

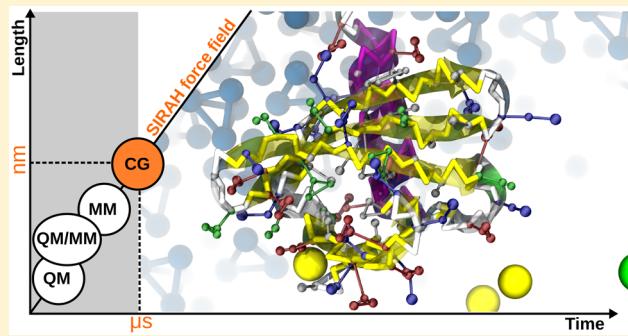
# SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics

Leonardo Darré,<sup>†,‡</sup> Matías Rodrigo Machado,<sup>†</sup> Astrid Febe Brandner,<sup>†</sup> Humberto Carlos González,<sup>†</sup> Sebastián Ferreira,<sup>†</sup> and Sergio Pantano<sup>\*,†</sup>

<sup>†</sup>Institut Pasteur de Montevideo, Montevideo, Uruguay

<sup>‡</sup>Department of Chemistry, King's College, London, United Kingdom

**ABSTRACT:** Modeling of macromolecular structures and interactions represents an important challenge for computational biology, involving different time and length scales. However, this task can be facilitated through the use of coarse-grained (CG) models, which reduce the number of degrees of freedom and allow efficient exploration of complex conformational spaces. This article presents a new CG protein model named SIRAH, developed to work with explicit solvent and to capture sequence, temperature, and ionic strength effects in a topologically unbiased manner. SIRAH is implemented in GROMACS, and interactions are calculated using a standard pairwise Hamiltonian for classical molecular dynamics simulations. We present a set of simulations that test the capability of SIRAH to produce a qualitatively correct solvation on different amino acids, hydrophilic/hydrophobic interactions, and long-range electrostatic recognition leading to spontaneous association of unstructured peptides and stable structures of single polypeptides and protein–protein complexes.



## INTRODUCTION

The exponential growth of computer power added to the development of faster algorithms has contributed to make molecular simulations a reliable tool for the study of biomolecular systems. Nevertheless, direct comparison with experimental data is often difficult owing to the large size and long time scales needed for a proper description of the complex biological environment. These difficulties have motivated the development of simplified methods aimed to bridge the gap between experiments and simulations. A large number of coarse-grained (CG) molecular representations have been described in the literature for the simulation of the most common biological species.<sup>1–18</sup> In general, the microscopic details are coarsened following either top-down or bottom-up approaches. In bottom-up schemes, a given Hamiltonian function is chosen and parametrized to fit fine-grained (FG) simulations taken as a reference. Several strategies to derive CG potentials have been developed on the basis of mining degrees of freedom from FG simulations through force matching techniques, Boltzmann inversion, thermodynamic integration, etc.<sup>19,20</sup> In top-down approaches, force fields are often tailored on the basis of physicochemical intuition and/or trial and error simulations, and interaction parameters are fitted to match available experimental data.

Bottom-up strategies can produce very accurate potentials and are very well suited for the description of uniform systems. However, it may be difficult to derive a general and transferable CG force field for highly heterogeneous macromolecules as proteins.<sup>20</sup> On the other hand, the accuracy of top-down

models may be strongly related to the availability of experimental data but may provide potentials that are more easily transferable.<sup>21</sup> For recent reviews on different CG approaches, see Ingolfsson et al.<sup>22</sup> and Brini et al.<sup>23</sup>

Recently, our group has undertaken the initiative to develop a CG force field for biomolecules named SIRAH (<http://www.sirahff.com>). We followed a top-down approach fitting structural properties of macromolecules using a standard pairwise Hamiltonian common to most MD simulation packages. So far, the SIRAH force field includes parameters and topologies for simulating DNA using an implicit solvation scheme<sup>24,25</sup> or embedded in an explicit CG representation of aqueous solvation.<sup>26</sup> Our CG model for water (named WatFour or WT4 for shortness) is composed by four linked beads, each carrying a partial charge. This confers to WT4 the capacity to create its own dielectric permittivity, while the use of CG electrolytes helps to account for ionic strength effects and osmotic pressure.<sup>26</sup> The WT4 model has been recently shown to be suitable for hybrid or dual-resolution simulations, where regions of interest within molecular systems can be treated at full atomic detail, while bulk regions of the solvent are simulated at the CG level without perturbing the structure and dynamics of the atomistic part.<sup>27–29</sup> Along this line, we have also expanded our force field to consider a dual-resolution version of double stranded DNA<sup>30</sup> compatible with the FG AMBER99 force field.<sup>31</sup>

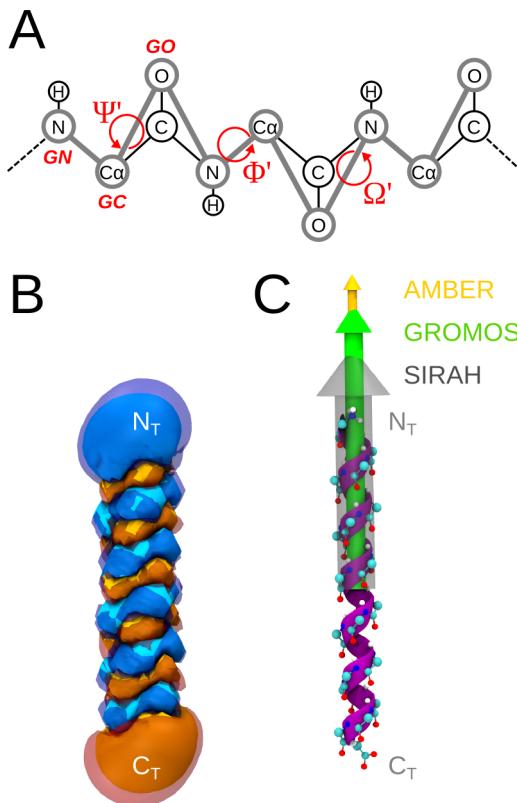
**Received:** August 26, 2014

**Published:** December 17, 2014

Here, we present a novel CG model for proteins and peptides, which works in combination with the WT4 model for explicit solvation. The set of parameters presented here aims to overcome some common limitations of CG force fields as the use of uniform dielectric constant, lack of long-range interactions, use of topological information to maintain the secondary structure, implicit or no ionic strength effects, etc. The performance of our model is illustrated on a series of simulations on different peptides, proteins, and protein–protein complexes chosen as working examples to highlight particular characteristics of the force field. These examples include sequence, temperature, and ionic strength-dependent conformational changes of helical peptides, spontaneous formation  $\beta$ -strands, reproduction of stable structures of proteins with different folds, and protein–protein complexes.

**Derivation of the Model.** The conception of the protein model follows the general structure-based philosophy used for DNA and aqueous solvent. The FG-to-CG mapping scheme is based on physicochemical intuition and uses the position of real atoms to place CG beads. Interaction parameters are empirically fitted to reproduce structural features of target molecules. In particular, the internal coordinates of our tetrahedral CG water model are based on the transient configuration of water clusters, and interaction parameters between water and free electrolytes are fitted to match the structure of the second solvation shell as determined by neutron scattering experiments.<sup>26</sup> Similarly, our CG model of DNA contains two beads at the position of the phosphorus and C5' carbon, representing the 5'-3' polarity of the backbone. A point charge of  $-1e$  is placed on the phosphate, and the sizes of the beads represent the excluded volume of their atomistic counterpart.<sup>24</sup> In contrast, the beads representing the base moiety present an atomistic size to allow for proper stacking between bases and hydrogen bond-like interactions in the Watson–Crick region. This nonuniform granularity is used to represent nearly atomistic interactions at the bases (i.e., hydrogen bonds and stacking) and coarser (less specific) interactions at the backbone. It is worth noting that this mapping combined with a uniform mass distribution, bond and angular stretching constants leads to nearly atomistic reproduction of DNA structure and dynamics, including ionic strength conformational effects.<sup>32</sup> Fitting of the WT4 bead's mass to water density resulted in 50 au, which is very close to the value obtained by summing the FG masses of the four nucleotides and taking the average on the six CG beads composing one of them (i.e., 51.53 au<sup>26</sup>). Therefore, in this version of the force field, we decided to adopt a uniform value of 50 au for all the protein beads. This choice is also computationally convenient because in combination with bond-stretching parameters it allows for time steps of 20 fs.

For the CG mapping of proteins, the peptide bonds are treated with a relatively low granularity, keeping the positions of the nitrogen (N),  $\alpha$  carbon (C $\alpha$ ), and oxygen (O) (Figure 1A), while side chains are modeled at lower degree of detail (see below). Similar strategies have been previously presented, as for instance refs 10, 11, and 33. This choice might be advantageous in relation to more uniform CG mappings presented in the literature. In particular, the correspondence between FG and CG conformational space is univocal, leading to a straightforward physical interpretation of the conformational space explored by peptides/proteins. Additionally, backbone representations considering only one bead on the C $\alpha$  may need specific constraints to reproduce secondary



**Figure 1.** Structure and electrostatics of the SIRAH protein backbone. (A) CG mapping scheme. Gray circles indicate the position of the atoms used to define the CG beads in the backbone, while thick gray lines indicate their connectivity. The main dihedral angles at the CG level are shown. (B) Comparison between electrostatic profiles in the neighborhood of a  $\alpha$ -helix generated by SIRAH and AMBER99. Solid and semi-transparent surfaces correspond to AMBER99 and SIRAH, respectively. Positive (blue) and negative (red/orange) isosurfaces are traced at  $\pm 10$  mV. (C) Dipole moments calculated on the same helix using the charge distribution of AMBER99 (yellow), Gromos96/45a3 (green), and SIRAH (gray, semi-transparent). The origin of the dipoles is displaced in the vertical direction to the center of the helix for visualization purposes.

structure elements depending on the peptide conformation.<sup>34</sup> In contrast, our choice of backbone beads with atomistic sizes is conformation independent and allows for the formation of the optimally compacted topology of  $\alpha$ -helices.<sup>35</sup> Moreover, the use of partial charges on each bead can roughly account for the formation of hydrogen bond-like interactions, stabilizing the formation of the  $\alpha$ -helices and  $\beta$ -sheets without imposing ad hoc constraints. On the other hand, this mapping based on effective interaction points within functional groups may not be easy to generalize. Hence, simple mapping procedures as the “four heavy atoms to one CG bead” rule used by the popular Martini force field<sup>9</sup> may be difficult to derive in our case.

The connectivity between N, C $\alpha$ , and O beads (named GN, GC, and GO, respectively) is presented in Figure 1A. Equilibrium distances and angles for bonded parameters between all backbone beads are adopted from the minimum energy conformation of FG glycine tripeptides using the AMBER99 force field.<sup>31</sup> Notice that these quantities are independent of the conformation or the entity of the amino acid used and the actual connectivity in the FG systems. As an initial guess, we used the same bond and angular stretching force constants used for the DNA model, i.e., 41840 kJ/mol

$\text{nm}^2$  and  $627.6 \text{ kJ/mol rad}^2$ . As shown in the Results section, this choice leads to stable MD trajectories of peptides and proteins without further refinement.

