

Coarse Point Charge Models For Proteins From Smoothed Molecular Electrostatic Potentials

Laurence Leherte* and Daniel P. Vercauteren

Laboratoire de Physico-Chimie Informatique, Groupe de Chimie Physique Théorique et Structurale, University of Namur (FUNDP), Rue de Bruxelles 61, B-5000 Namur, Belgium

Received April 21, 2009

Abstract: To generate coarse electrostatic models of proteins, we developed an original approach to hierarchically locate maxima and minima in smoothed molecular electrostatic potentials. A charge-fitting program was used to assign charges to the so-obtained reduced representations. Templates are defined to easily generate coarse point charge models for protein structures, in the particular cases of the Amber99 and Gromos43A1 force fields. Applications to four small peptides and to the ion channel KcsA are presented. Electrostatic potential values generated by the reduced models are compared with the corresponding values obtained using the original sets of atomic charges.

I. Introduction

The design of protein coarse-grain (CG) models and their corresponding interaction potential functions are nowadays an active field of research, especially for solving problems such as protein folding and docking through, e.g., molecular mechanics (MM) and molecular dynamics (MD) methods.¹ Indeed, all-atom simulations may be out of practical computational resources for macromolecules, and a strategy to consider large size systems and long time scales in a simulation consists in limiting the number of interacting particles. Among the essential parameters involved in all-atom and CG potentials, electrostatic interactions are of crucial importance since they govern local and global properties, e.g., their stability, flexibility, etc. Various approaches to evaluate electrostatic interactions are, e.g., reviewed by Dong et al.² Nevertheless, evaluating the adequacy of a particular method is not straightforward; a presentation of this problem can, e.g., be found in the work of Schutz and Warshel³ who discussed the choice of dielectric constants.

Common approaches used to design a CG description of a protein consist in reducing groups of atoms into single interaction sites. For example, in the work by Skepö et al.,⁴ each amino acid (AA) is represented by a single spherical site, with unit or nul electric charge. The authors studied a

proline-rich protein PRP-1 interacting with a mica surface using Monte Carlo simulations. Curcó et al.⁵ developed a CG model of β -helical protein fragments, where the AAs are represented by two, three, or four blobs depending upon the AA type, in accordance with a best fitting between Monte Carlo (MC) all-atom and CG energies. In their work, the AAs are depicted by the amide hydrogen atom, the oxygen atom, the geometric center of the side chain (except for Gly), and a fourth blob whose position depends on the AA type (except for Gly, Ala, and Val). In the Basdevant et al. paper,⁶ each AA residue is modeled using one sphere located on the geometric center of the backbone and one or two spheres located on the geometric centers of the side chain fragments (except for Gly). Differently, Pizzitutti et al.⁷ represented each AA of a protein sequence by a charged dipolar sphere. For each AA, one CG sphere is located on the center-of-mass (com) of uncharged residues, while for charged residues, one CG is assigned to the com of the neutral part of the AA, and one CG is assigned to the com of the charged part. Charged residues are Arg, Asp, Glu, Lys, and terminal AAs. The authors show that, in protein association, their model provides a good approximation of the all-atom potential, if the distance between the protein surfaces is larger than the diameter of a solvent molecule. In a recent work, Zhang et al.⁸ proposed a method to define CGs that reflect the collective motions computed by a principal component analysis of an atomistic MD trajectory. Each CG site is the com of a domain, i.e., a group of contiguous C α atoms that

* Corresponding author. Telephone: +32-81-724560. Fax: +32-81-725466. E-mail: laurence.leherte@fundp.ac.be.

move in a highly correlated fashion. Very recently, Bereau and Deserno⁹ presented a generic CG model for proteins with one grain located on each of the N, C α , and C atoms and a fourth grain on C β . Such a simple side chain description was aimed at facilitating the parametrization of the corresponding CG potential function.

The development of CG interaction potential functions is generally made either from atomistic interaction potential¹⁰ or MD results^{11–13} via experimental data, such as B-factors,¹⁴ or from the fitting of a potential function achieved by matching CG and atomistic distributions.^{13,15} For example, Lyman et al.¹⁶ presented a new method for fitting spring constants to mean square CG–CG distance fluctuations computed from atomistic MD. One can also cite the inverse MC approach¹⁷ used for iteratively adjusting an effective CG potential function until it matches a target radial distribution function. Noid et al.^{18,19} proposed a statistical mechanics theory to check the consistency between CG and all-atom models. More specifically, the parametrization of the well-known MARTINI force field (FF), dedicated to MD simulations of biomolecular systems, is based on the reproduction of partitioning free energies between polar and apolar phases of a large number of chemical systems.^{20,21} In that model, groups of four heavy atoms are represented by a single interaction center, except for small ring-like fragments. AAs, thus, consist of one to four side chain beads and one backbone bead.²¹ Only four main types of interaction sites are defined: polar (P), nonpolar (N), apolar (C), and charged (Q). Each particle type has a number of subtypes, which allow for a more accurate representation of the chemical nature of the underlying atomistic structure. In MARTINI, only AA residues Arg, Asp, Glu, and Lys are charged. Such a description was, e.g., applied to protein channels embedded in a lipid membrane environment.²² In the UNRES model,²³ a peptidic chain is represented by a sequence of backbone beads located at peptide bonds, while side chains are modeled as single beads attached to the C α atoms, which are considered only to define the molecular geometry. In the so-called SimFold CG description and energy function, a mixed representation is used.^{24,25} Residues of aqueous proteins are represented by backbone atoms N, C α , C, O, and H and by one side chain centroid. In UNRES and SimFold, electrostatic interactions are not explicitly calculated using the Coulomb term like they are in the MARTINI FF for charged AAs.

Basdevant et al.⁶ proposed an approach to determine charges for their reduced models built on the com of backbone and side chain groups of AA residues. Prior to these studies, Gabdoulline et al.²⁶ used a model that consisted of a small number of point charges (monopoles) suitable for the description of the intermolecular electrostatic interactions. As later applied by Basdevant et al.,⁶ these charges were derived from a fitting procedure applied to reproduce the molecular electrostatic potential (MEP) obtained by solving the Poisson–Boltzmann equation. In their example, the charges are located at the geometric centers of the head groups of the charged residues. The mimick of all-atom electrostatic interactions using a limited set of point charges was also proposed by Berardi et al.,²⁷ who applied a genetic

algorithm to determine the location and the values of a given number of charges for molecules involved in liquid-crystalline materials. Extended multipolar models are also reported, such as the one described by Golubkov et al.²⁸ In that approach, illustrated for small molecules, such as water, methanol, and benzene, the charge distribution is represented by a point multipole expression with charges (nul for the hereabove examples), by a dipole, and by quadrupole moments placed at the molecular com. These dipole and quadrupole moments were set equal to the corresponding average moments obtained from analyses of all-atom MD simulations. In the framework of proteins, Cascella et al.²⁹ presented a method to parametrize an AA reduced model that allows reproduction of all-atom electrostatic properties evaluated as averages during MD simulations for the side chains and statistically for the backbones. Reviews on the progresses of CG dynamical models can also be found in additional references.^{30,31}

Multiscale methods, that combine several levels of description, are also appealing since they allow to model limited regions of space with details while representing the outer regions by coarser models.^{32,33} The consideration of outer influences, such as external stresses³⁴ or solvent effects,³⁵ can also be treated with CG approaches.

In the present paper, we propose a method to elaborate coarse point charge models for protein structures from smoothed MEPs. The quality of such models is approached by comparing CG-based MEP and dipole values with the corresponding all-atom properties. In a previous work,³⁶ a protein structure was decomposed into separate molecular fragments that were determined through a merging/clustering procedure of atom trajectories generated in progressively smoothed electron density (ED) distribution functions. This was followed by a second study³⁷ where atoms were clustered according to their trajectories defined in a smoothed MEP function. That procedure allowed to locate the corresponding MEP local maxima (peaks) and minima (pits). A fitting algorithm was applied to evaluate the peak and pit charges. Results, presented for the twenty AAs, were derived from the all-atom Amber charges reported in Duan et al.³⁸

With respect to that second paper, we have extended, refined, and automated our approach, which now consists of the three following steps. First, extrema corresponding to the AA backbone are located in the smoothed MEP of a β -pentadecapeptide Gly₁₅ model. For the AA side chains, the CGs are identified as extrema in the smoothed MEPs of each of the 20 natural AA residues in their isolated state from both the Amber99,³⁹ as available in PDB2PQR,^{40,41} and Gromos43A1⁴² sets of charges. Gromos charges were taken from the files provided with the software SwissPDB-Viewer.^{43,44} Second, charges are assigned to each of the CGs through a charge-fitting procedure applied to reproduce unsmoothed MEP grid values and dipole moments. Third, a library of the resulting AA point charge templates, including CG locations and their charge values, is built for further modeling of proteins.

In Section II, we present a brief overlook of the theoretical background. In Section III, we describe the methodology to design the CG templates of the AAs from the pentadecapep-

tide $\beta\text{-Gly}_{15}$ and the isolated AA models as well as from their associated atom charges. Finally, in Section IV, we detail applications to four small peptides and to the ion channel KcsA. Let us finally note that further in the text: (i) we will use the expressions “reduced” and “CG” indifferently, and (ii) that all three-dimensional (3D) illustrations were generated with OpenDX⁴⁵ unless otherwise stated.

II. Theoretical Background

In this section, we present the mathematical formalism that was used to design a protein-reduced representation and its corresponding point charges. First, the smoothing algorithm is described. This description is followed by the mathematical expressions that are specific to the Coulomb electrostatic interaction function. Finally, we detail the approach applied to calculate the CG point charges.

A. Smoothing Algorithm. To follow the trajectories of the local maxima and minima in a MEP function, as a function of the degree of smoothing, we implemented an algorithm initially described by Leung et al.⁴⁶ The authors initially proposed a method to model the blurring effect in human vision. This was achieved by filtering a digital image $p(x)$ through a convolution product with a Gaussian function and by assigning each data point of the resulting $p(x, t)$ image to a cluster via a dynamical equation built on the gradient of the convoluted image:

$$x(n+1) = x(n) + h \nabla_x p(x, t) \quad (1)$$

where h is defined as the step length. We adapted this idea to 3D images, such as ED and MEP functions, f , such as:

$$\mathbf{r}_{f(t)} = \mathbf{r}_{f(t-\Delta t)} + \frac{\Delta}{f(t)} \nabla f(t) \quad (2)$$

where \mathbf{r} stands for the location vector of a point in a 3D function, such as a MEP field.

The various steps of the resulting merging/clustering algorithm are as follows: First, at scale $t = 0$, each atom of a molecular structure is considered as either a local maximum (peak) or minimum (pit) of the MEP function. All atoms are consequently taken as the starting points of the merging procedure. Second, as t increases from 0 to a given maximal value t_{\max} , each point moves continuously along a gradient path to reach a location in the 3D space where $\nabla f(t) = 0$. From a practical point of view, this consists of following the trajectory of the peaks and pits on the MEP distribution surface calculated at t according to eq 2. The trajectory search is stopped when $|\nabla f(t)|$ is lower or equal to a limit value, grad_{\lim} . Once all peak/pit locations are found, close points are merged if their interdistance is lower than the initial value of $\Delta^{1/2}$. The procedure is repeated for each selected value of t . If the initial Δ value is too small to allow convergence toward a local maximum or minimum within the given number of iterations, then its value is doubled (a scaling factor that is arbitrarily selected), and the procedure is repeated until final convergence.

B. Molecular Electrostatic Potentials. The electrostatic potential function generated by a molecule A is simply calculated as a summation over its atomic contributions:

$$V_A(\mathbf{r}) = \sum_{a \in A} \frac{q_a}{|\mathbf{r} - \mathbf{R}_a|} \quad (3)$$

where \mathbf{R}_a is the position vector of atom a , and q_a is the electric charge. A smoothed version can be expressed as:

$$V_{A,t}(\mathbf{r}) = \sum_{a \in A} \frac{q_a}{|\mathbf{r} - \mathbf{R}_a|} \text{erf}\left(\frac{|\mathbf{r} - \mathbf{R}_a|}{2\sqrt{t}}\right) \quad (4)$$

where the error function erf can be calculated using the analytically derivable expression:⁴⁷

$$\text{erf}(x) = 1 - (a_1 T + a_2 T^2 + a_3 T^3 + a_4 T^4 + a_5 T^5) e^{-x^2} \quad \text{with } T = \frac{1}{1 + px} \quad (5)$$

The values of the parameters p and a are: $p = 0.3275911$, $a_1 = 0.254829595$, $a_2 = -0.284496736$, $a_3 = 1.421413741$, $a_4 = -1.453152027$, and $a_5 = 1.061405429$.⁴⁷ Equation 4 is identical to the expression found in the potential smoothing approach, a well-known technique used in MM applications.⁴⁸

C. Calculation of Point Charges. Charge values were obtained using the charge-fitting program QFIT.⁴⁹ Among the approaches that are reported in the literature, e.g., either excluding the MEP grid points that are located too close or too far from the molecular structure under consideration or including grid points located at large distances up to 30–45 Å from the molecular center,²⁶ we selected the first approach to modulate the influence of the neighborhood of the AA under interest. Indeed, we wished to establish AA CG charges that are as independent as possible on the selected models. All MEP grids were built using either the Amber99³⁹ or Gromos43A1^{42,43} point charges, assigned using the software PDB2PQR,^{40,41} with a grid step of 0.5 Å. Fittings were achieved by considering points located at distances between 1.4 and 2.0 times the van der Waals (vdW) radius of the atoms. These two limiting distance values were selected after the so-called Merz–Singh–Kollman scheme.⁵⁰

In all fittings presented, the total electric charge and the magnitude of the molecular dipole moment were constrained to be equal to the corresponding all-atom Amber99 or united-atom Gromos43A1 MEP values. The quality of the fittings was evaluated by two root-mean-square deviation (rmsd) values, i.e., the rmsd V determined between the MEP grid values obtained using the fitted charges and the reference unsmoothed MEP grid values and the rmsd μ evaluated between the dipolar value calculated from the fitted CG charges and the reference dipole moment of the molecular structure. All dipole moment components were calculated with the origin of the atom coordinates set to (0. 0. 0.).

