JCTC Journal of Chemical Theory and Computation

Topologically Based Multipolar Reconstruction of Electrostatic Interactions in Multiscale Simulations of Proteins

Michele Cascella,*,†,⊥ Marilisa A. Neri,† Paolo Carloni,§,|| and Matteo Dal Peraro‡

Laboratory of Computational Chemistry and Biochemistry and Laboratory for Biomolecular Modeling, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, International School for Advanced Studies (SISSA/ISAS) and CNR-INFM-DEMOCRITOS, I-34014 Trieste, Italy, and Italian Institute of Technology, Italy

Received April 07, 2008

Abstract: We present a new method to incorporate electrostatic interactions in coarse-grained representations of proteins. The model is based on a topologically reconstructed multipolar expansion of the all-atom centers of charge, specifically of the backbone dipoles and the polar or charged side chains. The reliability of the model is checked by studying different test cases, namely protein−cofactor/substrate interactions, protein large conformational changes, and protein−protein complexes. In all cases, the model quantitatively reproduces the all-atom electrostatic field in both a static and a dynamic framework. The model is of general applicability and can be used to improve both full coarse-grained simulations and hybrid all-atom/coarse-grained multiscale approaches.

1. Introduction

In past years, atomistic-detailed molecular dynamics (MD) simulations proved to be a reliable method for description of events occurring in the $10^{-12}-10^{-9}$ s time scales.[1] Although the increase in computational power allows application of such technique to larger systems, and for longer times, it remains in general too expensive to investigate dimensional and dynamical scales that are critical to most of the biological processes both *in vitro* and *in vivo*. In fact, all fundamental biological processes necessary to life (e.g., protein folding, signal transduction, DNA transcription), which are triggered by interactions at atomistic dimensionality, occur at very different time scales (from femtoseconds to seconds and even longer) and span over different sizes (from few tens to millions of atoms).[2] Therefore, vast dimensional and temporal scales should be taken into account.

To address such size/time scale issues various coarse-grained (CG) Hamiltonians have been recently developed for systems like membranes,[3–8] proteins,[9–16] and DNA[17–19] and were successfully applied to problems of great biological relevance, from membrane self-assembly and dynamics [5,7,8] to protein folding.[20–22] The drawback of CG approaches lies in the lack of an atomistic-detailed description, which is crucial when studying phenomena involving molecular recognition (e.g., receptor−ligand binding). These interactions are of fundamental importance for establishing functional protein networks entitled to control complex metabolic pathways.[2] On the experimental side, proteomics techniques are progressing very fast in enabling the characterization of such challenging systems; however, what is still missing for a complete comprehension of the biological function is the atomistic picture of these assemblies. The only accessible way to get to this level, having the opportunity to develop

* Corresponding author e-mail: michele.cascella@epfl.ch.

† Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne.

⊥ Present Address: Universität Bern, Department für Chemie und Biochemie, Freiestrasse 3, CH-3012 Bern, Switzerland.

§ International School for Advanced Studies (SISSA/ISAS) and CNR-INFM-DEMOCRITOS.

‡ Laboratory for Biomolecular Modeling, Ecole Polytechnique Fédérale de Lausanne.

|| Italian Institute of Technology.

10.1021/ct800122x CCC: $40.75
Published on Web 07/12/2008

strategies to interfere with these networks, is through current structural biology techniques. For this reason, the possibility to computationally predict the molecular interactome is very desirable for the advance of life sciences.[23–25]

Recently, considerable efforts have been put into development of novel multiscale simulation techniques that are able to couple atomistic Hamiltonians to CG models (MM/CG).[8,26–41] In particular, it has been shown that hybrid schemes, where only a portion of the system is treated at the atomistic-detailed level, can be implemented and successfully applied to proteins[30,42–44] and protein-DNA complexes.[29,46] A key issue when building hybrid multiscale models lies in retaining an accurate description of long-range interactions. Such long-range interactions can significantly influence the polarization and the structural orientation at an atomistic level and, thus, can be functional to the protein biological activity.[47] Therefore, hybrid MM/CG schemes should somehow incorporate such potentials between the CG part and the all-atom portion. In particular, long-range electrostatic interactions, crucial in many biological processes, are the most difficult to model. In fact, the electrostatic field of biopolymers is dominated by their total charge, which is usually nonzero, and varying with pH and ionic strength, and by a dipolar component, which is a function of both the residue sequence and its instantaneous spatial organization. Specifically, the secondary structure elements of a folded protein can induce strong dipolar fields by aligning the backbone dipoles (e.g., in helical structures). This feature is of great importance, for example, when assessing the topology of a helix-bundle or when determining the correct orientation of a molecule during dynamic ligand−receptor binding studies. Therefore, CG long-range interactions should be modeled to have a structure-dependent asymptotic behavior. Such behavior is in general nontrivial to reproduce by force field-like CG models.[9] For example, CG Hamiltonians based on effective potentials (typically, potentials of mean force) vanish as a power of six with the distance.[3,4]

Recently developed coarse-grained potentials, where specific polar regions are introduced (see e.g. refs 48 and 49), have been proposed. Interactions among polar groups still take a radial form, and, therefore, they are not able to reproduce the directionality of the backbone dipolar field. Therefore, to represent the different behavior of the backbone in different structural local environments, the polarity of the backbone centroid is defined according to the secondary structure of the amino acid. Such an approach is able to reproduce well the secondary structure elements of a protein but, nonetheless, biases the potential at the beginning of the run, forcing each amino acid to keep the initial secondary structure geometry and thus reducing its universality. A RESP[50] expansion of the all-atom electrostatic field on all the CG centroids represents an alternative approach to such a problem.[16] This scheme is particularly suited to improve CG-based protocols for rigid docking. Unfortunately, such an approach cannot be straightforwardly applied when treating hybrid multiscale systems, where different dimensionalities are present and interacting in the simulation. In particular, MD schemes based on RESP-based expansions of all-atom based electrostatic potentials on CG centroids suffer some inconsistencies, as shown later in this paper.