A combination of dihedral angles is imposed to the quadruplets defined by the backbone beads (Figure 1A and Table 1). This representation keeps a correspondence between

**Table 1. Torsional Parameters in SIRAH Force Field for Proteins<sup>a</sup>**

	<i>m</i>	<i>k<sub>m</sub></i> (kJ/mol)	<i>n<sub>m</sub></i>	$\alpha_{0m}$ (deg)
backbone				
$\Psi'(\text{GN}_i\text{-GC}_i\text{-GO}_i\text{-GN}_{i+1})$	1	1.8	1	160
	2	15.0	2	-270
	3	4.8	1	-130
$\Phi'(\text{GO}_{i-1}\text{-GN}_i\text{-GC}_i\text{-GO}_i)$	1	11.5	4	60
	2	22.4	3	60
	3	5.1	4	180
	4	16.0	5	-90
$\Omega'(\text{GC}_{i-1}\text{-GO}_i\text{-GN}_i\text{-GC}_i)$	1	60.0	1	0
amino acid "L" chirality				
side-chain-GC-GN-GO <sup>b</sup>	1	100	1	35
tryptophan side chain (improper)				
BPE-BCG-BCZ-BCE <sup>c</sup>	1	250.0	1	180
BCZ-BCG-BPE-BNE <sup>c</sup>	1	100.0	1	0

<sup>a</sup>The torsional potential in GROMACS is defined as  $V = \sum_m k_m [1 + \cos(n_m \alpha + \alpha_{0m})]$ . <sup>b</sup>Applies to the first bead in the side chain. <sup>c</sup>See Figure 2 for naming.

FG and CG dihedral angles  $\Psi$ ,  $\Phi$ , and  $\Omega$ , defining the secondary structure (see Calculated Properties). Hence, we set these dihedral angles by a polynomial fitting forcing the existence of minima in the two more stable conformations, i.e.,  $\alpha$ -helices ( $\Psi = -57^\circ$ ,  $\Omega = -47^\circ$ ) and  $\beta$ -strands ( $\Psi = 150^\circ$ ,  $\Omega = -80^\circ$ ). The suitability of the functions obtained was initially tested on peptides in  $\alpha$ -helical and  $\beta$ -sheet conformations and further refined on alpha beta (a + b) proteins. The final set of functions containing three and four terms for  $\Psi$  and  $\Phi$ , respectively, is shown in Table 1. At this stage of development, only peptides in *trans* configuration are considered. Amino acids in *cis* conformations, which are much less frequent in nature, will be incorporated in further versions of the force field. Hence,  $\Omega$  is represented with a single cosine function with one minimum in the *trans* region (Table 1).

The partial charges on the backbone's beads are set to roughly reproduce the electrostatic potential generated by a fully atomistic  $\alpha$ -helix, also known as the helix macropole. Neutral and charged termini are available in the current version of SIRAH. Charged termini are simply constructed by adding a  $\pm 1e$  charge on the N and/or O terminal beads of the chain, respectively.

To compare the similarity between the electrostatic potentials generated by popular FG force fields and SIRAH, we considered a 24 residues long polyglycine in canonical  $\alpha$ -helical conformation. This helical extension was chosen, as it is the approximate length of a typical transmembrane helix. The electrostatic potential is calculated by solving the linearized Poisson–Boltzmann equation on a box with 12 nm edges and grid spacing of 0.05 nm imposing a zero value for the potential at the borders using APBS.<sup>36</sup> The qualitative likeliness between the electrostatic potential generated by SIRAH and AMBER99 can be observed in Figure 1B. A more quantitative evaluation can be acquired by comparing the dipole moments generated

by different force fields (Figure 1C). Considering neutral terminals results in nearly collinear vectors with modules of 65 D, 76.5 D, and 84.6 D for SIRAH, Gromos96/45a3, and AMBER99, respectively. If zwitterionic terminals are used, we obtain dipole moments of 285.1 D, 267.3 D, and 259.5 D for SIRAH, Gromos96/45a3, and AMBER99, respectively. This suggests that long-range interactions reproduced by our CG scheme at the backbone level are comparable to those of popular FG force fields.

Finally, van der Waals (vdW) interactions within backbone beads are set to the same values of AMBER99, ensuring a correct degree of compaction upon formation of  $\alpha$ -helices and  $\beta$ -sheets.

The CG topology of the side chains follows the philosophy of representing interaction points according to the characteristics of each residue (Figure 2). Hydrophobic residues (Val, Ile, Leu, and Met) are represented by one single bead. Aromatic side chains are mapped to three (Phe, His, and Tyr) or five beads on a plane (Trp). The beads of polar and charged side chains are mapped on charged groups or acceptors/donors of hydrogen bonds.

Following the same procedure described for the backbone, bond and angular equilibrium positions are taken from the AMBER99 force field, while DNA values are used as initial guess of force constants. While these values are well suited for the backbone beads, iterative sampling revealed that the values reported in Table 2 perform better for side chains. Improper dihedral angles are imposed on the backbone and side-chain beads to ensure the "L" chirality of the amino acids. Additionally, two improper dihedrals are used on the side chain of tryptophan to force its planarity (Table 1).

Point charges are assigned under the general hypothesis that functional groups bearing more hydrogen bond acceptors/donors at the FG level should carry higher charges in the CG scheme to establish stronger electrostatic interactions. In this line, hydrophobic beads carry a zero charge, and polar moieties within aromatic side chains carry a charge of  $\pm 0.1e$ . Hydroxyl and amide groups are charged  $\pm 0.2 e$  and  $\pm 0.4 e$ , respectively, and charged amino acids present a unitary charge spread on the side-chain beads (Figure 2). In the present version, all protonation states are considered at pH 7.

The vdW parameters are treated as free variables to modulate interactions. As an initial guess, we started with sigma and epsilon corresponding to those of the WT4 beads (Figure 2). The values of the radii (sigma) underwent a progressive adjustment on static structures until no steric clashes are present. This is performed by varying the radii to obtain nonpositive values when calculating the vdW component of the potential energy pairwise on each couple of amino acids. After setting the sigma values on static structures, the epsilon parameters are iteratively varied to obtain stable trajectories. Aimed to limit the search of optimal combinations, these parameters are assigned by residue type (hydrophobic, aromatic, etc.). Ad hoc modifications are introduced to single beads of proline, methionine, lysine, and alanine and CG beads mapped on hydrogen atoms (Figure 2).

van der Waals interactions are, in general, calculated according to the Lorentz–Berthelot combination rules. However, owing to the high granularity of the CG solvent, it may not properly intercalate between polar moieties, introducing spurious conformational effects. Similarly, we found that the atomistic parameters used for backbone–backbone interactions result in poorly balanced interaction with

	FG	CG	SIRAH name	$q$ (e)	$\sigma$ (nm)	$\epsilon$ (kJ/mol)		FG	CG	SIRAH name	$q$ (e)	$\sigma$ (nm)	$\epsilon$ (kJ/mol)
G			1: GC 2: GN 3: GO	0,10 0,125 -0,225	0,40 0,40 0,40	0,55 0,55 0,55	A			1: GC 2: GN 3: GO	0,10 0,125 -0,225	0,41 0,40 0,40	2,00 0,55 0,55
S			4: BOG 5: BPG	-0,20 0,20	0,41 0,40	0,35 0,01	I			4: BCG	0	0,41	3,20
T			4: BOG 5: BPG	-0,20 0,20	0,41 0,40	0,35 0,01	V			4: BCB	0	0,41	3,20
N			4: BCG 5: BOD 6: BND	0 -0,40 0,40	0,40 0,40 0,40	0,35 0,55 0,55	L			4: BCG	0	0,41	3,20
Q			4: BCD 5: BOD 6: BND	0 -0,40 0,40	0,40 0,40 0,40	0,35 0,55 0,55	C			4: BSG 5: BPG	-0,20 0,20	0,41 0,40	0,35 0,01
Y			4: BCG 5: BCE1 6: BCE2	0 0,10 -0,10	0,35 0,35 0,35	1,70 1,70 1,70	M			4: BSD	0	0,45	3,20
He			4: BCG 5: BNE 6: BND	0 0,10 -0,10	0,35 0,35 0,35	1,70 1,70 1,70	P			4: BCG	0	0,43	0,60
K			4: BCG 5: BCE	0,40 0,60	0,40 0,55	0,55 0,55	F			4: BCG 5: BCE1 6: BCE2	0 0 0	0,35 0,35 0,35	1,70 1,70 1,70
R			4: BCG 5: BCZ 6: BNN1 7: BNN2	0 0,30 0,35 0,35	0,40 0,40 0,45	0,55 0,35 0,55	W			4: BCG 5: BNE 6: BPE 7: BCZ 8: BCE	0 -0,10 0,10 0 0	0,35 0,35 0,35 0,35 0,35	1,70 0,10 0,10 1,70 1,70
D			4: BCG 5: BOE1 6: BOE2	-0,30 -0,35 -0,35	0,40 0,45 0,45	0,35 0,55 0,55	E			4: BCD 5: BOE1 6: BOE2	-0,30 -0,35 -0,35	0,40 0,45 0,45	0,35 0,55 0,55
K <sup>+</sup> and 6 water molecules			1: KW	1,00	0,645	0,55	WT4 11 water molecules			1: WN1 2: WN2 3: WP1 4: WP2	-0,41 -0,41 0,41 0,41	0,42 0,42 0,42 0,42	0,55 0,55 0,55 0,55
Na <sup>+</sup> and 6 water molecules			1: NaW	1,00	0,58	0,55	Cl <sup>-</sup> and 6 water molecules			1: CIW	-1,00	0,68	0,55

**Figure 2.** SIRAH representation of amino acids and solvent. The one-letter code, FG heavy atoms, CG representation, bead's names, partial charges, and vdW parameters are presented for each amino acid. Only hydrogen atoms used for the CG mapping are shown. Numbers near the FG atoms indicate the position for the corresponding CG beads. Backbone mapping is indicated only for glycine and alanine. FG atoms are colored by name, while CG beads of amino acid and WT4 are colored by charge range (negative, red; positive, blue). The sizes of all CG beads are at scale and correspond to their actual vdW radii.

CG beads. To ameliorate this flaw, we set few specific corrections to the vdW interactions that avoid over stabilization of salt bridges, hydrogen bond-like interactions, or spurious protein–ion contacts (Figure 3). These corrections have been added case-by-case following an iterative trial and error procedure.