III. Results and Discussion

This section is dedicated to the elaboration of coarse point charge models of proteins based on the local maxima and minima observed in their smoothed MEP functions. After selection of the best smoothing degree to work at, the first two steps of our strategy rely on the CG description of the protein backbone and the development of side chain CG

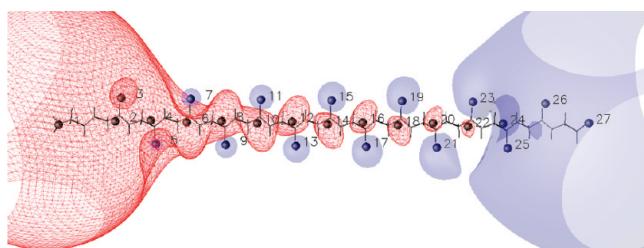


Figure 1. Amber99 MEP isocontours (blue plain surface: -0.03 ; red mesh: $0.03 \text{ e}^-/\text{bohr}$) of $\beta\text{-Gly}_{15}$ with charged NH_3^+ and COO^- ends, as obtained by smoothing the original MEP at $t = 1.25 \text{ bohr}^2$. Local maxima and minima (black spheres) were obtained using the hierarchical merging/clustering algorithm applied to the all-atom Amber99 MEP function.

models. Each stage involves the determination of CG locations and corresponding electrostatic point charges. The final part of the section focuses on the application of our CG models to four small peptides (PDB access codes 2EVQ, 1BXX, 1BC5, and 2RD4) and to the tetrameric ion channel KcsA (PDB access code 1BL8).

As mentioned earlier, to determine the backbone-reduced representation, we limited our study to a fully extended peptide model made of 15 amino acids, i.e., $\beta\text{-Gly}_{15}$. That particular peptide sequence was chosen to minimize the interference between the central Gly residue Gly8 and the whole peptide structure as well as to get a nul charge on Gly8. The concept of “interference” is solely based on the CG description that can be obtained for various secondary structures. We indeed showed that the MEP-based clustering results are highly dependent on the peptide conformation.³⁷ For the studied pentadecapeptide $\beta\text{-Gly}_{15}$, end residues were not charged. At first, this may sound artificial, but the presence of a large negative or positive charge in the structure strongly affects the homogeneity of the CG distribution along the peptide chain.³⁷ Figure 1 illustrates the local extrema observed in the MEP of $\beta\text{-Gly}_{15}$ characterized by charged ends built from the Amber99 charges and smoothed at $t = 1.25 \text{ bohr}^2$ using eq 4. In this figure, one notices the presence of point charges in the close neighborhood of the C and O atoms of all residues but the two end ones. The two terminal AAs involve point charges on NH_3^+ and COO^- only. The volume embedded by the negative and positive isocontours is also varying along the chain, increasing or decreasing toward COO^- , respectively. This reflects variations in the corresponding CG point charge values.

The structures of the isolated AAs involved the $(\text{C}\alpha-\text{C=O})_{\text{AA}}(\text{N}-\text{H})_{\text{AA+1}}$ backbone atoms so as to allow the merging of the $(\text{C=O})_{\text{AA}}$ and $(\text{N}-\text{H})_{\text{AA+1}}$ atoms, as observed in $\beta\text{-Gly}_{15}$. The consideration of isolated AAs is part of a strategy to favor CG models to approximate all-atom representations. That strategy was selected to reduce the mutual influence of the backbone atoms on the side chain descriptions. It was indeed observed, in a previous study on $\text{Gly}_7\text{-AA-Gly}_7$ structures,³⁷ that for AAs like Asp and Phe, the side chain CG representation is dependent on its conformation and on the presence of the backbone, respectively. Let us also mention that treating separately backbone and side chain CG descriptions was, e.g., applied by Casella et al.²⁹ in their method to evaluate protein electrostatic

potentials as summations over backbone dipolar and side chain multipolar contributions.

To generate the 3D structure of all AAs studied in this work, the simulated annealing (SA) procedure implemented in the program SMMP05^{51,52} was applied to pentadecapeptide models, i.e., $\text{Gly}_7\text{-AA-Gly}_7$ structures, with Ω , Φ , Ψ , and χ dihedrals constrained to predefined values. The ECEPP/3 FF⁵³ and SA default running parameters were selected. Each SA run consisted in a first 100-step equilibration MC Metropolis stage carried out at 1 000 K. Then the procedure was continued for 50 000 MC Metropolis iterations until the final temperature, 100 K, was reached. The lowest potential energy structure generated during each run was kept. Isolated AA structures were then obtained by pruning the optimized pentadecapeptides.

The hierarchical decompositions of the molecular structures from MEP functions were carried out with the following parameters: $t = 0.05\text{--}3.0 \text{ bohr}^2$, $\Delta_{\text{init}} = 10^{-4} \text{ bohr}^2$, $\text{grad}_{\text{lim}} = 10^{-6} \text{ e}^-/\text{bohr}^2$.

A. Selection of the Smoothing Degree. As illustrated in Figure 2 for residue Trp, the CG description of an AA is dependent on the smoothing value t . At $t = 0.05 \text{ bohr}^2$, peaks and pits observed in the MEP are closely located on the atoms of the molecular structure. Starting at $t = 0.3 \text{ bohr}^2$, the extrema begin to move away from the atomic centers and their number decreases. At $t = 2.5 \text{ bohr}^2$, there are only three extrema left on the side chain of the AA.

To select the optimal smoothing degree for the building of the reduced models, we used the charge-fitting algorithm QFIT⁴⁹ and applied it, with the same conditions as reported in Section II, to each set of peaks and pits obtained for the $\beta\text{-Gly}_{15}$ structure at various smoothing levels. The resulting minimal objective function (MOF) values are reported in Figure 3. The MOF function is built on the rmsdV and $\text{rmsd}\mu$ values defined in Section II. The best fittings, corresponding to a dipolar description of each AA backbone (Figure 4), were obtained at $t = 1.25$ and 1.3 bohr^2 for Amber99 and Gromos43A1, respectively, i.e., $\text{MOF} = 1.76$ and 0.48 . For Amber99, the loss of one CG between $t = 1.25$ and 1.3 bohr^2 involves a steep rise in the MOF value, followed by a slower decrease observed up to $t = 1.9 \text{ bohr}^2$. Between $t = 1.5$ to 1.9 bohr^2 , the better fit is due only to a more adequate arrangement of peaks and pits, their number being constant, i.e., equal to 30 (Figure 3). For Gromos43A1, the MOF values are well below the corresponding values obtained with the Amber99 FF. This is due to the fact that Gromos43A1 is already a united-atom FF. Indeed, most of the atoms in alkyl groups, for instance, have a nul electric charge. Beyond $t = 1.3 \text{ bohr}^2$, the fitting is less and less efficient due to a progressive change in the location of the CGs with respect to the original structure. Models obtained for $\beta\text{-Gly}_{15}$ at $t = 1.25$ and 1.3 bohr^2 for Amber99 and Gromos43A1, respectively, contain 32 and 31 CGs (Figure 4 and Table 1). In this sense, the application of the smoothing algorithm to the MEP function levels out the differences between the all-atom and united-atom FFs, but the CG charge values differ (Table 1), as explained in the next paragraph.

B. Protein Backbone Modeling. As announced hereabove, to generate a regular point charge distribution for the

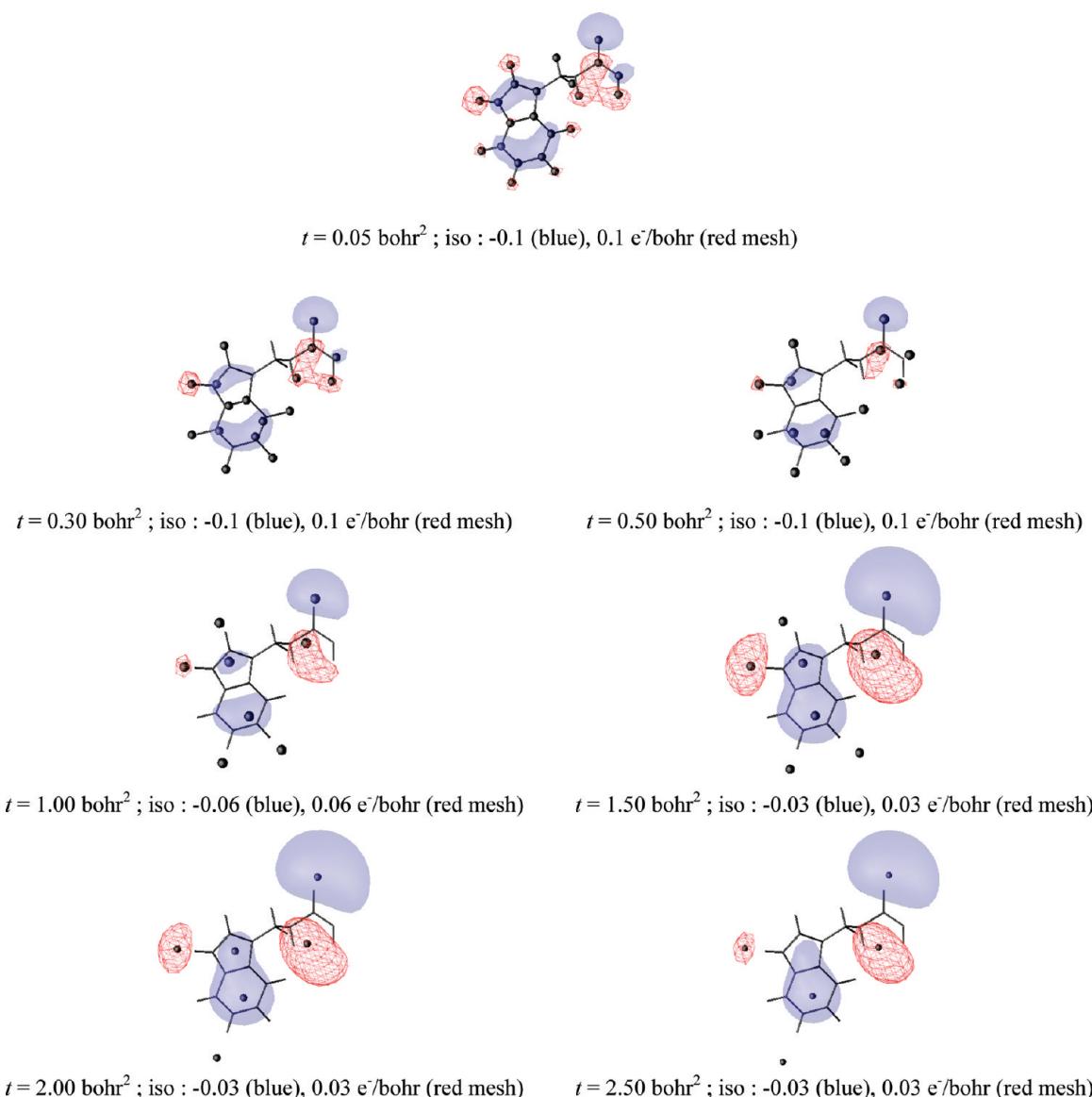


Figure 2. Amber99 MEP isocontours of Trp in a conformational state corresponding to the $g-g$ -rotamer, smoothed at various values of t . Local maxima and minima (black spheres) were obtained using the hierarchical merging/clustering algorithm applied to the all-atom Amber99 MEP function.

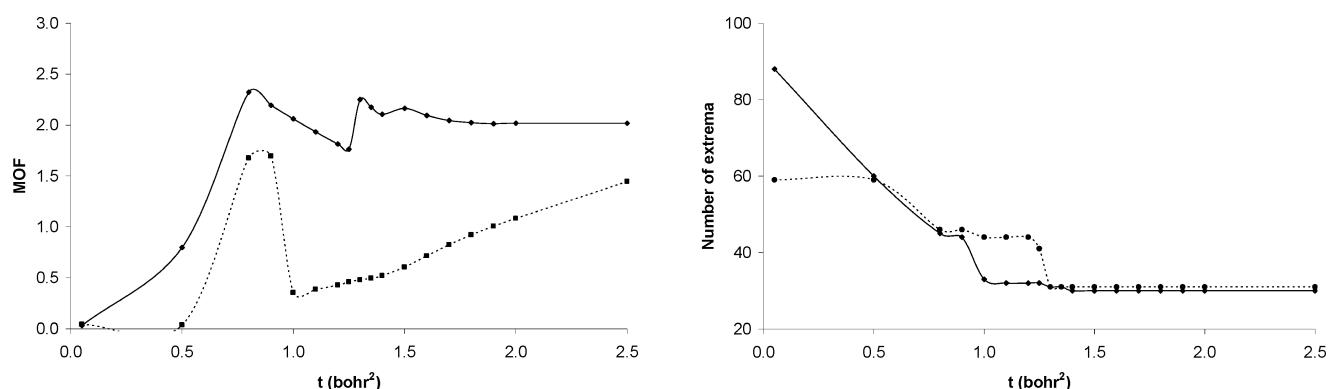


Figure 3. (Left) The MOF of the charge fittings of $\beta\text{-Gly}_{15}$ CG points vs the unsmoothed Amber99 (plain line) and the Gromos43A1 (dashed line) MEP values. (Right) Number of local minima and maxima observed in the smoothed MEPs, as a function of the smoothing degree t .

backbone, an extended geometry characterized by $\Omega = 180^\circ$, $\Phi = -139^\circ$, and $\Psi = 135^\circ$ was considered. Indeed, for MEP analyses, the conformation of the peptide appeared to be

extremely important on the results of the merging/clustering algorithm applied to MEP functions.³⁷ Fitted CG charges of structure $\beta\text{-Gly}_{15}$, depicted in Figure 4, are reported in Table

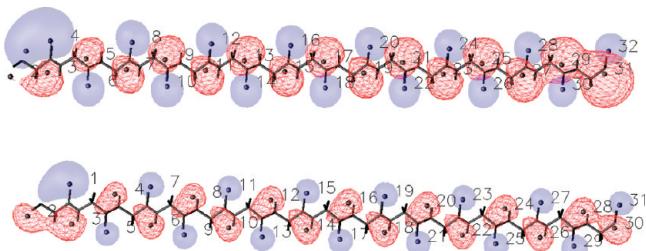


Figure 4. MEP isocontours (blue plain surface: $-0.03\text{ e}^-/\text{bohr}$; red mesh: $0.03\text{ e}^-/\text{bohr}$) of $\beta\text{-Gly}_{15}$ smoothed (top) at $t = 1.25\text{ bohr}^2$ using Amber99 and (bottom) at $t = 1.3\text{ bohr}^2$ using Gromos43A1. Local maxima and minima (black spheres) were obtained using the hierarchical merging/clustering algorithm applied to the original MEP functions. CG points are numbered as in Table 1.

