JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Modeling Protein−Protein Recognition in Solution Using the Coarse-Grained Force Field SCORPION

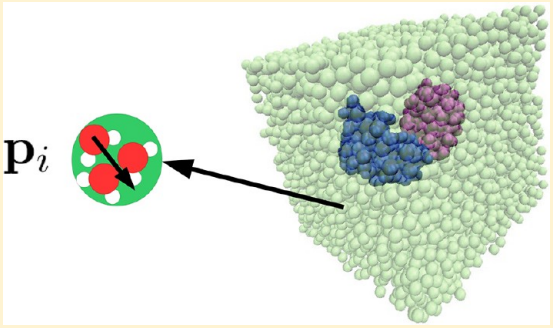
Nathalie Basdevant,*[,†] Daniel Borgis,[‡] and Tap Ha-Duong*[,†,§]

[†]Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, UMR8587 CNRS-UEVE-CEA, Université d'Evry-Val-d'Essonne, Bd François Mitterrand, 91025 Evry Cedex, France

[‡]Ecole Normale Supérieure, Département de Chimie, UMR 8640 CNRS-ENS-UPMC, 24 rue Lhomond, 75005 Paris, France

Supporting Information

ABSTRACT: We present here the SCORPION—Solvated COaRse-grained Protein interactION—force field, a physics-based simplified coarse-grained (CG) force field. It combines our previous CG protein model and a novel particle-based water model which makes it suitable for Molecular Dynamics (MD) simulations of protein association processes. The protein model in SCORPION represents each amino acid with one to three beads, for which electrostatic and van der Waals effective interactions are fitted separately to reproduce those of the all-atom AMBER force field. The protein internal flexibility is accounted for by an elastic network model (ENM). We now include in SCORPION a new Polarizable Coarse-Grained Solvent (PCGS) model, which is computationally efficient, consistent with the protein CG representation, and yields accurate electrostatic free energies of proteins. SCORPION is used here for the first time to perform hundreds-of-nanoseconds-long MD simulations of protein/protein recognition in water, here the case of the barnase/barstar complex. These MD simulations showed that, for five of a total of seven simulations starting from several initial conformations, and after a time going from 1 to 500 ns, the proteins bind in a conformation very close to the native bound structure and remain stable in this conformation for the rest of the simulation. An energetic analysis of these MD show that this recognition is driven both by van der Waals and electrostatic interactions between proteins. SCORPION appears therefore as a useful tool to study protein−protein recognition in a solvated environment.



1. INTRODUCTION

Protein recognition is the key event to many cellular processes, such as signalization, transcription, or gene regulations. Whereas experiments can yield a lot of structural and kinetic information on protein recognition, molecular modeling methods are very useful, not only to predict associations of proteins but also to understand the physical factors driving the recognition. In the field of molecular modeling, although all-atom (AA) models have proven to be very efficient and accurate, simplified or coarse-grained (CG) models are becoming increasingly popular, not only because they can overcome the computational limit of AA models, but also because they smooth the potential energy surface of the system, facilitating the sampling from one conformation to another. CG protein models were originally developed for protein folding studies,[1,2] generally representing amino acids by a few pseudoatoms for the backbone and one for the side chain, and with an important effort made on intramolecular potentials for flexibility.[3,4] Later, CG models were developed with the aim of phospholipid simulations.[5−10] More recently, CG models were designed to study proteins within membranes[11−13] and explored dynamics of protein interactions in a membrane environment. In the field of protein−protein association, considering proteins as rigid-body can be used as a first approximation, as direct intermolecular interactions between the two partners usually govern the recognition. To represent these nonbonded interactions, a CG model has to describe accurately both electrostatic and van der Waals contributions, as well as the solvation effects. In order to tackle this issue, different strategies have emerged: the first strategy is to do a statistical analysis of all available protein structures in the Protein Data Bank (PDB)[14] and extract an effective pairwise potential energy using a Boltzmann inversion.[15,16] The drawbacks of these knowledge-based potentials is that they depend on the variety of available structures in the PDB and that they make it difficult to discern the different contributions to the recognition process. Another strategy is to attempt to separate the different contributions using physicochemical principles, usually using a Lennard-Jones-type[12,17] or a Morse-type potential[18] for van der Waals interactions, whereas electrostatic interactions are generally taken into account with natural point charges[12,17] or dipoles,[2] associated with some effective hydrophobic interactions,[2] or a distance-dependent dielectric (DDD) function to mimic the solvent screening effect.[17,18]

Received: September 18, 2012
Published: December 10, 2012

The important role of solvent in protein recognition is well-known, and its representation is therefore a key to the accuracy of CG models. Regarding protein−protein docking predictions, the ATTRACT CG force field was successfully used to predict protein−protein complexes for the CAPRI competition[19] using an implicit DDD solvent model which allows efficient and fast calculations. As for CG molecular dynamics simulations of protein complexes, the MARTINI force field for proteins was combined with a very simple representation of the solvent using a screening function associated[13] or not[20] with CG Lennard-Jones neutral molecules. The role of water in protein−protein recognition is known to be more than electrostatic screening, especially for dynamic processes, and more elaborate CG solvent models than implicit models tend to emerge. A review on CG water models was recently published,[21] comparing the different properties and limitations of the available CG water models. The W model[5] was developed for phospholipid CG simulations and consists of a neutral CG molecule representing three water molecules, with a 6−4 Lennard-Jones (LJ) type potential consistent with the lipid LJ potential. A multiscale representation of liquid systems, such as water and methanol, was derived from atomistic simulations, using a force-matching procedure, yielding different CG water models that either reproduced electrostatics or structural properties of water.[22] The original solvent model in the MARTINI force field was a 12−6 LJ particle representing four water molecules, combined with a DDD screening function $\varepsilon(r) = 15r$ for direct electrostatic interactions.[9] Recently, a CG polarizable solvent model was developed to be used with the MARTINI CG force field, adding on each solvent molecule, which still represents four water molecules, two sites carrying opposite partial charges that interact by pure Coulomb interactions and rigidly bound to the center.[23] As the bending angle between the three sites is flexible, the CG bead is therefore polarizable, and it is combined with an empirical DDD function $\varepsilon(r) = 2.5r$ for direct electrostatic interactions. Also based on the MARTINI CG water model, the BMW—Big Multipole Water—solvent model is a modification of the original neutral four-to-one CG water bead by changing the van der Waals potential to a softer non-LJ-like one, and by adding three charged sites into the bead, so that it reproduces several electrostatic and thermodynamic properties of water,[24] such as bulk water permittivity and surface tension. Another CG polarizable water model was recently developed[25] representing five water molecules in one CG bead with two opposite charges linked by an adjustable bond embedded in a classical 12−6 LJ sphere. It was parametrized to reproduce various thermodynamic properties of liquid water, such as self-diffusion coefficient, viscosity, and surface tension. This water CG model was designed to be used in combination with an AA model for proteins,[26] in a hybrid manner analogous to QM/MM simulations. However, a recent review on the properties of CG water models[27] showed that a CG bead for water should not contain more than three water molecules in order to keep correct liquid state properties. On the other hand, the coarse-grained model WT4[28] represents roughly 11 water molecules by four tetrahedrally connected beads carrying partial charges. It was shown to satisfactorily reproduce both several liquid water properties (diffusion coefficient, surface tension, etc.) at a temperature range close to 300 K and electrostatic properties around ions, and it was combined to a CG model of DNA. Later, it was mixed with an all-atom SPC representation of water.[29]

We present here the SCORPION force field, which combines our previously described CG protein model[30] and a consistent CG solvent model (a short preliminary version was presented in ref 31). It will be applied to several protein−protein recognition issues. The outline of the paper is as follows. In section 2, we briefly summarize the protein model, the parameters of which were derived from a consistent bottom-up approach that clearly separates electrostatic from van der Waals interactions. The resulting van der Waals potential is softer than a Lennard-Jones potential, and CG point charges are optimized *in vacuo* so that the protein model can be combined with any solvent representation. This SCORPION model for proteins was successfully used for the redocking of several protein complexes using a simple DDD model for the solvent,[30] and a few other examples will be described here. In section 3, the SCORPION solvent model is presented. It is an extension of our previous Polarizable Pseudo-Particles (PPP) solvent model, developed for all-atom biomolecular simulations,[32,33] that is made consistent with the protein CG model. This Polarizable Coarse-Grained Solvent model (PCGS), which consists in van der Waals particles carrying an induced dipole, is aimed at reproducing both the hydrophobic and electrostatic properties of water. In section 4, we will present an application of SCORPION to molecular dynamics simulations of the barnase/barstar complex recognition in the PCGS solvent.

