JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Electrostatic-Consistent Coarse-Grained Potentials for Molecular Simulations of Proteins

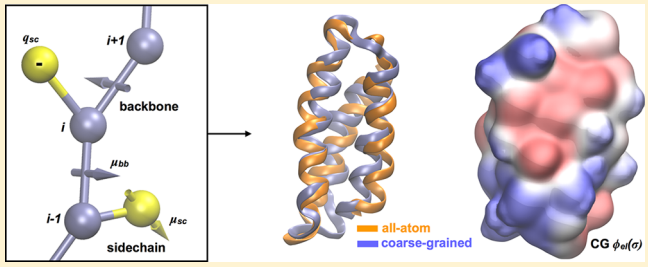
Enrico Spiga,[†] Davide Alemani,[†] Matteo T. Degiacomi,[†] Michele Cascella,[‡] and Matteo Dal Peraro[†,*]

[†]Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne-EPFL, Lausanne, CH-1015, Switzerland

[‡]Departement für Chemie und Biochemie, Universität Bern, Freiestrasse 3, Bern, CH-3012, Switzerland

Supporting Information

ABSTRACT: We present a new generation of coarse-grained (CG) potentials that account for a simplified electrostatic description of soluble proteins. The treatment of permanent electrostatic dipoles of the backbone and polar side-chains allows to simulate proteins, preserving an excellent structural and dynamic agreement with respective reference structures and all-atom molecular dynamics simulations. Moreover, multiprotein complexes can be well described maintaining their molecular interfaces thanks to the ability of this scheme to better describe the actual electrostatics at a CG level of resolution. An efficient and robust heuristic algorithm based on particle swarm optimization is used for the derivation of CG parameters via a force-matching procedure. The ability of this protocol to deal with high dimensional search spaces suggests that the extension of this optimization procedure to larger data sets may lead to the generation of a fully transferable CG force field. At the present stage, these electrostatic-consistent CG potentials are easily and efficiently parametrized, show a good degree of transferability, and can be used to simulate soluble proteins or, more interestingly, large macromolecular assemblies for which long all-atom simulations may not be easily affordable.



INTRODUCTION

Fundamental biological processes at the molecular level involve macromolecular assemblies of the most different sizes and can occur in a broad spectrum of time and length scales.[1] The most straightforward and accurate way to study these processes by computational approaches is to develop and use models at an atomistic level of resolution. Atomistic models for biomolecules have been successfully applied in the past decades[2] and are today at the most advanced frontier of biomolecular simulations.[3−6] To date, the boundaries for atomistic simulations have reached the millisecond time scale and passed the million of atoms.[7−10] Despite the current progress and success of all-atom simulations, the computational cost for this resolution remains challenging for the routine study of larger systems and for longer time scales. Following this objective, a variety of simplified models have been proposed in the last decades.[11,12] In particular, coarse-grained (CG) Hamiltonians have been introduced to describe macromolecular systems. The first CG models focused on simple hydrophobic-polar interactions, which led to a series of models for surfactant−water or lipid−water mixtures.[13−16] In more recent years, the same approach has been adopted for the development of multiscale methods and CG models for proteins and nucleic acids are used for the investigation of a large variety of processes.[17−29]

CG models for proteins are based on structural topologies that map the atomistic dimensionality to a given CG resolution, and on effective potentials able to reproduce the interactions of the original atomistic representation. The different strategies to derive CG potential parameters discussed in the literature are mostly based on mining degrees of freedom through Boltzmann inversion techniques, thermodynamic integration, force matching or cumulant-based descriptors.[13,30−32] In fact, it is extremely difficult to derive a general, multiresolution rigorous coarse-graining theory that would be able to generate a consistent and transferable CG force field at any given level of resolution. In the recent past, several steps toward this goal have been reported in the literature.[31,33,34] Force matching strategies have been particularly successful in determining effective coarse-grained potentials from atomistic simulations.[35,36] This approach has been applied to the study of several problems ranging from the folding of small peptides[37] to the simulation of immature HIV-1 virion.[38] Moreover, tentatives to overcome the problem of transferability of the potential parameters have been done checking the correspondence of parameter values among systems[39] or adopting a "host and guest parametrization strategy" with consequent CG MD simulations to quantify the transferability of the parameters.[40] With the same objective, using the Yvon−Green−Born integral equation, it has been possible to treat many-body structural correlations with the aim to determine more transferable

Received: February 21, 2013

   dx.doi.org/10.1021/ct400137q | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

potentials for folded proteins.[41] Using the concept of relative entropy to guide the parametrization procedure, promising results have been recently obtained for the study of large-scale fibrillar assembly.[42] Still, many issues afflict current CG schemes, which limit their general applicability to a large class of relevant biological problems. The functional form at the CG level is not univocally defined, and in principle, it should explicitly treat many body effects[43] or polarization terms. Also, optimal mapping schemes able to ensure the accuracy of the potentials in reproducing particular properties of interest[44] should be applied. In practice, effective schemes able to overcome some of these problems producing CG force fields adapted to tackle specific biological problems are present in the literature. For example, the MARTINI force field for proteins and lipids, which is characterized by good transferability, has been developed using a combination of free energy based calculations and Boltzmann inversion.[30,45] This model was successfully applied to membrane simulations and showed great potential for membrane proteins investigations.[46−49] One drawback of the MARTINI force-field lies in the requirement of external biases to preserve secondary structure elements, limiting the possibility to explore phenomena associated to secondary structure transitions. Another transferable coarse-grained model with dipolar backbone contributions has been applied to small folded proteins, showing promising results for the description of structural fluctuation properties at this level of resolution.[50]

The investigation of large conformational changes has been successfully addressed by cumulant-based approaches for the definition of effective multibody potentials. Using such an approach, it has been possible to quantify correlations between local and nonlocal interactions, creating an united-residue force-field.[32,51] This force-field has been applied to the folding of α- and α/β-proteins[52] and the opening and closing of Hsp70 chaperones.[53] Preserving an atomistic description of the backbone, while only the side chains are coarse-grained, is an effective solution to explore and stabilize the secondary structure elements. Not surprisingly, CG models obtained using this strategy are widely used to study the aggregation of amyloidoigenic peptides, the folding of small peptides and refinement of protein structures.[26,54−59] On the other side, while improving on secondary structure stability, this approach departs from a uniform CG mapping and introduces additional degrees of freedom at the backbone level to create an almost atomistic representation.

We have recently proposed a strategy to describe the backbone contribution while preserving a single bead representation. The introduction of the explicit backbone dipole defined by three consecutive C_α beads along with the treatment of nonradial dipole−dipole interactions in dynamics allowed to obtain stable secondary structure elements of unspecific poly peptides and to sample conformational transitions.[60] In this work, we introduce a description of side chains electrostatics and extend this model to real proteins. The electrostatic contribution is explicitly considered to the second order of the multipole expansion, so that the remaining part of the nonbonded interactions are responsible only for short-ranged contributions. As previously demonstrated,[61] this approach has the benefit to better describe the actual electrostatic field at a CG level with implications to the treatment of molecular recognition in protein−protein interactions (PPIs). Moreover, as shown for short peptides,[60] within this approach the secondary structure of a variety of folding motifs is naturally maintained.

Following a parametrization protocol that combines Boltzmann inversion schemes and force-matching methods, we tuned the bonded and nonbonded terms of the backbone and side-chain beads for a broad range of protein folds using a novel algorithm based on particle swarm optimization.[62−66] The resulting set of electrostatic-consistent potentials allowed simulating large soluble proteins and protein−protein complexes in the microsecond scale, preserving a very good structural and dynamic agreement with respective all-atom simulations and experimental reference structures. Moreover, the proposed scheme of parametrization provides an inexpensive way to quickly derive CG parameters for protein systems and can eventually contribute to the generation of more transferable parameters for a general CG force field .

METHODS

Coarse-Graining the Atomistic Structure and Electrostatics of Proteins. Our CG model is based on an approximately four-to-one atoms-to-bead mapping, consistent with that used by other CG force fields (e.g., MARTINI[30,45,67]). On average, four heavy atoms are represented by a single interaction center, with the exception of aromatic side chains, where we used a higher resolution to map their geometric specificity (Figure 1). All the amino acids are composed by a bead representing the backbone and placed on top of the C_α atom, whereas one or more beads are used for the side chain, which are placed at the center of mass of their constituent heavy atoms. Alanine and glycine amino acids constitute the only exception, each being composed by a single bead. The mass of each bead is constituted by the total mass of all the respective atoms (Figure 1, Table S1 in the Supporting Information). Amino- and carboxy- terminal backbone beads are described bringing their respective zwitterionic charge and with the corresponding different total mass with respect to normal backbone beads.