Table 1. CG Charges q (in e^-) of $\beta\text{-Gly}_{15}$ Fitted vs the Unsmoothed Amber99 and Gromos43A1 MEP Grids Using the Program QFIT^a

no.	Amber99			Gromos43A1				
	closest atom	d	q	closest atom	d	q		
1	N	Gly1	0.978	-0.1401	O	Gly1	0.522	-0.2470
2	H	Gly1	1.038	0.0762	H	Gly1	0.698	0.0885
3	C α	Gly1	1.104	0.2616	H	Gly2	0.555	0.1695
4	O	Gly1	0.538	-0.2753	H	Gly3	0.441	0.1959
5	C	Gly2	0.769	0.2543	O	Gly2	0.555	-0.1772
6	O	Gly2	0.545	-0.2612	H	Gly4	0.461	0.1750
7	C	Gly3	0.803	0.2479	O	Gly3	0.568	-0.1828
8	O	Gly3	0.560	-0.2452	H	Gly5	0.463	0.1794
9	C	Gly4	0.794	0.2421	O	Gly4	0.565	-0.1778
10	O	Gly4	0.556	-0.2427	H	Gly6	0.464	0.1769
11	C	Gly5	0.798	0.2408	O	Gly5	0.567	-0.1781
12	O	Gly5	0.559	-0.2413	H	Gly7	0.465	0.1786
13	C	Gly6	0.796	0.2428	O	Gly6	0.566	-0.1773
14	O	Gly6	0.558	-0.2425	H	Gly8	0.465	0.1766
15	C	Gly7	0.797	0.2430	O	Gly7	0.567	-0.1783
16	O	Gly7	0.558	-0.2435	H	Gly9	0.465	0.1785
17	C	Gly8	0.796	0.2440	O	Gly8	0.567	-0.1771
18	O	Gly8	0.558	-0.2439	H	Gly10	0.466	0.1771
19	C	Gly9	0.797	0.2450	O	Gly9	0.567	-0.1782
20	O	Gly9	0.559	-0.2445	H	Gly11	0.465	0.1777
21	C	Gly10	0.796	0.2417	O	Gly10	0.567	-0.1773
22	O	Gly10	0.558	-0.2433	H	Gly12	0.468	0.1776
23	C	Gly11	0.796	0.2479	O	Gly11	0.567	-0.1777
24	O	Gly11	0.558	-0.2426	H	Gly13	0.466	0.1754
25	C	Gly12	0.796	0.2343	O	Gly12	0.567	-0.1764
26	O	Gly12	0.558	-0.2448	H	Gly14	0.468	0.1766
27	C	Gly13	0.796	0.2588	O	Gly13	0.567	-0.1729
28	O	Gly13	0.560	-0.2454	H	Gly15	0.436	0.1586
29	C	Gly14	0.786	0.2580	O	Gly14	0.576	-0.1802
30	O	Gly14	0.565	-0.2574	C	Gly15	0.542	0.1637
31	C	Gly15	0.585	0.3158	O	Gly15	0.665	-0.1673
32	O	Gly15	0.666	-0.2406				
rmsdV			1.33			0.69		
rmsd μ			0.16			0.15		

^a Local maxima and minima at $t = 1.25$ and 1.3 bohr^2 , respectively, were obtained using the hierarchical merging/clustering algorithm applied to the original Amber99 and Gromos43A1 MEP functions. For each point, the distance vs the closest atom, d , is given in Å. RmsdV and rmsd μ are given in kcal/mol and D, respectively. Point numbers (no.) refer to Figure 4.

1. For Amber99, positive and negative charges located near the C and O atoms of the central residue Gly8 are equal to $\pm 0.244\text{ e}^-$ and are separated by a distance of 2.52 \AA . rmsdV and rmsd μ values are equal to 1.33 kcal/mol and 0.16 D .

For Gromos43A1, the two-site CG description of each AA backbone differs. Rather than being located along the C=O axis of a residue, as in the case of Amber99, it is displaced such as the positive charge is closer to the H atom of the neighboring residue (Table 1), and the two opposite charges, equal to $\pm 0.177\text{ e}^-$, are separated by a distance of 3.66 \AA .

Our CG models are, thus, of an intermediate description level between representations that involve only one grain per AA backbone, like in the MARTINI^{20,21} and the Basdevant's⁶ models, and finer descriptions that allow to more precisely account for the various secondary structure elements of a protein.⁹ A dipolar representation of the backbone of AAs will appear to be useful in applications where the dipolar character of AAs is important, as further illustrated in the KcsA case.

C. Protein Side Chain Modeling. CG representations of each of the 20 AA side chains were obtained by considering the AAs in specific conformational states. Except for AA = Gly and Ala, most recurrent rotamers were generated by taking into account the angular constraints given in Table 2. These rotamers were selected according to their occurrence degree in protein structures as reported in the Structural Library of Intrinsic Residue Propensities (SLIRP).^{54,55} As already mentioned, from the pentadecapeptide chains $\beta\text{-Gly}_7\text{-AA-Gly}_7$ generated using SMMP05,^{51,52} only the central AA residue was kept with backbone atoms ($\text{C}\alpha-\text{C=O}_{\text{AA}}(\text{N}-\text{H})_{\text{AA+1}}$). This was achieved to avoid the generation of side chain CGs that might depend on a particular secondary structure motif. As already specified above, we considered the following protonation states: Lys(+1), Arg(+1), Glu(-1), and Asp(-1). For Gln, it appeared that both specific conformations first selected to represent classes $g-, t, \text{Ng}+$ and $g-, t, \text{Og}+$ led, through the program SMMP05, to an identical 3D structure. We, thus, kept only one structure, $g-, t, \text{Og}+$ and summed over the two initial weights reported in SLIRP to get a value of 28.6. Similarly, Gln conformations representing classes $g-, t, \text{Og}-$ and $g-, t, \text{Ng}-$ led to only one rotamer, with a total weight of 33.2. This occurred for another AA, His, for which two conformers, depicting classes $g-, \text{Ng}-$ and $g-, \text{Cg}-$, are characterized by a total weight of 35.8.

In a further step, we determined the charge values for the CG descriptions of each AA through a fitting procedure carried out using QFIT⁴⁹ vs unsmoothed all-atom MEP grids. In this procedure, and for each of the AAs, all rotamer descriptions in terms of peaks and pits observed in the Amber99 and Gromos43A1 MEPs, smoothed at $t = 1.25$ and 1.3 bohr^2 , respectively, were considered according to their occurrence probability (Table 2). This step was carried out in four stages. First, isolated AA structures were assigned atom charges using PDB2PQR.^{40,41} Side chain extrema were located using our merging/clustering algorithm. Second, the corresponding charge values were fitted vs the all-atom MEP generated from the side chain atoms only. Third, the backbone CGs were added in accordance with the motif found for Gly8 in $\beta\text{-Gly}_{15}$, and fourth, a second charge-fitting procedure, now carried out vs the MEP calculated using all the AA atoms, was achieved to determine the charge values

Table 2. Geometrical Parameters and Occurrence Probability of the Selected AA Side Chain Rotamers^{54,55} with the Exception of Ala and Gly^a

	conformation	χ_1 (°)	χ_2 (°)	χ_3 (°)	χ_4 (°)	occurrence (%)
Arg	$g-, t, g-, g-$	300	180	300	300	9.5
	$g-, t, g-, t$	300	180	300	180	11.9
	$g-, t, g+, t$	300	180	60	180	12.2
	$g-, t, t, t$	300	180	180	180	12.2
Asn	t, Nt	180	0			11.1
	$t, Og-$	180	300			21.3
	$t, Og+$	180	60			23.6
Asp	$t, g+$	180	60			62.8
Cys	$g-$	300				56.3
	$g+$	60				15.1
	t	180				28.7
Gln	$g-, t, Nt$	300	180	0		11.2
	$g-, t, Og-$	300	180	300		33.2
	$g-, t, Og+$	300	180	60		28.6
Glu	$g-, t, g-$	300	180	120		29.9
	$g-, t, g+$	300	180	60		25.3
His	$g-, Ng-$	300	300			35.8
	$t, Ng+$	180	60			15.0
Ile	$g-, g-$	300	300			22.7
	$g-, t$	300	180			28.3
Leu	$g+, t$	60	180			42.5
	$t, g+$	180	60			24.1
Lys	$g-, g-, t, g-$	300	300	180	300	8.5
	$g-, g-, t, g+$	300	300	180	60	6.5
	$g-, t, t, g-$	300	180	180	300	21.7
Met	$g-, t, t, g+$	300	180	180	60	14.3
	$g-, g-, g-$	300	300	300		15.5
	$g-, g-, t$	300	300	180		11.6
	$g-, t, g-$	300	180	300		19.4
	$g-, t, g+$	300	180	60		16.4
Phe	$g-, t, t$	300	180	180		15.4
	$g-, g-$	300	300			37.8
Pro	$t, g+$	180	60			31.5
	$g+$	0				66.8
Ser	$g-$	300				73.1
	$g+$	30				24.8
Thr	$g-$	300				51.6
	$g+$	30				46.3
Trp	$g-, g-$	300	90			28.2
	$g-, t$	300	0			16.5
	$t, g-$	180	60			11.6
	$t, g+$	180	300			13.8
	t, t	180	0			11.2
Tyr	$g-, g-$	300	120			38.3
	$t, g+$	180	60			31.7
Val	$g-$	300				46.4
	t	180				51.9

^a g and t stand for *gauche* and *trans*, respectively.

of the two backbone CGs, while preserving the side chain CG charge values first obtained.

It is to be specified that, for some AA residues, the initial peak/pit-based CG representation obtained for their side chain was replaced by a simpler model consisting of one point centered on a selected atom, as detailed below. This was achieved as a first stage in the easy design of a CG protein model from its atom coordinates retrieved from the PDB,^{56,57} as in ref 37.

In Figures 5 and 6, we report the so-obtained original or simplified CG representations for the 20 AA residues as derived from the results of our hierarchical merging/clustering algorithm applied to the Amber99 and Gromos43A1 MEP functions, smoothed at $t = 1.25$ and 1.3 bohr²,

respectively. Corresponding CG charges and deviations of the electrostatic properties vs the all-atom ones are reported in the tables provided in the Supporting Information. With Amber99 (Figure 5), all noncyclic C–H based residues, i.e., Ala, Ile, Leu, and Val, have no side chain points. This was chosen because of the low charge values obtained initially for their CGs and was an easy way to model those specific residues in possible MM applications. For Lys, we also simplified the model by setting the positive charge exactly on the $N\delta$ atom (point three). For all other AAs, the original point locations observed in the smoothed MEP functions were kept for the charge-fitting procedures. As illustrated in Figure 5, we note that for hydroxyl-containing residues, i.e., Ser, Thr, and Tyr, there are two charges located near but not exactly on the O and H atoms (points three and four for Ser and Thr; points six and seven for Tyr). A similar representation is obtained for the sulfur-containing residues with a charge close to the S atom and a charge in the neighborhood of the H atom (point four for Cys) or CH₃ group (point four for Met). For the negatively charged residues, i.e., Asp and Glu, each carboxylate functional group leads to two negative charges located near the O atoms (points three and four). Positively charged residues, Arg and Lys, present different behaviors. While the side chain of Lys leads to only one positive charge value (point three), the Arg side chain is characterized by a four-point motif (points three to six), wherein each charge is somewhat symmetrically located on the bisectors of each of the three N–C–N angles of the guanidinium group.

Regarding side chain descriptions obtained using Gromos43A1 (Figure 6), alkyl chains such as Ala, Ile, Leu, and Val do not involve any CG. This is also observed for Met. Indeed, for these AAs, the only charged atoms in the united-atom model are the backbone N, H, C, and O atoms. The CG description of Arg differs from the Amber99-based representation, as there is only one positive charge initially located in the neighborhood of the atom C δ . We have simplified the Arg CG model by fixing that CG point exactly on C δ (point three).

Let us additionally mention that for Asp, Glu, and Phe, an identity in the charge values was imposed such as $q_3 = q_4$, $q_3 = q_4$, and $q_4 = q_6$, respectively, as reported in Tables 3 and 4. One also directly notices that some charges are displaced toward the outer part of some AAs, as in Phe, where H atoms seem to be associated with charges located away from the side chain. Even if this, at first, looks unnatural, it was nevertheless decided to keep such original charge distributions, as they correspond to real topological features of the smoothed MEP functions.