2. THE SCORPION PROTEIN MODEL

2.1. SCORPION Parameters for Proteins. We briefly recall here the SCORPION Coarse-Grained (CG) protein model previously described in ref 30. The coarse-graining procedure maps each amino acid residue into one particle for the backbone and one or two for the side chain according to their size, yielding 29 CG types. The position of all spherical pseudoatoms is located at the geometric center of the heavy atoms that form the coarse grains.

As in classical all-atom force fields, the nonpolar and electrostatic energy contributions were separated and expressed as sums of pairwise van der Waals (vdW) and Coulombic energy functions depending on the grain to grain distances r_{ij}:

$$V_{\mathrm{nb}} = \sum_{i<j} V_{\mathrm{vdw}}(r_{ij}) + \frac{1}{4\pi\varepsilon_0} \sum_{i<j} \frac{Q_i Q_i}{\varepsilon r_{ij}} \tag{1}$$

where ε is the medium dielectric constant. The coarse grains van der Waals potentials were obtained by numerical integration of the nonpolar mean force between all-atom amino acid molecules. Twenty all-atom molecular dynamics (MD) simulations were performed for all homologue pairs of amino acids. This was done *in vacuo*, using fully flexible molecules, and with *vanishing charge* in order to capture the purely nonelectrostatic interaction. The classical potential of mean force (PMF) between any two homologue groups of atoms, that we defined as grains, was extracted from those MD trajectories, as a function of the geometric center distance. All the computed PMFs were fitted with a unique mathematical function, with three adjustable parameters. The function that best fits all 29 potentials is composed of a very smooth repulsive part in $1/r^6$ and a Gaussian attractive part:

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

dx.doi.org/10.1021/ct300943w | J. Chem. Theory Comput. 2013, 9, 803−813

The values of the energy parameters ε_{ii} and the two lengths λ_{ii} and σ_{ii} for the self-van der Waals interactions of each CG type are given in ref 30. To get cross-interactions, the parameters between two pseudoatoms of different types were determined using empirical Lorentz−Berthelot mixing rules: $\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$, $\lambda_{ij} = (\lambda_{ii} + \lambda_{jj})/2$, and $\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2$.

As for the electrostatic potential, the CG point charges, which are located at the center of every coarse grain, are optimized for each protein to generate a potential which best fits, in a least-squares sense, the *vacuum* electrostatic potential created by the partial charges of the all-atom model, on a tridimensional grid outside the protein. The least-squares function was completed by two quadratic penalty terms imposing that the CG effective charges respect the total charge and the permanent dipole of the protein. Without any other constraint, this optimization procedure yields CG charges that do not necessarily respect the charge per residue. Some examples of optimized CG charges of a protein can be found in ref 30. Since these charges were optimized *in vacuo*, they can be combined with any implicit or explicit solvent model.

The SCORPION CG force field for proteins was implemented in the molecular dynamics software ORAC.[34]

2.2. Application to the Redocking of Protein Complexes. Using the same protocol as the one described in ref 30, and very similar to the one used by Zacharias,[17] we first tested the SCORPION protein model to predict the structure of protein−protein complexes, knowing the bound structures of the two partners, therefore performing "redocking." From the experimental structure of each complex, thousands of initial structures were generated with different relative positions and orientations of the two proteins with respect to each other. The relative position is characterized by the two latitude and longitude angles, and the different orientations are described by the three Euler angles. Each initial conformation was created so that the intermolecular distance between the two closest pseudoatoms is 4 Å. All angles were sampled regularly by 30°, yielding more than 46 000 initial structures. The proteins were maintained rigid.

The electrostatic solvation effects were taken into account implicitly using a sigmoidal distance-dependent dielectric function,[35] which screens the Coulombic interactions between the CG point charges, going continuously from 1 for short distances to $\varepsilon = 80$ at long distances. This approximate solvation model is very fast and thus adequate for such rather exhaustive docking procedures. All the generated CG complexes were energy-minimized using a conjugate gradient algorithm. In order to evaluate our scoring function to discriminate between the correct and incorrect redocked complexes, we plotted the minimized energy of each CG complex versus the root-mean-square deviation (RMSD) from the experimental structure. The RMSD was calculated over all the pseudoatoms of one of the two partners after fitting the other one on the experimental structure. Figure 1 represents these plots for three different protein complexes: the barnase/barstar complex (PDB code: 1BRS[36]), which will be studied below with MD simulations, and two new complexes with respect to our previous work,[30] the cohesin/dockerin complex (PDB code: 1OHZ)[37] and a complex formed by a trihelical domain associated with another α helix (PDB code: 1T6O).[38]

On all these protein−protein complexes, the lowest energy CG structures are the ones with the lowest RMSD from experimental structures, and with always a significant energy gap with respect to the first false complex structures. There are several complexes in the cluster of the lowest energy structures, eight for 1BRS, nine for 1OHZ, and 19 for 1T6O, respectively, with an energy gap with respect to the first false complex structures of 30 kJ/mol for 1BRS, 40 kJ/mol for 1OHZ, and 35 kJ/mol for 1T6O, respectively.

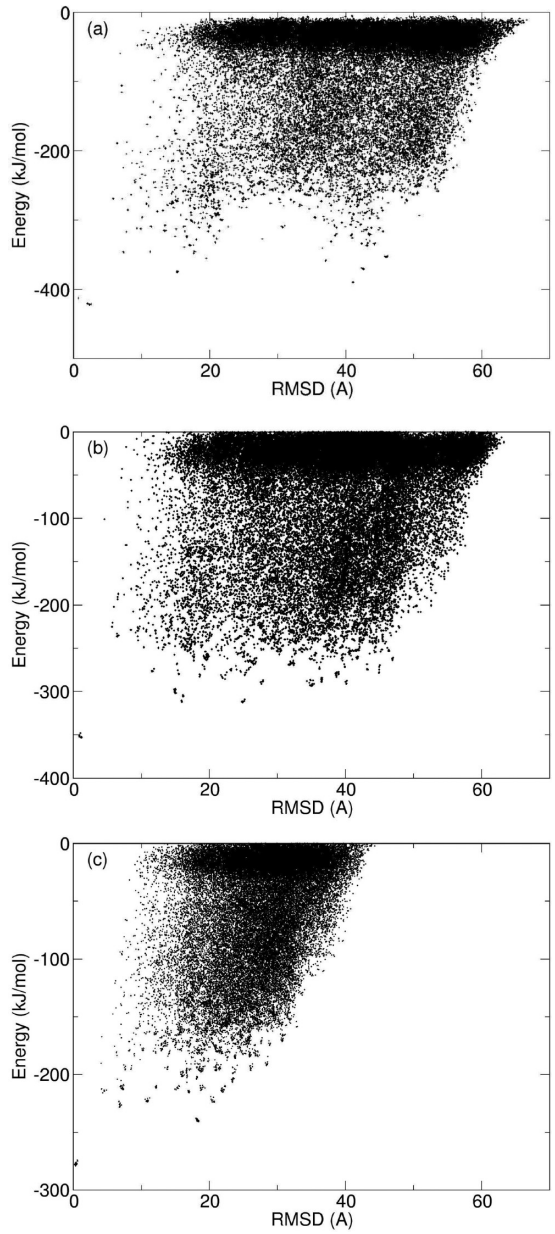


Figure 1. Interaction energies of all the minimized CG protein−protein complexes versus the RMSD from the crystal of one of the molecules when the other one is fitted on the experimental structure. (a) barnase/barstar complex (1BRS), (b) cohesin/dockerin (1OHZ), (c) α helix complex (1T6O).