In the past few years, Voth and co-workers have shown that coarse-grained potentials can be rigorously determined from atomistic scale interactions through a force-matching approach.[51] Such a procedure implies an initial all-atom classical MD run, over which the CG potential is fitted to reproduce all-atom forces.[33,34] The CG Hamiltonian obtained by such a scheme is not based on thermodynamic quantities and, therefore is, in general, unbiased toward the specific simulated thermodynamic conditions, and it is in principle able to correctly describe long-range interactions. The method has been successfully applied to construct models of lipid bilayers,[33,52] cholesterol-lipid interactions,[34] monosaccharides,[53] and distinct peptide structural moieties.[54,55] In practice, the quality of such a coarse-graining procedure is determined by the efficacy of the underlying all-atom MD to sample the relevant phase space of the system.[34] This may be more stringent when studying large macromolecular complexes, which present long characteristic time scales.

In this work, we propose an accurate and robust scheme to evaluate electrostatic interactions in hybrid MM/CG simulations of proteins.[30] The proposed model uses topological information coming from experimental structures and from computational molecular models to reconstruct the electrostatic field for both contributions coming from the backbone and the side chains of a protein. Our method is not computationally expensive and is consistent with respect to protein motion. Moreover, it allows a reliable reconstruction of the electrostatic interactions both at an atomistic and a CG scale, allowing applications to different biological systems for the efficient investigation of crucial processes as ligand binding recognition or protein and nucleic acids network dynamics.

2. Computational Methods

The electrostatic field of a biomolecule is univocally determined by its tridimensional structure. In proteins, the electrostatic potential can be further divided into two separate contributions, coming from the backbone peptide bonds (V_{bb}) and from the polar or charged side chains (V_{sc}):

$$V_{el} = V_{bb} + V_{sc} \tag{1}$$

The presented model reconstructs the all-atom electrostatic field of a protein using the tridimensional coordinates of the C_α atoms only, which is some structural information present in most of the CG models for proteins. It will be referred to as *topologically based multipolar reconstruction* (TBMR) hereafter.

2.1. Backbone Contribution. The electrostatic potential generated by the backbone units at point $\vec{R}$ can be expressed as a sum of dipolar contributions

$$V_{bb}(\vec{R}) = \frac{1}{4\pi\varepsilon_0}\sum_i \frac{\vec{\mu}_i \cdot \vec{R}}{|\vec{R}|^3} \tag{2}$$

where the *-i* index runs over the backbone units of the protein, and $\{\vec{\mu}_i\}$ are the electrostatic dipoles associated with them. The position of the C_β of an amino acid can be

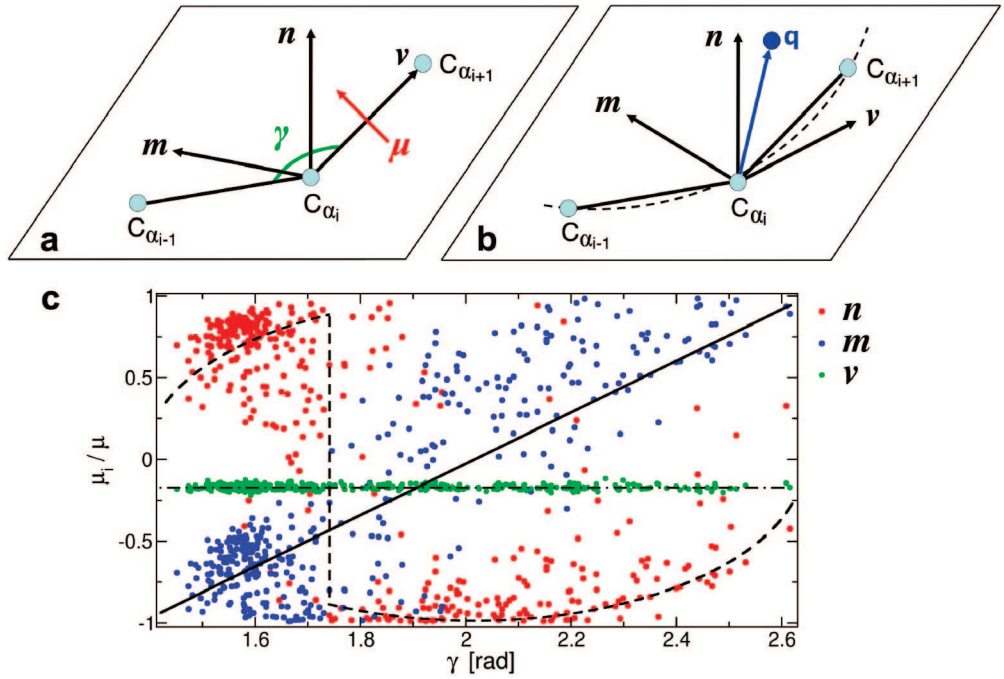

Figure 1. Definition of the internal reference systems. Panels **a** and **b**: internal reference systems used for the backbone and side-chain reconstruction, respectively. Panel **c**: statistical distribution (extracted from the PDB databank) of the projection of the backbone dipole over the vectors *v* (in green), *n* (in red), and *m* (in blue). The black lines are representing the proposed fits for the distributions.