Apart from the massive beads, our CG structure presents multiple electrostatic centers bringing a multipolar expansion of the corresponding all-atom electrostatic potential arrested to the dipolar term. In particular, we introduce monopolar charges and/or permanent electrostatic dipoles at all charged/polar side-chains as well as at each peptide-bond of the backbone (Figures 1 and 2). The backbone dipoles are embedded in the structure of the polypeptide chain following our previous work.[60]

Coarse-Graining the Potential Function. We adopted for our CG protein representation an additive potential function as typically used in all-atom Hamiltonians. This approach provides the best compromise between reasonable accuracy and computational efficiency for CG simulations.[22] The explicit introduction of electrostatic dipolar terms following[61] adds to the potential minimal many-body contributions, which enhance the stability of secondary structure elements without the use of *ad hoc* bias potentials.[60] The total potential function is given by

dx.doi.org/10.1021/ct400137q | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

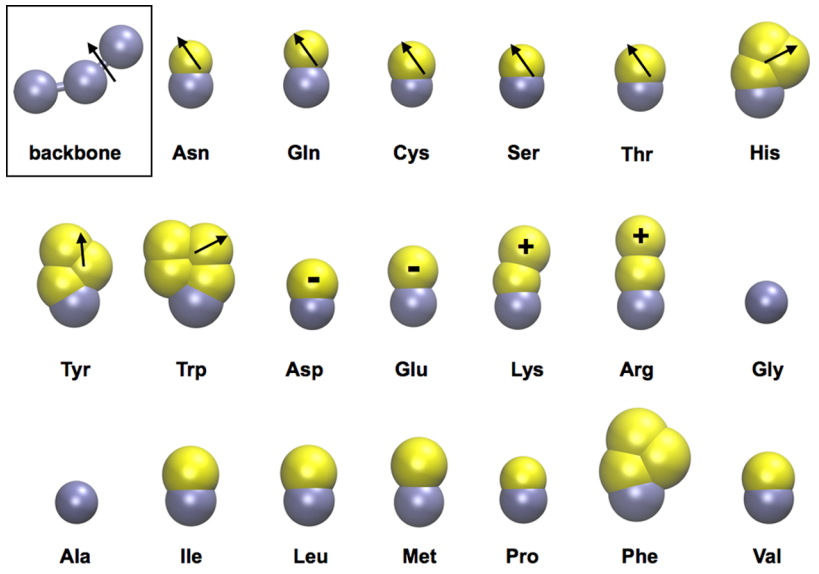


Figure 1. Coarse-grained representation of amino acids used in this work. Backbone beads are represented in ice-blue; side-chain beads are in yellow; arrows represent the electrostatic dipole moments associated with polar side-chains and the backbone. Acidic and basic amino acids carry net unitary charges.

$$V_{\text{total}} = \sum_{\text{bonds}} k_{\text{b}}(|\vec{r}_{ij}| - r_0)^2 + \sum_{\text{bendings}} \sum_{n=2}^{4} k_{a,n}(\theta_{ijk} - \theta_0)^n + \sum_{\text{dihe}} \sum_{n=1}^{3} k_{d,n}[1 + \cos(n\phi_{ijkl} - \phi_{0,n})] + \sum_{\text{improper}} \sum_{n=2}^{4} k_{i,n}(\psi_{ijkl} - \psi_0)^n + \sum_{\text{pairs}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{|\vec{r}_{ij}|}\right)^{12} - \left(\frac{\sigma_{ij}}{|\vec{r}_{ij}|}\right)^{6}\right] + V_{\text{el}}(\vec{r}_{ij}) \quad (1)$$

where the first four terms describe bonded interactions and the remaining are used to describe nonbonded ones. In particular, the first term describes pseudobonds between backbone beads and beads that belong to the same residue using a simple harmonic approximation, where r_0 is the equilibrium value and k_{b} is the constant force. The second term accounts for pseudobending for backbone and side chain beads,[68] with θ_0 the equilibrium value and $k_{a,n}$ the constant forces. The third for torsion potential of pseudodihedrals[68] for backbone and multibead side chains, with ϕ_0 as the equilibrium values and $k_{d,n}$ as the constant forces. The final term of the bonded potential describes improper torsion potentials used to force the L-chirality to the side chains or to force the planarity of aromatic side-chains where ψ_0 is the equilibrium value and $k_{i,n}$ the force constant.

The last two terms represent the nonbonded part of the total potential function: a common 6−12 Lennard-Jones potential is used to account for effective nonbonded interactions not explicitly included in the electrostatics potential term. The electrostatic potential, instead, reads

$$V_{\text{el}}(\vec{r}_{ij}) = C(|\vec{r}_{ij}|)[V_{q_iq_j} + V_{q_i\mu_j} + V_{\mu_iq_j} + V_{\mu_i\mu_j}] \quad (2)$$

where all charge−charge, charge−dipole, and dipole−dipole interactions are considered, where

$$C(|\vec{r}_{ij}|) = \frac{1}{4\pi\varepsilon_0\varepsilon(|\vec{r}_{ij}|)} \quad (3)$$

depends on the distance-dependent dielectric function:

$$\varepsilon(|\vec{r}_{ij}|) = 1 + k_{ij}(|\vec{r}_{ij}|) \quad (4)$$

with $k_{ij} = 4$.[69] This dielectric model has been initially chosen on the basis of its simplicity. However, more accurate implicit screening treatments can be easily coupled to our models.[70]

Along with this simple distance-dependent screening, we also implemented and tested an explicit solvent model following the works of Warshel[71] and Borgis.[72] In our model, four water molecules are mapped into one single water bead. The electrostatics problem of a dipolar or charged molecular solute immersed in a dielectric medium is described by a local nonequilibrium solvation free energy ΔF_{pol}, which is numerically integrated by discretizing the solvent region in "water" grains. Each water grain is associated with an induced dipole $\vec{p}_i$ and the electrostatics field $\vec{E}_{0i}$, generated by the solute. In this framework, the free-energy of solvation is given by

$$\Delta F_{\text{pol}} = \sum_{i=1}^{N} \frac{\vec{p}_i^{\,2}}{2\alpha} - \sum_{i=1}^{N} \vec{p}_i \vec{E}_{0i} \quad (5)$$

Thus, the local solvent model is obtained by minimizing the functional relative to all the $\vec{p}_i$, obtaining the equilibrium dipole moment, $\vec{p}_{i,\text{eq}}$

$$\vec{p}_{i,\text{eq}} = \frac{p_{\text{sat}}}{|\vec{E}_{0i}|} L\left(3\alpha\frac{|\vec{E}_{0i}|}{p_{\text{sat}}}\right)\vec{E}_{0i} \quad (6)$$

where

$$L(x) = \coth(x) - \frac{1}{x} \quad (7)$$

is the Langevin function, $\alpha = 2.3$ is the polarizability, and $p_{\text{sat}} = 1.8$ D is the dipole saturation, which are model parameters. As showed before, this model is computationally efficient, allowing

dx.doi.org/10.1021/ct400137q | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

application to simulation of nucleic acids, proteins and protein−protein complexes.[73−76]

Integrating Dipole Dynamics. Backbone dipoles are univocally defined by the C_α trace of the protein. The backbone dipole μ_i is associated to a triplet of consecutive C_α beads $(i-1,i,i+1)$, being located at the middle point between the second and the third bead, and its orientation determined by the angle of the bead triplet.[60,61] At each time step during the MD integration loop, the forces on each backbone dipole are calculated and distributed on the beads of the corresponding triplet, affecting their position and the amplitude of the relative triplet bending angle[60] (Figure 2A).