These encouraging results show that the SCORPION force field for proteins is accurate enough to discriminate native complex structures from the others, although still with an approximate solvent model. It can therefore already be used efficiently for protein−protein docking calculations when there is not too much deformation upon binding. It has been successfully recently used for protein−peptide structure predictions in conjunction with NMR experiments.[39]

On the other hand, such a model is lacking the notion of solvent density and polarization fluctuations, which allow

dx.doi.org/10.1021/ct300943w | J. Chem. Theory Comput. 2013, 9, 803−813

escape from metastable configurations. It lacks also the notion of hydrophobicity, which can be viewed as collective solvent phenomena[40,41] that cannot be modeled through pairwise solute−solute interactions. Although able to yield a satisfactory energy score for random docking, the model as such is not adequate to study the dynamics of recognition, including the possible exploration of many different contacts before the correct one and the random walk process between different metastable minima before reaching the optimal free-energy state. To substantiate this claim, we have carried out hundreds-of-nanoseconds-long MD simulations, including the coupling to a thermostat to mimic thermal fluctuations, for the prototypical barnase−barstar complex studied above, and with the same CG force field and implicit solvent model. We find that, starting from a number of independent unbound structures, the complex always quickly binds into a configuration close to the initial one and sticks there for the remaining simulations; see below for more details. Making the system escape this first contact and explore further the rough energy landscape in order to detect the best binding sites would require the forcing of phase-space exploration using an adaptive bias method.[42−44] For the problem of blind docking, these enhanced molecular dynamics methods are however very time-consuming and expensive in computer resources. We believe that this is a limitation that is intrinsic to many docking studies using dynamics approaches with implicit solvent presented so far. Brownian dynamics simulations of protein−protein association in continuous solvent are usually stopped at first encounter since they are subsequently faced with the problem just described.[45] Docking methodologies on the other hand are not able to describe any dynamical process occurring after immediate binding and are only focused on single-point energy minimizations.[17]

To study the dynamics of the recognition process with realistic solute−solvent interactions, we turn therefore to a particle-based CG solvent representation that is consistent with the CG protein representation and that, as will be seen, provides enough density and polarization fluctuations to make the direct, unbiased exploration of the whole recognition event possible.

3. THE SCORPION SOLVENT MODEL: THE POLARIZABLE COARSE-GRAINED SOLVENT (PCGS)

3.1. Presentation of the Model. In the same spirit as our protein model, we introduce a solvent model which separates the electrostatic interaction from the van der Waals one. This approach differs from previous CG water models which either neglect[5,9] or do not separate the electrostatic contribution.[22] It differs also from recent models that reintroduce electrostatics through point charged distribution:[23,24] our philosophy is rather to introduce the simplest particle-based model that reproduces the macroscopic dielectric properties of water, in addition to a correct representation of hydrophobic interactions. It relies on the observation that macroscopic electrostatics prevails in water at a CG resolution of 3−4 Å.[46]

Our SCORPION solvent model is a CG version of the Polarizable Pseudo-Particle (PPP) model, introduced in refs 32 and 33 to describe the aqueous medium surrounding all-atom models of proteins and nucleic acids. This upscaled model represents the solvent by pseudoparticles or grains of matter containing roughly three water molecules, as in the W model of Klein and colleagues.[5] The solvent particles carry an induced dipole at their center. As in the original development of the PPP model, whereas the particle induced dipoles should be coupled self-consistently through dipole−dipole interactions, it can be justified theoretically that those induced dipoles can be made sensitive to the solute electric field, but insensitive to each other, with the definition of an effective particle polarizability.[32,46] A dipole saturation value is introduced to account for dielectric saturation effects. More specifically, in the presence of a microscopic electric field $\mathbf{E}_0(\mathbf{r})$ created by a solute with an internal dielectric constant ε_i, the induced dipole of a solvent pseudoparticle located at $\mathbf{r}_k$ is calculated as

$$\mathbf{p}_k = \mu_s \mathcal{L}\left(\frac{3\alpha_e}{\mu_s} E_0(\mathbf{r}_k)\right)\frac{\mathbf{E}_0(\mathbf{r}_k)}{E_0(\mathbf{r}_k)} \tag{3}$$

where $E_0(\mathbf{r}_k) = |\mathbf{E}_0(\mathbf{r}_k)|$. $\mathcal{L}(x) = \coth(x) - 1/x$ is the Langevin function, and μ_s defines the dipole saturation value. The effective polarizability α_e is related to the solvent density ρ and dielectric constant ε through the relation $\alpha_e = (\varepsilon_0/\rho)(1/\varepsilon_i - 1/\varepsilon)$. For a given position of the solvent pseudoparticles, the resulting total electrostatic free energy can be written as

$$V_{\text{elec}} = \frac{1}{4\pi\varepsilon_0\varepsilon_i}\sum_{k<l}\frac{q_k q_l}{r_{kl}} - \frac{\mu_s^2}{3\alpha_e} \times \sum_k \ln\left(\frac{\sinh((3\alpha_e/\mu_s)E_0(\mathbf{r}_k))}{(3\alpha_e/\mu_s)E_0(\mathbf{r}_k)}\right) \tag{4}$$

This expression generalizes the one derived in refs 32 and 33 with $\varepsilon_i = 1$, which corresponds to the usual value for standard atomistic protein force fields. It corrects also for a misprint in our preliminary report,[31] since here the expression of α_e is divided by ε_i compared to the previous one. The whole demonstration of the case when $\varepsilon_i \neq 1$ can be found in detail in ref 47. The PPP model, with $\varepsilon_i = 1$, was shown to reproduce efficiently both structural and thermodynamic properties of water around proteins[32] and nucleic acids[33] and has also been successfully combined with a polarizable all-atom model of proteins.[48,49]

The PCGS van der Waals parameters have been adapted *mutatis-mutandis* from the CG water W model,[5] developed for representing the aqueous medium around membranes, and involving no Coulombic contribution. To be consistent with our SCORPION protein model, we have fitted their 6−4 Lennard-Jones with our van der Waals energy function formulated in eq 2, yielding the three parameters ε_{ii} = 1.96 kcal/mol, λ_{ii} = 3.97 Å, and σ_{ii} = 4.85 Å. This potential has a zero value for the distance σ_s = 4.62 Å.

Our preliminary parametrization based on hydrophobic forces[31] at contact yielded slightly different parameters that lead to a deeper well in the van der Waals potential. However, those parameters used in molecular dynamics were shown to be too attractive and to cause "water holes" at low temperatures, altering the heating procedure. We therefore decided to keep the parameters derived from the W model, since it was shown to reproduce the correct density of water, and to have a liquid phase for temperatures between 0 to 100 °C.[5] Here, we show furthermore that the W model is able to account for hydrophobic interactions. Following similar calculations in SPC/E water,[50] we have computed from short MD simulations in PCGS solvent the solvent-induced mean force between two purely hydrophobic spherical solutes as a function of their distance from contact. The numerical integration of the mean force yields the hydrophobic potential of mean force (PMF),

which is clearly attractive as can be seen in Figure 2. The hydrophobic PMF value at contact scales linearly with the square of the solute radius R, in agreement with results of the atomistic model:[50] using a liquid−vapor surface tension of water $\gamma = 0.43$ kJ/mol/Å^{-2}, our potential value at contact varies as $PMF_{contact} = -0.93\gamma R^2$ instead of $-1.19\gamma R^2$ when inferred from the MD simulations in SPC/E. This shows that the coarse-grained W model is able to describe qualitatively, and almost quantitatively, the solvent-induced attractive force between two hydrophobic solutes. A slight further adjustment of the W-model parameters is still needed for this purpose.