A comparison between two models reported in literature and our two MEP-based CG models of the AA side chains generated from the Amber99 and Gromos43A1 sets of charges is reported in Table 5. AA residues are listed according to their properties defined in the MARTINI FF,²¹ i.e., hydrophobic residues, mainly classified as apolar, polar residues with or without hydrogen-bond-forming characteristics and charged side chains. Such a description, known to lack an electrostatic contribution, is nevertheless interesting to compare with as it involves the concept of polarity. It is

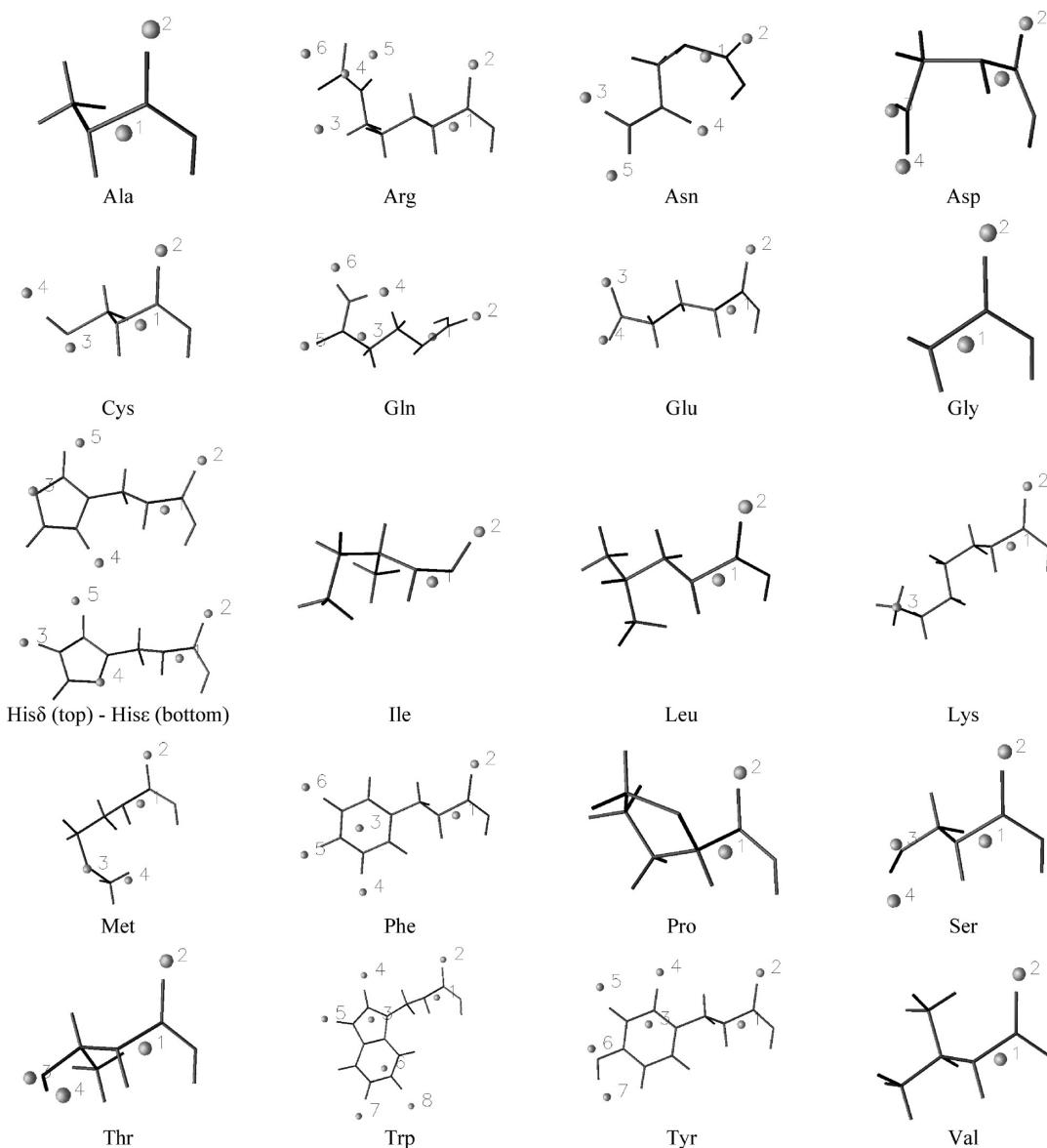


Figure 5. CG model for each of the 20 AA residues as established at $t = 1.25 \text{ bohr}^2$ from the hierarchical merging/clustering algorithm applied to the all-atom Amber99 MEP function. CG points are numbered as in the Supporting Information.

also one of the few descriptions that is easily available in the literature for all AAs. Parallelly, we report a description of the Basdevant's model,⁶ which is interesting as it is close to representations based on ED maxima described earlier.^{36,58} For all residues, the backbone CG representation consists either of one polar bead (MARTINI), one center (Basdevant), or two CGs with opposite charges (our models).

The total number of side chain CGs in each model is variable. The number of grains in the Basdevant's model is strongly dependent on the size of the side chains but does not exceed two. For MARTINI, it is higher than two only for ring-shaped side chains, i.e., Phe, His, Trp, and Tyr. In the case of our MEP-based CG representations, there are up to six CGs for Trp. For all small hydrophobic residues, the MARTINI CG representations involve only one apolar grain. Parallelly, in both MEP-based models, there is no side chain CGs. For Phe, our models involve a large number of points, i.e., four for both Amber99 and Gromos43A1. The charge brought by each of the side chain CGs of Phe stays low,

with $|q| < 0.10 \text{ e}^-$. Sulfur-containing residues, especially Cys, that are hydrophobic and do not form any H-bond, are however characterized by a dipole moment. In MARTINI, they are, thus, represented by one CG with the intermediate apolar/polar state. For the Amber99- and Gromos43A1-based models, there are two CGs with opposite charges. Regarding Asn and Gln, our MEP-based models provide a finer description of the side chains, with three grains located at the vicinity of the O and H atoms (Figures 5 and 6). In MARTINI, these side chains are represented using one grain characterized by a polar type with a hydrogen-bonding donor and acceptor character. For all residues containing an O–H group, i.e., Ser, Thr, and Tyr, our models include at least two opposite charges located in the neighborhood of O and H; they correspond to one polar group in MARTINI. The side chains of His and Trp not only contain hydrophobic rings but also hydrogen-bonding properties. In the framework of our MEP-based models, they are represented by CGs with a dipole occurring between $\text{HN}\delta$ and $\text{N}\epsilon$ in $\text{His}\delta$, between

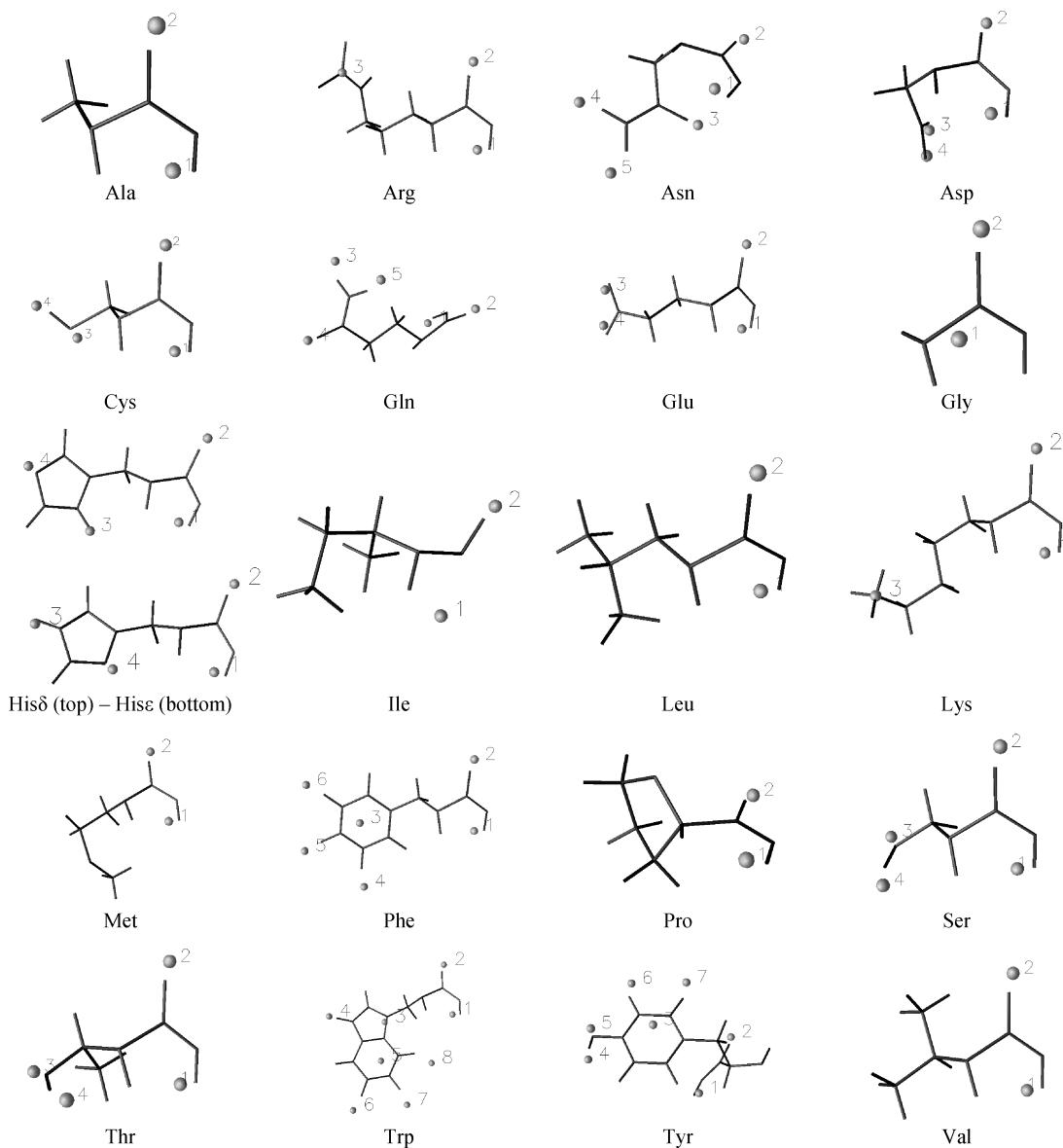


Figure 6. CG model for each of the 20 AA residues as established at $t = 1.3 \text{ bohr}^2$ from the hierarchical merging/clustering algorithm applied to the Gromos43A1 MEP function. CG points are numbered as in the Supporting Information.

HN ϵ and N δ in His ϵ , and between HN ϵ 1 and the rings in Trp. Details regarding point charge locations and values are given in the Supporting Information. The polarity property in MARTINI is, thus, expressed as a charge separation in our models. Finally, regarding the residues that are explicitly charged in the MARTINI FF, we observe a finer description of the negative Asp and Glu residues in our models, with two separate negative charges close to the O of the carboxyl group. The Amber99-based CG model of Arg is rather interesting and original as it involves three positive charges almost symmetrically spread around the atom C ζ , itself associated with a fourth CG charge (Figure 5). This might be seen as a description that is more consistent with a charge delocalization.

Thus, on the average, one can consider that there is a reduction ratio of about 4.5/1 between the CG and all-atom MEP-based models. Knowing that the calculation time evolves according to N^2 and $N \log N$, N being the number of particles for the Coulomb pair potential and the particle mesh

Ewald (PME) algorithm, respectively, one expects reduction ratios of CPU times of about 20 for the Coulomb potential and, for example, about 5 for the PME routine in the case of a protein with 100 000 atoms vs the all-atom representation.

D. Automated CG Generation Procedure. To study systems that are larger than oligopeptides, an automation stage was developed to avoid the lengthy generation of the CGs for each AA separately, as first carried out.³⁷ The resulting automated procedure was fully based on the application of a superimposition algorithm of CG motif templates of each AA onto the corresponding AA structures of the protein under study. We used the program QUATFIT^{59,60} to, first, superimpose a limited set of atoms from the template on the studied structure and then used the resulting transformation matrix to generate the corresponding CG coordinates.

The templates that were selected in this study are described in Tables 3 and 4 for the Amber99 and Gromos43A1 FFs, respectively. Their size consisted of at least three atoms so

Table 3. Template Coordinates (in Å) and Charges (in e⁻) as Used for the Amber99-Based CG Generation