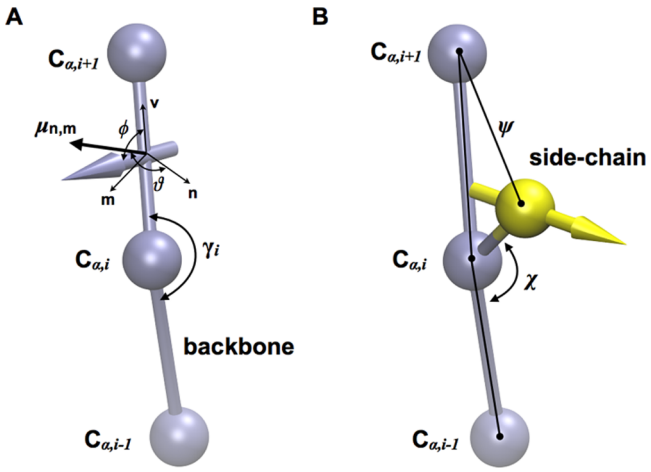


Figure 2. Representation of electrostatic dipole moments. (A) **v**, **n**, **m** are the vectors of the internal orthonormal basis used to reconstruct the backbone's dipole; $\mu_{m,n}$ is the projection of the dipole in the plane **m**, **n**; ϕ is the angle between the backbone dipole and the vector **v**; θ is the angle between the projection $\mu_{m,n}$ and the vector **n**; γ is the bending angle of the three consecutive C_α used to reconstruct the backbone dipole.[60] (B) Side-chain bead with the associated dipole drawn as an arrow; χ is the bending angle used to describe reorientation of the side chain; ψ is the improper torsion used to force the chirality of L-amino acids.

The orientation of the dipoles associated with the polar side-chains (Figure 2B) are updated by solving the classical equation:

$$\frac{d\vec{\mu}}{dt} = \vec{\omega} \times \vec{\mu} \tag{8}$$

where $\vec{\mu}$ is the dipole moment and $\vec{\omega}$ is the angular velocity of the side-chain. The angular velocity of the side-chain is determined by its inertia tensor $\bar{I}$, derived from the respective all-atom representation, and the electrostatic torque experienced by the dipole (Table S2, SI), following the equation:

$$\vec{\tau} = \bar{I}\frac{d\vec{\omega}}{dt} \tag{9}$$

For accuracy reasons, eqs 8 and 9 are implemented in the MD loop using the quaternion formalism.[77] The side chain electrostatic dipole moments are treated as rigid bodies that can rotate around a fix point (the center of mass of the bead which they belong). The module of the electrostatic dipole moment represents the actual value as obtained from quantum chemical calculations.[78] From an analysis of a nonredundant subset of NMR structures in the Protein Data Bank, we observed that, once CG-mapped, there is no strong preferential orientation for the electrostatics dipole moment of the side chains like asparagine, glutamine, serine, threonine, and cysteine. In the case of tyrosine, instead, we have two preferential orientation with respect to the plane of the ring. For histidine and tryptophan, the dipole remains rigidly linked to the plane of the ring. We make use of ellipsoids of rotation calculated from all-atom molecular dynamics simulations[79] to describe the reorientation of the electrostatic dipole moment of the beads. In our model, the electrostatics dipole moment is assumed to be aligned along the direction of the inertia axes with highest eigenvalue of the inertia tensor. The procedure for the backbone[60] and side chain dipolar dynamics has been implemented in the molecular dynamics code Lammps.[80]

Parameterizing the Coarse-Grained Potentials. A two-step procedure was adopted to derive reliable values of the parameters for the potential function discussed. A first general set of bonded parameters is derived using a Boltzmann inversion approach[13] on structural ensembles extracted from the PDB and from MD simulations; then, bonded and nonbonded parameters are refined for a given protein using a force matching procedure.[31,81]

Boltzmann Inversion. Based on the adopted CG mapping scheme (Figure 1), from every atomistic degree of freedom, the relative CG conformational distributions of bonded terms are obtained. In particular, the individual CG distributions for bond lengths $\{r\}$, bending angles $\{\theta\}$, and torsions $\{\phi\}$, $P^{CG}(\chi_i,T)$ are Boltzmann inverted[13,82] to obtain the corresponding potentials for the generic degree of freedom $\chi_i = r_i,\theta_i,\phi_i$, using the equation:

$$V^{CG}(\chi_i, T) = -k_B T \ln\left[\frac{P^{CG}(\chi_i, T)}{f(\chi_i)}\right] + C_{\chi_i} \tag{10}$$

where $f(\chi_i)$ is a function that takes into account the components of the Jacobian determinant.

This procedure was performed on a nonredundant subset of the PDB in order to identify all the possible coarse-grained degrees of freedom for the adopted mapping and possible preferential orientations of side-chain dipoles. The NMR part of the subset has been analyzed to identify possible preferential orientations of polar side chains. Since not all amino acids are equally represented in the PDB and some degrees of freedom could have poor statistics (e.g., degrees of freedom for tryptophan), we also extracted additional probability distributions from all-atom MD simulations of all possible homononapeptides in explicit solvent in both α- and β-conformation. We observed that the conformational distributions obtained from the two sets, at least for the most represented degrees of freedom, are in qualitative agreement (e.g., position of the minima for the potentials).

This initial seeding set was optimized using a force matching procedure on a given protein to further tune the bonded and parametrize the nonbonded interatomic potentials based on the trajectories obtained from all-atom MD simulations. It is important to notice that the Coulomb term for charge and dipole interactions is not affected by the parametrization procedure. Intramolecular electrostatic and Lennard-Jones interactions between charges and/or dipoles separated by one or two bonds (1−3 electrostatic interactions) have been excluded from our potential energy function.

Force Matching. The force matching procedure has been widely discussed in several publications.[35−37,81] Here, we briefly recall its main steps. Let $\{\omega\}$ indicate the entire set of L

dx.doi.org/10.1021/ct400137q | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

parameters $\{\omega_1,\ldots\ \omega_L\}$ used to define the potential function adopted for the coarse-grained representation. The optimal $\{\omega\}$ set defining the CG potential function is the one minimizing the fitness function $Z_F(\omega)$:

$$Z_F(\omega) = \sqrt{(3\sum_{k=1}^{M} N_k)^{-1} \sum_{k=1}^{M}\sum_{i=1}^{N_k} |F_{ki}(\omega) - F_{ki}^0|^2} \quad (11)$$

where M is the number of sets of atomic configurations available, N_k is the number of beads in configuration k, $F_{ki}(\omega)$ is the force on the i-th bead in set k obtained with parametrization ω, and F_{ki}^0 is the reference force acting on the bead as given by the following formula:

$$F_{ki}^0 = \sum_{j=1}^{L_i} F_{jki}^0 \quad (12)$$

which is the sum of the forces acting on the atoms that belong to the i-th bead. All quantities are averaged for a large set of different configurations, sampled from a preceding all-atom MD run.

Particle Swarm Optimization. The set of parameters $\{\omega\}$ that minimizes the fitness function $Z_F(\omega)$ are obtained using a Particle Swarm Optimization (PSO) heuristic method.[83] To do so, an ensemble of solutions (also called particles p) have their position $\omega(p)$ and velocity $v(p)$ randomly initialized in the multidimensional search space identified by boundaries. Along the whole optimization process, every particle will keep track of the position $\omega(p)$ associated with the best objective (fitness) function value $Z_F(\omega(p))$. At the beginning of every discrete step, particles are updated about the swarm status (i.e., the current position of all particles), as well as their respective best found solution value and position. Subsequently, they will independently update their own velocity, which will be used to update their position. Velocity update is affected by three factors. The first, inertia, determines how a particle's trajectory is preserved along time. The second, personal best, attracts particles toward their own best solution. The third, global best, attracts particles toward the best solution found by neighboring particles. Once velocity has been updated, a new position in which to evaluate the objective (fitness) function can be computed.

The boundaries associated with each parameter in the search space are guided by previous values obtained by Boltzmann inversion (as in the case of side chain bonded terms) or physically reasonable quantities preliminarily calculated, as for the case of the backbone bonded terms. In particular, the PSO approach was used to define effective bending potential parameters, able to correctly describe secondary structure conformations of an unspecific polypeptide (e.g., poly alanine) arranged as α-helix and in β-turn conformations. For the purpose of tuning the nonbonded potential terms the adopted boundaries for the Lennard-Jones terms in the following range: 4.0 Å $< \sigma_{ij} <$ 5.0 Å and 0.4 kcal mol^{-1} $< \varepsilon_{ij} <$ 1.3 kcal mol^{-1}.