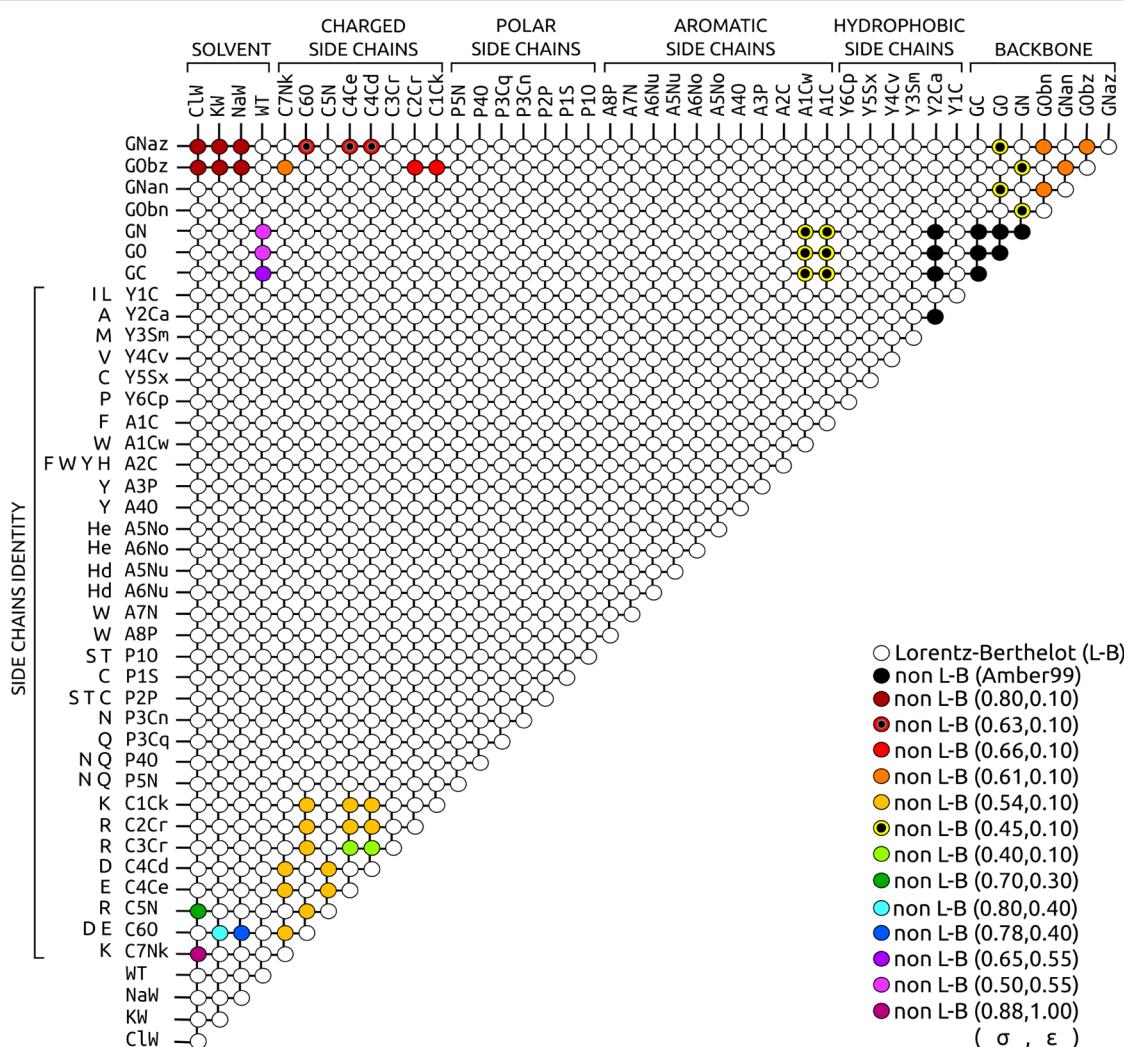
A comprehensive view of the topologies, sizes, names, charges, and vdW parameters of the 20 amino acids, water, and electrolytes available in the SIRAH force field is presented in Figure 2.

**Computational Details.** The simulation protocols applied in this work are essentially the same used for any plain molecular dynamics (MD) simulation: (i) Initial coordinates

**Table 2. Force Constants of Angular Parameters<sup>a</sup>**

position of the beads	k (kJ/mol rad <sup>2</sup> )	applies to
backbone–backbone–backbone	627.6	all residues
backbone (terminal)–backbone–side-chain	60.0	all terminal residues
backbone–backbone–side-chain (not aromatic)	100.0	Ser, Thr, Asn, Gln, Lys, Arg, Asp, Glu, Ile, Leu, Val, Met (not terminal)
backbone–backbone–side-chain (aromatic)	150.0	Tyr, His, Phe, Trp (not terminal)
backbone–side-chain–side-chain	50.0	Ser, Thr, Cys, His, Tyr, Phe, Trp
backbone–side-chain–side-chain	10.0	Arg, Lys, Asn, Gln, Asp, Glu,
side-chain–side-chain–side-chain	0.0	Asp, Gln, Tyr, His, Arg, Asp, Glu, Phe, Trp

<sup>a</sup>Equilibrium values are taken from AMBER99. The angular potential in GROMACS is defined as  $V = k/2(\sigma - \sigma_0)^2$ .



**Figure 3.** Scheme of vdW interactions in SIRAH. All the atom types currently defined in the SIRAH force field for proteins, ions, and water are arranged in a diagonal matrix. Interactions calculated with the Lorentz–Berthelot combination rules are indicated with empty circles. Colored circles correspond to interactions computed outside Lorentz–Berthelot combination rules using specific values indicated in the legend.

are taken from the PDB and protonated at pH 7. (ii) The FG structures are converted to CG by an ad hoc script. (iii) The CG model is solvated using a prestabilized box of WT4 molecules. (iv) Ionic strength or electroneutrality is set by substituting WT4 molecules by electrolytes (CG versions of  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Cl}^-$  are currently available). As the densities of WT4 and FG water are coincident, and considering the molecular weight of WT4 (200 g/mol), a proportion of one ionic couple per 50 WT4 molecules roughly represents a 0.1 M concentration. The simulation protocols also include (v) energy minimization, (vi) equilibration MD, and (vii) production run.

Starting structures are protonated using pdb2pqr,<sup>37</sup> converted to CG and solvated in an octahedral box with a solute–box distance of 1.5 nm. The ionic strength and target temperature are set to values reported in each case. Simulations are performed using GROMACS 4.5.5 (<http://www.gromacs.org>) with a time step of 20 fs updating the neighbor list every 10 steps. Electrostatic interactions are calculated using Particle Mesh Ewald (PME)<sup>38</sup> with a direct cut off of 1.2 nm and a grid spacing of 0.2 nm. The same cut off is used for vdW interactions. Energy minimization is carried out by 1000 iterations of the steepest descent algorithm. Equilibration

dynamics is accomplished by 5 ns of MD with positional restraints of 1000 kJ/mol nm<sup>2</sup> applied to all the protein beads. Production runs are performed by 1  $\mu$ s (unless otherwise stated) in the NPT ensemble using v-rescale thermostat<sup>39</sup> and Parrinello–Rahman barostat.<sup>40</sup>

### Calculated Properties.

- (1) Hydrogen bonds (H-bonds) are operationally defined to be formed if two beads of opposite charge are at a distance  $\leq$  0.4 nm. Similarly, a salt bridge is considered to exist if a couple of beads of opposite charge belonging to ionic residues are within a distance  $\leq$  0.6 nm.
- (2) Secondary structure is calculated as a function of the dihedral angles along the backbone beads and defined

$$\text{H: } -180^\circ \leq \Phi \leq 10^\circ \text{ and } -120^\circ \leq \Psi \leq 45^\circ$$

$$\text{SS}_i(\Phi, \Psi): \text{E: } (-180^\circ \leq \Phi \leq 0^\circ \text{ or } \Phi > 135^\circ) \text{ and } (-180^\circ \leq \Psi \leq -120^\circ \text{ or } 45^\circ \leq \Psi \leq 180^\circ)$$

$$\text{C: otherwise}$$

where  $\text{ss}_i$  is the conformation assigned to residue  $i$ . Additionally, for residues to belong to the H category, either the GO or the GN beads must be within the H-bond definition with a residue separated by four consecutive positions in the chain. Similarly, for residues to be considered in E conformation, GO, GN, or GC beads must form H-bonds with backbone neighbors separated by more than two consecutive positions in the chain and belonging to the same category. Residues not assigned to H or E are considered, for the sake of simplicity, as members of category C. Notice that the secondary structure is dynamic and can change during the trajectory.

- (3) Contacts between residues are considered to exist if two  $\text{C}\alpha$  beads are within 0.8 nm. Native contacts are defined on the experimental structure, and its conservation is reported as a percentage along the trajectory. Global contacts are calculated on all the residues in the chain, while noncontiguous contacts are computed excluding residue pairs within five neighboring positions in sequence. The accuracy is evaluated as  $\text{Acc} = \text{TP}/(\text{TP} + \text{FP})$ , where TP (true positives) is the number of correctly reproduced contacts, and FP (false positives) are new contacts during the simulation.
- (4) Peptide aggregates are considered to exist only if the distance between centers of mass of amino acids is within 0.5 nm, otherwise peptides are considered to belong to different aggregates. GROMACS' g\_clustersize tool is used to analyze the coordinates of the peptides every 1 ns during 5  $\mu$ s (a total of 5000 simulation frames).
- (5) Protein accessible surfaces (SAS) are calculated with GROMACS' g\_sas with a probe radius of 0.21 nm (i.e., the radius of a WT4 bead). Beads are considered to be hydrophobic if they carry a zero charge. Notice that specific vdW radii of SIRAH reported in Figure 2 have to be used.
- (6) Protein–protein interfaces are calculated as  $0.5(\text{SAS}_{\text{Prot1}} + \text{SAS}_{\text{Prot2}} - \text{SAS}_{\text{Complex}})$
- (7) RMSD, B-factors, and gyration radii are calculated on the  $\text{C}\alpha$  beads using the standard GROMACS' tools.

between  $\pm 180^\circ$ . Owing to the absence of the carbonyl carbon in the CG backbone, the  $\varphi$  and  $\psi$  dihedral angles defining the Ramachandran plot are calculated as (see Figure 2 for naming)

$$\Phi = \Phi'(\text{GO}_{i-1} - \text{GN}_i - \text{GC}_i - \text{GO}_i) - 20^\circ$$

$$\Psi = \Psi'(\text{GN}_i - \text{GC}_i - \text{GO}_i - \text{GN}_{i+1}) + 13^\circ$$

From this definition the conformation of each residue is assigned to the categories: helical (H), extended (E), or coil (C) using the following criteria

## RESULTS AND DISCUSSION

In this section, we test our CG model on a series of peptides and protein systems. The main goal of this set of simulations is not that of furnishing new biophysical insights for each particular system but highlighting specific features of our force field in different possible scenarios.

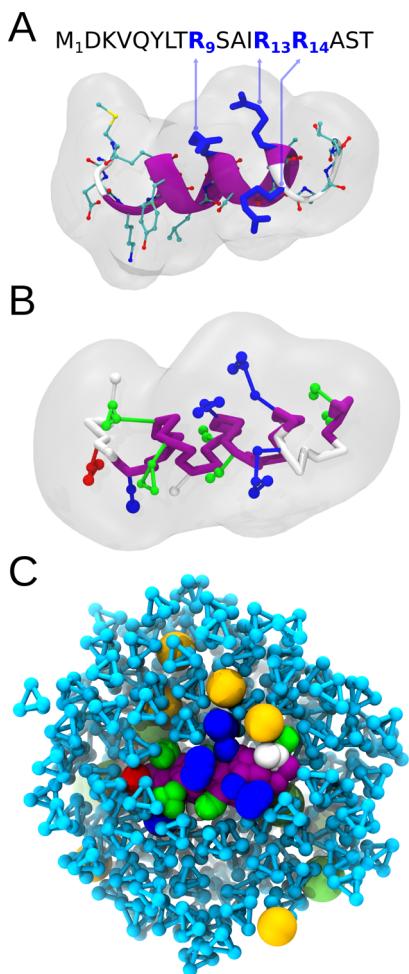
### Test Case 1: Cytoplasmic Domain of Phospholamban.

Phospholamban (PLB) is a small membrane protein that regulates the activity of the calcium ATP-ase in the cardiac, slow-twitch, and smooth muscle sarcoplasmic reticulum.<sup>41</sup> It consists of a cytoplasmic moiety subdivided in domains Ia (residues 1–17), Ib (residues 18–30), and transmembrane (residues 31–52). Several NMR studies have shown that domain Ia adopts an  $\alpha$ -helical configuration, independent of the rest of the protein.<sup>42,43</sup> Early attempts for structural determination also demonstrated the importance of added salts in the solution to stabilize the structure of this protein segment.<sup>42</sup>

The small size of this domain and the availability of structural experimental data make it suitable as an initial test case for our force field on a biologically relevant system. Moreover, the simplicity of this motif and the presence of polar, hydrophobic, and charged amino acids, all exposed to the solvent (Figure 4A), allows us to illustrate characteristic protein–solvent interactions disregarding the structural context.