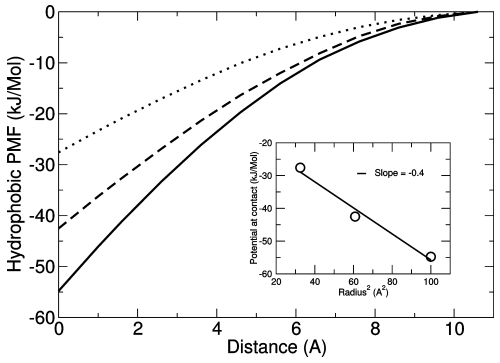


Figure 2. Potential of mean force between two hydrophobic spherical solutes in the SCORPION solvent model as a function of their surface-to-surface distance. Results are shown for solute radii $R = 5.7$ Å (dot lines), $R = 7.8$ Å (dashed lines) and $R = 10.0$ Å (solid lines). The inset shows the correlation of the hydrophobic potential at contact with respect to solute radii.

Regarding the electrostatic parameters of the PCGS model present in the electrostatic energy expression of eq 4, the internal dielectric constant was fixed to $\varepsilon_i = 4$, which is a typical value to account for electronic and nuclear polarization effects in proteins, and we took $\varepsilon = 80$ for the solvent. The value of the saturation dipole μ_s was fixed by keeping the same reduced value as in our previous microscopic version of the PPP model: $\mu^* = \mu_s/(\varepsilon_s\sigma_s^3)^{1/2} = 1.333$, where ε_s is the minimum value of the van der Waals potential energy between two solvent particles; this yields $\mu_s = 1.806$ D.

3.2. Solvation Free Energies of Peptides with SCORPION. With the SCORPION model being at this stage fully defined for proteins and solvent, we tested its ability to yield "on the fly" the correct electrostatic solvation free energies, by comparing for a set of CG peptides averaged electrostatic solvation free energies with the numerical solution of the Poisson−Boltzmann equation (PB). These free energies (last term in eq 4) can be understood as a preliminary Boltzmann average over orientations for given particle positions.[46] These two models are used here to estimate the energy for transferring peptides, considered as rigid bodies, from a low dielectric medium $\varepsilon_i = 4$ to a high dielectric solvent $\varepsilon = 80$.

The 17 tested peptides, including beta hairpins, alpha helices, and coils, have been downloaded from the Protein Data Bank[14] and slightly relaxed with a short minimization using the all-atom AMBER force field.[51] The list and PDB codes of the used peptides can be found in ref 31. Then, the CG peptide models were built on these all-atom structures, and their CG charges were optimized as previously described in section 2.1. MD simulations of all the rigid peptides embedded in a box of around 2000 PCGS particles were performed over 120 ps, and the averaged electrostatic solvation free energies were evaluated over the runs. The results are correlated to direct PB calculations on the AA representation of peptides, using the APBS software[52] in Figure 3. Those results are virtually identical if the CG representation is used instead. The agreement is seen to be quite satisfactory, with a correlation coefficient very close to one and a slope slightly over one. This is an encouraging sign that our rather simple particle-based CG water model is able to reproduce correctly fine electrostatic effects.

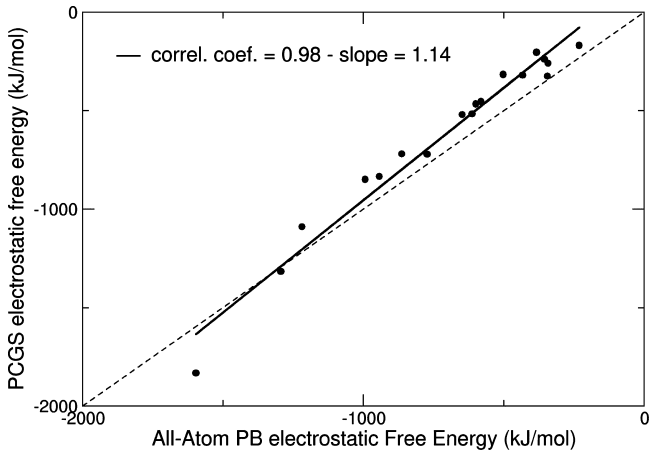


Figure 3. Correlation between the electrostatic solvation free energies of 17 peptides computed along short MD with the SCORPION model for both proteins and solvent, and using a Poisson−Boltzmann solver on AA representations of peptides.

4. MD SIMULATIONS OF PROTEIN RECOGNITION WITH SCORPION: THE BARNASE/BARSTAR COMPLEX

4.1. Introduction. The SCORPION protein and solvent model has been shown to reproduce both the specificity of protein−protein recognition and the fine solvent effects well. The combination of the two now makes it possible to perform molecular dynamics simulations of proteins in a solvent environment. The systematic study of protein association dynamics in solution is at the limit of the scope of atomistic simulations and is well suited to CG modeling, which allows for larger time-scale exploration and smooths the potential energy surface, while keeping a reasonable degree of molecular specificity. In addition to the final, most favorable structure of the complex, it provides insight into the different possible association pathways. There are many works devoted to the dynamics of protein/membrane interactions at a CG level, but yet only a limited number devoted to the dynamics of protein−protein association in solution.[20] Having decomposed in a systematic way van der Waals interactions (that drive hydrophobic effects) from electrostatic ones, our model makes it possible to pinpoint the physical factors that drive the recognition.

We chose to apply our SCORPION model to the MD of the barnase−barstar complex, which is a typical benchmark system for protein recognition studies since it has been well studied both by experiments and in theory. Several experiments showed that these proteins bind fast and with a high affinity, and that the recognition is electrostatically driven.[36,53] The presence of

many polar and charged residues at the interface suggests that electrostatics should play an important role in the recognition, and many theoretical studies emphasized the role of electrostatic interactions in the barnase/barstar binding, although the solvent role in binding is still under debate. Brownian dynamics simulations emphasized the importance of long-range electrostatic steering on binding.[45] Calculations with a continuum solvent model showed that the charge distribution of barstar is electrostatically optimized to bind to barnase.[54] Calculations with a continuum solvent model showed that the desolvation electrostatic cost was almost canceled by the direct electrostatic attraction.[55] Moreover, a theoretical study on binding energetics on several mutated barnase/barstar complexes confirmed that the total electrostatic interaction free energy was the dominant contribution to binding and that direct electrostatic interactions can compensate for the desolvation cost.[56] The importance of the solvation treatment in the binding calculations was pointed out both in a static theoretical study using the Poisson−Boltzmann theory[57] and in molecular dynamics docking simulations using several implicit solvent models.[58] In the latter all-atom MD study, it was shown that complexation of barnase and barstar occurred using a DDD solvent model or a surface-area dependent model, whereas it did not occur using a generalized Born model for solvent, emphasizing the sensitivity of protein recognition upon the solvent model. A recent constraint-biased MD simulation with explicit SPC water showed that the displacement of water at the complex interface creates an energy barrier for binding.[59] However, another unbiased MD study in explicit water showed that water in the interfacial gap plays an adhesive role and stabilizes early intermediates before native contacts are formed.[60] It should be noted that in those two all-atom MD simulations,[59,60] the starting structures were generated from the crystal structure of the complex by displacing barstar along the vector connecting the centers of mass of the two proteins and sometimes rotating slightly one protein around this vector. In the unrestrained all-atom MD of Ahmad et al.,[60] five simulations out of nine resulted in the native complex structure after hundreds of nanoseconds of MD. Simulation of the recognition dynamics of barnase/barstar in solution with all-atom MD seems therefore not to be straightforward. Due to the large amount of information for that system and the fact that we already handled it using the redocking procedure in implicit solvent presented before, this complex appears as an interesting test case to study the dynamics of recognition with our SCORPION protein and solvent CG model.