Multiple runs of PSO showed how some parameters converged sooner that others; for instance, bonded parameters for the side chains invariantly converged to the same values, permitting to fix them on following optimization cycles for tuning more fluctuating parameters such as bonded terms for the backbone and the general nonbonded term. The backbone bonded and nonbonded parameters converged roughly to identical values (Table S3, Figure S1 (Supporting Information)). This already hinted to a partial set of parameters that can be transferable and used for a general CG force field.

Reference All-Atom Simulations. The atomistic reference forces, used for the force-matching procedure on the set of proteins studied in this work, have been extracted from all-atom simulations carried out using NAMD simulation package[3] in explicit solvent and periodic boundary conditions, using a Langevin dynamics for the thermostat and a Nosé–Hoover–Langevin piston for the barostat. Simulations were carried out using smooth particle-mesh Ewald (SPME)[84] for the calculation of electrostatic interactions. All simulations were carried out using all-atom force field Amber99SB[85,86] for the protein and TIP3P model[87,88] for the water. We ran 100 ns long MD simulations for five proteins belonging to different structural families, including both single molecules and protein–protein complexes in water solution. Namely, α-, β-, α/β-proteins and small protein–protein complexes were selected to test the performance of our CG approach. The protein α_3W, with PDB entry 1lq7, is a *de novo* α-protein composed by 67 amino acids arranged as a clockwise bundle of three helices, whose structure has been obtained by NMR.[89] The Cox11 protein is the β-protein, PDB entry 1sp0, which structure has been obtained by NMR and is composed by 131 amino acids.[90] The LysM Domain, with PDB entry 1e0g, is the α/β-protein: it has been obtained by NMR and is composed by 48 residues.[91] The coiled-coil protein is the engineered water-soluble phospholamban, which is composed by four helical monomers of thirty amino acids in an antiparallel arrangement and the structure of which has been obtained by X-ray crystallography.[92,93] Finally, the barnase–barstar complex solved by X-rays crystallography is composed by a total of 189 residues.[94]

For all the proteins, we extracted 1000 structures from the corresponding MD run to be used for the force-matching. For our purpose, we decided to run PSO using a setup of 20 particles with 3 consecutive repetitions of 300 optimization steps each. For all the PSO runs, the difference between reference forces and the calculated one was in the order of 1 kcal·mol^{-1}·Å^{-1} for degree of freedom (Figure S1 (Supporting Information)). For a protein of ∼50 residues such a setup permits to have a refined set of parameters in less than two days on four CPUs. Some parameters obtained are listed in the Supporting Information.

We explored the possibility to reduce the number of structures and the length of the required all-atom MD trajectory to be used for the particle swarm optimization runs. Using the Jarvis–Patrick clusterization method,[95] as implemented in Gromacs,[96] we saw that it is possible to reduce the number of structures of another order of magnitude. Preliminary force-matching calculations and consequent results, obtained from CG simulations, gave the same qualitative results showed in the present article (data not shown). For the α_3W protein, we also ran a simulation in explicit solvent without tuning the water–protein Lennard-Jones parameters of interaction and setting them to the same value that are ε_{ij}= 0.8 kcal/mol and σ = 4.7 Å, whereas the water–water Lennard-Jones parameters of interaction are ε_{ij}= 0.8 kcal/mol and σ = 4.6 Å.

We also explored the possibility to define fully transferable nonbonded parameters as extracted from the specific parametrization of the five structurally different proteins studied in this work. To do so, a simple averaging of the values of nonbonded ε and σ obtained for each pairs of beads were

Table 1. Summary of the Structural and Dynamic CG Properties[a]

protein	RMSD	R_g	SI_{RMSF}	SI_{S^2}	SI_{bend}	SI_{dihe}
α_3W	2.6 ± 0.2	11.0 ± 0.2	0.95	0.94	0.99	0.88
	[2.5 ± 0.2]	[12.2 ± 0.2]				
Cox11	3.2 ± 0.2	15.1 ± 0.1	0.80	0.93	0.98	0.71
	[3.0 ± 0.3]	[16.8 ± 0.2]				
LysM domain	2.6 ± 0.2	8.7 ± 0.3	0.91	0.91	0.98	0.70
	[2.7 ± 0.4]	[9.9 ± 0.2]				
water-soluble phospholamban	5.4 ± 0.6	14.8 ± 0.4	0.91	0.98	0.99	0.93
	[2.7 ± 0.5]	[16.4 ± 0.2]				
barnase−barstar	3.5 ± 0.2	15.7 ± 0.1	0.95	0.99	0.99	0.83
	[1.1 ± 0.2]	[17.2 ± 0.1]				

[a]RMSD: root mean square displacement in Ångstrom; R_g: gyration radius in Ångstrom; SI_{RMSF}, SI_{S^2}, SI_{bend}, and SI_{dihed} are respectively the similarity indexes between all-atom and CG representation for RMSF, S^2, bending and dihedral angles quantities. In square brackets are reported the all-atom values for RMSD and R_g in Å. See Figures S2−5, S12 and S13, and Table S6 (Supporting Information) for additional details.

calculated and the resulting structural features from CG MD simulations were compared with the previous specific CG parametrization. We used such a parametrization also to simulate protein structures that do not belong to the training set, namely L25 and B1 immunoglobulin-binding domain protein,[97,98] with PDB entries 1b75 and 1pgb, respectively.

Coarse-Grained Simulations and Structural Observables. The coarse-grained molecular dynamics simulations for all the proteins were performed with the MD suite of programs Lammps, in the canonical NVT ensemble using the Langevin thermostat and an integration time step of 5 fs. The values of the harmonic spring constants of the CG models dictate the most convenient time steps.[17,99] Calculating the ratio between the highest bond frequencies between the CG and atomistic potentials, we estimated a convenient value for the CG time step $\delta t_{CG,}$ which is up to is ≃4 times bigger than for all-atom MD. Therefore, not using any constraints on the bonded degrees of freedom,[100] we could conservatively integrate the equations of motion with a time step of 5 fs. The use of an algorithm like SHAKE on all bonds potential terms will presumably allow us to increase the time step to 10 fs. All systems were first progressively heated from 100 to 300 K for 0.5 ns, then equilibrated at this temperature for an additional 1 ns, and finally simulated for a production trajectories lasting between 100 ns and 1 μs. For Lennard-Jones interactions we used a cutoff of 15 Å, whereas for electrostatics interactions we used a cutoff of 50 Å.

On average, for the proteins under study, the computational gain using our CG model is in the order of 200 times, without considering further optimization of our routines. The computational gain has been calculated dividing the total number of hours needed to run 100 ns with the all-atom force-field by the total number of hours need to run 100 ns with our coarse-grained force-field. For the explicit solvent, the computational gain is in the order of 10 times.

For each simulated protein, we monitored structural fluctuations and electrostatics properties and compared the results from CG and atomistic MD simulations. Among the structural properties, we considered values of backbone bending and torsional angles, RMSD (root-mean-square deviation), gyration radius (R_g). Structural fluctuations were also considered as RMSF (root-mean-square fluctuations) and S^2 order parameters of backbone's dipoles. The S^2 order parameter quantifies the angular amplitude of N−H dipole internal motions, in our case quantification has been done for the backbone dipoles. We calculated the S^2 order parameter using the formula:[101]

$$S_i^2 = \frac{1}{2}[3\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3} <\mu_{i,\alpha}\mu_{i,\beta}>^2 -1] \quad (13)$$

where $\mu_{i,\alpha}$ (α = 1, 2, 3) are the x, y, and z components of the normalized backbone's dipole moment at all-atom and coarse-grained level. The presence of coarse-grained monopoles and dipoles allows for the comparison the electrostatic features at the two levels of resolution. The electrostatics potentials of each protein have been calculated using APBS,[102] and the results compared using the PIPSA 3.0 package.[103,104] We compared the all-atom results of RMSF, S^2, bending and dihedral curves with coarse-grained results model using the cosine similarity to which we will refer as a similarity index as it has been done in PIPSA,[103] calculated as the inner product space of two vectors. The cosine similarity is a measure of the similarity of two vectors of a inner product space.