**Protein–Solvent Interactions in Peptide Context.** We first analyze the solvation structure around amino acids with different physicochemical characteristics in domain Ia of PLB. With this aim, we take the first 17 amino acids of the structure 1FJK (Figure 4A) and map them into our CG representation (Figure 4B). The CG peptide is then embedded in a solvation box containing 287 WT4 molecules and 9 and 13 CG ions ( $\text{Na}^+$  and  $\text{Cl}^-$ , respectively), which ensure electro-neutrality and represent a salt concentration of  $\sim$ 0.2 M (Figure 4C). After energy minimization and equilibration, MD simulations are carried out at 300 K and 1 atm for 1  $\mu$ s.

The collected data is first used to compute the solvation structure around different moieties in terms of the radial distribution function (RDF) of water. Calculation of solvent's RDF around the backbone increases smoothly after 0.5 nm (Figure 5A) as a consequence of specific backbone–WT4 vdW interactions (Figure 3). Consequently, the solvent distribution



**Figure 4.** Molecular representations of domain Ia of PLB. (A) Sequence and starting MD conformer corresponding to the first 17 residues of the NMR structure 1FJK. The cartoon traced on the backbone atoms indicates the secondary structure (helix, purple; coil, white). Heavy atoms are colored by element, except for Arg9, 13, and 14, which are shown in blue. The semi-transparent surface corresponds to the SAS traced with a probe radius of 0.14 nm. (B) SIRAH mapping of the peptide presented in panel A. Amino acids are colored by residue type (positive, blue; negative, red; polar, green; and hydrophobic, white). In this case, the solvent-accessible surface is traced with a probe radius of 0.21 nm, which corresponds to the dimension of a WT4 bead. (C) Typical SIRAH simulation box. The domain Ia of PLB is placed within an octahedron box and solvated with WT4 and CG salt. The sizes of the solute (colored as in panel B) and ions ( $\text{Na}^+$ , yellow;  $\text{Cl}^-$ , green) correspond to their vDW radii, while WT4 molecules (light blue tetrahedrons) are shown with balls and sticks for visualization sake.

around the main chain results in a hydrophobic-like profile. Hence, the global shape of this distribution function is roughly comparable to that measured around the hydrophobic side chain of Ile12, which is taken as a representative hydrophobic bead (Figure 5A). Discriminating between negative or positive beads of WT4 (Figure 2) does not significantly affect the RDFs neither for the  $\text{Ca}$  bead nor for the hydrophobic residues (not shown).

In stark contrast, the solvent organization around charged side chains presents qualitative and quantitatively different profiles. The organization of the positive beads of WT4 around BOE1 in D2 (see Figure 2 for naming) shows the typical shape of a hydrophilic profile (Figure 5B). We found a marked peak

at 0.42 nm, followed by a minimum, indicating the passage from a region of high population of positive beads to a depletion caused by electrostatic interactions. In agreement with this electrostatics-driven solvent organization, the distribution of positive WT4 beads around BCE at K3 is shifted to the right and follows an opposite trend to that shown for D2 (Figure 5B). The role of electrostatics in the solvent organization is confirmed if we consider the arrangement of negative beads around the same moieties. The partial charge placed on the BCE bead results in a significantly sharp solvation peak at 0.45 nm, while a minimum is observed in the close neighborhood of BOE1 (Figure 5C). After nearly 1.2–1.5 nm, all the distributions flatten, converging to bulk values with oscillations presenting a periodicity corresponding to the size of WT4 molecules.<sup>26</sup>

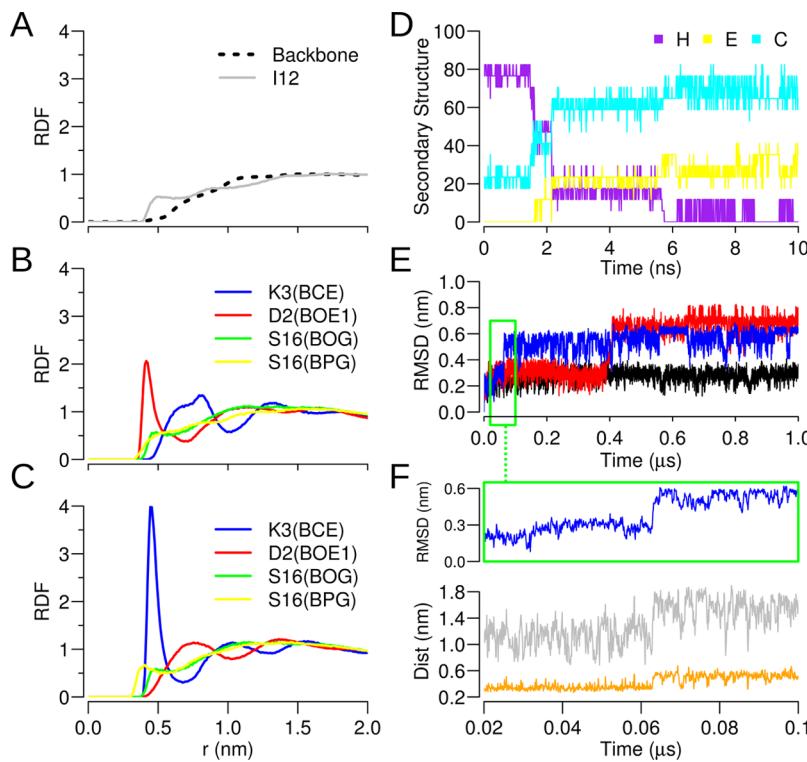
As representative of a polar residue, we consider the side chain of S16. In this case, we see that the CG beads placed at the positions of the gamma oxygen (BOG) and hydrogen (BPG) of S16 generate a less marked organization of the surrounding solvent than charged amino acids, being actually more similar to the hydrophobic moiety of I12 (Figure 5). The impaired capacity of serine to induce a marked hydrophilic-like solvation structure can be ascribed to the small charges placed on this side chain and to the separation between the beads (i.e., bond distance in a hydroxyl group). However, even these small partial charges are able to create differences in solvent organization. This is more evident in the distribution of negative beads, which exhibits a right-shift for the BOG bead (Figure 5C).

**Sequence, Temperature, and Ionic Strength Conformational Effects.** A further step in assessing the performance of our force field is to explore the conformational response of SIRAH to variations introduced in the molecular system.

To define a reference state, we briefly describe the results of the simulation introduced in the previous section. Along the trajectory, M1, D2, S16, and T17 very frequently visit disordered conformations. However, these transitions do not compromise the global structural stability of the peptide within the  $\mu\text{s}$  time scale. Indeed, the RMSD oscillates around 0.27 nm with a mean helical content of 73%. These slight conformational changes are in good agreement with the variation observed in the family of conformers of the cytoplasmic domain of PLB reported on the basis of NMR studies<sup>42</sup> and previous FG simulations on a shorter time scale.<sup>44</sup>

On this simple system, we sought to analyze the response of SIRAH to variations in (i) sequence, (ii) temperature, and (iii) ionic strength.

- (i) To rule out possible conformational biases in the parametrization of the backbone beads favoring secondary structures, we used the same starting conformation of PLB but mutating all amino acids to glycine. As shown in Figure 5D, the helical content of the polyglycine falls rapidly arriving to nearly 50% already within the first 2 ns and continues to decrease until the 6 ns to then oscillate between 0% and 17%. These changes are accompanied by an increase in the coil content and to a minor extent the extended configurations. Conformational transitions may happen one residue at a time or more than one simultaneously. These transitions, which imply the rupture of H-bond interactions between backbone beads and rotations along dihedral angles, happen

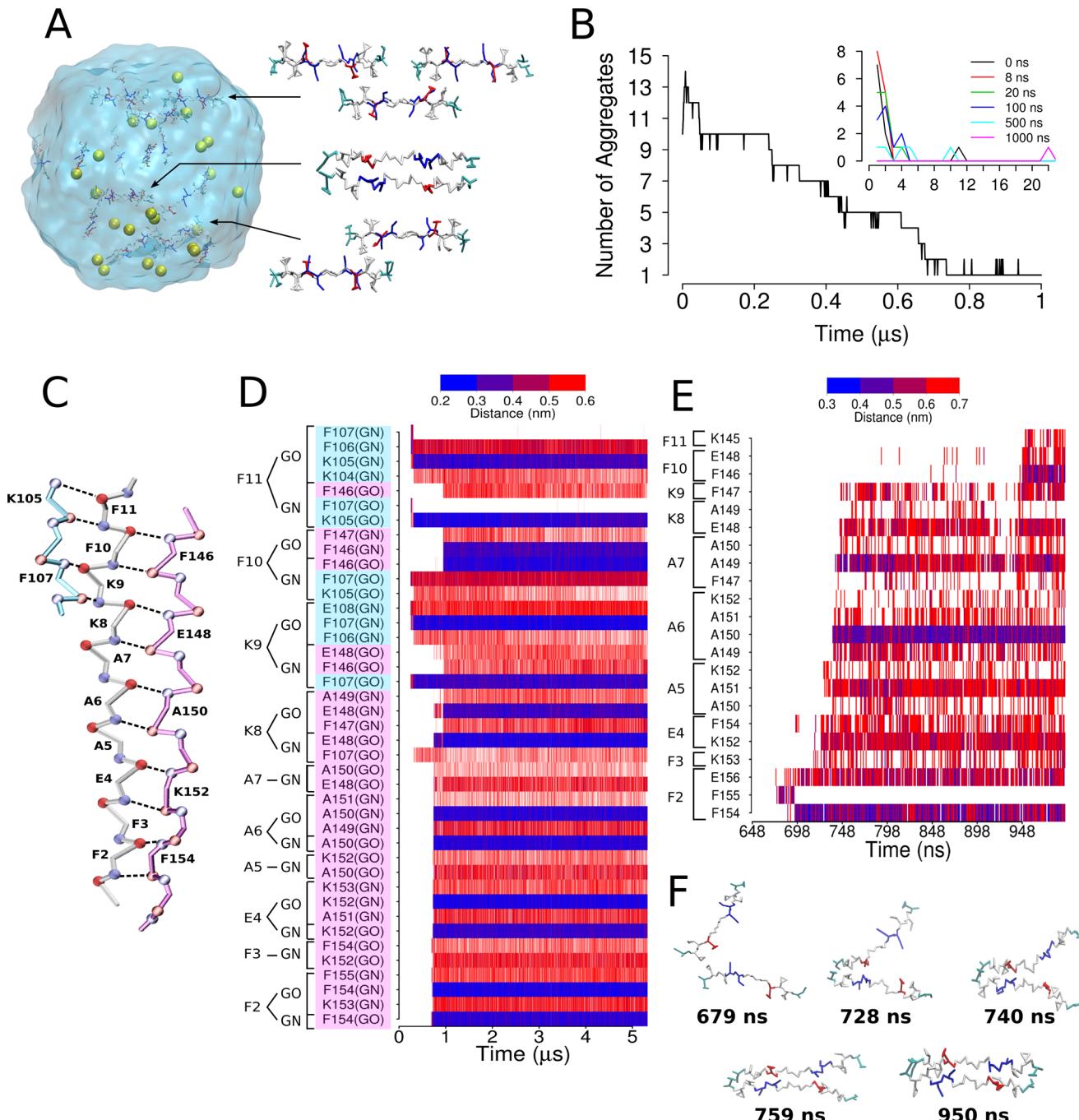


**Figure 5.** MD simulations of domain Ia of PLB. (A) RDF of WT4 around backbone and hydrophobic beads. The dashed line represents the distribution of WT4 beads around backbone beads corresponding to residues 2 to 16, while the continuous line is calculated for the side-chain bead of Ile12. (B) RDF of positive beads of WT4 around BOE1 in Asp2 (red), BCE in Lys3 (blue), BOG in Ser16 (green), and BPG in Ser16 (yellow). (C) Same as panel B for negative WT4 beads. (D) Secondary structure content of a 17-mer poly glycine started from a helical configuration. (E) RMSD calculated on systems at 300 K and 0.2 M NaCl (black), 330 K and 0.2 M NaCl (red), and 300 K with no added salts (blue, only 3 Cl<sup>-</sup> ions are added to ensure electroneutrality). (F) Top: Close up on the RMSD trajectory of PLB at low ionic strength in the moment of the conformational jump. Bottom: Distances between BCZ beads of Arg9 and 14 (gray) and backbone's GO and GN beads of Arg9 and 13, respectively (orange).