4.2. MD Simulation Method. We performed molecular dynamics on the barnase/barstar complex (PDB: 1BRS[37]) using the SCORPION model for protein and solvent. We used the ORAC software[34] in which we implemented all SCORPION features. The main addition made in the code of ORAC is a subroutine that calculates the solvent dipoles using eq 3 at each time step of the MD trajectory, and which can be easily implemented into any other MD codes than ORAC. Since the solvent dipoles are insensitive to each other, this calculation does not need iterations to converge to a self-consistent solution and is thus straightforward and fast. The performance of the coarse-grained model was evaluated by comparing the CPU times needed to perform two short simulations of the barnase/barstar complex, one at a CG level and the other with an all-atom model. Using the same simulation box size and cutoff distance (see below) and a time step of 10 and 2 fs for, respectively, the CG and AA models, the CPU time of the CG simulation is 130 times faster than the AA one.

The initial crystal structure of the complex (chains A and D) was first shortly relaxed using the all-atom AMBER force field[51] before the CG model was built on this all-atom structure. The CG point charges were optimized for each protein separately using the protocol detailed previously in this paper. Then, we generated seven different initial structures with various latitude, longitude, and Euler angles, with a minimum interprotein grain-to-grain center distance of 8 Å. The seven complex initial structures are presented in Figure 4 and will be later referred to as sim. a to sim. g. RMSDs from the crystal structure of all seven initial structures are presented in Table 1.

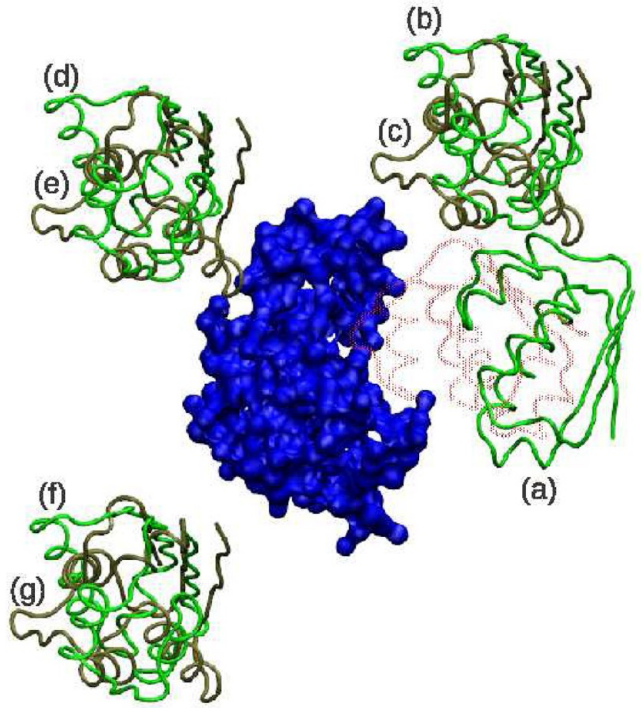


Figure 4. Initial positions of barstar molecules (in green and tan, Cα trace) relative to barnase (in blue, surface representation). Barstar from the crystal structure is represented in transparent red Cα trace. Image realized using VMD.[61]

Table 1. RMSD from 1BRS for the Seven Initial Complex CG Structures, Calculated on Barstar CG Sites after Fitting on the Barnase Molecule

initial structures	sim. a	sim. b	sim. c	sim. d	sim. e	sim. f	sim. g
RMSD from 1BRS (Å)	13.7	38.2	34.9	50.5	47.1	55.1	53.5

Starting from these seven initial structures of the complex, we first performed 500-ns-long MD simulations using the same implicit sigmoidal distance-dielectric function as the one used in the docking simulations described in section 2.2, using a thermostat at 500 K. However, in all seven simulations, the two proteins rapidly stuck to each other in a wrong conformation and did not move for the rest of the simulation. Although we were able to reproduce the native structure of the complex with exhaustive docking calculations, this result suggests that, as for MD simulations, the implicit representation of the solvent lacks

the solvent entropy and the electrostatic fluctuations necessary to explore dynamically the phase space.

The complexes, for which the total charge is $-3e$ (barnase is charged $+2e$ and barstar $-5e$), were therefore solvated in a cubic box of 94 Å size by roughly 6300 PCGS particles. The MD simulations were performed in the NVE ensemble, at a temperature of 500 K in order to accelerate phase-space sampling, and without any constraints on the relative protein positions. The protein fold was maintained using a quasi-rigid elastic network model,[62] with a cutoff between particles of 10 Å and a force constant of 5 kcal/mol/Å^2. The cutoff for long-range interactions was fixed to 21 Å, and a time step of 10 fs was used. The lengths of the simulations go from 200 to 550 ns, depending on the system. Along the MD simulations, we analyzed the time evolution of the complexes' RMSD from the crystal structure. The RMSD was calculated for each complex on all CG sites of the barstar molecule after fitting on the barnase molecule. To explore in more detail the interface between the two proteins, we also calculated the fraction of correct interface residues (FIR, percentage of correct residues at the interface with respect to the total number of residues at the interface), as defined in the similar cohesin-dockerin CG MD reported in ref 20. Residues were considered to be at the interface if one of their coarse grains was inside a 7.5 Å cutoff of another coarse grain from the other protein. There are therefore 47 residues considered to be at the interface in the native barnase/barstar complex. We also computed the time evolution of the total solute Coulombic energy, the total van der Waals solute energy, and the solute−solvent interaction energies, in order to see which part of the energy was driving the recognition. Since the energy fluctuations were relatively large, we computed the running averages of all energies over 20 ns time windows for clarity.

4.3. Results. *4.3.1. Structural Analysis of the Seven MD Simulations.* Time evolutions of RMSD for the seven simulations are shown in Figure 5, and the tridimensional final structures are shown in Figure 6. It can be seen on the top panel of Figure 6 that five of seven simulations, sims. a, b, c, e, and g, tend to adopt a conformation very similar to the native complex structure with a RMSD from the crystal of around 0.5 to 2 Å. This occurs after various simulation times depending on the simulation. It took only one nanosecond for sim. a, which was initially the closest to the crystal structure (13.7 Å RMSD), to adopt the complex native structure, whereas it took 480 ns for sim. b, 120 ns for sim. c, 470 ns for sim. e, and 30 ns for sim. g. It should be noted that when the barstar molecule is attached to barnase as in the native bound structure, it stays in this conformation for the rest of the simulation, and sim. a, b, and g were therefore stopped before the others, respectively after 200, 300, and 200 ns. As for sims. d and f, which have not yet adopted the native conformation, the final structures are very similar to each other, as shown on the bottom panel of Figure 6, and very similar to the conformation of sim. b before it retrieves the crystal structure (represented in Figure 6 by b′, which is the barnase structure extracted after 450 ns of sim. b). The barstar molecule is already placed in the correct binding site at the barnase surface but is not quite correctly oriented. Since sim. b finally adopted the native configuration after 480 ns, it is very likely that sims. d and f will also adopt the native conformation after a certain time.

