Chemical Physics Letters 468 (2009) 79–82

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

A polarizable coarse-grained water model for coarse-grained proteins simulations

Tap Ha-Duong [a,*], Nathalie Basdevant [a], Daniel Borgis [a,b,*]

[a] *Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, Université d'Evry-Val-d'Essonne, Rue du Père André Jarlan, 91025 Evry Cedex, France*
[b] *UMR CNRS-Pasteur, Département de Chimie, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France*

ARTICLE INFO

Article history:
Received 16 October 2008
In final form 24 November 2008
Available online 6 December 2008

ABSTRACT

To describe solvation effects in coarse-grained molecular dynamics simulations, a numerically efficient coarse-grained water model is introduced. The solvent is represented by polarizable pseudo-particles embedding three water molecules. The particles carry induced dipoles that are made sensitive to the solute electric field, but not to each other. The solvent model is compatible with a coarse-grained proteins force field involving a reduced number of grains per residue and yields quantitative description of solvation properties, such as hydrophobic forces and electrostatic solvation free-energies. Those later quantities can be estimated 'on-the-fly' over short simulation windows.

1. Introduction

In the field of biomolecular modelling, coarse-grained (CG) approaches are becoming increasingly popular since they can overcome the computational limits of more accurate classical all-atom (AA) force fields and they can be used to simulate macromolecule dynamics on biological relevant time-scales (about a few microseconds). Such low resolution protein models have been historically developed to tackle the folding problem [1,2]. In these models, each amino-acid was represented with a few pseudo-atoms (from one to four) for the backbone and generally one bead for the side-chain. Important efforts have been made to develop effective intramolecular potentials for the protein backbone (bonds, valences and torsions) in order to model its internal flexibility and reproduce its secondary structures propensities [3,4].

In the study of protein–protein associations, however, one can consider as a first approximation that each molecule is a rigid body. In this case, the specificity of the biomolecular complexes comes from the direct non-bonded interactions between the atoms or pseudo-atoms of the molecules, that is the Coulomb and van-der-Waals potentials, and also from the solvent screening properties. Following recent mesoscale approaches to phospholipids dynamics [5–7] or to proteins docking [8–10], we have presented recently a new coarse-grained non-bonded force field for protein–protein interactions [11], derived in a consistent bottom-up procedure from the higher resolution all-atom AMBER force field [12]. The van-der-Waals potential is extracted from classical all-atom molecular dynamics simulations of amino-acid pairs in vacuo and is clearly separated from the Coulomb interactions between coarse grains charges. The latter are determined independently by fitting the vacuum electrostatic potential created around the isolated protein by an all-atom charge distribution. Indeed, this is a standard procedure used in theoretical chemistry to go from a quantum to an atomic molecular representation [12,13]. A serious advantage of such a procedure is that the same parametrization for radii and charges can be used in various solvent models, such as a simple distance-dependent dielectric function [14], or a more accurate Poisson–Boltzmann/solvent-accessible-surface-area (PB/SA) model [15].

We introduce here a new coarse-grained water solvent model, that we call the polarisable coarse-grained solvent model (PCGS), and that is fully consistent with the CG description of the proteins themselves. As for most water CG models proposed in the literature, each grain embeddes about three water molecules. With respect to those, the originality of the model is to incorporate electrostatic screening effects with the correct dielectric constant for water in addition to a consistent hydrophobic behavior fitted on the effective interaction between hydrophobic spheres. This model is a rescaled version of the polarisable pseudo-particle (PPP) solvent model introduced in Refs. [16,17] to describe the solvation of biomolecules in an all-atom molecular dynamics simulation context. The natural idea is that each solvent grain carries an induced dipole at its center. It was justified theoretically that those induced dipoles can be made sensitive to the solutes electric field but insensitive to each other [16–19]. The model is thus computationally very efficient since electrostatic interactions are limited to solute–solvent, no self-consistency is needed for the induced dipoles, and solvent–solvent interactions are limited to short-range, Lennard-Jones-like potentials. This Letter describes the rationale of the model, its first implementation with the coarse-grained protein model

---

* Corresponding authors. Address: Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, Université d'Evry-Val-d'Essonne, Rue du Père André Jarlan, 91025 Evry Cedex, France.