spontaneously within the ns time scale. These characteristic times are in good agreement with ns dynamics reported by single-molecule FRET and photoinduced electron transfer fluorescence experiments on glycine-rich peptides of similar length.<sup>45</sup>

- (ii) As a second test, we attempt the thermal unfolding of domain Ia of PLB. To this goal, we use the starting configuration of the first simulation but coupled it to a reference temperature of 330 K. The dynamics of this system during the first 0.4  $\mu$ s is comparable to that of the system at 300 K. However, after 0.4  $\mu$ s of simulation, the helix undergoes a major break. As a consequence, the RMSD increases steeply undergoing large oscillations between 0.5 and 0.85 nm from the initial conformer (Figure 5E). This effect is accompanied by a marked reduction of the helical content. The average secondary structure content calculated during the last 0.6  $\mu$ s is 40%, 13.5%, and 52.5% for helix, extended, and coil conformations, respectively.
- (iii) A more challenging test for our force field is to assess its ability to capture ionic strength effects. We have previously shown that WT4 can adequately reproduce electrolytic properties as screening, osmotic pressure, Bjerrum and Debye lengths, etc., which result in correct concentration profiles, ion specificity, and local conformational changes observed in high-resolution X-ray structures of DNA.<sup>26,32</sup> The cytoplasmic domain of PLB seems a suitable test case as its stability has been shown to depend on the electrolytic content of the solution.

Therefore, we set up a simulation of domain Ia of PLB at 300 K using the same starting conformer but in the absence of added salts. During the first 60 ns, the conformational sampling of the peptide is indistinguishable from its counterpart with RMSD peaks below 0.3 nm. After this period, the peptide helix experiences a major break, which is evident from the RMSD (Figure 5E). This produces a partially unfolded conformation with only a ~40% of average helical content. Analysis of the trajectory indicates that the molecular cause of the helix rupture resides in the strong electrostatic repulsion between neighboring charged residues. In a helical context, the side chains of Arg9, 13, and 14 (Figure 4A) experience a strong reciprocal electrostatic repulsion. In the absence of proper electrostatic screening, this repulsion is hardly compatible with a helical structure. Because Arg13 and 14 are consecutive in sequence, the high electrostatic potential energy is relaxed by breaking the helix and separating the positive moieties. Indeed, the observed rise in RMSD is concomitant with the rupture of the H-bond interaction between backbone beads of Arg9 and 13, which leads to a separation of the side chains of Arg9 and 14 (Figure 5F). Conversely, the distances between the side chains of Arg9 and 14 during the simulation conducted at higher ionic strength are maintained at 0.9 nm (s.d. 0.2 nm). Taking into account that the Bjerrum length measured at 300 K in a solution of 0.2 M of NaCl in WT4 is of 0.57 nm,<sup>26</sup> it can be concluded that the electrostatic screening imposed by the



**Figure 6.** Self-aggregation of microglobulin  $\beta$ -amyloid peptides. (A) Simulation box of test case 2 showing the initial distribution of peptides colored by residue type as in Figure 4 with terminal residues in cyan. The position of the preassembled aggregates and counterions is shown. (B) Number of aggregates as a function of simulation time during the first microsecond. The inset shows the number of aggregates vs the number of components of each cluster at different simulation times. (C) Snapshot showing the aggregation of a triplet of peptides. The backbone of each peptide is shown in different colors. H-bonds are indicated with dashed lines. (D) H-bond interactions between the central and flanking peptides are identified by background shading with colors as in panel C. The instantaneous distances between GN and GO beads defining the  $\beta$ -sheet are shown on the right side. White colors correspond to distances higher than 0.6 nm. (E) Close up on the zipper process between peptides 1–12 and 144–156. Distances are measured between the center of mass of each residue. (F) Series of snapshots along the zippering of peptides 1–12 and 144–156 at different simulation times.

simple electrolytes present in the solution is sufficient to overcome the electrostatic repulsion, leading to structural stabilization of the cytoplasmic domain of PLB.

Taken all together, the results presented for this test case suggest that the SIRAH force field can qualitatively capture the main physicochemical aspects of aqueous solvation and

interactions between functional groups in amino acids without any apparent conformational bias.

**Test Case 2: Self-Aggregation of Microglobulin  $\beta$ -Amyloid Peptides.** Amyloid fibrils are at the base of many lethal diseases, including Alzheimer's and spongiform encephalopathies (i.e., Creutzfeld–Jakob disease). In each disease, particular polypeptides form ordered, insoluble, and extended

fibrils that self-assemble in tissues, triggering different pathologies.<sup>46–48</sup> Simulating the aggregation processes that involves the formation of stable  $\beta$ -strand conformations and the unbiased self-assembly of  $\beta$ -sheets structures challenges the ability of our force field. For this task, we used the 12-mer amyloidogenic peptide KFFEAAAKKFFE, which has been studied by means of X-ray and electron diffraction.<sup>49</sup> In this peptide, the formation of ordered amyloid-like fibrils composed of antiparallel  $\beta$ -sheets is enhanced by  $\pi$ – $\pi$  stacking between adjacent phenylalanine rings and electrostatic interactions between charged residues (glutamic acid and lysine).

We set up a simulation box containing 10 isolated peptides plus three preassembled aggregates (one hexamer, one tetramer and one dimer), 4331 WT4 molecules, and 22 counterions (Figure 6A). The coordinates of the single peptides as well as those of tetramers and a hexamer are taken from the X-ray structure (PDB id: 2BFI).<sup>49</sup> Individual residues are numbered sequentially, so that peptide 1–12 contains residues from 1 to 12, etc.

The system shown in Figure 6A undergoes energy minimization, equilibration, and 5  $\mu$ s of production MD. Positional restraints are not used at any point of the simulation protocol. Already at the equilibration, the preassembled aggregates dissociate, perhaps owing to the lack of stabilization provided by the crystal environment. Nevertheless, single peptides rapidly engage interactions with other partners. The initial configuration of the production phase consists of 10 aggregates that evolve rapidly, toward 12–14 during the first 50 ns, approximately (Figure 6B). Throughout the first microsecond, the total number of aggregates decreases almost monotonically defining intermediate conformations with occurrence times ranging from a few ns to over 0.2  $\mu$ s. Small aggregates progressively fuse into higher size clusters, reaching saturation (one single aggregate of 22 peptides) after one microsecond of simulation.

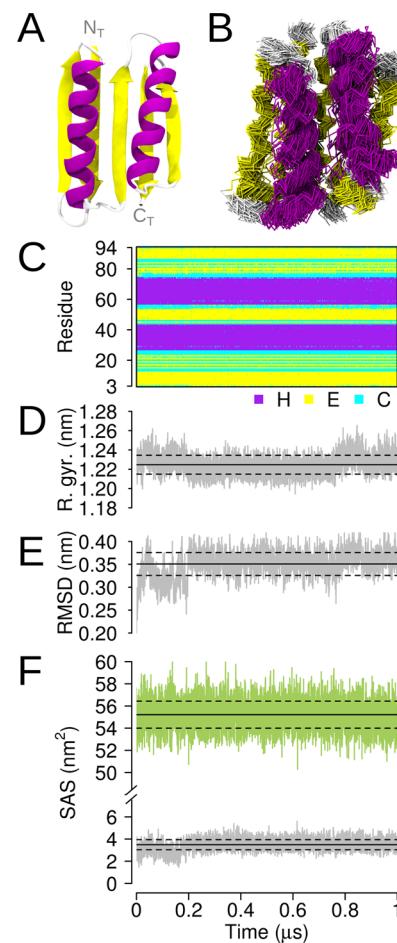
To gain further insight into the self-assembly process, we monitored the formation of interstrand H-bonds interactions. For shortness, we focus on the aggregation process around peptide 1–12. The first contact occurs at nearly 0.3  $\mu$ s between the backbone beads of residues K9 and F11 with F107 and K105, respectively (Figure 6C). After ~0.7  $\mu$ s, peptide 145–156 contacts peptide 1–12 initiating an antiparallel zippering to form a  $\beta$ -strand that remains stable for the rest of the simulation (Figure 6D). The alternation in distance profiles between backbone beads shown in Figure 6D is nicely coincident with that expected for an antiparallel  $\beta$ -sheet.

A more detailed analysis indicates that the first contact occurs at 679 ns between residues F2 and F155 (Figure 6E). Approximately 50 ns after such contact, a salt bridge forms between E4 and K152. Rapidly after that interbackbone H-bonds form between the three central alanines of each chain (740 ns). This is followed by the formation of a second salt bridge between K8 and E148 (759 ns). Finally, close to the end of the first microsecond, F10 pairs with F146 forming a second hydrophobic contact that “zips up” the  $\beta$ -strand. These interactions are maintained until the end of the simulation. As expected from solvent competition, salt bridges show more noticeable fluctuations than hydrophobic contacts (Figure 6F).

The stabilization provided by hydrophobic contacts between phenylalanine rings and salt bridge interactions between lysine and glutamate residues is consistent with previously reported data,<sup>49</sup> sustaining the ability of the model not only to describe the formation of backbone H-bonds that lead to  $\beta$ -sheets

configurations but also to feature selective interaction patterns according to the physicochemical properties of different residues.

**Test Case 3:Alpha/Beta Protein TOP7.** As a first test for the structural stability of a protein, we select the single domain alpha/beta protein Top7 (PDB id: 1QYS). This protein was originally designed using computational techniques and was showed to have an extremely stable folding.<sup>50</sup> Indeed, its structural stability has been recently proven by inserting different HIV-1 epitopes into the Top7 scaffold without perturbing the global folding.<sup>51</sup> This protein contains the most common and stable secondary structure motifs, i.e., two  $\alpha$ -helices and five  $\beta$ -strands arranged in an antiparallel  $\beta$ -sheet (Figure 7A). After protonation and mapping of FG coordinates to CG, we set up a system containing the protein solvated with 1033 WT4 molecules and 30 NaCl ionic couples representing a rough ionic strength of ~0.15 M. Because the first and last residues in the sequence are present in the crystal but not



**Figure 7.** MD simulations of Top7 protein. (A) Cartoon representation of the FG structure of Top7 showing the secondary structure elements (helix, purple; extended, yellow; coil, white). Amino and carboxyl termini (NT and CT) are indicated. (B) Backbone superposition of 100 snapshots taken every 10 ns during the trajectory colored as in panel A. (C) Secondary structure per residue as a function of time. (D and E) Radius of gyration and RMSD as a function of the simulation time. (F) Solvent accessible surface (SAS) of hydrophilic (top) and hydrophobic (bottom) residues. In panels D–F, black solid and dashed lines indicate the average and standard deviations values, respectively.

solved in the experimental electronic density, the capping amino acids are considered as neutral (not zwitterionic). Using this set up, we carry out a production run of 1  $\mu$ s. Visual inspection of the superposition of 100 frames equally separated along the simulation time suggests a good conservation of the secondary and tertiary structure of the protein upon CG dynamics, specially taking into account the absence of topological constraints in the force field (Figure 7B). In fact, it is immediate to recognize the helical and extended segments as a function of the sequence and simulation time (Figure 7C). As a result of thermal motion, the structure undergoes small fluctuations, which are more pronounced near the extremities of structural elements. In particular, the strands spanning residues 15–24 and 77–84 show a more labile character, which is originated by the topology of the protein. Inspection of Figure 7A indicates that these two segments are indeed at the flanks of the  $\beta$ -sheet and hence less stabilized by the protein scaffold. Quantitatively, the secondary structure calculation on the initial coordinates results in H: 38.0%, E: 43.5% and C: 18.5%. These values experience small variations over the dynamics with averages of H, 37.4% (s.d. 0.8); E, 39.2% (s.d. 2.2); and C, 23.4% (s.d. 2.3).