Time evolutions of the FIR, which reveals how much the interface between the two partners is correct, are represented in Figure 7. In almost all simulations, FIR increases very rapidly

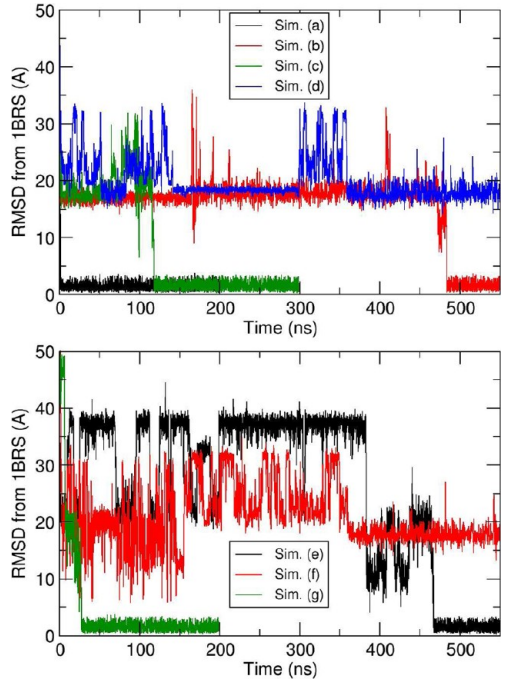


Figure 5. Time evolutions of RMSD from crystal 1BRS, calculated on all CG sites of barstar after fitting on barnase for the seven simulations.

since the barstar molecule goes quickly close to the native binding site on barnase. Then barstar rotates until it finds the native conformation. In sim. b, whereas the FIR first increases very quickly to about 60%, since barstar goes close to the binding site on barstar in a wrong orientation, it decreases to around 30% after 170 ns for almost 300 ns, before increasing to around 100% after 470 ns, showing that barstar moved away from barnase before adopting the orientation of the native complex. On the contrary, in sim. e, the FIR stays low at the beginning of the simulation, around 25%, until around 480 ns of dynamics, since the barstar molecule is far from the binding site of barnase, before it goes close to the binding site and rotates to find the correct orientation.

The variety of pathways to form the native bound conformation of the complex confirms that there is no unique way in the dynamics of recognition of this complex, which is consistent with previous all-atom MD simulations.[59] Nevertheless, there is clearly a strong basin of attraction at the barnase surface which guides barstar into the binding site, where barstar rotates until finding the correct position. It should be noted that, except when the correct complex is formed and becomes extremely stable, the molecules are not fixed in a certain conformation, and although the RMSD looks globally stable, both partners are relatively mobile with respect to each other.

Movies of the seven molecular dynamics simulations can be found in the Supporting Information, without the solvent molecules and after fitting on barnase for clarity.

4.3.2. Energetic Analysis of the Seven MD Simulations. Time evolutions of the running averages of the total solute electrostatic and van der Waals energies are plotted in, respectively, Figures 8 and 9. Both energies decrease significantly with the formation of the native complex and are on the same order of magnitude, showing that the native complex structure is more favorable both for electrostatics and for van der Waals interactions. However, the van der Waals

dx.doi.org/10.1021/ct300943w | J. Chem. Theory Comput. 2013, 9, 803−813

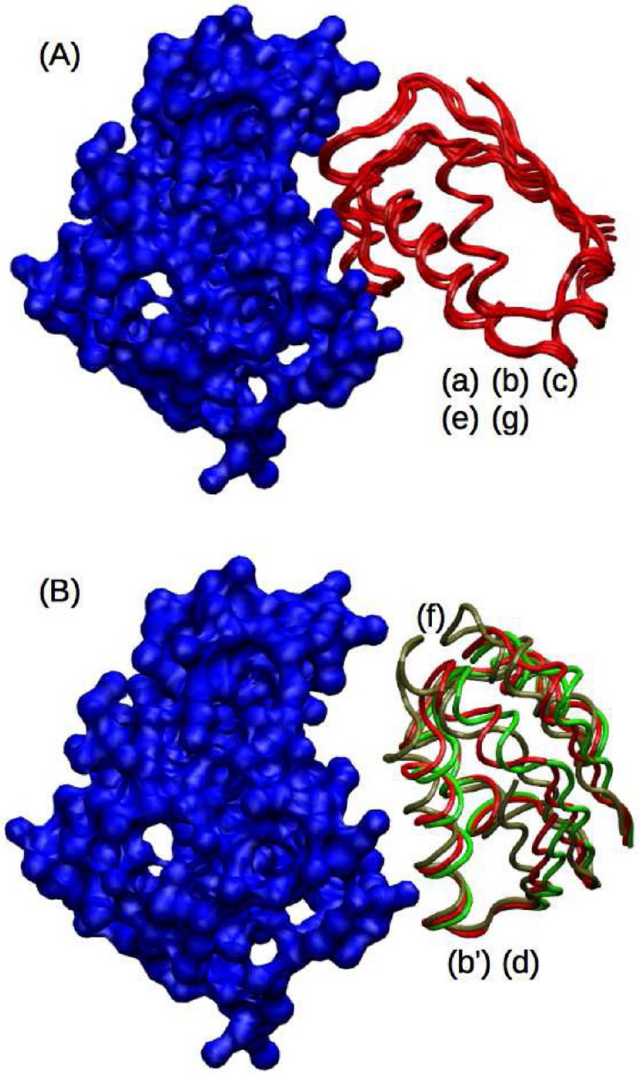


Figure 6. Final positions of barstar molecules (Cα trace) relative to barnase (in blue, surface representation): top panel (A), barstar for sims. a, b, c, e, and g in red; bottom panel (B), barstar for sim. d in green, sim. f in tan, and b′, which is sim. b after 450 ns, in red. Image realized using VMD.[61]

energy is the one that decreases the most when the native complex is formed, decreasing by more than 200 kJ/mol for all simulations, whereas the electrostatic energy decreases by about 100 kJ/mol. These results show that the dynamics of recognition of barnase/barstar is both electrostatically and sterically driven. Interestingly, as for sim. b for which barstar moves away from barstar in order to find the correct orientation of the native complex, we can see that, whereas the van der Waals energy increases when barstar moves away, the electrostatic energy decreases, suggesting that this recognition pathway should be guided by electrostatics at long range and van der Waals interactions at short range. This long-range electrostatic steering effect is consistent with previous Brownian dynamics simulations.[45]

The solute−solvent electrostatic and van der Waals energies are represented respectively on Figures 10 and 11. Both energies slightly increased for about 100 kJ/mol (this effect can be seen mostly for sims. b and e) upon binding, showing that the desolvation cost of binding is both electrostatic and sterical. The electrostatic desolvation cost is globally compensated by

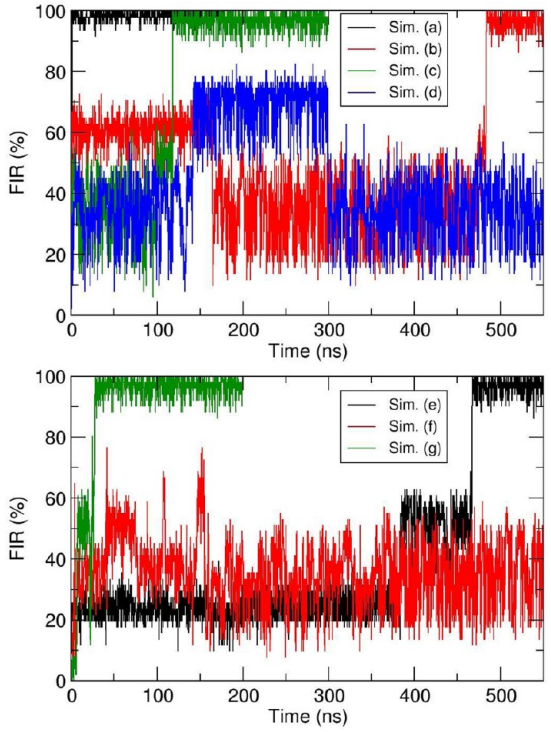


Figure 7. Time evolution of the FIR percentage (percentage of correct residues at the interface over the total number of residues at the interface) for sim. a to g.