RESULTS AND DISCUSSION

Structural and Electrostatic Coarse-Grained Properties for Different Protein Families. We tested our coarse-graining procedure using a set of proteins representative of distinct SCOP families. In particular, we simulated α_3W as an α-helical protein, Cox11 as a representative of full β proteins, and the LysM domain as a mixed α/β fold. The latter was also previously investigated by other CG approaches,[105] allowing for a cross-comparison with our method.

All the proteins simulated at the CG level conserved their fold in the microsecond time scale, showing a very good agreement between all-atom and coarse-grained description (Table 1, Figure 3). The secondary structures were conserved without the use of any additional *ad hoc* biases on the bending and torsional potential terms. Only minor discrepancies are observed on the loop regions connecting secondary structure elements like in α_3W and Cox11 (Figure 3A and B). The agreement between the backbone bending and torsional angles calculated at the two levels of resolution is very good (Table 1 and Figures S2 to S4 (Supporting Information)). The cosine similarities between all-atom and coarse-grained values for backbone bending angles are between 0.98 and 0.99, while for backbone dihedrals are between 0.70 and 0.93 (Table 1).

The RMSD values reach convergence in around 5−10 ns (similarly to atomistic MD), fluctuating to values as low as 3 Å for a simulation time up to 1 μs (Table 1 and Figure S5 A, B

dx.doi.org/10.1021/ct400137q | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

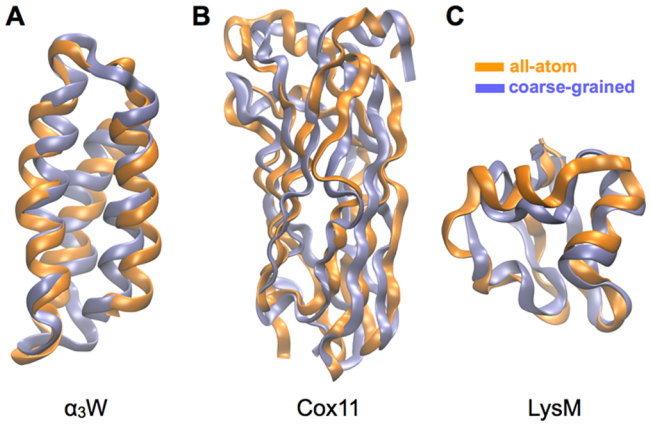


Figure 3. Structural comparison between all-atom and CG simulations of soluble proteins. Backbone superpositions of the structure obtained after 100 ns from all-atom (in orange) and CG (ice-blue) MD simulations for (A) the α_3W protein, (B) Cox11 protein, and (C) LysM domain protein. Relative RMSD and gyration radius values are reported in Table 1.

(Supporting Information)). The absolute values observed for RMSD are in line or lower than results reported using other CG models.[105] For instance, the LysM domain protein shows using our CG representation an RMSD as low as 2.6 Å, while simulations with the force-field OPEP 4.0[105] obtained a RMSD of 3.6 Å. The gyration radius is systematically slightly higher at all-atom level with respect to the coarse-grained representation, being the difference however in the order of 1 Å (Table 1). This slight collapse is likely to be intrinsically dependent on the coarse-grained representation, because the adopted mapping is not able to completely reproduce the steric effects of all the side chains, and buried cavities accommodating few water molecules cannot be filled by water beads having larger hindrance at CG granularity.

The general dynamic features are also in good agreement with the atomistic simulations. The RMSF calculated at CG level is systematically lower than for the all-atom one (see Figure 4), as already observed using other models.[50] The major differences are again on the loop regions. For example in the case of the α_3W protein the loops are composed by glycines that are very flexible, whereas our coarse-grained representation of the bending potential does not take into account in the current state a specific bending for glycines. RMSF peaks are not always quantitatively well reproduced but our model correctly reproduces the trends of the fluctuations. The similarity cosines of the RMSF calculated at all-atom and coarse-grained level are 0.95 for α_3W, 0.80 for the Cox11 and 0.91 for the LysM domain. The decrease in flexibility observed for the RMSF is confirmed also when calculating the S^2 order parameter of the backbone. Anyway, the agreement between the two levels of resolution is good: the similarity cosines of the S^2 calculated at all-atom and coarse-grained level are 0.94 for α_3W, 0.93 for the Cox11, and 0.91 for the LysM domain. We attribute the difference in flexibility observed for RMSF and S^2 (i) to the simple fact that at the coarse-grained level the lower number of degrees of freedom does not intrinsically allow the complete description of the structural fluctuations and (ii) to the potential form of the bending terms is not parametrized to be sequence-dependent but has a single general form α, β, and coil structures.

The monopole and dipolar terms for the backbone and side chains are able to well reproduce the electrostatic potential of the proteins. In fact, when comparing the CG and atomistic values of the electrostatic potential using PIPSA, we obtained quite high similarity indexes (Table 2, Figures S6–S8 (Supporting Information)).

Table 2. Summary of the Electrostatics Properties[a]

protein	$\|P_{AA}\|$	$\|P_{CG}\|$	SI_{ele}
α_3W	88.6	119.3	0.97
Cox11	50.6	55.4	0.99
LysM	39.0	44.2	0.99
water-soluble phospholamban	630.7	630.2	0.95
Barnase–Barstar	200.2	218.8	0.93
L25	47.3	48.3	0.97
B1 immunoglobulin-binding	63.8	94.6	0.93

[a] P (in eÅ) is the total electrostatic dipole moment for the entire protein, SI_{ele} is the similarity index between the all-atom and CG electrostatic potential as calculated with PIPSA. Compare with Figures S6, S7, S8, S14, S15, S22, and S23 (Supporting Information) for a 3D visualization of the electrostatic potential at the molecular surface calculated using APBS.

Along with the results using the distance-dependent model of implicit solvent we also obtained results with the explicit CG

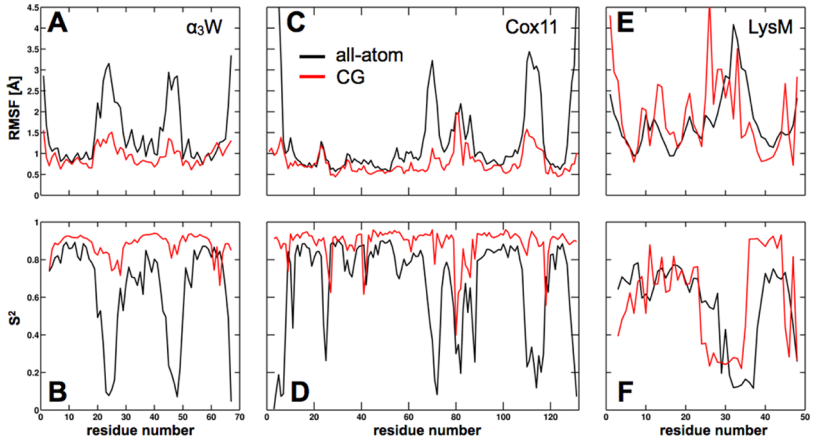


Figure 4. Comparison between dynamic properties of all-atom and CG MD simulations. RMSF and S^2 are reported for α_3W protein (A, B), Cox11 protein (C, D), LysM domain protein (E, F).

dx.doi.org/10.1021/ct400137q | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