To acquire a more comprehensive assessment of the quality of SIRAH, we adopted a series of descriptors to characterize the structure and dynamics simulated for this and other systems.

As a first indicator of the global molecular shape, we calculate the radius of gyration, which after minor variations within the first 0.2  $\mu$ s adopts a value very close to the original (1.21 nm) oscillating around 1.22 nm. (Figure 7D). In line with these results, the RMSD suggests that the overall folding of the protein is well maintained, although the RMSD experiences a steep increase during the first 50 ns, followed by smaller fluctuations during the next 0.15  $\mu$ s to then reach a plateau with an average value of 0.36 nm (s.d. 0.02) (Figure 7E).

Despite the good behavior of these descriptors, it is still possible that the protein remains in a compact state with  $C\alpha$  relatively close to the initial conformation but with flipped amino acids side chains exposing hydrophobic parts to the solvent. To exclude this possibility, we calculated the SAS of hydrophilic and hydrophobic residues. This is an important control as significant variations in SAS would be indicative of a poorly reproduced hydrophobic/hydrophilic balance in our CG water model. As shown in Figure 7F, hydrophilic and hydrophobic surfaces are well maintained with a modest increase in the hydrophobic SAS, which stabilizes after  $\sim$ 0.2  $\mu$ s. Average values calculated before and after that point differ 0.8 nm<sup>2</sup> from the initial value. This represents a very small difference considering that the hydrophobic surface of an atomistic alanine is 1.1 nm<sup>2</sup>, suggesting a good reproduction of the hydrophobic effect in WT4.

Finally, we sought to monitor the conservation of native contacts along the dynamics as a sensible indicator of the faithful conservation of the protein topology. If the analysis is performed skipping the first five neighboring residues to avoid contacts directly related with the secondary structure (non-contiguous contacts), we obtain a conservation of 75% (s.d. 3) with an accuracy of 64% (s.d. 3). However, considering all possible contacts (global contacts), the conservation rises to 89% (s.d. 1) with an accuracy of 84% (s.d. 1).

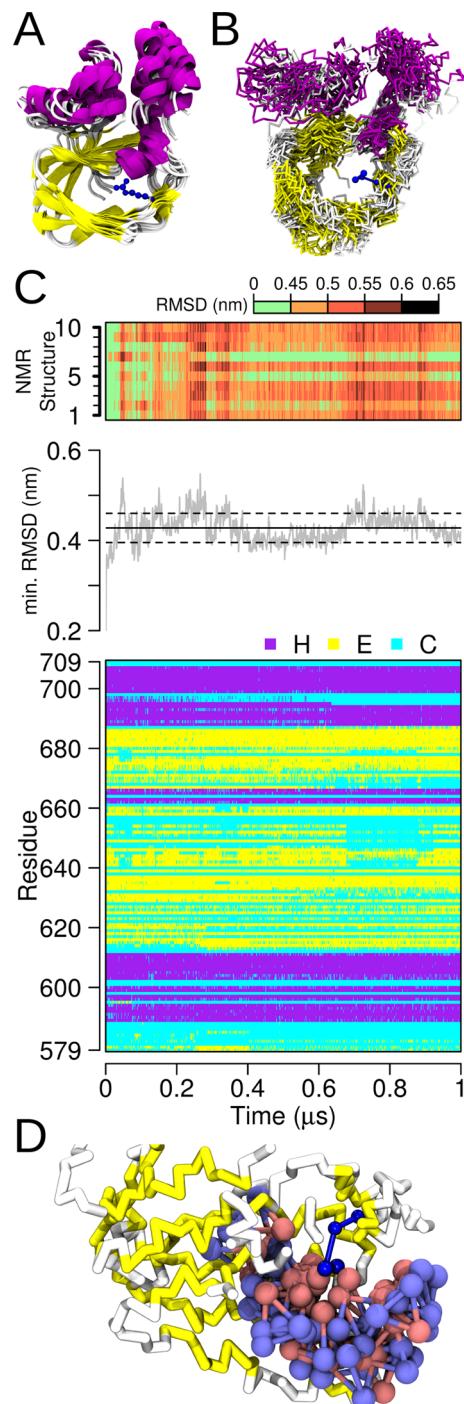
All together, the conformational and dynamical descriptors analyzed indicate good behavior of the force field for this very stable and well-folded protein.

**Test Case 4: Cyclic Nucleotide Binding Domain of HCN Channel.** The cyclic nucleotide binding domain (CNBD) is a conserved protein module of nearly 120 amino acids, comprising helical elements and an eight-stranded  $\beta$ -barrel, which serves as a binding site for cyclic nucleotides. A conserved arginine is present within the binding site to establish a salt bridge interaction with the exocyclic phosphate of the cyclic nucleotide upon binding.<sup>52</sup> This arginine is partially buried but accessible to the solvent in the apo state (Figure 8A). Structural determinations of these protein domains in absence of cAMP remained elusive to experimental methods for many years owing to its intrinsic flexibility. Recently, the structure of the CNBD of the HCN channel has been solved by a combination of spectroscopic techniques (PDB id: 2MNG).<sup>53</sup> This structure was chosen as a test case because of the challenging topology of the  $\beta$ -barrel and to ascertain the capability of the force field to reproduce a flexible fold and the local solvation of a partially buried Arginine.

In close similarity with the previous case, the secondary and tertiary structure of this protein is well conserved, as shown from the comparison between Figure 8A and B. The variation in the radius of gyration remains below 5% of its initial value, and the conservation of global contacts scores 84% (s.d. 3) with an accuracy of 88% (s.d. 1). Calculation of the RMSD indicates that the structure stabilizes within the first 0.1  $\mu$ s (Figure 8C). Along the trajectory, the protein explores conformations that are almost 0.6 nm far from the initial state (model 1 in the NMR family). However, to establish a proper contrast with experimental data, we compare the structural plasticity of the protein along the MD trajectory with all the conformations in the NMR ensemble.<sup>53</sup> As shown in Figure 8C top, the protein visits a series of states that are alternatively closer to one or more NMR conformers. After 0.3  $\mu$ s, the CNBD separates from the initial conformer by more than 0.5 nm but samples conformations closer to NMR models 5 and 7. Calculating the minimum RMSD to any of the NMR conformers indicates that the protein remains at nearly 0.4 nm from at least one of the experimental structures (Figure 8C middle). This puts forward the capability of SIRAH to sample different energetic states on a potential energy surface, roughly reproducing the experimental ensemble of conformations.

Analysis of the secondary structure during the simulation time suggests a good global conservation with some marked local fluctuations (Figure 8C, bottom). In particular, we notice that the region spanning residues 640–660 shows poorer secondary structure conservation. However, it is interesting to notice that this stretch contains the  $\beta$ 4–5 loop, which is the least conserved region in the entire protein family.<sup>54</sup> Moreover, NMR studies report a local minimum in the HN NOE profile for residues 645–655,<sup>53</sup> and atomistic simulations have shown high fluctuations for the same residues.<sup>55,56</sup>

Finally, we sought to investigate the solvation of R669, which is conserved in this protein module and is responsible for the affinity to cyclic nucleotides. Given the relatively large dimensions of WT4, it is not obvious that this arginine partially buried in a hydrophobic environment can be solvated without perturbing the protein scaffold. However, monitoring the arrangement of WT4 molecules around R669, we observe that WT4 can penetrate into the cavity and interact with this arginine (Figure 8D). This interaction is clearly specific and driven by electrostatics as negative beads of WT4 present a preferential orientation toward the side chain of R669. This reinforces the role of the electrostatics in the proper orientation



**Figure 8.** MD simulations of HCN’s CNBD. (A) Superposition of all conformers in the NMR structure (PDB id: 2MNG) colored by secondary structure. R669 within the  $\beta$ -barrel binding site is shown in blue balls and sticks. (B) Backbone superposition of 100 snapshots taken every 10 ns during the trajectory. R669 is represented as in panel A. (C) Top: Color matrix of RMSD values along the CG trajectory against the NMR ensemble. Middle: Minimum RMSD from any of the NMR conformers. Continuous and dashed lines indicate the average and standard deviations. Bottom: Secondary structure assignment. (D) Superposition of WT4 molecules around 0.5 nm of the BCZ bead of R669 colored by charge (positive, blue; negative, red). The  $\beta$ -barrel and R669 are shown in the initial state, while WT4 snapshots are taken every 100 ns.

of the solvent observed for charged residues shown in Figure S8 and C for test case 1.

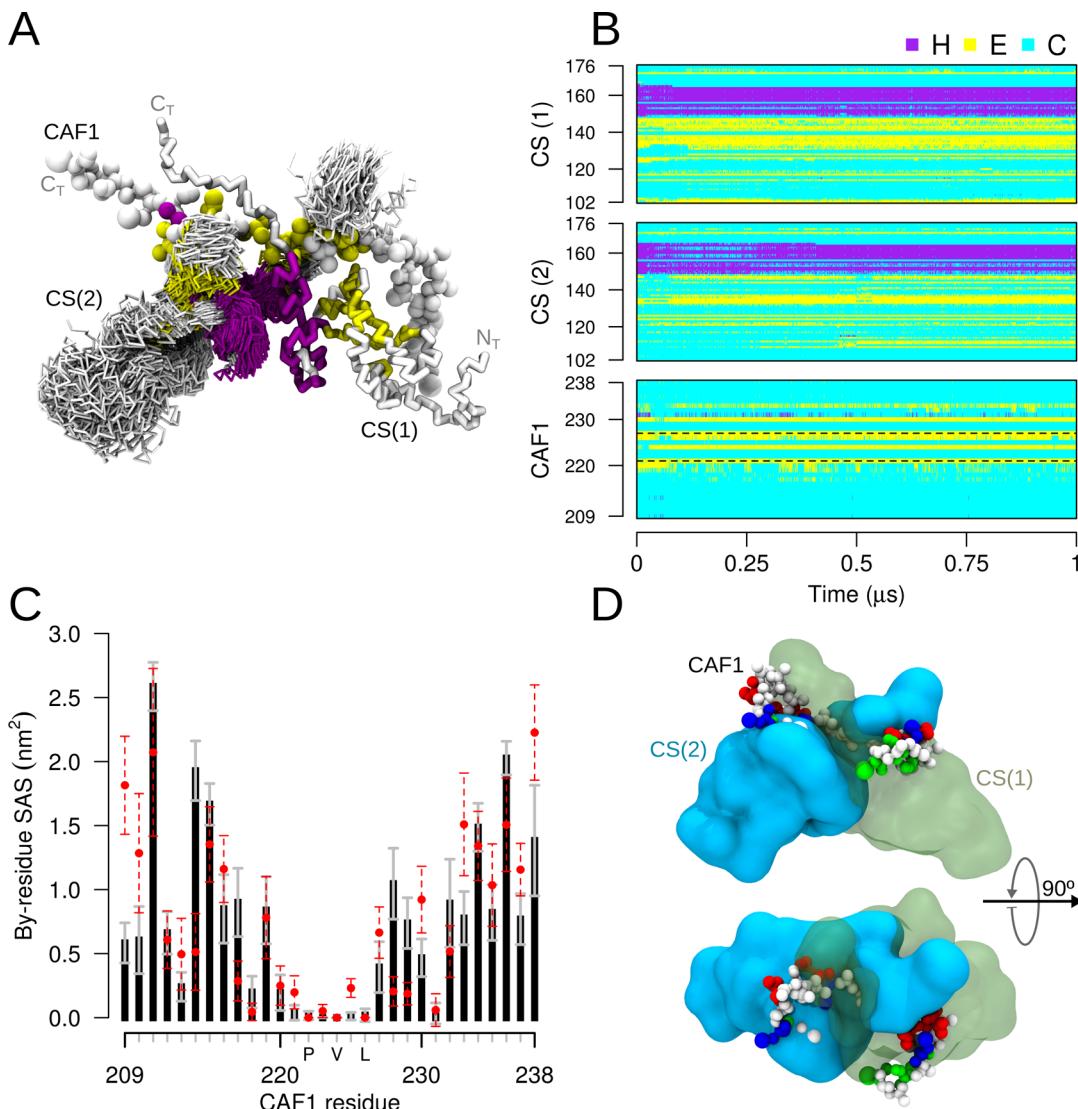
**Test Case 5: HP1 Chromo-Shadow Dimer in Complex with CAF1 Peptide.** As a more challenging test case, we undertake the study of a trimeric complex as that constituted by the chromo-shadow (CS) domain of the heterochromatin protein 1 (HP1) bound to an extended peptide from the chromatin assembly factor-1 (CAF1) determined by NMR spectroscopy.<sup>57</sup> The CS domain is present at the C-terminal of the HP1 and is named after its high homology to the chromo (CHROmatin MOdifier) domain.<sup>58</sup> Upon dimerization, the CS homodimer creates a binding site that recognizes small peptides carrying the consensus motif PXVXL. These peptides bind with high affinity in an extended conformation to a cleft formed by C-terminal residues of the CS dimer forming a trimeric  $\beta$ -sheet (Figure 9A). Each component of this trimeric complex contains a rigid core with highly flexible and solvated terminals.