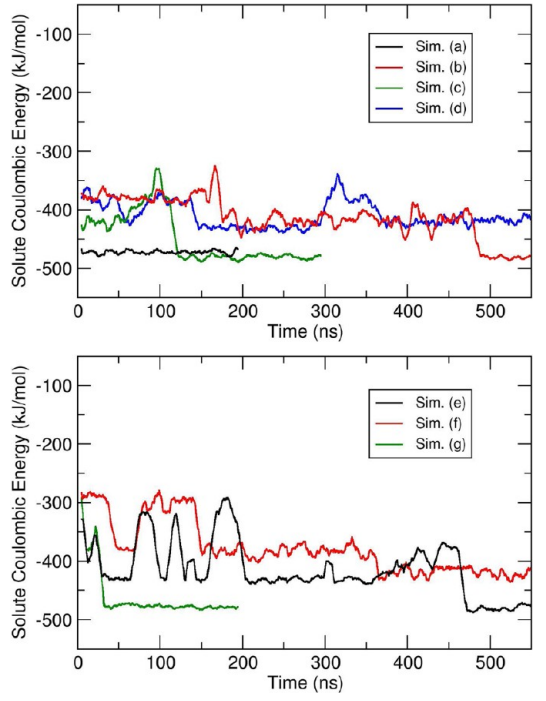


Figure 8. Time evolution of the running average, over 20 ns time windows, of the total solute electrostatic energy of the complex for the seven simulations.

the direct electrostatic gain shown in Figure 8, which is consistent with previous studies on barnase−barstar energetics.[55,56] Regarding the van der Waals desolvation cost, it is globally outranged by the direct van der Waals gain shown in Figure 9.

dx.doi.org/10.1021/ct300943w | J. Chem. Theory Comput. 2013, 9, 803−813

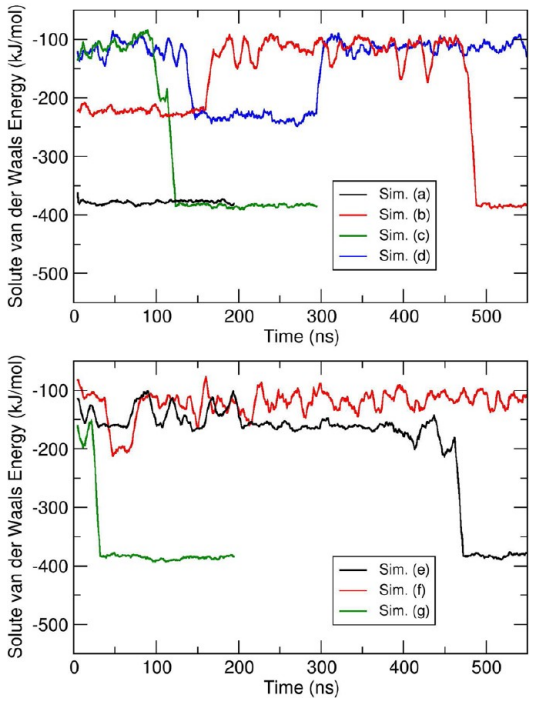


Figure 9. Time evolution of the running average, over 20 ns time windows, of the total solute van der Waals energy of the complex for the seven simulations.

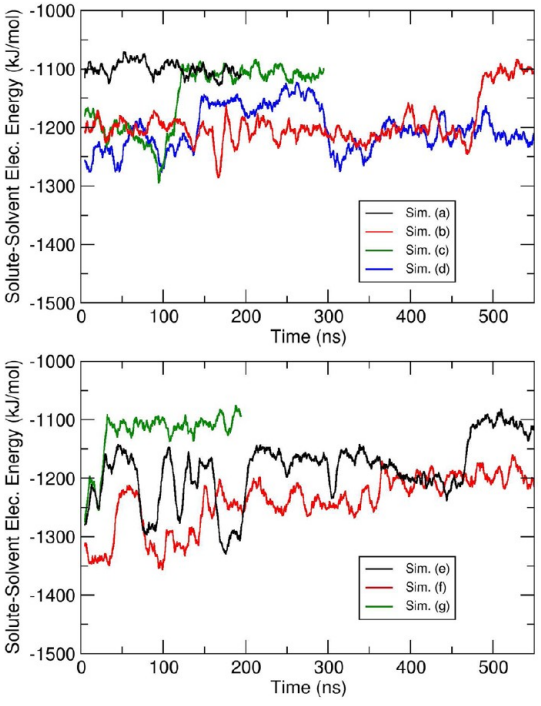


Figure 10. Time evolution of the running average, over 20 ns time windows, of the solute−solvent electrostatic energy of the system for the seven simulations.

4.3.3. Discussion. The SCORPION force field for protein and solvent was successfully used to perform MD of association of the complex barnase/barstar. Starting from different initial structures, for five over seven simulations, complexation in the exact native bound state occurred on a limited time scale going from 1 to 500 ns. Our PCGS model proved to be accurate

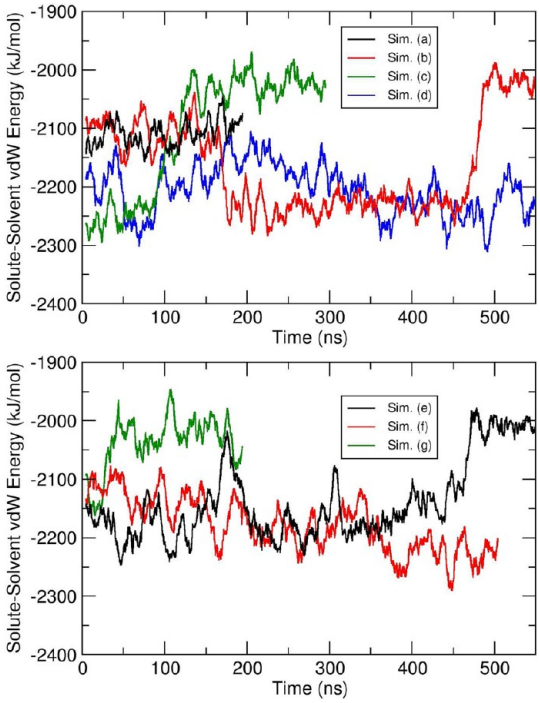


Figure 11. Time evolution of the running average, over 20 ns time windows, of the solute−solvent van der Waals energy of the system for the seven simulations.

enough to allow the complexation in the native form, although the initial structures were relatively far from the native complex, with RMSD going from 13.7 Å to 55.1 Å. Different intermediate structures were observed before correct binding, showing that there are multiple pathways for the complexation of this system, which is consistent with an all-atom constraint-biased MD simulation in explicit solvent.[59] However, part of the interface was always very quickly formed, as shown by the fraction of correct residues at the interface that increased right from the beginning of the simulations. This confirms the fact that the preformation of this complex is electrostatically driven. The SCORPION force field allowed us to study binding energetics of recognition and confirmed that electrostatic interactions play a role in recognition and compensate the electrostatic desolvation cost. Our study emphasized also the favorable role of van der Waals interactions in recognition.

5. CONCLUSION

We have here completed our systematic derivation of the coarse-grained protein force field SCORPION from an underlying all-atom one. This CG representation can be used to study protein−protein interactions in conjunction with any implicit solvent method, allowing the performance of efficient rigid docking predictions of protein−protein complexes. At the same level of coarse-graining, we have thus introduced a polarizable particle-based coarse-grained water model which is able to reproduce the macroscopic dielectric behavior of water, and to provide "on-the-fly" electrostatic solvation free energies of proteins. This model is consistent with the CG protein representation, yielding a complete force field for protein and solvent. Using a quasi-rigid body approximation, we demonstrated that SCORPION is able to simulate efficiently the dynamics of recognition for a protein complex starting from different initial structures. Moreover, it allowed an energetic

dx.doi.org/10.1021/ct300943w | J. Chem. Theory Comput. 2013, 9, 803−813

study of the recognition process which is consistent with previous all-atom theoretical studies. Note again that, along the same directions, Ahmad et al.[60] have used unbiased all-atom MD simulations to analyze the mechanism and energetics of the barnase−barstar association, but in contrast to our CG simulations, all their initial configurations were all chosen quite close to the final bound structure.