	X	Y	Z	charge		X	Y	Z	charge
backbone									
C	22.575	13.923	2.131						
O	23.021	13.167	2.993						
N	23.318	14.688	1.345						
PT1	21.839	14.199	1.699	<i>q1^a</i>					
PT2	23.280	12.855	3.318	<i>q2^a</i>					
side chain									
ARG									
N _ε	18.561	15.333	5.213		HIS_ε				
C _ζ	18.087	14.123	5.542		C _γ	18.921	15.161	2.698	
NH1	17.049	13.606	4.871		Nδ1	18.437	15.752	1.543	
NH2	18.651	13.432	6.542		Cε1	17.116	15.812	1.626	
PT3	16.526	14.616	3.638	0.2780	Nε2	16.744	15.267	2.820	
PT4	18.195	14.221	5.594	0.0555	Cδ2	17.833	14.874	3.469	
PT5	19.530	14.832	6.483	0.4811	PT3	15.177	15.203	3.248	0.1803
PT6	17.112	12.282	5.831	0.2215	PT4	18.512	15.718	1.511	-0.2142
ASN					PT5	17.525	14.157	5.209	0.0699
C _γ	21.154	16.268	3.071		PHE				
Oδ1	22.355	16.412	3.224		C _γ	18.926	14.982	2.956	
Nδ2	20.298	17.275	2.917		Cδ1	18.399	15.648	1.894	
HNδ21	20.637	18.192	2.709		Cε1	16.993	15.816	1.788	
HNδ22	19.315	17.114	3.009		C _ζ	16.173	15.310	2.748	
PT3	18.536	16.868	2.689	0.1470	Cε2	16.700	14.643	3.810	
PT4	22.863	16.544	3.245	-0.2340	Cδ2	18.106	14.476	3.916	
PT5	20.335	19.138	2.987	0.0820	PT3	17.237	15.111	2.852	-0.0753
ASP					PT4	16.307	16.902	0.127	0.0425
C _γ	21.094	16.293	3.152		PT5	14.140	15.609	2.572	0.0093
Oδ1	20.959	17.047	2.164		PT6	15.393	13.922	5.289	0.0425
Oδ2	21.670	16.583	4.223		SER				
PT3	21.800	16.833	4.393	-0.4290	C _β	20.443	14.916	2.987	
PT4	21.037	17.390	2.124	-0.4290	O _γ	19.047	15.112	2.779	
CYS					H _γ	18.754	14.579	1.988	
C _β	20.432	14.938	2.963		PT3	18.804	15.397	3.193	-0.0997
S _γ	18.650	15.161	2.610		PT4	18.739	14.433	1.036	0.1547
H _γ	18.127	14.540	3.187		THR				
PT3	18.679	15.325	2.115	-0.0705	C _β	20.358	14.870	2.971	
PT4	17.875	13.550	3.798	0.0515	O _γ 1	19.001	14.913	2.536	
GLN					H _γ 1	18.880	14.363	1.701	
C _γ	18.930	15.114	2.699		PT3	18.660	15.285	2.856	-0.1157
C _δ	18.288	16.002	3.767		PT4	19.161	13.795	1.024	0.1682
O _ε 1	17.102	16.285	3.744		TRP				
N _ε 2	19.135	16.423	4.701		C _γ	18.914	15.167	2.725	
HN _ε 21	19.656	15.761	5.240		Cδ1	17.884	14.443	3.185	
HN _ε 22	19.252	17.403	4.866		Nε1	16.676	14.963	2.769	
PT3	18.923	15.585	3.173	0.1768	Cε2	16.950	16.087	1.998	
PT4	20.178	15.240	5.492	0.0958	C _ζ 2	16.065	16.958	1.352	
PT5	16.632	16.505	3.781	-0.3215	CH2	16.638	18.008	0.644	
PT6	19.738	18.140	5.034	0.0818	C _ζ 3	18.021	18.142	0.611	
GLU					Cε3	18.919	17.280	1.251	
C _δ	18.288	16.002	3.767		Cδ2	18.322	16.220	1.965	
O _ε 1	17.754	17.063	3.377		PT3	17.891	15.125	2.830	-0.1380
O _ε 2	18.345	15.599	4.949		PT4	18.246	12.882	4.318	0.0963
PT3	18.270	15.634	5.269	-0.4410	PT5	15.128	14.269	3.159	0.1002
PT4	17.622	17.347	3.562	-0.4410	PT6	17.745	17.637	0.978	-0.0640
HIS_δ					PT7	15.309	19.357	-0.386	0.0284
C _γ	18.921	15.161	2.698		PT8	18.696	19.982	-0.699	0.0068
Nδ1	18.437	15.752	1.543		TYR				
Cε1	17.116	15.812	1.626		C _ζ	16.164	15.242	2.786	
Nε2	16.744	15.267	2.820		OH	14.815	15.390	2.691	
Cδ2	17.833	14.874	3.469		HH	14.590	15.887	1.852	
PT3	16.613	15.212	2.909	-0.2306	Cδ1	18.097	14.402	3.947	
PT4	19.443	16.282	0.260	0.1863	Cε1	16.669	14.558	3.847	
PT5	18.531	14.300	4.885	0.0523	Cε2	16.946	15.786	1.815	
MET					Cδ2	18.374	15.629	1.915	
C _γ	18.937	15.046	2.800		PT3	17.673	14.741	3.465	0.0406
S _δ	18.639	15.869	1.245		PT4	18.866	13.489	5.441	0.0498
C _ε	19.506	17.402	1.535		PT5	15.831	13.569	5.556	0.0115
PT3	18.579	15.903	1.151	-0.0654	PT6	14.650	15.117	3.194	-0.1535
PT4	20.535	18.072	2.000	0.1154	PT7	14.781	16.345	1.047	0.1610

^a Values of q1 and q2 depend on the AA type (Table SI1 in the Supporting Information).

Table 4. Template Coordinates (in Å) and Charges (in e⁻) as Used for the Gromos43A1-Based CG Generation

	X	Y	Z	charge		X	Y	Z	charge
backbone									
C	22.575	13.923	2.131						
O	23.021	13.167	2.993						
N	23.318	14.688	1.345						
PT1	22.478	15.192	0.748	<i>q1^a</i>					
PT2	23.357	12.902	3.371	<i>q2^a</i>					
side chain									
ASN					PHE				
C γ	21.154	16.268	3.071		C δ 1	18.399	15.648	1.894	
O δ 1	22.355	16.412	3.224		C ϵ 1	16.993	15.816	1.788	
N δ 2	20.298	17.275	2.917		C ζ	16.173	15.310	2.748	
HN δ 21	20.637	18.192	2.709		C ϵ 2	16.700	14.643	3.810	
HN δ 22	19.315	17.114	3.009		C δ 2	18.106	14.476	3.916	
PT3	22.823	16.442	3.455	-0.1910	PT3	17.269	15.144	2.896	-0.0936
PT4	18.479	17.146	2.969	0.1104	PT4	16.392	16.836	0.108	0.0386
PT5	20.388	19.047	2.427	0.0840	PT5	14.145	15.597	2.521	0.0163
ASP					PT6	15.379	13.928	5.206	0.0386
C γ	21.094	16.293	3.152		SER				
O δ 1	20.959	17.047	2.164		C β	20.443	14.916	2.987	
O δ 2	21.670	16.583	4.223		O γ	19.047	15.112	2.779	
PT3	21.645	16.781	4.010	-0.5000	H γ	18.568	14.235	2.826	
PT4	21.142	17.151	2.618	-0.5000	PT3	20.331	14.350	4.599	-0.1466
CYS					PT4	18.819	12.975	3.200	0.1466
C β	20.432	14.938	2.963		THR				
S γ	18.650	15.161	2.610		C β	20.358	14.870	2.971	
H γ	18.127	14.540	3.187		O γ 1	19.001	14.913	2.536	
PT3	18.642	15.763	2.474	-0.0299	H γ 1	18.880	14.363	1.701	
PT4	17.761	14.191	3.407	0.0299	PT3	18.739	15.245	3.017	-0.1459
GLN					PT4	19.133	13.940	0.949	0.1459
C δ	18.288	16.002	3.767		TRP				
O ε 1	17.102	16.285	3.744		C γ	18.914	15.167	2.725	
N ε 2	19.135	16.423	4.701		C δ 1	18.147	14.690	1.736	
HN ε 21	19.656	15.761	5.240		N ε 1	16.840	15.119	1.858	
HN ε 22	19.252	17.403	4.866		C ϵ 2	16.769	15.917	2.994	
PT3	19.737	18.118	5.180	0.0850	C ζ 2	15.668	16.584	3.544	
PT4	16.715	16.386	3.841	-0.2031	CH2	15.904	17.316	4.702	
PT5	20.244	15.453	5.740	0.1182	C ζ 3	17.183	17.352	5.245	
GLU					C ϵ 3	18.294	16.690	4.708	
C δ	18.288	16.002	3.767		C δ 2	18.038	15.953	3.534	
O ε 1	17.754	17.063	3.377		PT3	18.602	15.283	2.908	-0.1232
O ε 2	18.345	15.599	4.949		PT4	15.743	14.810	0.939	0.1553
PT3	18.190	15.902	4.841	-0.5000	PT5	17.181	16.729	4.389	-0.1409
PT4	17.773	16.922	3.755	-0.5000	PT6	14.277	18.300	5.454	0.0466
HISδ					PT7	17.248	18.565	6.949	0.0319
C ε 1	17.116	15.812	1.626		PT8	19.915	17.376	5.998	0.0303
N ε 2	16.744	15.267	2.820		TYR				
C δ 2	17.833	14.874	3.469		C ζ	21.930	18.777	3.647	
PT3	18.980	16.087	0.718	0.2623	OH	22.401	20.040	3.824	
PT4	16.431	15.172	3.065	-0.2623	HH	23.280	20.139	3.355	
HISε					C δ 1	20.825	17.132	2.281	
C γ	18.921	15.161	2.698		C ε 1	21.324	18.469	2.469	
N δ 1	18.437	15.752	1.543		C ε 2	22.086	17.879	4.656	
C ε 1	17.116	15.812	1.626		C δ 2	21.587	16.541	4.468	
PT3	15.789	15.256	2.998	0.2729	PT3	21.369	17.439	4.089	0.0148
PT4	18.719	15.839	1.336	-0.2729	PT4	24.024	20.132	2.845	0.1523
					PT5	21.930	20.166	4.149	-0.1988
					PT6	23.204	17.862	6.329	0.0153
					PT7	21.719	15.562	6.013	0.0165

^a Values of *q1* and *q2* depend on the AA type (Table SI2 in the Supporting Information).

as to generate unique superposition results, i.e., CG coordinates. For rigid side chains, such as His, Phe, and Trp, more than three atoms were also used to better fit the whole side chain plane. For Arg, more than three atoms were also used to generate, at once, all CGs, within the frame of the Amber99 FF. For Gln, points four and six of the Amber99-based CG representation were generated using the template formed by atoms N ε 2, HN ε 21, and HN ε 22, while points

three and five were determined using atoms C γ , C δ , and N δ 2. The Gromos43A1 CG model of the Gln side chain contained only three points. Points three and five were generated using the template formed by N ε 2, HN ε 21, and HN ε 22, while the location of point four was based on atoms C γ , C δ , and N δ 2. Similarly, for Asn, points three and five for Amber99 (or points four and five for Gromos43A1) were generated using the template formed by atoms N δ 2, HN δ 21,

Table 5. Descriptions^a of Protein Side Chain CG Models, as Defined in MARTINI, in Basdevant's model, and As Obtained from the Hierarchical Merging/Clustering of MEP Functions^b

	MARTINI ^{20,21}	Basdevant ⁶	Amber99	Gromos43A1
Gly	—	—	—	—
small hydrophobic residues				
Ala	—	1	—	—
Ile	1 apolar	1	—	—
Leu	1 apolar	1	—	—
Pro	1 apolar	1	—	—
Val	1 apolar	1	—	—
large hydrophobic residue				
Phe	3 apolar	2	4 $ q < 0.08 \text{ e}^-$	4 $ q < 0.10 \text{ e}^-$
sulfur-containing residues				
Cys	1 apolar/polar	1	2	2
Met	1 apolar/polar	2	2	—
polar amide-containing residues with H-bond property				
Asn	1 polar	1	3	3
Gln	1 polar	2	4	3
small hydrophilic residues with OH group				
Ser	1 polar	1	2	2
Thr	1 polar	1	2	2
ring-shape hydrophobic residues with H-bond property				
His	1 apolar, 2 polar	2	3	2
Trp	3 apolar, 1 polar	2	6	5
Tyr	2 apolar, 1 polar	2	5	5
charged residues				
Arg	1 apolar/polar, 1 charged	2	4	1
Asp	1 charged	1	2	2
Glu	1 charged	2	2	2
Lys	1 apolar, 1 charged	2	1	1

^a Descriptions are in terms of number and property. ^b At $t = 1.25 \text{ bohr}^2$ using Amber99 and $t = 1.3 \text{ bohr}^2$ using Gromos43A1.

and HN δ 22, while point four for Amber99 (or point three for Gromos43A1) was obtained using C γ , O δ 1, and O ε 1. In the case of Tyr, points three to five were located using the template formed by ring atoms C ζ , C ϵ 1(or C ϵ 2), and C δ 1(or C δ 2) in the opposite direction to the O–H bond, while points six and seven were generated using atoms C ζ , OH, and HH. A similar procedure is valid for the Gromos43A1-based models of Tyr. For the AA residues that are not reported in Tables 3 and 4, the CG coordinates were directly obtained from the side chain atom coordinates as specified in the tables reported in the Supporting Information.

Thus, from Tables 3 and 4, it is clear that CG points and charges can be directly obtained from a high-resolution structure/map of a protein, more precisely, from 3D atomic coordinates. For some AAs, like Asn, Cys, Gln, Ser, Thr, and Tyr, templates involve the knowledge of H atomic coordinates. Presently, these atoms were not defined in the PDB files but were added through the use of a software, such as PDB2PQR^{40,41} and SwissPDBViewer.^{43,44} In the case of lower crystallographic resolution maps, only a limited number of ED maxima could be located. In previous papers,^{36,58} we showed how regular motifs of ED peaks still characterize AA backbone and side chains at resolution values close to 3 Å. A deeper study would be needed to relate the topology-based properties of these ED maxima, i.e., location, main ED curvatures, and local eigenvectors, to the positioning of the CG charges reported in Tables 3 and 4. Indeed, at the location of each ED maximum, a so-called Hessian matrix, built on the second derivatives of the density function vs the position, can be calculated. The

diagonalization of such a matrix provides three eigenvalues, which physically define the main curvatures of the density function at the peak location and the three corresponding eigenvectors. Their orientation can help in locating CG charges.