This domain dimerizes through a nontrivial interface that involves hydrophobic, aromatic, polar, and charged residues in  $\alpha$ -helix and  $\beta$ -sheet conformations. Hence, the correct representation of the intramolecular interactions shown in the previous test cases may not be enough to properly capture the noncovalently bonded forces holding the trimeric complex. Moreover, this requires a fine specificity in the formation of interchain H-bonds and hydrophobic protein–protein interfaces in the correct relative position.

The CG simulation shows a stable trajectory with an average radius of gyration of the dimer of 1.8 nm, which compares very well with the 1.9 nm calculated on the NMR set. The average RMSD of the dimer calculated against the NMR ensemble results in 0.47 nm (s.d. 0.02). Analogously, we found a comparable structural variability for each CS monomer with RMSD values of 0.45 nm (s.d. 0.02) and 0.40 nm (s.d. 0.03). Indeed, the global shape is well preserved as pointed out by snapshots taken along the trajectory (Figure 9A). Consistently, the secondary structure calculated along the trajectory is conserved in all three components of the complex. Both monomers present the same global pattern of secondary structure along the simulation with unstructured N- and C-tails, a  $\beta$ -sheet, and two short helical elements, which mediate the dimerization interface. Additionally, the binding peptide remains in an extended  $\beta$ -strand strand conformation (Figure 9B).

Analysis of global contacts at the interface residues reveals a conservation of 55% (s.d. 4) with an accuracy of 26% (s.d. 2.6), suggesting that lumping the side chain of hydrophobic amino acids into one single bulkier CG bead (Figure 2) may weaken the specificity of some interactions. To further assess this, we calculate the protein–protein interface. The NMR family is characterized by an average total protein–protein interface of  $10.7 \text{ nm}^2$  (s.d. 0.5) in comparison with an average of  $15.2$  (s.d. 0.8) obtained along the CG simulation. This effect is, however, limited to the dimerization interface, as calculation of global contacts lead to 80% of conservation with an accuracy of 76%. Hence, although fine details of the protein–protein interface can be missed, the global topology is well conserved. In agreement, the average total SAS of the complex for the NMR family is  $115.1 \text{ nm}^2$  (s.d. 1.9), while its counterpart in the simulation is  $112.1 \text{ nm}^2$  (s.d. 2.2).

Regarding the interaction of the CS dimer with the CAF1 peptide, the by-residue SAS and the secondary structure analysis of the CAF1 peptide shows that the PXVXL region remains buried in the binding site cavity and interacting in an



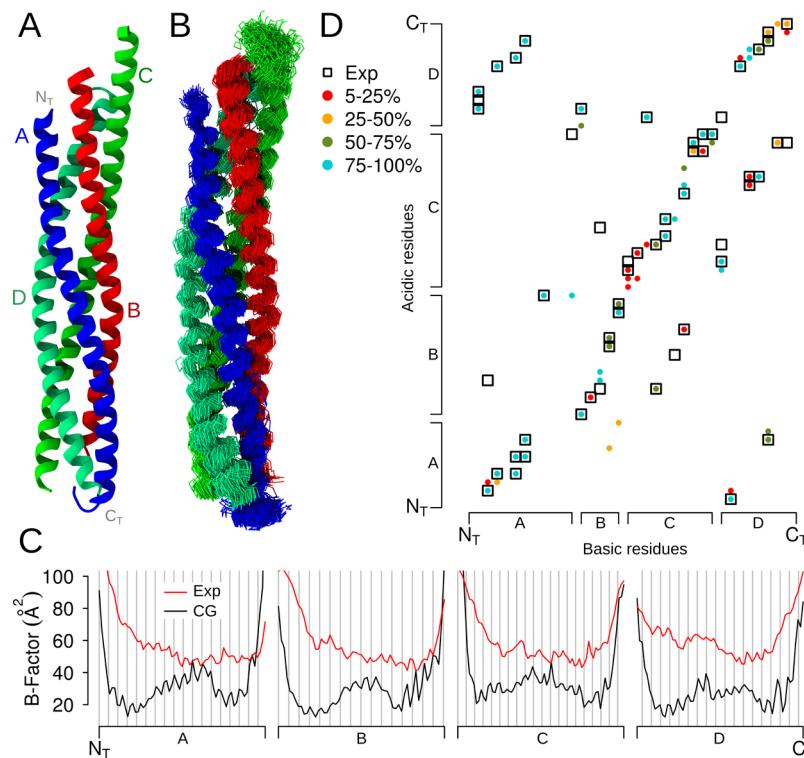
**Figure 9.** HP1 CS dimer in complex with CAF1 peptide. (A) The backbone's initial conformation of monomer CS(1) is represented by thick tubes, while an ensemble of conformations for monomer CS(2), taken one each 10 ns along the of trajectory, is represented by thin tubes. The initial conformation of the CAF1 peptide is shown as balls and sticks. All residues are colored according to the secondary structure. Amino and carboxy terminal residues (NT and CT) are indicated. (B) Variation of secondary structure along 1  $\mu\text{s}$  of trajectory. The location of the PXVXL motif within the CAF1 sequence is indicated by dashed lines. (C) By-residue SAS calculation on the CAF1 peptide. Columns and solid line error bars correspond to average and standard deviation values during the CG simulation. Circles and dashed error bars indicate the analogous quantity measured from the NMR family of conformers. Residues belonging to the signature motif (PXVXL) are indicated. (D) Molecular representation of the final coordinate from the CG trajectory. The CS monomers are represented as cyan and green surfaces, while the CAF1 peptide is shown in balls and sticks, and it is colored by residue type (basic, blue; acidic, red; polar, green; and nonpolar, white).

extended beta conformation, in good agreement with the experimental data<sup>57</sup> (Figure 9B and C, respectively). Moreover, we retrieve a quasi-quantitative agreement between the exposed surfaces of the portion of the CAF1 peptide outside the signature motif. This region is mostly unstructured and exposed to the solvent with the C-terminal tails of the CS monomers embracing the CAF1 peptide (Figure 9D) as it is needed for these regions to actively participate in the ligand recognition.<sup>59</sup>

**Test Case 6: The SNARE Complex.** The SNARE proteins are the main constituents of the synaptic vesicle fusion machinery. The cytoplasmic domains of the three SNARE proteins, VAMP (vesicle-associated membrane protein, also known as synaptobrevin), syntaxin, and SNAP-25, spontaneously form a parallel coiled-coil four-helix bundle, providing sufficient free energy to drive membrane fusion.<sup>60–62</sup>

Continuous helical segments of about 60 residues (one from VAMP, one from syntaxin, and two from SNAP-25) form the fully zippered SNARE complex (Figure 10A).

This system is chosen as a final test case because if small systematic errors are present in our force field they should accumulate and become evident in the very long helices of the SNARE complex, impairing the stability of the heterotetramer upon dynamics. This system is constructed from the PDB structure 1KIL<sup>63</sup> removing the chain E corresponding to complexin. This simulation results in a very stable trajectory with a RMSD that stabilized already after the first 10 ns oscillating around values of  $\sim 0.3$  nm. Along the simulation, the secondary structure is very well maintained, and the whole stability of the 4-helix bundle is not compromised as shown by the superposition of snapshots in Figure 10B.



**Figure 10.** MD simulations of the SNARE complex. (A) Cartoon representation of the SNARE complex colored by chain (A, VAMP; B, syntaxin; C, SNAP-25 N-terminal helix, and D, SNAP-25 C-terminal helix). (B) Superposition of 100 snapshots taken every 10 ns during the trajectory colored as in panel A. (C) B-factors on the  $\text{Ca}$  beads calculated along the dynamics plotted against residue number for the four helices in the complex. The vertical gray lines are drawn every 3.6 residues to indicate the helical periodicity (see text). Experimental values are shown in red. (D) Global pattern of salt bridges in the SNARE complex. The vertical axis indicates the position of acidic residues (aspartic and glutamic acid) sequentially ordered versus basic residues (arginine and lysine). Dots represent the existence of salt bridges at any point of the simulation and are color coded by the percentage of occurrence along the trajectory. The experimentally observed salt bridges are presented as empty squares.

Calculation of the B-factors on the  $\text{Ca}$  beads measured along the simulation indicates that as expected termini are more mobile than the central portion of the protein (Figure 10C). Moreover, each chain presents a fine structure with a serrated shape. A deeper consideration of this pattern indicates that the alternation of values nicely coincides with the 3.6 periodicity of a  $\alpha$ -helix (gray lines, Figure 10C) and is imposed by the quaternary structure. Exposed residues show higher B-factors, while buried ones do the opposite. Although a direct comparison with the crystal structure would not be strictly correct because of packing effects, temperature, buffer conditions, etc., it is still possible to find a rough correlation with the experimental B-factors (Figure 10C).

The conservation of native contacts along the simulation considering noncontiguous interactions is 67% (s.d. 3), while the global contacts results in 92% (s.d. 1). Considering the predominantly helical topology of the complex, noncontiguous contacts are essentially related to the protein–protein interface. Hence, the difference between global and noncontiguous conservation is in line with the conclusion drawn in the previous test case. There, we observe a certain degree of promiscuity within the protein–protein interface. However, the protein–protein interfaces between each helix and the rest of the complex calculated along the simulation, differ at most in a 5% from the experimental values.