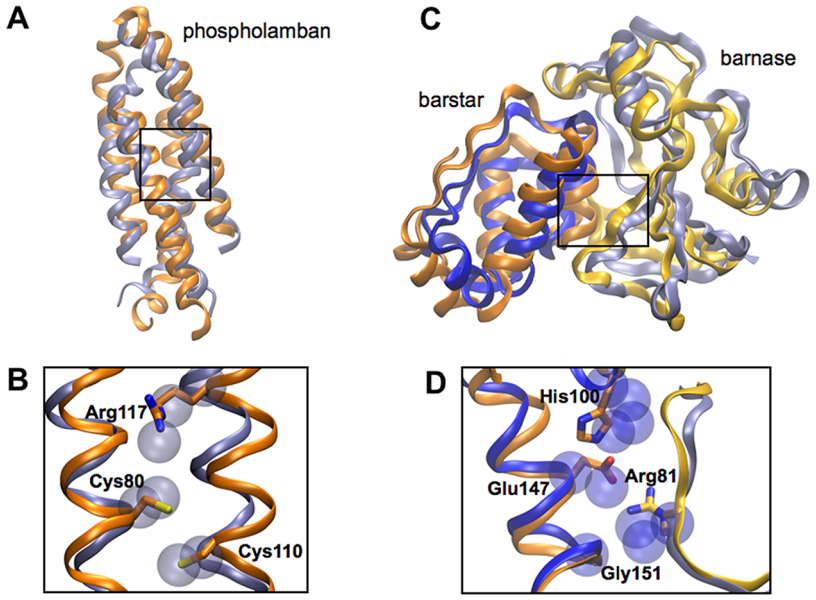


Figure 5. Structural comparison between all-atom and CG simulations of molecular complexes. Backbone superpositions of the structure obtained after 100 ns from all-atom and CG MD simulations for (A) water-soluble phospholamban and (C) barnase−barstar complex (color code as in Figure 3 apart for all-atom barnase in light orange, CG barnase in orange, all-atom barstar in blue, and CG barstar in ice-blue). (B) Interface of water-soluble phospholamban with all-atom residues in licorice representation and relative CG beads in transparent van der Waals representations. (D) Interface of barnase−barstar with all-atom residues in licorice representation and CG ones in transparent van der Waals representations. Relative RMSD and gyration radius values are reported in Table 1.

water model for the α_3W protein that are reported in the Supporting Information (Figures S9−S11). The RMSD is 2.5 ± 0.3 Å, whereas the gyration radius is 11.7 ± 0.1 Å, results that are in line with the all-atom ones. For the gyration radius, we did not observe the same type of collapse observed with the implicit solvent (Figure S9 B (Supporting Information)). A good agreement between all-atom and coarse-grained simulations with explicit solvent has been observed also for the others properties, namely, RMSF, S^2, backbone bending and dihedral (Figures S10−11 (Supporting Information)). This indicates that some of the drawbacks observed using the implicit solvent could be partially solved by using an explicit CG model for water.

Summarizing, we showed that it is possible to obtain, with a minimal amount of investment in terms of CPU time, a tailored parametrization for any soluble protein. The CG simulations produced results in very good agreement with the atomistic simulations and similar to results obtained using force fields adopting a comparable mapping topology;[105] moreover, the secondary structure elements are preserved during a microsecond time scale.

Structural and Electrostatic Coarse-Grained Properties for Protein−Protein Complexes. First, we studied the engineered soluble phospholamban engineered protein−protein complex, chosen because it is among the simplest not covalently bonded helix bundles.[92] The second complex investigated was the barnase−barstar complex,[94] which is in principle more challenging because it is composed by different secondary structure elements with different reciprocal arrangements.

As already obtained for the single proteins, the secondary structures were strongly conserved during CG simulations with only some discrepancies on the loop regions. Comparing the last structures obtained at the two levels of resolution, the percentage of conserved secondary structures was on average around 70%. The RMSD reached convergence after about 10 ns, in line with the atomistic simulations. The difference on the RMSD calculated at the two levels of resolution differ of about 2 Å (Table 1). In the case of the water-soluble phospholamban, the main structural differences are in correspondence of the protein termini, whereas in the case of the barnase−barstar complex similar differences are observed for both proteins composing the dimer. This is likely due to two main reasons: (i) the coarse-grained representation does not perfectly reproduce the steric effect of the side chains at the interface and (ii) the implicit solvent model does not allow for optimal solvation of the regions at the interface between the two proteins. This can be particularly relevant in the case that interstitial waters localize in the area (e.g., CG water models would have the same problem), and it can lead to the exploration of slightly different conformations at the interface. The gyration radius results confirm this hypothesis because the difference between the all-atom and the coarse-grained values are in the order of 1.5 Å (Table 1).

Nonetheless, the protein−protein interfaces are well conserved for both complexes and key electrostatic interactions are mainly preserved. For example, in the water-soluble phospholamban−electrostatics interactions that stabilize the complex such as Cys80−Cys110 and Cys80−Arg117 are maintained (Figure 5B). In the case of the barnase−barstar complex interactions at the interface are also well preserved reproducing the majority of the contacts (e.g., Arg81 (barnase)−Asp147 (barstar), His100 (barnase)−Asp147 (barstar), Arg81 (barnase)−Gly151 (barstar), see Figure 5D and Table S4−S6 (Supporting Information)).

The CG RMSF and S^2 values are in good agreement with all-atom MD results (Figure 6). The structural fluctuation properties are preserved, and also the secondary structure elements, as seen from the superposition of the last structures obtained at the two levels of resolution (Figure 5). This is

dx.doi.org/10.1021/ct400137q | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

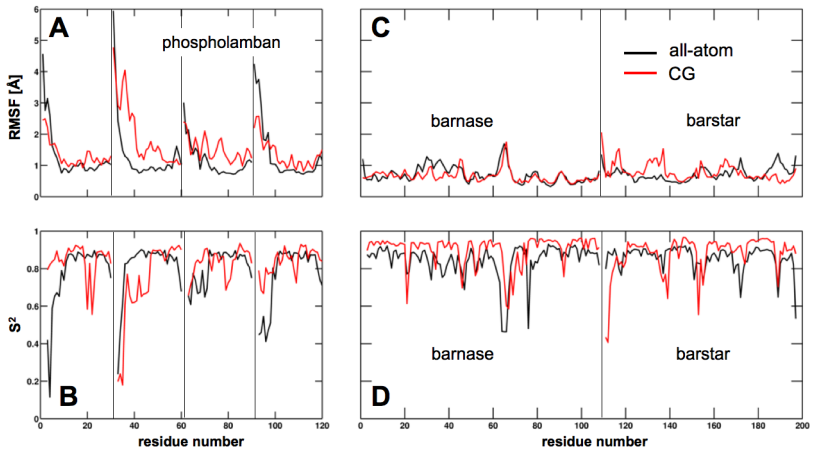


Figure 6. Comparison between dynamic properties of all-atom and CG MD simulations of molecular complexes. RMSF and S^2 are respectively reported for water-soluble phospholamban (A, B) and barnase−barstar complex (C, D).

Table 3. Summary of the Structural and Dynamic CG Properties Using a Generalized CG Parameterization[a]

protein	RMSD	R_g	SI_{RMSF}	SI_{S^2}	SI_{bend}	SI_{dihe}
α_3W	2.4 ± 0.2	11.1 ± 0.2	0.92	0.93	0.99	0.85
	[2.5 ± 0.2]	[12.2 ± 0.2]				
Cox11	3.0 ± 0.4	15.1 ± 0.2	0.81	0.94	0.98	0.73
	[3.0 ± 0.3]	[16.8 ± 0.2]				
LysM domain	2.7 ± 0.3	8.8 ± 0.1	0.95	0.93	0.99	0.83
	[2.7 ± 0.4]	[9.9 ± 0.2]				
water-soluble phospholamban	4.5 ± 0.4	16.3 ± 0.2	0.93	0.98	0.99	0.97
	[2.7 ± 0.5]	[16.4 ± 0.2]				
barnase−barstar	3.4 ± 0.2	15.9 ± 0.1	0.93	0.98	0.99	0.82
	[1.1 ± 0.2]	[17.2 ± 0.1]				
L25	4.2 ± 0.3	11.8 ± 0.2	0.87	0.95	0.98	0.66
	[3.5 ± 0.8]	[13.1 ± 0.3]				
B1 immunoglobulin-binding	3.4 ± 0.2	9.7 ± 0.1	0.88	0.97	0.98	0.79
	[1.1 ± 0.3]	[10.4 ± 0.1]				

[a]RMSD: root mean square displacement in Ångstrom; R_g: gyration radius in Ångstrom; SI_{RMSF}, SI_{S^2}, SI_{bend}, and SI_{dihe} are respectively the similarity indexes between all-atom and CG representation for RMSF, S^2, bending and dihedral angles quantities. In square brackets are reported the all-atom values for RMSD and R_g in Å. See Figures S16-21 and Table S7 (Supporting Information) for additional details.

strengthened by the agreement of the backbone bending and torsional angles for the two complexes (Table 1, Figures S12−13 (Supporting Information)). Also in this case, we find a very good agreement between the electrostatic potentials calculated at the atomistic and CG levels of resolution: the similarity indexes calculated with PIPSA are 0.95 and 0.93 for the water-soluble phospholamban and the barnase−barstar complex, respectively (Table 2, Figures S14−15 (Supporting Information)).