determined, within good approximation, starting from the positions of three consecutive C_α's in the amino acids sequence.[56] Such an idea can be extended with little effort to the orientation of the peptide bond linking two amino acids. In fact, the positions of three consecutive C_α's, their chirality, and the planarity of the peptide moiety provide, in principle, a sufficient number of geometrical constraints to determine the peptide bond geometry. In practice, a good orientation of the backbone peptides can be conveniently achieved by mining statistical orientations from the PDB databank.[57] First, one has to define an internal reference system (Figure 1a) related to three consecutive amino acids. In particular, the following system is chosen: the origin is placed at C_{α_i}, the α-carbon of the second amino acid of the triplet; the three axes are defined by the unitary vectors: $\hat{\mathbf{v}}$ is the direction C_{α_i}-$C_{\alpha_{i+1}}$, $\hat{\mathbf{n}}$ is the normal to the plane defined by the three consecutive $C_{\alpha_{i-1}}$-C_{α_i}-$C_{\alpha_{i+1}}$ C_α's, and $\hat{\mathbf{m}}$ is perpendicular to the first two (Figure 1a). In such a reference system, the electrostatic dipole $\vec{\mu}_i$, associated with the peptide bond between amino acids *i* and *i*+1 is well defined: in fact, the projection of $\vec{\mu}_i$ over $\hat{\mathbf{v}}$ is, by construction, constant; the projection of $\vec{\mu}_i$ over $\hat{\mathbf{m}}$ can be fitted by a linear function of the angle γ formed by the $C_{\alpha_{i-1}}$-$C_{\alpha_{i+1}}$ atoms (Figure 1); and finally, the projection of $\vec{\mu}_i$ over $\hat{\mathbf{n}}$ is determined according to the cosine rule

$$\mu_n = \pm|\vec{\mu}_i|\left[1-\left(\frac{\mu_v}{|\vec{\mu}_i|}\right)^2-\left(\frac{\mu_m}{|\vec{\mu}_i|}\right)^2\right]^{1/2} \tag{3}$$

where $\mu_{n,m,v}$ are the projections of the dipole along the $\hat{\mathbf{n}}$, $\hat{\mathbf{m}}$, and $\hat{\mathbf{v}}$ axes, respectively. The sign of μ_n can be derived by the statistical distribution in the PDB databank (Figure 1c). In our model, it is defined as positive for $\gamma < 1.72$ *rad* and negative for $\gamma \geq 1.72$ *rad*. Following such a scheme, the electrostatic field produced by the backbone peptides in a CG region is easily reconstructed by simply determining the angles formed by all triplets of consecutive C_α's in the protein sequence.

2.2. Side-Chain Contribution. CG models usually lack structural information about the side-chain position and orientation. Therefore, it is necessary to use a serial multiscale procedure to build it. Starting from an initial information (e.g., X-ray or NMR data), we make the plausible assumption that, in a folded protein, the close packing forces the side-chain orientations to remain, on average, unchanged over time.

In this case, one can determine the side-chain positions at any time by (i) imposing that the motion of the side chain is solidal to a properly chosen internal coordinate system (defined below) and (ii) calculating the instantaneous rototranslations undergone by such an internal reference system.[58] Similarly to what is done for the backbone dipoles, the internal reference system (Figure 1b) for the side chain of an *i*th amino acid is defined according to the positions of the $C_{\alpha_{i-1}}$-C_{α_i}-$C_{\alpha_{i+1}}$ triplets. In this case, the unitary vector $\hat{\mathbf{v}}$ is defined as the vector tangent in C_{α_i} to the circle defined by the three consecutive C_α's[58] (Figure 1b).

The electrostatic potential produced by each side chain can be expressed by a multipolar expansion

$$V_{sc}(\mathbf{R}) = \frac{1}{4\pi\varepsilon_0}\left[\frac{q}{|\mathbf{R}|}+\frac{\vec{\mu}\cdot\mathbf{R}}{|\mathbf{R}|^3}+O(quadrupole)+...\right] \tag{4}$$

where $q = \sum_i q_i$ is the total charge of each side chain. The geometrical center of the single multipolar expression is defined as the center of charge $\mathbf{C}_q$ of the side chain

$$\mathbf{C}_q = \frac{\sum_i |q_i| \mathbf{r}_i}{\sum_i |q_i|} \quad (5)$$

where $\mathbf{r}_i$ are the coordinates of the side-chain atoms, and q_i are their respective MM charges. The single centers of charge and the respective multipolar contributions to the electrostatic potential are then defined with respect to their relative internal coordinates ($\hat{\mathbf{v}}$, $\hat{\mathbf{n}}$, $\hat{\mathbf{m}}$). All these data have to be evaluated on the initial all-atom structure. In practice, it is more convenient to get this set of data from an average of structures, e.g. generated from a MD run. In such a case, the multipolar expansion will be averaged on the typical mobility of each side chain. Thus, it will represent the electrostatic mean-field produced by the side chains.