E. Application to Small Peptides. Four small peptidic structures with electrostatic properties reported in the literature were selected. The first structure, a 12-residue β -hairpin HP7 was retrieved from the PDB^{56,57} (PDB code 2EVQ) following the work of Basdevant et al.⁶ The primary structure of that peptide is Lys–Thr–Trp–Asn–Pro–Ala–Thr–Gly–Lys–Trp–Thr–Glu, with a global net charge of +1. It is an interesting reference structure because a fragment-based description, as well as the corresponding point charges, were provided.⁶ In that representation, each pseudoatom is defined as the geometric center of the heavy atoms of a protein fragment. Structure of two other peptides, i.e., the Tgn38 internalization peptide Dyqrln, with sequence Asp–Tyr–Gln–Arg–Leu–Asn, (PDB code 1BXX) and the C-terminal fragment of the chemotaxis receptor, with sequence Asn–Trp–Glu–Thr–Phe, (PDB code 1BC5) were studied following the work of Exner and Mezey.⁶¹ Additionally, we selected the structure of a phospholipase inhibitor, with sequence Leu–Val–Phe–Phe–Ala, (PDB code 2RD4) involved in the A β 7 structure studied by Pizzitutti et al.⁷

For each of those peptides, CG models were obtained by applying the automated procedure specified above. End charges were considered by including two additional charges, one on each of the terminal atoms N and OXT. By default, the corresponding charge values were set equal to ± 1 . The

Table 6. Electrostatic Properties of the Amber99-based CG Model of Small Peptides vs Their Corresponding All-Atom Version^a

		charge fitting			
		none ^c	q_{end} only ^d	all CG ^e	Basdevant ^d
2EVQ	197 atoms				
q	1.0				
μ (all-atom) ^b	4.78, -2.21, -66.43				
no. of CGs	51	51	51	28	
rmsdV	5.98	3.63	1.54	5.27	
rmsd μ	9.29	2.00	0.25	1.37	
μ^b	1.50, 2.46, -73.76	3.54, -2.99, -67.81	4.67, -2.01, -66.33	4.89, -2.49, -65.10	
q_{end}	± 1.0000	± 0.7660			
1BXX	110 atoms				
q	0.0				
μ (all-atom) ^b	11.34, -0.96, -15.73				
no. of CGs	32	32	32	15	
rmsdV	6.03	3.52	1.90	7.09	
rmsd μ	12.65	1.79	0.34	2.47	
μ^b	21.06, 6.64, -12.96	12.31, -1.51, -14.33	11.62, -0.82, -15.60	13.55, -0.14, -14.99	
q_{end}	± 1.0000	± 0.8532			
1BC5	90 atoms				
q	-1.0				
μ (all-atom) ^b	-310.92, -287.53, 7.06				
no. of CGs	29	29	29	13	
rmsdV	5.73	3.02	1.72	8.59	
rmsd μ	9.02	1.49	0.09	4.82	
μ^b	-308.68, -296.26, 6.90	-312.11, -286.69, 6.75	-310.98, -287.55, 7.12	-311.29, -282.99, 5.50	
q_{end}	± 1.0000	± 0.8589			
2RD4	88 atoms				
q	0.0				
μ (all-atom) ^b	35.12, 22.63, -44.04				
no. of CGs	20	20	20	12	
rmsdV	4.40	3.25	1.62	8.74	
rmsd μ	5.80	2.60	0.60	3.51	
μ^b	39.38, 22.10, -47.94	35.71, 20.10, -43.84	34.66, 22.26, -44.14	34.98, 22.72, -40.53	
q_{end}	± 1.0000	± 0.9162			

^a RmsdV and rmsd μ are given in kcal/mol and D, respectively. Electric charges are given in e⁻. ^b X, y, and z components of μ . ^c No charge-fitting applied. ^d Charge-fitting applied to end charges q_{end} only. ^e Charge-fitting applied to all CG charges. ^f Charge-fitting applied to Basdevant's model.

quality of the Amber99-based CG model is evaluated vs the all-atom one in Table 6. It is achieved in terms of the rmsdV and rmsd μ deviation values. When no charge fitting is applied, models 1BC5 and 2RD4 approximate fairly well the all-atom electrostatic properties, while models 2EVQ and 1BXX are less well reproduced, especially at the level of the dipole moment values. For example, the sign of μ_y is inverted. While keeping all charges constant but the two end ones, we then applied a charge-fitting procedure that led to better models, with end charges q_{end} lower than unity. Models, and especially their dipole approximation, can, thus, be largely improved by fixing the end charges to absolute values lower than 1 e⁻. However, we consider this improvement as the reflect only of altered end charges, i.e., modifications that depend on the particular protein structure under study. It is rather artificial to modify two AA models to approximate a global protein property, such as its dipole. The fitting of all CG charges obviously leads to largely better approximations with rmsdV and rmsd μ values ranging between 1.54–1.90 kcal/mol and 0.09–0.60 D, respectively. For comparison, we also fitted the charge values of the Basdevant's representation, with less efficiency, i.e., rmsdV = 5.27 to 8.74 kcal/mol and rmsd μ = 1.37 to 4.82 D. A

similar discussion is valid for the Gromos43A1-based CG model (results are provided in the Supporting Information). However, let us mention that keeping q_{end} to unit values is observed to be a good choice for that particular set of charges. Indeed, for each of the four peptides studied, the fitting of the end charges led to absolute q_{end} values ranging between 1.02 and 1.06 e⁻. That model, thus, appears to be characterized by more robust transferability properties than that of the Amber99-based one. It is assumed that Amber99-based CG models would provide better approximations of the all-atom representations if established at a less drastic smoothing degree. In conclusion, one observes, from the rmsdV and rmsd μ values, that the use of our model provides a good approximation of the all-atom MEPs, especially when end charges are fitted. A point charge model based on the com of the AA side chains and backbones is efficient too but requires a charge-fitting step that is not needed in our case. Additionally, the charge values that would be obtained using a full charge-fitting procedure are strongly dependent on the 3D conformation of the molecule. As illustrated in Figure 7 for Amber99, wherein CG-based MEP isocontours and projected values of the MEP onto the 0.0002 e⁻/bohr³ ED isosurface are compared vs the corresponding all-atom

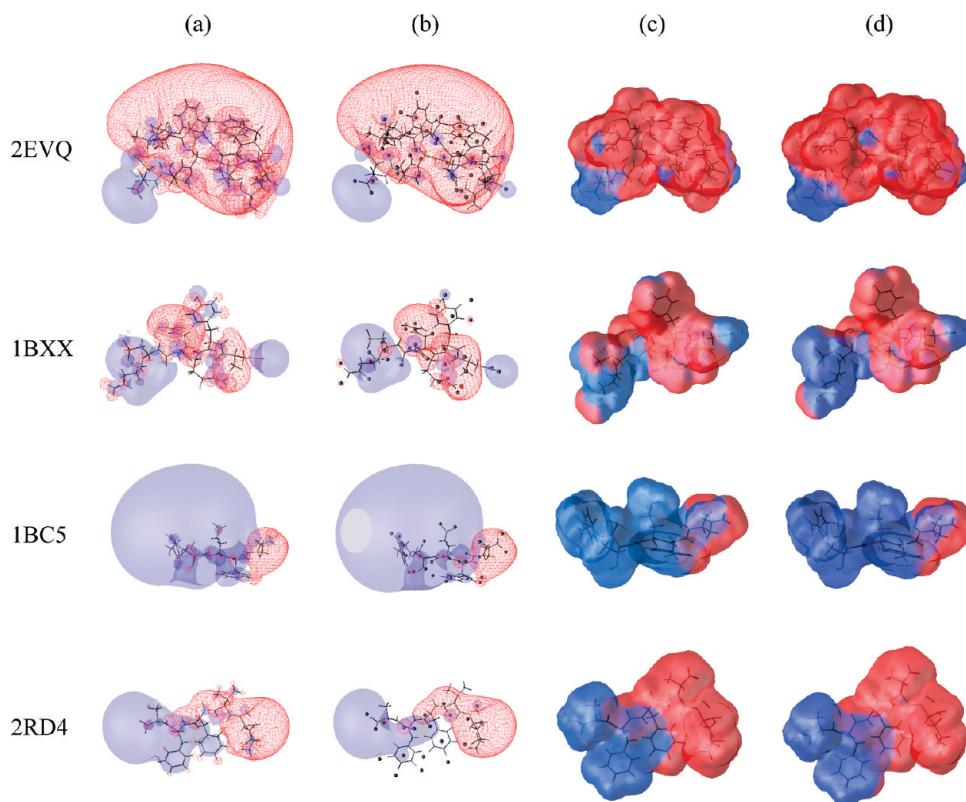


Figure 7. Amber99 MEP isocontours (blue plain surface: -0.07 , red mesh: $0.07 \text{ e}^-/\text{bohr}$) and MEP projected on the ED surface defined at $0.0002 \text{ e}^-/\text{bohr}^3$ (blue: negative, red: positive) of peptides 2EVQ, 1BXX, 1BC5, and 2RD4. (a) Unsmoothed all-atom MEP, (b) CG with fitted q_{end} MEP with CGs (black spheres), (c) all-atom MEP on ED isocontour, and (d) CG with fitted q_{end} MEP on ED isocontour.

properties, only local differences are clearly visible, especially at the close proximity of the molecular structure. There is a good correspondence between the CG and all-atom MEP 3D properties. A similar discussion is valid for the Gromos43A1 results provided in the Supporting Information.

F. Application to the Potassium Ion Channel KcsA. The protein structure selected to test our automated procedure was the KcsA potassium channel (Figure 8), a transmembrane protein structure that is commonly used to model biological ion channels^{22,62–64} as well as to evaluate computational approaches in the study of protein electrostatics.^{65–67} It is formed by four identical chains, each chain containing two α -helices connected by a loop located in the channel region (Figure 8). The channel consists of the so-called selectivity filter, that is about 18 \AA long, pointing to the extracellular region, a larger cavity of about 10 \AA and a 15 \AA long narrow gating pore opened toward the intracellular region. The gating pore and the cavity are hydrophobic regions, while the selectivity filter, mainly formed by five residues Thr74–Thr75–Val76–Gly77–Tyr78, is covered by in-line carbonyl O atoms of the protein backbone, which build a structure that is similar to a water solvation shell around a K^+ ion.

In the present work, the 3D model of the protein was prepared according to the X-ray crystal structure of the KcsA K^+ channel (PDB access code 1BL8) by adding missing side chain atoms using the program SwissPDBViewer.^{43,44} The design of the His residues into a His_e configuration was achieved with the program VEGA ZZ.^{68,69} The three K^+

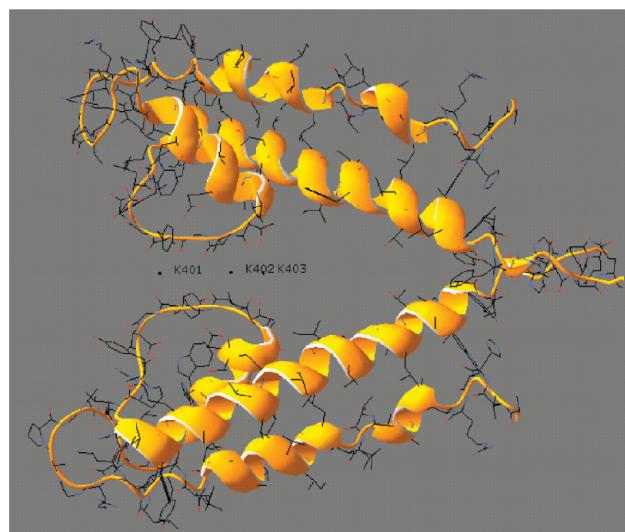


Figure 8. 3D conformation and secondary structure of the potassium channel KcsA (PDB code 1BL8). Two monomers only, chains A and C, are displayed. Figure was generated using SwissPDBViewer.^{43,44} Ions K401 and K403 are separated by a distance of 10.62 \AA .

ions, labeled K401, K402, and K403 (Figure 8), were not considered. Atom charges were assigned using PDB2PQR.^{40,41} From an original structure of 5 888 atoms, the application of our automated procedure, completed by the addition of unit charges on the N and OXT atoms of the end residues of each of the four monomers, led to the generation of 1 284 and 1 204 CGs in the frameworks of the Amber99 and

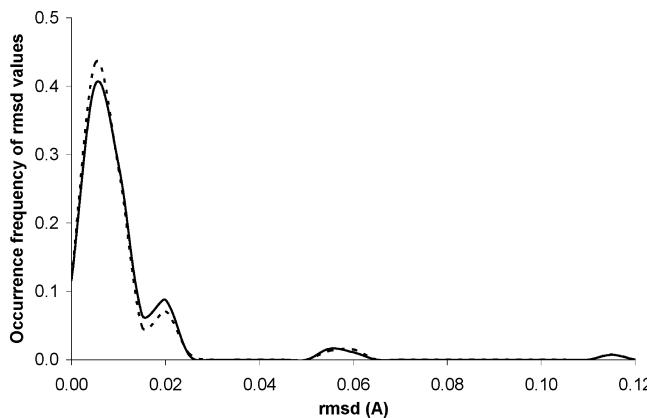


Figure 9. Occurrence frequency of the rmsd values calculated between the atom positions of the AA template motif and the atom positions of the actual AA backbones or side chains, over all superimpositions achieved for the generation of the Amber99-based (plain line) and Gromos43A1-based (dashed line) CGs of protein structure KcsA.

Gromos43A1 FFs, respectively. The obtained reduction ratios, slightly larger than 4.5/1, is close to the 4/1 value reported by Bond and Sansom,^{70,71} who studied the interaction of membrane proteins with lipid molecules through MD simulations. A visualization of the rmsd values obtained between the atoms of the AA templates and the corresponding atoms of the protein crystal structure, for each of the superimpositions achieved using QUATFIT^{59,60} during the CG generation, is presented in Figure 9. The largest rmsd values, i.e., beyond 0.1 Å, correspond to a less efficient fit of the four end residues Gln119 required to design the Amber99-based CG model due to the terminal OXT atoms (Figure 10, left). The lowest rmsd values, around 0.01 Å, characterize the superimpositions of the backbone templates, while all larger rmsd values, from 0.02 to 0.06 Å, characterize the superimpositions of the side chain templates. Particularly, rmsd values around 0.05–0.06 Å originate from the superimpositions of the Tyr side chains. For example, Tyr82 of chain A that led to rmsd = 0.057 Å is illustrated in Figure 10 (right), where one can see that it nevertheless corresponds to a rather good superimposition of the three template atoms C_α, OH, and HH.

The resulting full KcsA CG models are characterized by dipole moments and total charges that are reported in Table

7, both for the Amber99 and Gromos43A1 FFs. In the case of the Amber99-based model, as the number of CGs is too large to allow any charge fitting procedure, we simply modified the end charge values q_{end} and observed that $q_{\text{end}} = 0.5 \text{ e}^-$ provided a model characterized by deviation values rmsdV and $\text{rmsd}\mu$ that are equal to 8.28 and 0.58 D, respectively. It is to be compared to the values of 7.38 kcal/mol and 81.60 D obtained when $q_{\text{end}} = \pm 1 \text{ e}^-$ is used (Table 7). The original MEP grid values were best approached when $q_{\text{end}} = 0.8 \text{ e}^-$ with a lower $\text{rmsdV} = 6.13 \text{ kcal/mol}$ but this, however, led to $\text{rmsd}\mu = 48.94 \text{ D}$, a value that is acceptable considering the magnitude of the dipole moment, i.e., 1411.36 D.