In the SCORPION force field presented here, the full internal flexibility of proteins was not taken into account. A flexible version was previously developed from the same CG protein model by one of us.[63] It implies the addition of CG sites necessary to maintain the secondary structures of proteins, which adds non-negligible computational cost. In order to take into account protein flexibility in docking simulations, mutliple-copy methods associated with CG models have been shown to improve the docking predictions.[64] The multiple-copy algorithm can either be at a residue-level,[19] allowing side chains to adapt upon binding, or at a larger amplitude level with loop optimization,[65] allowing global loop movements upon binding. An implementation of the SCORPION force field in the very efficient PTools library,[66] which includes a multiple-copy algorithm, is currently ongoing and will make it possible to apply SCORPION to various nonrigid protein−protein docking predictions on a large scale.

■ ASSOCIATED CONTENT

Supporting Information

Movies of the seven molecular dynamics simulations a to g, generated with VMD,[61] without the solvent molecules and after fitting on barnase for clarity. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: nbasdeva@univ-evry.fr; tap.ha-duong@u-psud.fr.

Present Address

§BIOCIS, UMR 8076, Faculté de Pharmacie, Université Paris Sud, 5 rue Jean-Baptiste Clément, 92296 Châtenay-Malabry, France

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

N.B. and T.H.-D. acknowledge the financial support of ANR-08-JCJC-0081-01 from the French Agence Nationale de la Recherche.

■ REFERENCES

(1) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59−107.
(2) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715−1731.
(3) DeWitte, R. S.; Shakhnovich, E. I. *Protein Sci.* **1994**, *3*, 1570−1581.
(4) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.
(5) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B.* **2001**, *105*, 4464−4470.
(6) Loison, C.; Mareschal, M.; Schmid, F. *J. Chem. Phys.* **2004**, *121*, 1890−1900.
(7) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B.* **2005**, *109*, 2469−2473.
(8) Larini, L.; Lu, L.; Voth, G. A. *J. Chem. Phys.* **2010**, *132*, 164107−164110.
(9) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750−760.
(10) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(11) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. *J. Struct. Biol.* **2007**, *157*, 593−605.
(12) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(13) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. *J. Chem. Theory Comput.* **2009**, *5*, 2531−2543.
(14) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(15) Miyazawa, S.; Jernigan, R. L. *Macromolecules* **1985**, *18*, 534−552.
(16) Zhang, Z.; Shi, Y.; Liu, H. *Biophys. J.* **2003**, *84*, 3583−3593.
(17) Zacharias, M. *Protein Sci.* **2003**, *12*, 1271−1282.
(18) Chang, C.-E.; Shen, T.; Trylska, J.; Tozzini, V.; McCammon, J. A. *Biophys. J.* **2006**, *90*, 3880−3885.
(19) Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 252−256.
(20) Hall, B. A.; Sansom, M. S. P. *J. Chem. Theory Comput.* **2009**, *5*, 2465−2471.
(21) Darré, L.; Machado, M. R.; Pantano, S. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 921−930.
(22) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105−134113.
(23) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *Plos Comput. Biol.* **2010**, *6*, e1000810.
(24) Wu, Z.; Cui, Q.; Yethiraj, A. *J. Chem. Theory Comput.* **2011**, *7*, 3793−3802.
(25) Riniker, S.; van Gunsteren, W. F. *J. Chem. Phys.* **2011**, *134*, 084110−084112.
(26) Riniker, S.; Eichenberger, A. P.; van Gunsteren, W. F. *Eur. Biophys. J.* **2012**, *41*, 647−661.
(27) He, X.; Shinoda, W.; DeVane, R.; Klein, M. L. *Mol. Phys.* **2010**, *108*, 2007−2020.
(28) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. *J. Chem. Theory Comput.* **2010**, *6*, 3793−3807.
(29) Darré, L.; Tek, A.; Baaden, M.; Pantano, S. *J. Chem. Theory Comput.* **2012**, *8*, 3880−3894.
(30) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Phys. Chem. B.* **2007**, *111*, 9390−9399.
(31) Ha-Duong, T.; Basdevant, N.; Borgis, D. *Chem. Phys. Lett.* **2009**, *468*, 79−82.
(32) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Comput. Chem.* **2004**, *25*, 1015−1029.
(33) Basdevant, N.; Ha-Duong, T.; Borgis, D. *J. Chem. Theory Comput.* **2006**, *2*, 1646−1656.
(34) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J. Comput. Chem.* **1997**, *18*, 1848−1862.
(35) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 7666−7669.
(36) Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8878−8889.
(37) Carvalho, A. L.; Dias, F. M. V.; Prates, J. a. M.; Nagy, T.; Gilbert, H. J.; Davies, G. J.; Ferreira, L. M. A.; Romão, M. J.; Fontes, C. M. G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13809−13814.
(38) Kingston, R. L.; Hamel, D. J.; Gay, L. S.; Dahlquist, F. W.; Matthews, B. W. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8301−8306.
(39) Clément, M.-J.; Jourdain, I.; Lachkar, S.; Savarin, P.; Gigant, B.; Knossow, M.; Toma, F.; Sobel, A.; Curmi, P. A. *Biochemistry* **2005**, *44*, 14616−14625.
(40) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570−4577.
(41) Varilly, P.; Patel, A. J.; Chandler, D. *J. Chem. Phys.* **2011**, *134*, 074109−074115.
(42) Barkema, G.; Mousseau, N. *Phys. Rev. Lett.* **1996**, *77*, 4358−4361.
(43) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(44) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169−9183.

(45) Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997**, *72*, 1917−1929.
(46) HaDuong, T.; Phan, S.; Marchi, M.; Borgis, D. *J. Chem. Phys.* **2002**, *117*, 541−556.
(47) Ha-Duong, T.; Basdevant, N.; Borgis, D. In *Modeling Solvent Environments*; Feig, M., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2010; Chapter 11, pp 251−272.
(48) Masella, M.; Borgis, D.; Cuniasse, P. *J. Comput. Chem.* **2008**, *29*, 1707−1724.
(49) Masella, M.; Borgis, D.; Cuniasse, P. *J. Comput. Chem.* **2011**, *32*, 2664−2678.
(50) Dzubiella, J.; Hansen, J.-P. *J. Chem. Phys.* **2004**, *121*, 5514−5530.
(51) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(52) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037−10041.
(53) Schreiber, G.; Fersht, A. R. *J. Mol. Biol.* **1995**, *248*, 478−486.
(54) Lee, L. P.; Tidor, B. *Protein Sci.* **2001**, *10*, 362−377.
(55) Sheinerman, F. B.; Honig, B. *J. Mol. Biol.* **2002**, *318*, 161−177.
(56) Wang, T.; Tomic, S.; Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **2004**, *87*, 1618−1630.
(57) Dong, F.; Vijayakumar, M.; Zhou, H.-X. *Biophys. J.* **2003**, *85*, 49−60.
(58) Wang, T.; Wade, R. C. *Proteins: Struct., Funct., Bioinf.* **2003**, *50*, 158−169.
(59) Hoefling, M.; Gottschalk, K. E. *J. Struct. Biol.* **2010**, *171*, 52−63.
(60) Ahmad, M.; Gu, W.; Geyer, T.; Helms, V. *Nat. Commun.* **2011**, *2*, 261.
(61) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(62) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.
(63) Ha-Duong, T. *J. Chem. Theory Comput.* **2010**, *6*, 761−773.
(64) May, A.; Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 774−780.
(65) Bastard, K.; Prévost, C.; Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2006**, *62*, 956−969.
(66) Saladin, A.; Fiorucci, S.; Poulain, P.; Prévost, C.; Zacharias, M. *BMC Struct. Biol.* **2009**, *9*, 27.