In the current work, only highly polar and charged side chains are explicitly considered, and their multipolar expansion is truncated to the dipole term. This is due to the particular form of the electrostatic potential of the GROMOS96 force field,[59] which is the one implemented in the MM/CG Hamiltonian developed and used in this work. In fact, the charge-fitting protocol used in GROMOS96 makes use of charge-groups, which, in turn, make the all-atom electrostatic potential dominated by dipolar terms. Moreover, in the GROMOS96 force field, apolar/aliphatic side chains do not contribute to the electrostatic potential.

The reconstruction protocol is anyway general and can be applied with little effort to other force fields, although expansion to higher order terms (e.g., quadrupoles) may be necessary to accurately describe the electrostatic field.

The computational advantage of our scheme is directly related to the reduced number of interactions to be computed. In fact, within such a protocol, the total number of dipoles corresponds to the number of peptide bond moieties (≈number of amino acids) plus the number of polar side chains (statistically ≈0.5 times the number of amino acids).[57] Given that the total number of atoms in a protein is statistically ≈15.7 times larger than its number of amino acids,[57] our scheme is at least ≈9.5 times faster in evaluating solute−solute interactions. Such an estimate is done considering electrostatic interaction evaluated in both cases via commonly used *N*log *N* schemes (where *N* is the number of particles), like Particle Mesh Ewald algorithms.[60] Moreover, coarse-grained schemes are usually coupled to implicit solvent methods or zero-charge CG water models,[5,36,48] while all-atom simulations suffer for solvent-solvent electrostatic interaction evaluations, which can typically account for twice the CPU time of solute−solute calculations. Therefore, in a rough estimation, TBMR can be up to 30 times faster than all-atom MD. Importantly, we finally note that the routines needed to reconstruct the spatial topology of the dipoles require negligible CPU time with respect to the electrostatic calculations.

3. Results and Discussion

TBMR is applied to a set of systems, representing different test situations in which long-range electrostatics interactions can be relevant for the overall energetics. Namely, we address the performance and accuracy of TBMR in describing the following: (1) a protein-induced electrostatic field at binding/active sites, (2) an electrostatic field tuning upon conformational changes in the secondary structure, and (3) long-range electrostatic interactions at the interface of protein complexes.

3.1. Protein-Induced Electrostatic Field at Binding/Active Sites. Two representative proteins have been chosen−actin[61,62] and azurin[63]−to test TBMR using a hybrid all-atom/coarse grained model (MM/CG). As proposed in the original formulation of MM/CG,[30] the (re)active center can be described using a classical force field, whereas the rest of the system is treated using a CG scheme.

Actin is a globular protein that polymerizes to form helical filaments constituting the cytoskeleton. ATP hydrolysis drives the filament dynamics inducing the structural change in the protein.[61,64] ADP/ATP have a well-known binding site (Figure 2a) located in the core of the protein,[61] which has peculiar electrostatic interactions with actin residues.[61,62] This specific protein is meant to be representative of the vast class of proteins and enzymes where the ligand-protein interactions are the crucial ingredient for ligand recognition and catalytic activation.[47]

Azurin from *Pseudomonas aeruginosa*[63] is known to be shaped in such a way that the long-range electrostatic contribution of the protein frame is crucial for the correct tuning of the redox properties of the enzyme.[65] On a general basis, the secondary/tertiary structure motifs often have a direct influence on the electronic properties of the reactive centers in metalloenzymes.[66,67] Here, azurin is chosen as a representative target for the broad class of metalloproteins, where long-range electrostatic interactions play a crucial functional role.

The crystallographic structures of actin and azurin (PDB codes 2HF3[68] and 4AZU,[63] respectively) are used as initial models to calculate the all-atom electrostatic potential (V_{aa}) in implicit solvent (PB scheme[69,70]), which is used as a reference. Hydrogen atoms and terminal and missing side-chain atoms are reconstructed following standard protocols (see e.g. ref 71). The GROMOS96 force field is used to describe the system in the all-atom calculations. In the MM/CG calculations, the (re)active centers along with neighbor residues are described at the MM level; an interface region linking MM and CG, defined as all the residues within a 6.5 Å radius from the MM part, is described at the intermediate layer following ref 30; and the remaining part of the protein is treated at the CG level. The calculated electrostatic potential (V_{hybrid}) produced by the model on the active site agrees well with the reference as shown in Figure 2b,d. From a quantitative standpoint, the weighted deviation between the all-atom and the MM/CG field $[(|\mathbf{E}|_{hybrid} - |\mathbf{E}|_{aa})/|\mathbf{E}|_{aa})]$ is calculated for all the values of the electric field, and the covariance is used for characterizing the correlation between the vectors $\mathbf{E}_{aa}$ and $\mathbf{E}_{hybrid}$ $[\mathbf{C} = (\langle \mathbf{E}_{aa} \cdot \mathbf{E}_{hybrid} \rangle)/(\langle \mathbf{E}_{aa} \rangle \langle \mathbf{E}_{hybrid} \rangle)]$ for the possible values of the electric field in the different systems (Figure 2b,d). It is clear that the error on the module of the electric field made by TBMR is non-negligible only for very low absolute values of the field. The same trend is found for the vector covariance: except for low values of the field, the covariance is very close to one, showing also