Toward a Transferable Coarse-Grained Force Field for Proteins. In the presented simulations, specific sets of parameters were optimized each time for each protein under study. Although our results show very good accuracy with respect to all-atom simulations and parametrization is very efficient, such approach lacks full transferability, as it requires reoptimization of some of the force-field terms every time a new system is studied. In order to test the possibility of extending our approach toward the definition of a fully transferable potential, we implemented an averaged force-field potential. This new set of parameters was then used to run CG molecular dynamics of the five systems in the training set for 100 ns, and L25 and B1 immunoglobulin-binding proteins, two additional systems not included in the set employed for calibration, for 1 μs.

We obtained a good agreement for the CG simulations of the proteins from the training set (Table 3). In particular, the percentage of preserved secondary structure elements for the proteins that belong to the training set are in line with what has been obtained with specific parametrizations (Table 3). The systems not included in the training set were instead described with slightly lower accuracy but in line with the results obtained for the others proteins (Table 3). Secondary structure elements are preserved, the main differences being on loop regions (Figures S16−17 (Supporting Information)).

Superposition of the last structures obtained at the two levels of resolution gives RMSD values for L25 and B1 immunoglobulin-binding of 4.2 Å and 3.9 Å, respectively, whereas the percentage of preserved secondary structure elements is around 70% for both for the last structures for the two level of resolution (Figures S18−19 (Supporting Information)). RMSD and gyration radius (Table 3) obtained using the general CG force field are very similar to the relative atomistic values.

Both L25 and B1 immunoglobulin-binding proteins have been recently used to test two CG force-fields, that is, the one

dx.doi.org/10.1021/ct400137q | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

developed by Ha-Duong[50] and OPEP 4.0.[105] An RMSD of 6 Å and 2.9 Å when using the former and the latter model, respectively, was reported for L25. For the OPEP 4.0 model, it should be noted that only some parts of the protein have been selected for the calculation of the RMSD. B1 immunoglobulin-binding protein showed instead an RMSD of around 4 Å and 3.3 Å, respectively. For both these proteins, our coarse-grained approach performs reasonably well in the microsecond time scale, presenting always a discrepancy of around 1.5 Å with respect to the atomistic value for the RMSD and of around 1 Å for the gyration radius (Table 3).

The agreement of structural fluctuations with all-atom values is lower than for specific parametrizations (Table 3 and Table S7 (Supporting Information)); however, the similarity index for the properties calculated at the two levels of resolution for L25 and B1 immunoglobulin-binding proteins are quite high (Tables 2 and 3, Figures S20−23 (Supporting Information)), showing that this preliminary version of a general set of electrostatic-consistent CG potentials goes in the right direction toward the development of a reliable transferable CG force field.

CONCLUSIONS

In this work, we presented a coarse-graining procedure to generate potentials for molecular simulation of soluble proteins incorporating an explicit description of electrostatics. The adopted strategy for the parametrization of the coarse-grained potentials is based on Boltzmann inversion and a force-matching scheme relying on high-resolution protein structures and atomistic simulations. The derivation of the parameters is obtained using a new and robust global optimization algorithm based on particle swarm optimization that handles the assignment of several hundreds parameters in a relatively short amount of time, allowing the CG parametrization of large molecular systems.

The combination of electrostatic terms at the backbone and polar side-chains to terms accounting for the steric hindrance of the CG beads produces stable protein tertiary structures, and maintains the global fold of a variety of soluble proteins. Our approach produces also dynamically stable quaternary complexes, as in the case of the phospholamban and barnase−barstar systems. Despite the intrinsic limitations of any coarse-grained representation, the CG potentials generated by our procedure preserve a very good and consistent agreement with all-atom simulations, well reproducing the main structural, dynamic, and electrostatic properties. Importantly, these CG models are also able to describe the main interface interactions, producing stable protein complexes. Although not explicitly tested for all the existing folding families, we expect that the derivation of parameters obtained using this strategy would be as accurate for other proteins at this level of granularity.

These results are promising and suggest that electrostatic-consistent CG potentials can be efficiently used to explore protein−protein molecular recognition using molecular dynamics sampling. Our results are, in fact, in good agreement with all-atom simulations and, when directly compared with previously reported CG force fields, showed similar or better performances in describing structural and dynamic determinants of soluble proteins. Moreover, our procedure can be straightforwardly extended for the parametrization of any protein. The extension of this optimization procedure to a larger data set may prelude to the generation of a fully transferable CG force field that will be applied in principle to any protein or, more interestingly, any large macromolecular assembly for which direct, long all-atom simulations may not be easily affordable.

ASSOCIATED CONTENT

Supporting Information

Additional details are reported for the CG model, convergence of PSO-based parametrization, electrostatic maps, superpositions of CG and atomistic structures, comparison between atomistic and CG RMSD, S^2, bending, dihedral values. This information is available free of charge via the Internet at http://pubs.acs.org/

AUTHOR INFORMATION

Corresponding Author

*E-mail: matteo.dalperaro@epfl.ch.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Dr. Luciano Abriata for proofreading the manuscript and discussions; Dr. Marco Stenta and Dr. Francesca Collu for useful discussions; Thomas Lemmin and Christophe Bovigny for technical support. This work was supported by the Swiss National Science Foundation (SNSF) (Grant No. 200021_122120, 200020_138013 (M.D.P.), and PP00P2_139195 (M.C.)).

REFERENCES

(1) Russel, D.; Lasker, K.; Phillips, J.; Schneidman-Duhovny, D.; Velazquez-Muriel, J. A.; Sali, A. *Curr. Opin. Cell Biol.* **2009**, *21*, 97−108.
(2) Schlick, T.; Collepardo-Guevara, R.; Halvorsen, L. A.; Jung, S.; Xiao, X. *Q. Rev. Biophys.* **2011**, *44*, 191−228.
(3) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; K., S. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(4) Shaw, D. E. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC09)*, Portland, OR, 2009.
(5) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. *J. Comput. Chem.* **2007**, *28*, 2618−2640.
(6) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, C. W. *J. Chem. Theory Comput.* **2012**, *8*, 2997−3002.
(7) Shaw, D. E.; et al. *Science* **2010**, *330*, 341−346.
(8) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.
(9) Sanbonmatsu, K. Y.; Tung, C. S. *J. Struct. Biol.* **2007**, *157*, 470−480.
(10) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. *Structure* **2006**, *14*, 437−449.
(11) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 94−98.
(12) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59−107.
(13) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. *Acta Polym.* **1998**, *49*, 61−74.
(14) Goetz, R.; Gompper, G.; Lipowsky, R. *Phys. Rev. Lett.* **1999**, *82*, 221−224.
(15) Venturoli, M.; Smit, B. *PhysChemComm* **1999**, *10*, 45−49.
(16) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464−4470.
(17) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423−12430.
(18) Klein, L. M.; Shinoda, W. *Science* **2008**, *321*, 798−800.
(19) Kamerlin, S. C. L.; Warshel, A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10401−10411.