Visualizations of 3D MEP isocontours, generated from MEP maps built with a grid step of 0.5 Å (Figure 11), do not permit to clearly differentiate the MEPs calculated using the original sets of charges (Figure 11, left) from those calculated using the CG models (Figure 11, right). Finer and more quantitative comparisons were, thus, achieved. MEP profiles were calculated using the original atom charges along the channel axis, defined by the Cartesian coordinates of ions K401 and K403 (Figure 12). As illustrated, the channel axis region of the selective filter region is characterized by two MEP minima, followed by a large energy barrier which covers the hydrophobic cavity and narrow pore regions. The calculation of the corresponding MEP profiles using the Amber99- and Gromos43A1-based CG models generate similar behaviors, very close to their all-atom version. In that sense, the models presented in this paper led to better approximations than those obtained in a previous approach,³⁷ wherein AA CG models were generated using pentadecapeptide structures rather than isolated structures. It seems that decoupling backbone and side chain contributions in the elaboration of a CG model is interesting for reproducing all-atom electrostatic properties.

IV. Conclusions and Perspectives

In this work, we applied a hierarchical merging/clustering algorithm to molecular scalar fields, like molecular electrostatic potential (MEP) functions, to generate coarse point charge representations of proteins. Through the use of such a procedure, the reduction of a molecular structure representation, particularly a protein structure, was achieved by

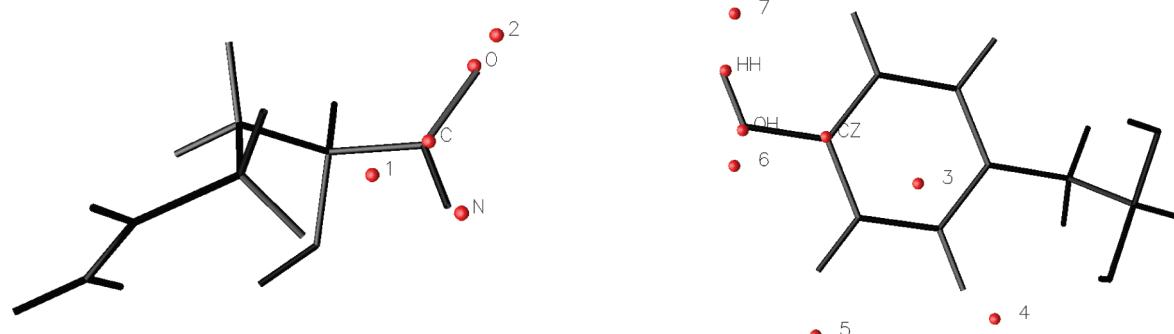


Figure 10. (Left) Amber99-based template motif (red spheres) of the Gln backbone as superimposed on Gln119 of chain A in protein KcsA. The three atoms C, O, and N are used to generate the transformation matrix that is further applied to CGs numbered 1 and 2. (Right) Amber99-based template motif of the Tyr side chain as superimposed on Tyr82 of chain A in protein KcsA. The three atoms C_α, OH, and HH are used to generate the transformation matrix that is further applied to CG points numbered 3 to 7.

Table 7. Electrostatic Properties of the Amber99- and Gromos43A1-Based CG Models of Structure KcsA vs Their Corresponding All-Atom Version^a

	Amber99	Gromos43A1
total charge	4.0	4.0
μ (D)	1411.36	1402.76
μ all-atom (D) ^b	1303.10, 511.49, 179.54	1295.8, 511.82, 163.31
no. of CG points	1 284	1 204
reduction factor	4.6/1	2.1/1 vs charged atoms
$q_{\text{end}} = \pm 1.0$	μ CG (D) ^b rmsdV (kcal/mol) rmsd μ (D)	1273.17, 512.15, 103.63 7.38 81.60
		1293.14, 512.66, 156.49 2.78 7.37

^a RmsdV and rmsd μ are given in kcal/mol and D, respectively. Electric charges are given in e⁻. ^b X, y, and z components of μ .

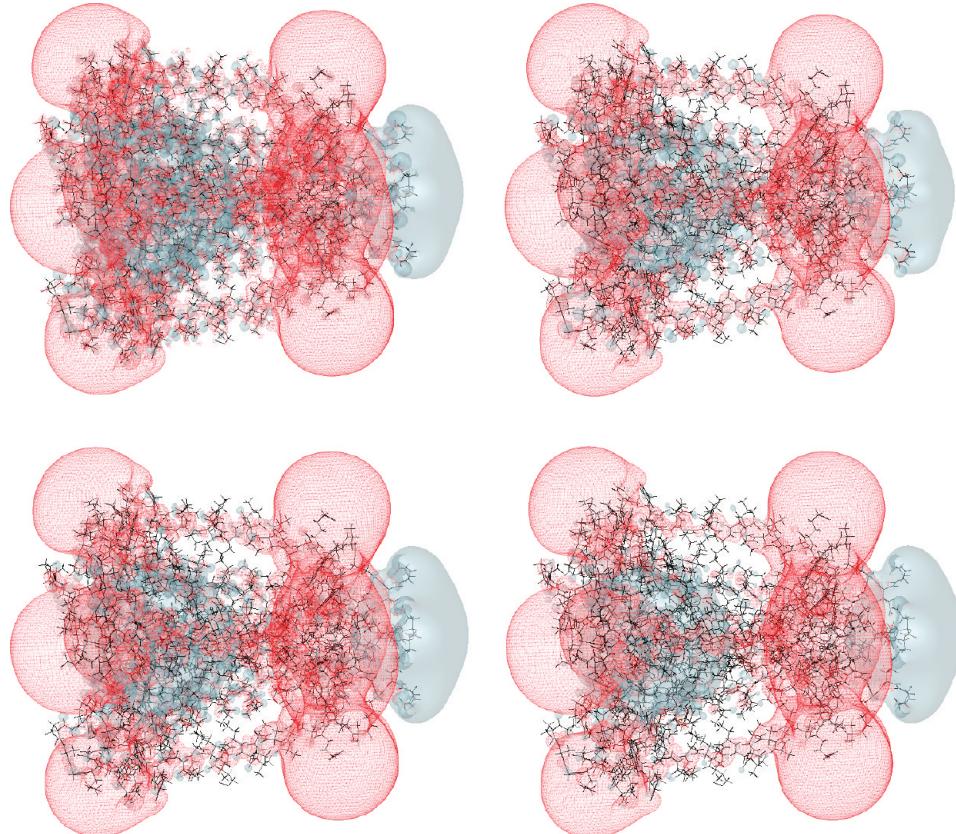


Figure 11. MEP isocontours (blue plain surface: -0.1 , red mesh: $0.1 \text{ e}^-/\text{bohr}$) of (top left) unsmoothed all-atom Amber99, (top right) Amber-based CG with $q_{\text{end}} = \pm 0.5 \text{ e}^-$, (bottom left) unsmoothed united-atom Gromos43A1, and (bottom right) Gromos43A1-based CG with $q_{\text{end}} = \pm 1.0 \text{ e}^-$, superimposed on the 3D structure of protein KcsA (sticks).

following the trajectories of its constituting atoms in its progressively smoothed three-dimensional (3D) molecular field. A protein structure can, thus, be described by a limited set of points, which correspond to the local extrema (peaks and pits) of the considered 3D MEP field. The aim of such calculations further consisted in the evaluation of electrostatic properties, such as point charges and dipole moments, of a protein using coarse-grain (CG) descriptions.

The present work especially focused on the use of the sets of charges of the all-atom Amber99 and the united-atom Gromos43A1 force fields (FF) but is readily applicable to other charge sets that are available in the literature. Reduced descriptions were obtained for each of the 20 natural amino acid (AA) residues with the following specific protonation states: Arg(+1), Lys(+1), Asp(-1), and Glu(-1). Each of the 20 AAs was modeled through various rotamers (except for Ala, Asp, Gly, and Pro). The first stage was to apply our

merging/clustering algorithm to determine the CG locations of the AA backbone and side chain, separately. In a second stage, charges were assigned to these AA CG representations through a charge-fitting algorithm and were further tabulated as reference values to be used for CG modeling of protein structures. MEP-based CG descriptions were shown to be sensitive to the molecular conformation. Additional studies, achieved at various levels of smoothing, showed that the optimal value of t is only slightly dependent on the selected FF charges. It is equal to 1.25 and 1.3 bohr² for Amber99 and Gromos43A1, respectively.

An automated procedure was implemented and tested on four small peptides (PDB access codes 2EVQ, 1BXX, 1BC5, and 2RD4) and on a larger system KcsA, a tetrameric potassium ion channel made of four 97-residue long monomers (PDB access code 1BL8). The generation of the CG representation of each residue was achieved through a

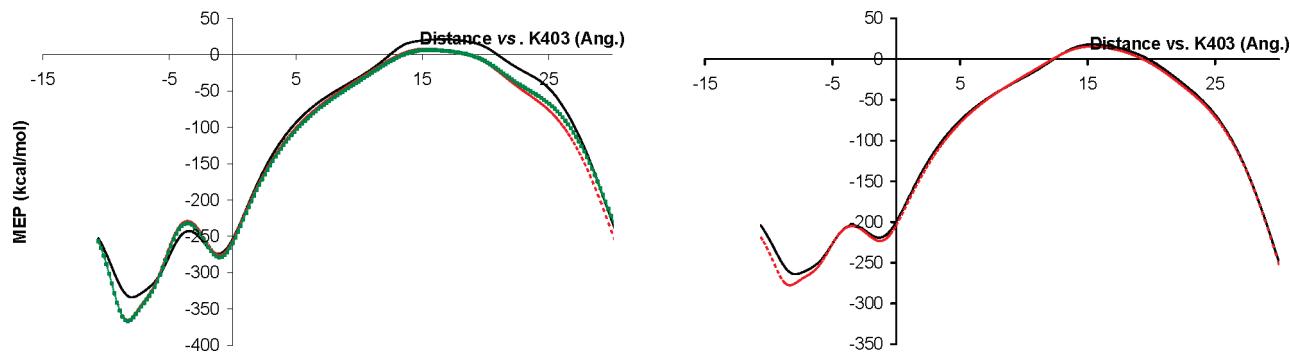


Figure 12. MEP profiles along the central axis of the KcsA potassium channel calculated using (left) the all-atom Amber99 set of charges (black plain line), the Amber99-based CG model with $q_{\text{end}} = \pm 1.0 \text{ e}^-$ (red dashed lines), the Amber99-based CG model with $q_{\text{end}} = 0.5 \text{ e}^-$ (green squares), and (right) the united-atom Gromos43A1 set of charges (black plain line), the Gromos43A1-based CG model with $q_{\text{end}} = \pm 1.0 \text{ e}^-$ (red dashed lines).

superimposition algorithm of CG template motifs on the 3D PDB structure. The study of the four peptides revealed that end charges should be lower than unity for the Amber99-based CG models and equal to unity for the Gromos43A1-based CG models. For Amber99, the variability in the end charge values is assumed to be the reflection of moderate transferability properties. For the larger system KcsA, the CG descriptions, consisting of 1 284 and 1 204 CG points and their tabulated charges, in the frameworks of Amber99 and Gromos43A1, respectively, allowed to well reproduce the trends observed in the unsmoothed all-atom MEP functions.

Our calculations suggest that decoupling backbone and side chain contributions in the elaboration of an AA CG model is interesting for reproducing all-atom electrostatic properties and that the location of CG steric centers, like those defined by ED peaks or by centers-of-mass of specific groups of atoms, differ from the location of CG electrostatic centers. This might be a point to consider in the further development of a CG FF.

During the elaboration of the MEP-based CG models, two points were considered to be important to favor transferability. First, AAs were studied in the isolated state to neglect the protein backbone conformation, and second, for each AA, CG charges were obtained by considering various side chain conformations. Though probably not sufficient to definitely demonstrate transferability robustness of our models, the results are encouraging, and they open an interesting extension to the present work, for example, in the comparison of MEP calculated using the Poisson–Boltzmann formalism.³ One can also imagine two more direct ways to test transferability of Coulomb potentials built from MEP-based CG models. The first one could consist in applying our procedure to a larger set of protein structures. The other would ask for a detailed comparison between MEP profiles calculated at the all-atom and CG levels, and this, for all possible AA-AA pairs. Finally, one could also elaborate for each AA type, CG representations and/or charges that depend on the rotamer class.

To directly link MEP and experimental ED distribution functions, one could use databases of transferable multipolar ED parameters for evaluating atom charges, as presented by

Zarichta et al.⁷² and preliminarily applied to the human aldose reductase system,³⁶ and then calculate MEP functions.

Acknowledgment. The authors thank the referees for extremely useful comments that led to an improvement in the CG models obtained for the AAs as well as Prof. E. Clementi and Prof. Cl. Lecomte and the members of his research group for very fruitful discussions. The FNRS-FRFC, the “Loterie Nationale” (convention no. 2.4578.02) and the Facultés Universitaires Notre-Dame de la Paix (FUNDP) are gratefully acknowledged for the use of the Interuniversity Scientific Computing Facility (ISCF) Center.

Supporting Information Available: Tables SI1 and SI2 contain a full description of the AA CG models in terms of charges and their rmsdV and $\text{rmsd}\mu$ values for Amber99 and Gromos43A1, respectively. Table SI6 for Gromos43A1 is equivalent to Table 6 for Amber99. Figure SI7 for Gromos43A1 is equivalent to Figure 7 for Amber99. 3D structure and charges of the all-atom and coarse point charge models generated for the four peptides (PDB codes 2EVQ, 1BXX, 1BC5, and 2RD4) and structure KcsA (PDB code 1BL8) can be downloaded from http://perso.fundp.ac.be/~lleherte/JCTC_SI/. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, USA, 2009.
- (2) Dong, F.; Olsen, B.; Baker, N. A. Computational Methods for Biomolecular Electrostatics. *Methods Cell Biol.* **2008**, *84*, 843–870.
- (3) Schutz, Cl. N.; Warshel, A. What Are the Dielectric “Constants” of Proteins and How To Validate Electrostatic Models? *Proteins* **2001**, *44*, 400–417.
- (4) Skepö, M.; Linse, P.; Arnebrant, T. Coarse-Grained Modeling of Proline Rich Protein 1 (PRP-1) in Bulk Solution and Adsorbed to a Negatively Charged Surface. *J. Phys. Chem. B* **2006**, *110*, 12141–12148.
- (5) Curcó, D.; Nussinov, R.; Alemán, C. Coarse-Grained Representation of β -Helical Protein Building Blocks. *J. Phys. Chem. B* **2007**, *111*, 10538–10549.