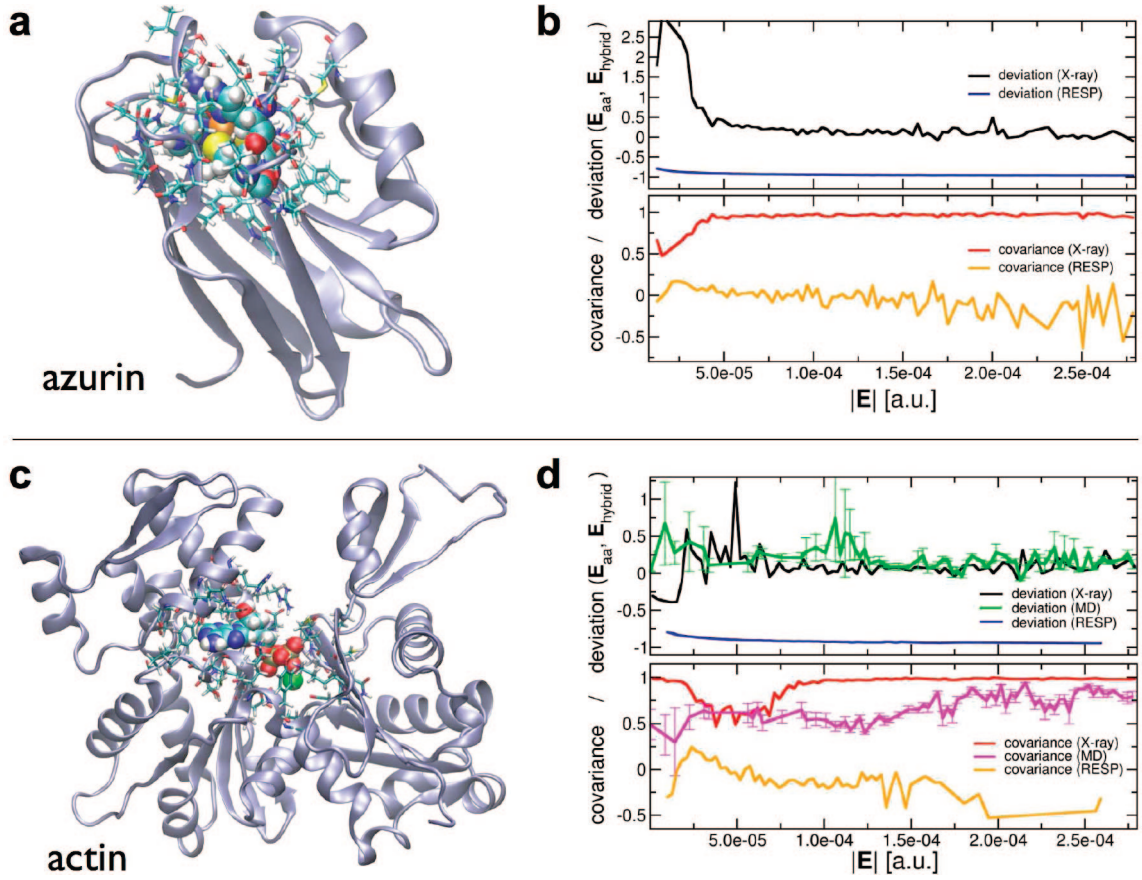

Figure 2. Reconstruction of the electrostatic field in the binding sites of azurin and actin. Panels **a** and **c**: cartoon representation of azurin and actin folding determinants, the active pockets are drawn in ball-and-stick and represent the all-atom portion in the MM/CG calculations. Panels **b** and **d**: the accuracy of the multipolar scheme is compared with all-atom simulations of azurin (b) and actin (d). Deviation and covariance between $\mathbf{E}_{aa}$ and $\mathbf{E}_{hybrid}$ are reported in the graphs for different values of the electrostatic field.

a striking match of the mutual orientation of the fields produced by the all-atom calculation and the TBMR model.

Our model is able to capture well the intensity and orientation of the all-atom electrostatic field. The usefulness of such model is straightforward for rigid docking studies, where the crystallographic structure of the protein is used as such. In fact, the TBMR model keeps its reliability also when the protein structure is allowed to relax and fluctuate, e.g. by MD. We have tested the performance of the model in such conditions by selecting snapshots along an equilibrated MD simulation of actin (one snapshot every 1 ns, averaged over 20 ns trajectory) and performing the same analysis on the electrostatic fields (Figure 2d). We find that the TBMR model produces, also in this case, good results both in terms of field intensity and orientation. The minor performance with respect to a fixed geometry relies on the fact that, in the case of moving structures, the side-chains contribution to the electrostatic field is generated by their average structure coming from an independent MD run. Therefore TBMR is expected to reproduce the average features of the electrostatic field but not to catch its instantaneous fluctuations as such.

Finally, the TBMR method is compared to the performance of a RESP-charges fitting. In this model, the all-atom electrostatic potential produced by the region of the protein that falls in the CG region is fitted by RESP-charges centered on each CG centroid. In particular, we have tested both RESP fittings assuming CG models comprise C_α centroids only or C_α and C_β centroids per amino acid. We find that the RESP fitting has an intrinsic inconsistency at different scales. In fact, when the RESP charges are fitted to reproduce the all-atom electrostatic potential on a large grid around the whole protein, we find that the electrostatic field in the smaller MM region is not even comparable to the all-atom one. If the RESP fitting is, on the contrary, performed on a grid centered at the MM region, the field there is more reliable. Nonetheless, from the graphs in Figure 2d, it is evident that even in this case RESP charges are unable to represent either the module or the orientation of the electric field at the static crystallographic structure and during the MD simulation. Such behavior (RESP systematically underestimates the field) is related to the nature of the RESP electrostatic potential. In fact, the electrostatic all-atom potential, which is generated by a series of electrostatic monopoles (the atomic charges), is fitted by a smaller number of electrostatic monopoles (the RESP centers). Moreover the spatial distribution of the RESP charges is not fully consistent with that of the all-atom charge groups. As a result, RESP-fitting can catch at best an average of the all-atom potential in a specific region, without being able to describe its local fluctuations, which lead to an underestimation of the electrostatic field and its nonaccurate alignment.