E-mail addresses: thaduong@univ-evry.fr (T. Ha-Duong), daniel.borgis@ens.fr (D. Borgis).

0009-2614/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2008.11.092

in a coarse-grained molecular dynamics code, and its ability to describe properly hydrophobic and electrostatic solvation effects.

2. The coarse-grained protein model

This section briefly summarizes the main features of our coarse-grained protein model, which was carefully presented in Ref. [11]. Each amino-acid residue is modelled by one pseudo-atom for the backbone and one or two beads for the side-chain according to their size. The glycine amino-acid is represented by a single sphere of the same type as the other backbone pseudo-atoms. For simplicity, the N-terminal and C-terminal backbone coarse grains do not differ from the others backbone group (except for their charge). The position of all spherical pseudo-atoms is located at the geometric center of the heavy atoms that form the coarse grains.

As in classical all-atom force fields, we assume that the non-polar and electrostatic energy contributions can be separated and expressed as sums of pairwise van-der-Waals and Coulombic energy functions depending on the grain to grain distances:

$$V_{nb} = \sum_{i<j} V_{vdw}(r_{ij}) + \sum_{i<j} \frac{Q_i Q_i}{r_{ij}} \tag{2.1}$$

The coarse grains van-der-Waals potential was obtained by numerical integration of the non-polar mean force between two amino-acid molecules. To do so, we have first performed 20 all-atom molecular dynamics (MD) simulations for all possible homologue pairs of amino-acids. This was done in vacuo, using fully flexible molecules, and with *vanishing charge* in order to capture the purely non-electrostatic interaction. Then we have extracted from those MD trajectories the classical Lennard-Jones potential of mean force (PMF) between any two homologue groups of atoms. Secondly, in order to have a numerically tractable expression of the van-der-Waals potentials, we have fit all the computed PMF with a unique mathematical function. The function that was found to best fit all the 29 potentials, whatever the size and the softness of the coarse grains, is composed of a repulsive part in r^{-6} and a gaussian attractive part:

$$V_{vdw}(r_{ij}) = \epsilon_{ij}\left[\left(\frac{\lambda_{ij}}{r_{ij}}\right)^6 - \exp\left(-\left(\frac{r_{ij}}{\sigma_{ij}}\right)^2\right)\right] \tag{2.2}$$

The values of the energy parameters ϵ_{ii} and the two lengths λ_{ii} and σ_{ii} for the self van-der-Waals interactions are given in Ref. [11]. To get cross-interactions, and keep the number of parameters as small as possible, we applied the empirical Lorentz–Berthelot mixing rules: $\epsilon_{ij} = \sqrt{\epsilon_{ii} \times \epsilon_{jj}}$, $\lambda_{ij} = (\lambda_{ii} + \lambda_{jj})/2$ and $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$.

To describe the electrostatic interactions, we have adopted a strategy that is reminiscent of the ESPF method in quantum chemistry, or of that of Gabdoulline and Wade for proteins in a dielectric continuum [20]. It consists in deriving a set of coarse-grained charges, located at the center of every grain, that generates the potential which best fits, in a least square sense, the *vacuum* electrostatic potential created by the partial charges of the all-atom model. The least squares function is completed by quadratic penalty terms imposing that the CG effective charges respect the permanent charge and dipole of the whole protein. We did show that the overall coarse-grained potential was able to reproduce accurately protein–protein interactions in aqueous solution when compared to all-atom calculations, using various implicit solvent models such as a distance-dependent dielectric function or the Poisson–Boltzmann/solvent-accessible-surface-area method [11].