- (6) Basdevant, N.; Borgis, D.; Ha-Duong, T. A. Coarse-Grained Protein-Protein Potential Derived from an All-Atom Force Field. *J. Phys. Chem. B* **2007**, *111*, 9390–9399.
- (7) Pizzitutti, F.; Marchi, M.; Borgis, D. Coarse-Graining the Accessible Surface and the Electrostatics of Proteins for Protein-Protein Interactions. *J. Chem. Theory Comput.* **2007**, *3*, 1867–1876.
- (8) Zhang, Z.; Lu, L.; Noid, W. G.; Krishna, V.; Pfaendtner, J.; Voth, G. A. A Systematic Methodology for Defining Coarse-Grained Sites in Large Biomolecules. *Biophys. J.* **2008**, *95*, 5073–5083.
- (9) Bereau, T.; Deserno, M. Generic Coarse-Grained Model for Protein Folding and Aggregation. *J. Chem. Phys.* **2009**, *130*, 235106/1–235106/15.
- (10) Paramonov, L.; Yaliraki, S. N. The Directional Contact Distance of Two Ellipsoids: Coarse-Grained Potentials for Anisotropic Interactions. *J. Chem. Phys.* **2005**, *123*, 194111/1–194111/11.
- (11) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
- (12) Liu, P.; Izvekov, S.; Voth, G. A. Multiscale Coarse-Graining of Monosaccharides. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.
- (13) Carbone, P.; Varzaneh, H. A. K.; Chen, X. Y.; Müller-Plathe, F. Transferability of Coarse-Grained Force Fields: The Polymer Case. *J. Chem. Phys.* **2008**, *128*, 064904/1–064904/11.
- (14) Kondrashov, D. A.; Cui, Q.; Phillips, G. N., Jr. Optimization and Evaluation of a Coarse-Grained Model of Protein Motion Using X-Ray Crystal Data. *Biophys. J.* **2006**, *91*, 2760–2767.
- (15) Fukunaga, H.; Aoyagi, T.; Takimoto, J.-I.; Doi, M. Derivation of Coarse-Grained Potential for Polyethylene. *Comput. Phys. Commun.* **2001**, *142*, 224–226.
- (16) Lyman, E.; Pfaendtner, J.; Voth, G. A. Systematic Multiscale Parametrization of Heterogeneous Elastic Network Models of Proteins. *Biophys. J.* **2008**, *95*, 4183–4192.
- (17) Lyubartsev, A. P.; Laaksonen, A. Calculation of Effective Interaction Potentials from Radial Distribution Functions: A Reverse Monte Carlo Approach. *Phys. Rev. E* **1995**, *52*, 3730–3737.
- (18) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The Multiscale Coarse-Graining Method. I. A Rigorous Bridge between Atomistic and Coarse-Grained Models. *J. Chem. Phys.* **2008**, *128*, 244114/1–244114/11.
- (19) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The Multiscale Coarse-Graining Method. II. Numerical Implementation for Coarse-Grained Molecular Models. *J. Chem. Phys.* **2008**, *128*, 244115/1–244115/20.
- (20) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Forcefield: Coarse-Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (21) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI Coarse-Grained Forcefield: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
- (22) Treptow, W.; Marrink, S.-J.; Tarek, M. Gating Motions in Voltage-Gated Potassium Channels Revealed by Coarse-Grained Molecular Dynamics Simulations. *J. Phys. Chem. B* **2008**, *112*, 3277–3282.
- (23) Liwo, A.; Czaplewski, C.; Oldziej, S.; Rojas, A. V.; Kazmierkiewicz, R.; Makowski, M.; Murarka, R. K.; Sheraga, H. A. Simulation of Protein Structure & Dynamics with the Coarse-Grained UNRES Force Field. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009; Chapter 8, pp 107–122.
- (24) Fujitsuka, Y.; Takada, S.; Luthey-Schulten, Z. A.; Wolynes, P. G. Optimizing Physical Energy Functions for Protein Folding. *Proteins* **2004**, *54*, 88–103.
- (25) Hori, N.; Chikenji, G.; Berry, R. S.; Takada, S. Folding Energy Landscape and Network Dynamics of Small Globular Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 73–78.
- (26) Gabdoulline, R. R.; Wade, R. C. Effective Charges for Macromolecules in Solvent. *J. Phys. Chem.* **1996**, *100*, 3868–3878.
- (27) Berardi, R.; Muccioli, L.; Orlandi, S.; Ricci, M. Zannoni, Cl. Mimicking Electrostatic Interactions with a Set of Effective Charges: A Genetic Algorithm. *Chem. Phys. Lett.* **2004**, *389*, 373–378.
- (28) Golubkov, P. A.; Ren, P. -Y. Generalized Coarse-Grained Model Based on Point Multipole and Gay-Berne Potentials. *J. Chem. Phys.* **2006**, *125*, 064103/1–064103/11.
- (29) Casella, M.; Neri, M. A.; Carloni, P.; Dal Peraro, M. Topologically Based Multipolar Reconstruction of Electrostatic Interactions in Multiscale Simulations of Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 1378–1385.
- (30) Yang, L.-W.; Chng, Ch.-P. Coarse-Grained Models Reveal Functional Dynamics - I. Elastic Network Models - Theories, Comparisons and Perspectives. *Bioinf. Biol. Insights* **2008**, *2*, 25–45.
- (31) Chng, Ch.-P.; Yang, L.-W. Coarse-Grained Models Reveal Functional Dynamics - II. Molecular Dynamics Simulation at the Coarse-Grained Level - Theories and Biological Applications. *Bioinform. Biol. Insights* **2008**, *2*, 171–185.
- (32) Clementi, C. Coarse-Grained Models of Protein Folding: Toy Models or Predictive Tools? *Curr. Opin. Struct. Biol.* **2008**, *18*, 10–15.
- (33) Sherwood, P.; Brooks, B. R.; Sansom, M. S. P. Multiscale Methods for Macromolecular Simulations. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630–640.
- (34) Eyal, E.; Bahar, I. Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models. *Biophys. J.* **2008**, *94*, 3424–3435.
- (35) Zhou, L.; Siegelbaum, S. A. Effects of Surface Water on Protein Dynamics Studied by a Novel Coarse-Grained Normal Mode Approach. *Biophys. J.* **2008**, *94*, 3461–3474.
- (36) Leherte, L.; Guillot, B.; Vercauteren, D. P.; Pichon-Pesme, V.; Jelsch, Ch.; Lagoutte, A.; Lecomte, C. Topological Analysis of Proteins as Derived from Medium and High-Resolution Electron Density: Applications to Electrostatic Properties. In *The Quantum Theory of Atoms in Molecules - From Solid State to DNA and Drug Design*; Matta, C. F., Boyd, R. J., Eds.; Wiley-VCH: Weinheim, Germany, 2007; Chapter 11, pp 285–316.
- (37) Leherte, L.; Vercauteren, D. P. Determination of Protein Coarse-Grain Charges from Smoothed Electron Density Distribution Functions and Molecular Electrostatic Potentials.

- In *Handbook of Computational Chemistry Research*; Collett, C. T.; Robson, C. D., Eds.; NovaScience Publishers, in press.
- (38) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (39) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (40) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.
- (41) An Automated Pipeline for the Setup, Execution, and Analysis of Poisson-Boltzmann Electrostatics Calculations. *PDB2PQR*; SourceForge: Mountain View, CA, 2007; <http://pdb2pqr.sourceforge.net/>. Accessed August 31, 2009.
- (42) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- (43) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer. An Environment for Comparative Protein Modeling. *Electrophoresis* **1997**, *18*, 2714–2723.
- (44) Guex, N.; Diemand, A.; Peitsch, M. C.; Schwede, T. *Swiss-PdbViewer DeepView*, Version 4.0, 2008; <http://spdbv.Vital-it.ch/>. Accessed August 26, 2009.
- (45) Open Source Software Project; Based on IBM's Visualization Data Explorer; IBM: Armonk, New York, 2007; <http://www.opendx.org/>. Accessed August 26, 2009.
- (46) Leung, Y.; Zhang, J.-S.; Xu, Z.-B. Clustering by Scale-Space Filtering. *IEEE T. Pattern Anal.* **2000**, *22*, 1396–1410.
- (47) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; Abramowitz, M., Stegun, I. A., Eds.; Dover Publications: New York, 1970.
- (48) Hart, R. K.; Pappu, R. V.; Ponder, J. W. Exploring the Similarities between Potential Smoothing and Simulated Annealing. *J. Comput. Chem.* **2000**, *21*, 531–552.
- (49) Borodin, O.; Smith, G. D. *Force Field Fitting Toolkit*; University of Utah: Salt Lake City, UT; <http://www.eng.utah.edu/~gdsmit/fbf.html>. Accessed August 26, 2009.
- (50) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (51) Eisenmenger, F.; Hansmann, U. H. E.; Hayryan, S.; Hu, C.-K. An Enhanced Version of SMMP-Open-Source Software Package for Simulation of Proteins. *Comput. Phys. Commun.* **2006**, *174*, 422–429.
- (52) Simple Molecular Mechanics for Proteins; Michigan Technological University: Houghton, MI; <http://www.smmp05.net/>. Accessed August 26, 2009.
- (53) Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. Energy Parameters in Polypeptides. 10. Improved Geometrical Parameters and Nonbonded Interactions for Use in the ECEPP/3 Algorithm, with Application to Proline-Containing Peptides. *J. Phys. Chem.* **1992**, *96*, 6472–6484.
- (54) Simms, A. M.; Toofanny, R. D.; Kehl, C.; Benson, N. C.; Daggett, V. Dymameomics: Design of a Computational Lab Workflow and Scientific Data Repository for Protein Simulations. *Prot. Eng. Des. Sel.* **2008**, *21*, 369–377.
- (55) DYNAMICOMICS; University of Washington: Seattle, WA, 2007; <http://www.dymameomics.org/>. Accessed August 26, 2009.
- (56) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weisssig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (57) RCSB PDB Protein Data Bank; Rutgers, the State University of New Jersey: Piscataway, NJ and San Diego Supercomputer Center (SDSC) and Skaggs School of Pharmacy and Pharmaceutical Sciences: La Jolla, CA, 2009; <http://www.rcsb.org/pdb>. Accessed August 26, 2009.
- (58) Leherte, L. Hierarchical Analysis of Promolecular Full Electron Density Distributions: Description of Protein Structure Fragments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1254–1265.
- (59) Heisterberg, D. J. Technical report. *Translation from FORTRAN to C and Input/Output*; Labanowski, J., Ed.; Ohio Supercomputer Center: Columbus, OH, 1990.
- (60) CCL quaternion-mol-fit; Computational Chemistry List, Ltd: Columbus, OH; <http://www.ccl.net/cca/software/SOURCES/C/quaternion-mol-fit/>. Accessed August 26, 2009.
- (61) Exner, Th. E.; Mezey, P. G. Ab Initio-Quality Electrostatic Potentials for Proteins: An Application of the ADMA Approach. *J. Phys. Chem. A* **2002**, *106*, 11791–11800.
- (62) Boiteux, C.; Kraszewski, S.; Ramseyer, Ch.; Girardet, Cl. Ion Conductance vs. Pore Gating and Selectivity in KcsA Channel: Modeling Achievements and Perspectives. *J. Mol. Model.* **2007**, *13*, 699–713.
- (63) Noskov, S. Y.; Roux, B. Importance of Hydration and Dynamics on the Selectivity of the KcsA and NaK Channels. *J. Gen. Physiol.* **2007**, *129*, 135–143.
- (64) Khavrutskii, I. V.; Fajer, M.; McCammon, J. A. Intrinsic Free Energy of the Conformational Transition of the KcsA Signature Peptide from Conducting to Nonconducting State. *J. Chem. Theory Comput.* **2008**, *4*, 1541–1554.
- (65) Gascon, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. A Self-Consistent Space-Domain Decomposition Method for QM/MM Computations of Protein Electrostatic Potentials. *J. Chem. Theory Comput.* **2006**, *2*, 175–186.
- (66) Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
- (67) Piccinini, E.; Ceccarelli, M.; Affinito, F.; Brunetti, R.; Jacoboni, C. Biased Molecular Simulations for Free-Energy Mapping: A Comparison on the KcsA Channel as a Test Case. *J. Chem. Theory Comput.* **2008**, *4*, 173–183.
- (68) Pedretti, A.; Villa, L.; Vistoli, G. VEGA - An Open Platform To Develop Chemo-Bio-Informatics Applications using Plug-In Architecture and Script Programming. *J. Computer. Aided Mol. Des.* **2004**, *18*, 167–173.
- (69) VEGA ZZ, 2009. Drug Design Laboratory-University of Milan. <http://www.ddl.unimi.it/> Accessed August 26, 2009.

- (70) Bond, P. J.; Sansom, M. S. P. Insertion and Assembly of Membrane Proteins via Simulation. *J. Am. Chem. Soc.* **2006**, *128*, 2697–2704.
- (71) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. Coarse-Grained Molecular Dynamics Simulations of Membrane Proteins and Peptides. *J. Struct. Biol.* **2007**, *157*, 593–605.
- (72) Zarychta, B.; Pichon-Pesme, V.; Guillot, B.; Lecomte, C.; Jelsch, C. On the Application of an Experimental Multipolar Pseudo-Atom Library for Accurate Refinement of Small-Molecule and Protein Crystal Structures. *Acta Crystallogr. A* **2007**, *63*, 108–125.

CT900193M