared with the corresponding 3 ns MD simulations with explicit water solvent.

Dynorphin:

Figure 1: NMR structure of Dynorphin, a 17-amino acid peptide; H-Tyr-Gly-Gly-Phe-Leu-Arg-Arg-Ile-Arg-Pro-Lys-Leu-Lys-Trp-Asp-Asn-Gln-OH; it binds selectively to κ-opioid receptors. Initial structure for both SCP-ISM and EW MD simulations.

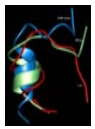

Figure 2: snapshots at three different times in the MD simulation with the SCP-ISM. Consistent with the MD simulation with EW, early in the simulation the helical portion of the peptide begins to open from an α-helical conformation. After 500 ps the helix is completely disrupted and remains open until the end of the simulation (t=3ns).

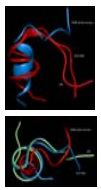

Figure 3: superposition of the initial structure (NMR, at t=0) and the final structures (t=3 ns) obtained with the SCP-ISM and EW MD simulations. Note the qualitative similarity of the peptide in the two simulations. The only difference appears in the N-terminus: Tyr1 is H-bonded to the peptide in the SCP-ISM MD, whereas the same residue is solvent exposed in the EW simulation. The quantitative similitude between the explicit and implicit simulation is also remarkable: about 2 Ang for the backbone atoms of the fragment 4-17.

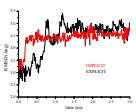

Figure 4: RMSD of backbone atoms of Dynorphin as a function of time for the SCP-ISM and EW MD simulations. The RMSD is measured with respect to the initial (t=0) NMR structure (see Fig.3). Note that conformational changes leading to the open helix occur earlier in the implicit model than in the explicit model simulation. Note also that the fluctuations are strikingly similar in both cases.

Protein G:

Figure 5: PDB structure of the immunoglobulin-binding domain of streptococcal Protein G, a 56-amino acid globular protein containing both α-helical and β-sheet motifs.

Figure 6: superposition of representative snapshots of the protein at the end of the 3 ns simulation (blue, SCP-ISM simulation; Green, EW simulation). The figure shows that all elements of secondary structure motifs (β-sheets and α-helix) are maintained throughout the simulation, as is the case in the MD simulation with explicit waters.

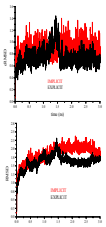

Figure 7: superposition of the RMSD of C_α atoms from the implicit and explicit simulation as a function of time. Both **average values** and **fluctuations** are **similar** in both simulations.

Figure 8: superposition of the RMSD of all the atoms in the system for the two simulations as a function of time. Although there is a split of the average RMSD between 1.5 and 3 ns (a maximum of 0.5 Angstroms is obtained at 2.3 ns), the overall trend is similar and the fluctuations are slightly larger in the implicit simulation. This discrepancy, although small, is due to the more movable side chains in the simulation with the SCP-ISM.

Computational Efficiency of the SCP-ISM:

1) **SCP-ISM requires only 3 x CPU times in vacuum** 2 times faster than GB approach
2) 3 ns simulation of protein G requires (in Alpha platform):
***) SCP-ISM: 5 days in ONE processor**
***) EW: 90 days in FOUR parallel processors**

This excellent performance obtained with the SCP-ISM is possible because the calculation of Born radius of an atom i in the macromolecule involves a pair-wise function of only few nearest neighbor atoms, as explained below.

Born Radius in a Macromolecular Environment

In the SCP-ISM, the Born radius $R_{i,B}$ is defined in terms of the solvent accessible surface area (SASA) of each particle in the form (see also Figure 1):

$$R_{i,B} = R_w \xi_i + R_p (1 - \xi_i)$$

where ξ_i is proportional to SASA (see caption to Figure 9).

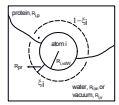

Figure 9: Schematic diagram showing the degree of exposure of an atom i to the solvent (ξ_i) and to the protein interior ($1-\xi_i$), used to define the effective Born radius of an atom in the protein environment. R_w and R_p are the Born radii of the atom in bulk water and in bulk protein interior, respectively.

The reasonableness of this approach based on SASA was already demonstrated in several applications (see Refs.[1-7]). However, since SASA is not a pair-wise quantity, its derivatives must be carried out numerically, dramatically reducing computational efficiency. To circumvent this drawback a new definition of "solvent exposure" is based on a contact solvent approach. The new expression for the Born radii is:

$$R_{i,B} = \alpha_i + \beta_i \sum_{j\neq i}^{N} \exp(-C_i r_{ij})$$

The derivative is now easily calculated analytically. Constants α, β and C are optimized by maximizing the correlation of $R_{i,B}$ between the SASA and contact model approach. This guarantees that the quality of all the results obtained so far (see Refs.[1-7]) is preserved.

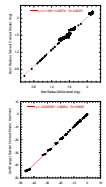

Figure 10: Upper panel: scatter plot of the Born radii of atoms in a small globular protein, calculated with the SASA approach and the new solvent contact model; Lower panel: scatter plot of the self-energies of atoms with the Born radii shown in previous panel. Note that the self energies calculated using SASA and the new approach are highly correlated. Therefore, the quality of the energetics (thoroughly tested in previous calculations) is preserved, as intended.

CONCLUSIONS: the SCP-ISM was extended for use in Molecular Dynamics simulations of proteins and peptides. The new approach for the calculation of Born radii is based on a solvent contact model that allows the energy and forces to be expressed by simple analytic forms that can be evaluated very efficiently. Thus, the SCP-ISM is only 3 times slower than vacuum calculation. The MD simulations reported here show that the average RMSD and fluctuations are well reproduced when compared to explicit water MD simulations, for both peptides and proteins. This demonstrates that the SCP-ISM is general enough to be applied to biomolecules without requiring *ad hoc* parametrizations and modification of the energy function depending on the size of the system. The SCP-ISM was already used in longer D trajectories of the order of fraction of μs, and the results will be reported elsewhere.

REFERENCES

[1] S. A. Hassan, F. Guarnieri and E. L. Mehler, J. Phys. Chem. **104**, 6478 (2000)
[2] S. A. Hassan, F. Guarnieri and E. L. Mehler, J. Phys. Chem. **104**, 6490 (2000)
[3] S. A. Hassan and E. L. Mehler, Proteins **47**, 45 (2002)
[4] S. A. Hassan and E. L. Mehler, Int. J. Quant. Chem. **83**, 193 (2001)
[5] S. A. Hassan, E. L. Mehler and H. Weinstein, *Lecture Notes in Computational Science and Engineering*, T. Schlick, Ed., Springer, New York (2002)
[6] M. Tartaglia, E. L. Mehler et al., *Nature Genetics* **29**, 465 (2001)
[7] I. Visiers, S. A. Hassan and H. Weinstein, Protein Eng. **14**, 409 (2001)