(20) Ayton, S. G.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192−198.
(21) Saunders, M. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2012**, *22*, 1−7.
(22) Tozzini, V. *Q. Rev. Biophys.* **2010**, *43*, 333−371.
(23) Cascella, M.; Dal Peraro, M. *CHIMIA* **2009**, *63*, 14−18.
(24) Voth, G. A. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca Raton, FL, 2009; pp 85−109.
(25) Roux, B. In *Multiscale Approaches to Protein Modeling*; Springer: New York, 2011; pp 85−109.
(26) Pasi, M.; Lavery, R.; Ceres, N. *J. Chem. Theory Comput.* **2013**, *9*, 785793.
(27) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. *J. Chem. Theory Comput.* **2010**, *6*, 1711−1725.
(28) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. *J. Chem. Theory Comput.* **2010**, *6*, 3793−3807.
(29) Nielsen, S. O.; Moore, P. B.; Ensing, B. *Phys. Rev. Lett.* **2010**, *105*, 237802.
(30) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(31) Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 1−11.
(32) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323−2357.
(33) Mullinax, J. W.; Noid, W. G. *Phys. Rev. Lett.* **2009**, *103*, 1−4.
(34) Scott Shell, M. *J. Chem. Phys.* **2008**, *129*, 1−7.
(35) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.
(36) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J. W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 1−20.
(37) Thorpe, I. F.; Zhou, J.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 13079−13090.
(38) Ayton, G. S.; Voth, G. A. *Biophys. J.* **2010**, *99*, 2757−2765.
(39) Hills, D. R. J.; Lu, L.; Voth, G. A. *PLoS Comput. Biol.* **2010**, *6*, 1−12.
(40) Thorpe, I. F.; Goldenberg, D. P.; Voth, G. A. *J. Phys. Chem. B* **2011**, *115*, 11911−11926.
(41) Mullinax, J. W.; Noid, W. G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19867−19872.
(42) Carmichael, S. P.; Scott Shell, M. *J. Phys. Chem. B* **2012**, *116*, 8383−8393.
(43) Larini, L.; Lu, L.; Voth, G. A. *J. Chem. Phys.* **2010**, *132*, 1−10.
(44) Sinitskiy, A. V.; Saunders, M. G.; Voth, G. A. *J. Phys. Chem. B* **2012**, *116*, 8363−8374.
(45) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(46) Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. *J. Am. Chem. Soc.* **2007**, *129*, 10126−10132.
(47) Louhivuoiri, M.; Risselada, H. J.; van der Giessen, E.; Marrink, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19856−19860.
(48) Schaefer, L. V.; de Jong, D. H.; Holt, A.; Rzepiela, A. J.; de Vries, A. H.; Poolman, B.; Killian, J. A.; Marrink, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1343−1348.
(49) van den Bogaart, G.; Meyenberg, K.; Risselada, H. J.; Amin, H.; Willing, K. I.; Hubrich, B. E.; Dier, M.; Hell, S. W.; Grubmüller, H.; Diederichsen, U.; Jahn, R. *Nature* **2011**, *479*, 552−555.
(50) Ha-Duong, T. *J. Chem. Theory Comput.* **2009**, *5*, 3211−3223.
(51) Scheraga, H. A. *Annu. Rev. Biophys.* **2011**, *40*, 1−39.
(52) Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga, H. A. *J. Chem. Theory Comput.* **2009**, *5*, 627−640.
(53) Gołas′, E.; Maisuradze, G. G.; Senet, P.; Ołdziej, S.; Czaplewski, C.; Scheraga, H. A.; Liwo, A. *J. Chem. Theory Comput.* **2012**, *8*, 1750−1764.
(54) Gopal, S. M.; Mukherjee, S.; Cheng, Y. M.; Feig, M. *Proteins* **2009**, *78*, 1266−1281.
(55) Derreumaux, P. *J. Chem. Phys.* **1999**, *111*, 2301−2310.
(56) Maupetit, J.; Tuffery, P.; Derreumaux, P. *Proteins* **2007**, *69*, 394−408.
(57) Barducci, A.; Bonomi, M.; Derreumaux, P. *J. Chem. Theory Comput.* **2011**, *7*, 1928−1934.
(58) Zacharias, M. *Protein Sci.* **2003**, *12*, 1271−1282.
(59) Zacharias, M. *Proteins* **2013**, *81*, 81−92.
(60) Alemani, D.; Collu, F.; Cascella, M.; Dal Peraro, M. *J. Chem. Theory Comput.* **2010**, *6*, 315−324.
(61) Cascella, M.; Neri, M. A.; Carloni, P.; Dal Peraro, M. *J. Chem. Theory Comput.* **2008**, *4*, 1378−1385.
(62) Besozzi, D.; Cazzaniga, P.; Mauri, G.; Pescini, D.; Vanneschi, L. *Evolutionary Computation, Machine Learning, and Data Mining in Bioinformatics*; Springer, 2009116127
(63) Elbeltagi, E.; Hegazy, T.; Grierson, D. *Adv. Eng. Inf.* **2005**, *19*, 43−53.
(64) Namasivayam, V.; Günther, R. *Chem. Biol. Drug Des.* **2007**, *70*, 475−484.
(65) Angeline, P. *Evolutionary Programming VII*; Springer: New York, 1998; pp 601−610.
(66) Abraham, A.; Liu, H. *Foundations of Computational Intelligence*, Volume 3; Springer: New York, 2009; pp 291−312.
(67) DeVane, R.; Shinoda, W.; Moore, P. B.; Klein, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 2115−2124.
(68) Tozzini, V.; Rocchia, W.; McCammon, J. A. *J. Chem. Theory Comput.* **2006**, *2*, 667−673.
(69) Rubinstein, A.; Sherman, S. *Biophys. J.* **2004**, *878*, 15441557.
(70) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem.* **2000**, *104*, 6478−6489.
(71) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583−5595.
(72) Ha-Duong, T.; Phan, S.; Marchi, M.; Borgis, D. *J. Chem. Phys.* **2002**, *117*, 541−556.
(73) Ha-Duong, T.; Basdevant, N.; Borgis, D. *Chem. Phys. Lett.* **2009**, *468*, 79−82.
(74) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Comput. Chem.* **2004**, *25*, 1015−1029.
(75) Basdevant, N.; Ha-Duong, T.; Borgis, D. *J. Chem. Theory Comput.* **2006**, *2*, 1646−1656.
(76) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Chem. Theory Comput.* **2013**, *9*, 803−813.
(77) Sutmann, G. *NIC Series* **2002**, *10*, 211−254.
(78) Chipot, C.; Ángyán, J. G.; Maigret, B.; Scheraga, H. A. *J. Chem. Phys.* **1993**, *97*, 9788−9796.
(79) Fogolari, F.; Esposito, G.; Viglino, P.; Cattarinussi, S. *Biophys. J.* **1996**, *70*, 1183−1197.
(80) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1−19.
(81) Ercolessi, F.; Adams, J. B. *Europhys. Lett.* **1994**, *26*, 583−588.
(82) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211−3223.
(83) Kennedy, J.; Eberhart, R. *IEEE International Conference on Neural Networks, Proceedings*; IEEE: New York, 1995; pp 1942−1948.
(84) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(85) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(86) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 712−725.
(87) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(88) Price, D. J.; Brooks, C. L. *J. Chem. Phys.* **2004**, *121*, 10096−10103.
(89) Dai, Q. H.; Tommos, C.; Fuentes, J. E.; Blomberg, M. R. A.; Dutton, P. L.; Wand, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 10952−10953.
(90) Banci, L.; Bertini, I.; Cantini, F.; Ciofi-Baffoni, S.; Gonnelli, L.; Mangani, S. *J. Biol. Chem.* **2004**, *279*, 34833−34839.
(91) Bateman, A.; Bycroft, M. *J. Mol. Biol.* **2000**, *209*, 1113−1119.
(92) Slovic, A. M.; Stayrook, S. E.; North, B.; DeGrado, W. F. *J. Mol. Biol.* **2005**, *348*, 777−787.
(93) Slovic, A. M.; Summa, C. M.; Lear, J. D.; DeGrado, W. F. *Protein Sci.* **2003**, *12*, 337−348.

(94) Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8878−8889.
(95) Jarvis, R. A.; Patrick, E. A. *IEEE Trans. Comput.* **1973**, *C-22*, 1025−1034.
(96) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindhal, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(97) Stoldt, M.; Wöinert, J.; Görlach, M.; Brown, L. R. *EMBO J.* **1998**, *17*, 6377−6384.
(98) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. *Biochemistry* **1994**, *33*, 4721−4729.
(99) Riniker, S.; van Gunsteren, W. F. *J. Chem. Phys.* **2011**, *134*, 1−12.
(100) Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(101) Smith, P. E.; van Schaik, R. C.; Szyperski, T.; Wüthrich, K.; van Gunsteren, W. F. *J. Mol. Biol.* **1995**, *246*, 356−365.
(102) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(103) Blomberg, N.; Gabdoulline, R. R.; Nilges, M.; Wade, R. C. *Proteins Struct. Funct. Bioinf.* **1999**, *37*, 379−387.
(104) Wade, R. C.; Gabdoulline, R. R.; De Rienzo, F. *Int. J. Quantum Chem.* **2001**, *83*, 122−127.
(105) Chebaro, Y.; Pasquali, S.; Derreumaux, P. *J. Phys. Chem. B* **2012**, *116*, 8741−8752.