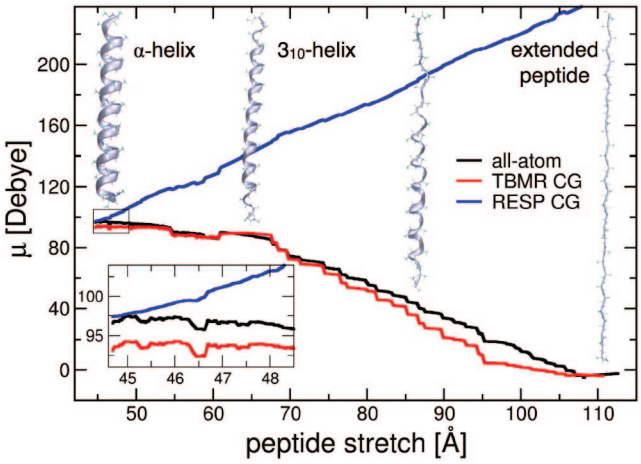

Figure 3. Change of the electrostatic dipole upon structural rearrangement. The component of the electrostatic dipole along the longitudinal direction of a poly ala α-helical peptide is reported upon progressive stretching. Snapshots of the structural determinants are superimposed indicating the progressive unfolding of the structure. The inset focuses on the initial fluctuations around the α-helix conformation.

The application of TBMR to these systems shows that the TBMR method is able to correctly describe the long-range contribution on the active center, demonstrating to be a promising solution for hybrid multiscale simulations of ligand binding recognition.

3.2. Electrostatic Field Tuning upon Conformational Changes in the Secondary Structure. An important requirement for any multiscale scheme is the need to be consistent and accurate upon extended structural rearrangements (such as secondary and/or tertiary structural changes, partial or total unfolding, etc.) that might occur during protein dynamics. The performance of our method in such conditions has been tested by studying the stretching of an α-helix (a 30 aa long poly-ala chain) using standard steering molecular dynamics. Upon mechanical stress, the polypeptide first refolds into a 3_{10} helical species and then unfolds progressively to form a β-sheet-like stretched peptide. We have calculated the projection of the total electrostatic dipole along the stretching direction produced by the TBMR model at different helical configurations, and we have compared it to the values of the all-atom system (Figure 3a). The dipole component obtained by the TBMR method matches very well the all-atom calculations. The model is able to reproduce qualitatively and quantitatively the increase in the electrostatic dipole at the conformational transition between the strained α and 3_{10} helical species. When the helix is completely unfolded, the electrostatic dipole along the stretching direction becomes negligible, as expected. Also in this case, we compare our results to those that would come from a RESP-based description of the electrostatics. In such a case, already in the regime of small oscillations, the change in the dipole is qualitatively not correct (Figure 3, inset). Further stretching of the peptide leads to a complete failure of the predicted values of the dipole. This is related to the fact that RESP charges are located at the C_α and C_β centroids, which do not correspond to the charge groups of the all-atom system and which generally follow qualitatively different spatial motions. While the helix is being stretched, the relative distances of the CG centroids increase, and this leads to an increase in the distances of the positive and negative centers of charge, that is, to an increase of the dipole. Rather, in the all-atom system, the stretching of the helix corresponds to a twist of the peptide bonds, which leads to a decrease of the total dipole along the stretching direction.

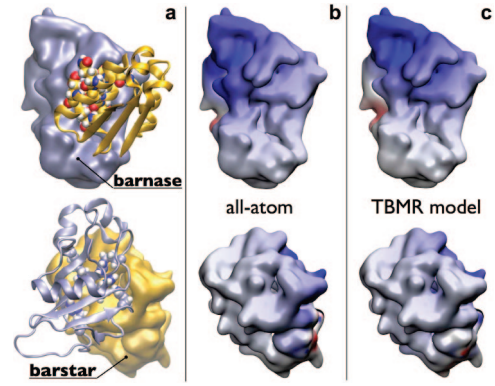

Figure 4. Protein−protein interactions in the barnase-barstar complex. Panel **a**: overall fold structure of the complex (PDB code: 1BRS), indicating the residues involved in the binding at the interface. Panels **b** all-atom and **c** hybrid electrostatic potentials are reported on a molecular surface drawn on top of barnase and barstar, respectively.

Thus, our method is able to describe well the twisting of the dipoles during helical unfolding, providing promising ground when coupled to nontopological coarse-grained force fields. In particular, the electrostatic description given by our method is general and consistent with incoming structural rearrangements of the secondary and tertiary determinants of a macromolecule. Thus, this contribution can be easily included in existing or future coarse-grained force fields developed to sample complex conformational spaces in order to address topics like protein folding and macromolecular assembly in biological networks.

3.3. Long-Range Electrostatic Interactions at Protein−Protein Interfaces. As a final test for the TBMR model, we have inferred its capability of reproducing long-range electrostatic interactions at the protein−protein interface in protein complexes. Thus, we apply the multipolar method to the barnase-barstar complex (PDB code: 1BRS[72]), which is known to be dominated by strong electrostatic interactions at the protein−protein interface[72] and which structure was determined by X-ray crystallography.