3. The coarse-grained water model

We have introduced recently a particle-based implicit solvent model, called the polarisable pseudo-particle (PPP) model, for describing the aqueous medium surrounding non-polarisable [16,17] or polarisable [21] all-atom models of proteins and nucleic acids. The solvent was represented by Lennard-Jones pseudo-particles, or grains, of roughly a water molecule size, and carrying induced dipoles at their center. In principle those induced dipoles should be coupled self-consistently through dipole–dipole interactions. It was justified theoretically that they can be made in fact sensitive to the solute electric field but insensitive to each other if the macroscopic polarisation fields is supposed to keep a longitudinal character [18]. A saturation value was introduced too to account for dielectric saturation effects. In the presence of the microscopic electric field $\mathbf{E}_0(\mathbf{r})$ created by the charges of a solute having an internal dielectric constant ϵ_i, the induced dipole of a solvent pseudo-particle located at $\mathbf{r}_k$ is defined as

$$\boldsymbol{\mu}_k = \mu_s \mathscr{L}\left[\frac{3\alpha_e}{\mu_s}E_0(\mathbf{r}_k)\right]\frac{\mathbf{E}_0(\mathbf{r}_k)}{E_0(\mathbf{r}_k)} \tag{3.1}$$

where $E_0(\mathbf{r}_k) = |\mathbf{E}_0(\mathbf{r}_k)|$, $\mathscr{L}(x) = \coth(x) - \frac{1}{x}$ is the Langevin function and μ_s defines the dipole saturation value. α_e is an effective polarizability that accounts for the local solvent orientational and electronic polarisability and also, in an effective way, for the solvent induced dipole-induced dipole interaction. It is related to the solvent density ρ and dielectric constant ϵ through the relation $\alpha_e = \frac{\epsilon_0}{\rho}\frac{\epsilon - \epsilon_i}{\epsilon}$. For a given position of the solvent pseudo-particles, the resulting total electrostatic free-energy can be written as

$$V_{elec} = \frac{1}{4\pi\epsilon_0\epsilon_i}\sum_{k<l}\frac{Q_k Q_l}{r_{kl}} - \frac{\mu_s^2}{3\alpha_e\epsilon_i}\sum_k \ln\left[\frac{\sinh((3\alpha_e/\mu_s)E_0(\mathbf{r}_k))}{(3\alpha_e/\mu_s)E_0(\mathbf{r}_k)}\right] \tag{3.2}$$

The first term represents the direct interaction between the solute charges in a uniform dielectric medium of constant ϵ_i, and the second one is the solute solvation free-energy (or solvent polarisation work) corresponding to the increase of the solvent dielectric constant from ϵ_i to ϵ. It should be noted that in our previous works in Refs. [16,17], all the preceding expressions were derived for $\epsilon_i = 1$, which corresponds to the usual convention for standard atomistic protein force fields. It was shown there that, when implemented in a molecular dynamics algorithm, the PPP solvent model does provide an efficient way to simulate proteins and nucleic acids structures and dynamics in water as well as relevant thermodynamic and structural aspects of their hydration. Solvation free energies can be estimated on-the-fly. Furthermore, despite the absence of explicit hydrogen bonds in the model, preferential hydration sites of proteins and nucleic acids can be correctly reproduced, therefore adding some useful information at a molecular level.

Compared to our original model introduced in an all-atom context, the coarse-grained version presented here is just an upscaled version with pseudo-particles of diameter 4.6 Å, rather than 2.9 Å, thus collecting roughly three water molecules instead of one. In a first version, the coarse-grained solvent van-der-Waals parameters have been adapted *mutatis-mutandis* from the 'W-model' of Klein et al. [5], developed for representing the aqueous medium around membranes, and involving no Coulombic contribution. To make it compatible with the CG protein model, we have simply fitted their 6-4 Lennard-Jones to our van-der-Waals energy function of Eq. (2.2), using the three parameters $\epsilon_{ij} = 1.96$ kcal/mol, $\lambda_{ij} = 3.97$ Å and $\sigma_{ij} = 4.85$ Å. This potential has a zero value for the distance $\sigma_s = 4.58$ Å. For reasons explained below, we have also considered an 'optimized' version with a deeper well (corresponding to the reduced temperature $T^* = k_B T/\epsilon_s = 1.3$ instead of 0.8), keeping the same van-der-Waals diameter σ_s and same solvent density (corresponding to the reduced density $\rho^* = \rho\sigma_s^3 = 0.8$). The optimized set of van-der-Waals parameters is $\epsilon_{ij} = 3.0$ kcal/mol, $\lambda_{ij} = 4.0$ Å and $\sigma_{ij} = 5.0$ Å.