Following the MM/CG protocol, the protein−protein interface (defined as the Leu26-Val45 loop-helix motif of barstar and the loops of barnase in contact with it) is described at all-atom level, while the rest of the barnase/barstar complex is described by CG centroids. The long-range electrostatic potential on the surface of the barnase-barstar contact appears to be qualitatively well reproduced by TBMR, as depicted in Figure 4. In particular, it is able to represent the strong positive potential produced by barnase, which is functional to the long-range recognition of the negatively charged residues present on the interacting portion of the surface of barstar. The long-range contribution to the

electrostatic potential of barstar seems to be of less importance, and such an effect is also well described by TBMR (Figure 4).

On a general note, the TBMR method is able to represent quite well the electrostatic potential in a coarse-grained scheme, and for this reason it can be used in a 2-fold manner in the field of protein−protein interaction studies: (i) in a hybrid MM/CG simulation to refine the late stage of a macromolecular complex formation, when the proteins are close together, and (ii) in the early stage of the same process, when the proteins are far apart and long-range electrostatics plays a role in determining the right orientation for the formation of the complex. The latter process can be studied, for example, with full CG calculations coupled to TBMR method.

4. Concluding Remarks

We propose a new scheme for introducing long-range electrostatic contributions in multiscale simulations of proteins.[27,30] The topologically based multipolar reconstruction (TBMR) scheme can be generally adopted to describe electrostatics in the coarse-grained region of hybrid multiscale simulations and can be further extended to full coarse-grained simulations of protein dynamics.

We show that a multipolar expansion of the coarse-grained residues, truncated in the present formulation to the dipolar term, is able to capture a large part of the electrostatic contribution in a variety of biological problems and gives results appreciably more accurately than alternative methods such as RESP charge fitting. The method enhances in fact the reliability of hybrid multiscale methods at localized active sites in enzymatic complexes where the electrostatic contribution is supposed to be crucial for ligand binding recognition and for catalytic activation. It has also been shown to describe with accuracy comparable to all-atom representation the evolution of the dipoles during helical unfolding. Moreover, the method is well suited for the application in the broad field of protein−protein interaction studies. It can be used to improve the reliability of molecular interactions at the interface of a macromolecular complex in the framework of hybrid multiscale simulations, whereas in full coarse-grained simulations it can contribute in identifying the protein−protein recognition pathway when the interacting interface is not known a priori.

The TBMR scheme can be easily incorporated within existing hybrid multiscale methods and can be naturally extended to higher orders in the multipolar expansion, when the adoption of different all-atom force fields is required. In particular, we think that our method is particularly fit to be used in combination with multiscale coarse-graining techniques,[33,34] which to date represent the most general and rigorous approach to coarse-graining.

Finally, we foresee that coupling our model to force field-like coarse-grained Hamiltonians will improve the reliability of molecular simulations dealing with large structural rearrangements that might happen in macromolecular assembly and protein networks dynamics. The description of nucleic acids and lipids constituting biological membranes is also a natural and desirable expansion of the present scheme, which would improve the accessibility and reliability of multiscale molecular simulations in the relevant and complex field of DNA−protein and membrane-protein interacting networks.

Acknowledgment. Authors thank Dr. Ivano Tavernelli for useful discussions. The presented work was supported by the Swiss National Science Foundation (Grant No. PP002-_118930). M.N. acknowledges Prof. Ursula Rothlisberger for mentoring her research work at EPFL.

References

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(2) Stryer, L. *Biochemistry;* W. H. Freeman and Company: New York, U.S.A., 1995.

(3) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464–4470.

(4) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 9785–9792.

(5) Saiz, L.; Klein, M. L. *Acc. Chem. Res.* **2002**, *35*, 482–489.

(6) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.

(7) Brannigan, G.; Lin, L. C. L.; Brown, F. L. H. *Eur. Biophys. J.* **2006**, *35*, 104–124.

(8) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045–15048.

(9) Tozzini, V. *Curr. Opin. Struct. Biol.* **2008**, *15*, 144–150.

(10) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694–698.

(11) Levitt, M. *J. Mol. Biol.* **1976**, *104*, 59–107.

(12) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 945–950.

(13) Go, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535–542.

(14) Bahar, I.; Atilgan, A. R.; Erman, B. *Fold. Des.* **1997**, *2*, 173–181.

(15) Micheletti, C.; Carloni, P.; Maritan, A. *Proteins* **2004**, *55*, 635–645.

(16) Basdevant, N.; Borgis, D.; Ha-Duong, T. *J. Phys. Chem. B* **2007**, *111*, 9390–9399.

(17) Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2005**, *122*, 124906.

(18) Becker, N. B.; Everaers, R. *Phys. Rev. E* **2007**, *76*, 021923.

(19) Voltz, K.; Trylska, J.; Tozzini, V.; Kurkal-Siebert, V.; Langowski, J.; Smith, J. *J. Comput. Chem.* **2008**, *29*, 1429–1439.

(20) Cieplak, M.; Hoang, T. X. *J. Biol. Phys.* **2000**, *26*, 273–294.

(21) Hoang, T. X.; Cieplak, M. *J. Chem. Phys.* **2000**, *112*, 6851–6862.

(22) Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 10712–10717.

(23) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, T.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Sali, A.; Rout, M. P. *Nature* **2007**, *450*, 684–700.

(24) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, T.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Rout, M. P.; Sali, A. *Nature* **2007**, *450*, 701–710.

(25) Trabuco, L. G.; Villa, E.; Mitra, K.; Frank, J.; Schulten, K. *Structure* **2008**, *16*, 673–683.

(26) Liu, P.; Voth, G. A. *J. Chem. Phys.* **2007**, *126*, 045106.

(27) Ayton, G. A.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.

(28) Delle Site, L.; Abrams, C. F.; Alavi, A.; Kremer, K. *Phys. Rev. Lett.* **2002**, *89*, 156103.

(29) Villa, E.; Balaeff, A.; Mahadevan, L.; Schulten, K. *Multiscale Model. Simul.* **2004**, *2*, 527–553.

(30) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. *Phys. Rev. Lett.* **2005**, *95*, 218102.

(31) Lyman, E.; Ytreberg, F. M.; Zuckerman, D. M. *Phys. Rev. Lett.* **2006**, *96*, 028105.

(32) Lyman, E.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2006**, *2*, 656–666.

(33) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.

(34) Izvekov, S.; Voth, G. A. *J. Chem. Theory Comput.* **2006**, *2*, 637–648.

(35) Ayton, G. S.; Voth, G. A. *J. Struct. Biol.* **2007**, *157*, 570–578.

(36) Praprotnik, M.; Delle Site, L.; Kremer, K. *J. Chem. Phys.* **2005**, *123*, 224106.

(37) Praprotnik, M.; Delle Site, L.; Kremer, K. *Phys. Rev. E* **2006**, *73*, 066701.

(38) Praprotnik, M.; Delle Site, L.; Kremer, K. *Phys. Rev. E* **2007**, *75*, 017701.

(39) Matysiak, S.; Clementi, C.; Praprotnik, M.; Kremer, K.; Delle Site, L. *J. Chem. Phys.* **2008**, *128*, 024503.

(40) Ensing, B.; Nielsen, S. O.; Moore, P. B.; Klein, M. L.; Parrinello, M. *J. Chem. Theory Comput.* **2007**, *3*, 1100–1105.

(41) Heyden, A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 217–221.

(42) Neri, M.; Cascella, M.; Micheletti, C. *J. Phys.-Condens. Matter* **2005**, *17*, S1581−S1593.

(43) Neri, M.; Anselmi, C.; Carnevale, V.; Vargiu, A. V.; Carloni, P. *J. Phys.-Condens. Matter* **2006**, *18*, S347−S355.

(44) Neri, M.; Baaden, M.; Carnevale, V.; Anselmi, C.; Maritan, A.; Carloni, P. *Biophys. J.* **2008**, *94*, 71–78.

(45) Villa, E.; Balaeff, A.; Schulten, K. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6783–6788.

(46) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210–3235.

(47) Marrink, S.-J.; Risselada, H. J.; Yefimov, S.; Tielemann, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(48) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tielemann, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.

(49) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* 199394, 10269–10280.

(50) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896.

(51) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

(52) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.

(53) Zhou, J.; Thorpe, I.; Izvekov, S.; Voth, G. A. *Biophys. J.* **2007**, *92*, 4289–4303.

(54) Ayton, G. S.; Noid, W. G.; Voth, G. A. *MRS Bull.* **2007**, *32*, 929–934.

(55) Park, B.; Levitt, M. *Proteins* **1996**, *258*, 367–392.

(56) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. E.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.

(57) Banavar, J. R.; Hoang, T. X.; Maritan, A.; Seno, F.; Trovato, A. *Phys. Rev. E* **2004**, *70*, 041905.

(58) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberg, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 manual and user guide*; Hochschulverlag AG an der ETH Zurich: Zurich, Switzerland, 1996.

(59) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(60) Otterbein, L. R.; Graceffa, P.; Dominguez, R. *Science* **2001**, *293*, 708–711.

(61) Dominguez, R.; Graceffa, P. *Biophys. J.* **2003**, *85*, 2073–2074.

(62) Nar, H.; Messerschmidt, A.; Huber, R.; van de Kamp, M.; Canters, G. W. *J. Mol. Biol.* **1991**, *221*, 765–772.

(63) Chu, J. W.; Voth, G. A *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13111–13116.

(64) Cascella, M.; Magistrato, A.; Tavernelli, I.; Carloni, P.; Rothlisberger, U. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19641–19646.

(65) Olsson, M. H. M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.

(66) Cascella, M.; Cuendet, M. L.; Tavernelli, I.; Rothlisberger, U. *J. Phys. Chem. B* **2007**, *111*, 10239–10247.

(67) Rould, M. A.; Wan, Q.; Joel, P. B.; Lowey, S.; Trybus, K. M. *J. Biol. Chem.* **2006**, *281*, 31909–31919.

(68) Fogolari, F.; Brigo, A.; Molinari, H. *J. Mol. Recognit.* **2002**, *15*, 377–392.

(69) Baker, N. A.; Sept, D.; Joseph, S.; Holst, H. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.

(70) Vidossich, P.; Cascella, M.; Carloni, P. *Proteins* **2004**, *55*, 924–931.

(71) Buckle, A. M.; Schreiber, G.; Fersht, A. R. *Biochemistry* **1994**, *33*, 8878–8889.

(72) Lee, L.; Tidor, B. *Nat. Struct. Biol.* **2001**, *8*, 73–76.

CT800122X