Finally, we turn our attention to the electrostatic interactions in the complex. To this aim, we calculated the salt bridges existing in the crystallographic structure and compare them with those conserved and/or arising during the simulation. A

global view of the salt bridge network on the SNARE complex is presented in Figure 10D. A comparison indicates that 92% of the salt bridges measured from the experimental data are conserved during the simulation with different occurrence times. We also observe the transient formation of additional interactions, mainly concentrated on helix C (SNAP-25 N-terminal). These interactions have in general low occurrence times and are concentrated in the diagonal region of the graph, being indicative of intramolecular bridges. Off-diagonal elements are conserved, supporting the good reproduction of the quaternary structure.

**Testing the Robustness of SIRAH.** Finally, aimed to acquire a more general vision of the performance of SIRAH, we carried out a set of simulations of different proteins. These structures are selected from the PDB with the following criteria: (i) Structures contain only a protein without any other small ligand or macromolecular species. (ii) There are no disulfide bonds or modified amino acids. (iii) Molecules are monomeric with one single chain per asymmetric unit. This minimally representative group of systems is protonated at pH 7 using the pdb 2pqr server ([http://nbcr-222.ucsd.edu/pdb2pqr\\_1.8/](http://nbcr-222.ucsd.edu/pdb2pqr_1.8/)). Without any further checking, proteins in their zwitterionic form are solvated in an octahedral box adding 1.5 nm of WT4 in each direction and simulated at 300 K and 1 atm in the presence of 0.15 M of NaCl for 1  $\mu$ s. For the sake of shortness, results are summarized in Table 3. In all the cases, the quantities are calculated on the experimental structure and compared against the average of the last 0.1  $\mu$ s of each trajectory.

Table 3. Robustness of SIRAH on Different Proteins<sup>a</sup>

PDB	N	RMSD (nm)	radius of gyration (nm)	SAS (nm <sup>2</sup> )	Hphob. Hphil.	secondary structure H, E, C	contacts conservation/accuracy
1TQG	105	0.38(0.04)	[1.38] 1.41(0.01)	[ 6.4] 80.3(1.3); [59.7] 54.0(1.1)	[86.7] 73.8(1.1); [ 1.9] 3.1(1.6); [11.4] 23.1(2.2)		86.8(0.6)/ 80.3(1.3)
1R69	63	0.87(0.10)	[1.01] 1.34(0.03)	[ 3.2] 10.3(0.7); [38.2] 42.1(1.2)	[69.8] 58.4(2.4); [ 0.0] 2.3(1.6); [30.2] 39.4(3.1)		83.8(1.0)/ 87.9(1.4)
2CKX	83	0.71(0.07)	[1.17] 1.31(0.01)	[ 8.2] 9.9(0.5); [46.5] 49.6(1.1)	[66.3] 58.2(2.0); [ 6.0] 10.7(1.7); [27.7] 31.0(2.7)		79.1(0.5)/ 72.9(1.2)
1BKR	108	0.60(0.08)	[1.24] 1.31(0.03)	[ 6.1] 11.4(1.1); [57.4] 56.2(1.5)	[58.3] 55.2(1.5); [ 3.7] 6.4(1.1); [38.0] 38.4(1.8)		77.2(1.0)/ 73.7(1.4)
1RA4	117	0.36(0.04)	[1.23] 1.24(0.01)	[ 6.2] 8.2(0.6); [59.7] 63.1(1.5)	[48.7] 45.6(1.3); [19.7] 15.1(1.3); [31.6] 39.3(1.8)		81.3(0.9)/ 76.4(0.9)
2KYR	108	0.58(0.06)	[1.27] 1.38(0.01)	[ 8.2] 13.4(0.6); [56.9] 64.2(1.5)	[46.3] 44.2(1.2); [32.4] 21.7(2.0); [21.3] 34.1(2.4)		74.8(1.0)/ 81.3(1.2)
2VIM	104	0.55(0.06)	[1.22] 1.35(0.02)	[ 5.6] 11.4(1.0); [52.7] 57.4(1.5)	[46.2] 40.6(0.8); [36.5] 29.6(2.2); [17.3] 29.8(2.3)		80.4(1.3)/ 78.2(1.8)
1CRN	46	0.76(0.08)	[0.97] 1.13(0.04)	[ 8.1] 13.1(0.8); [24.1] 28.7(1.0)	[45.7] 40.2(1.5); [23.9] 10.9(3.4); [30.4] 48.8(3.5)		80.4(1.4)/ 79.4(2.3)
1ORC	64	0.49(0.05)	[1.08] 1.10(0.02)	[ 4.9] 7.1(0.6); [43.3] 43.2(1.1)	[40.6] 33.4(1.6); [35.9] 21.5(2.7); [23.4] 45.1(3.1)		79.1(1.5)/ 73.1(2.4)
1PGB	56	0.41(0.06)	[1.03] 1.06(0.01)	[ 2.9] 3.2(0.4); [37.9] 41.6(1.1)	[26.8] 24.8(1.3); [51.8] 37.5(4.6); [21.4] 37.7(4.8)		88.5(1.9)/ 80.7(2.3)
2IUG	110	0.55(0.06)	[1.28] 1.39(0.02)	[ 8.0] 10.2(0.7); [57.1] 63.6(1.6)	[22.7] 21.1(1.1); [37.3] 18.6(2.7); [40.0] 60.4(3.0)		74.0(1.3)/ 77.3(1.2)
1QYO (E222N)	236	0.50(0.05)	[1.75] 1.80(0.01)	[ 8.2] 12.5(0.7); [105.1] 101.8(1.9)	[ 7.2] 8.7(0.6); [55.5] 55.0(1.7); [37.3] 36.3(1.8)		80.1(0.5)/ 77.3(0.6)
1OPS	64	0.56(0.06)	[1.01] 1.05(0.01)	[ 8.9] 10.2(0.5); [30.4] 35.4(1.1)	[ 4.7] 0.5(0.7); [56.2] 34.1(3.8); [39.1] 65.5(3.8)		69.6(1.1)/ 71.0(1.8)
1GYV	120	0.35(0.04)	[1.44] 1.50(0.01)	[ 9.1] 11.5(0.7); [60.1] 62.6(1.1)	[ 1.7] 1.0(0.6); [61.7] 52.8(2.7); [36.7] 46.2(2.8)		82.6(1.0)/ 81.0(0.9)
2E3H	76	0.55(0.06)	[1.12] 1.16(0.03)	[ 8.0] 9.7(0.8); [43.3] 44.1(1.7)	[ 0.0] 1.0(0.6); [48.7] 45.8(5.1); [51.3] 53.2(5.1)		73.7(1.3)/ 79.5(2.3)
1PWT	61	0.60(0.07)	[1.03] 1.15(0.02)	[ 5.1] 9.3(0.6); [43.5] 40.0(1.0)	[ 0.0] 0.6(0.8); [60.7] 34.9(3.2); [39.3] 64.6(3.4)		68.2(1.4)/ 81.1(1.4)
2O31	67	0.55(0.07)	[1.05] 1.26(0.03)	[ 9.5] 13.5(0.7); [40.1] 42.1(1.3)	[ 0.0] 0.2(0.5); [62.7] 36.3(4.2); [37.3] 63.5(4.3)		73.5(2.3)/ 83.5(1.5)
1XX8 (150 mM NaCl)	66	0.63(0.09)	[1.12] 1.14(0.03)	[ 3.7] 5.0(0.8); [49.2] 50.6(1.0)	[16.7] 10.7(1.7); [42.4] 34.8(3.9); [40.9] 54.5(4.3)		82.9(1.1)/ 81.7(2.0)
1XX8 (300 mM NaCl)	66	0.37(0.04)	[1.12] 1.06(0.01)	[ 3.7] 3.1(0.4); [49.2] 47.0(1.1)	[16.7] 13.6(1.1); [42.4] 42.0(3.1); [40.9] 44.3(3.3)		88.3(1.3)/ 81.3(1.7)

<sup>a</sup>N indicates the number of residues. Values reported correspond to the average calculated over the last 100 ns of each  $\mu$ s. Parentheses and square brackets indicate standard deviations and values calculated from the experimental structure, respectively.

In terms of RMSD, we observe that deviations from the initial structures range from nearly atomistic (0.35 nm for 1GYV) up to as high as 0.87 nm for 1R69 (see below). Nevertheless, a comprehensive comparison of the radius of gyration in all the cases indicates that the compaction of the proteins experiences less relevant variations. This is supported by relatively high conservation of global contacts for which the lowest value found is 68% for 1PWT. The low variations in the SAS values indicate a conservation of the hydrophobic core of the molecules with an acceptable balance between the hydrophobicity/hydrophilicity of amino acids with different physicochemical characteristics. Secondary structure elements show a variability that can be comparable with that expected from FG simulations, showing no apparent bias for any structural motif.

This set of simulations provides an estimation of the robustness of the force field because none of the systems was set up with optimal conditions such as temperature, ionic strength, protonation states, charges in the termini, etc. Similarly, gentler minimization/initialization protocols could significantly improve the outcome of the simulations. As a concrete example, we consider the simulation of the structure 1XX8 (last rows in Table 3). This structure corresponds to a

DNA binding protein and has a very high charge density with a net charge of +6 in 66 residues. Simply doubling the ionic strength in the solvation box significantly improves the stability in all the molecular descriptors (compare two last rows in Table 3).

It is also important to recognize that there are still situations in which the performance of our force field is not satisfactory. For instance, the protein 1R69 is the worst case reported in Table 3. This protein comprises five short  $\alpha$ -helices, two of these forming a helix-turn-helix motif. Analysis of the trajectory suggests that one of the causes of the partial unfolding is the solvation of a partially buried salt bridge between R10 and E35. Owing to the high granularity of WT4, it cannot occupy single-water cavities, failing to provide stabilizing interactions within small hydrophobic cavities. Additionally, the helix-turn-helix motif may present challenges to relax the tension in the reduced number of torsional degrees of freedom in CG models. These constitute clear limitations intrinsic of the CG approach and are difficult to capture without the use of specific constraints and have to be evaluated in each particular case. An example of possible solutions for this kind of problems is the simulation of the protein 1QYO, which is a variant of the green fluorescent protein that is not able to create a

chromophore. This protein contains nearly 30 single water molecules within the  $\beta$ -barrel and E222 pointing into the  $\beta$ -barrel. The impossibility of introducing WT4 molecules within the barrel results also in a badly unfolded protein (not shown). However, introducing the E222Q mutation to mimic a protonated glutamic acid, as suggested by PROPKA calculations,<sup>64</sup> produces a stable trajectory with well conserved overall molecular descriptors (Table 3).

## CONCLUSIONS

In this paper, we present an extension of the SIRAH force field for unbiased MD simulations of peptides and proteins at the CG level using explicit solvation and long-range electrostatics. In line with our previous developments, all the interactions of this residue-based CG model are represented within a classical Hamiltonian, which is common to most MD simulation packages. Comparative simulations with fully atomistic systems containing an equivalent number of particles as in the test cases presented indicate a speed up of 2 orders of magnitude. The FG-to-CG mapping using positions of real atoms to place CG beads allows for the nearly atomistic identification of functional groups establishing interactions as H-bonds and salt bridges. This mapping strategy will facilitate the future development of modified amino acids to study the influence of phosphorylation, methylation, different protonation states, etc. The use of long-range electrostatics, permittivity of the solvent, and explicit presence of ions in the solution allows for including ionic strength effects that are very difficult to capture for CG approaches.

Clearly, this contribution paves the way for the study of protein–DNA complexes using SIRAH. Although preliminary simulations show encouraging results, statistic analysis against protein–DNA interfaces reported in the PDB reveals low discrimination in the interaction with the positive side chains. This would call for an optimization of the interaction between phosphate groups and the side chains of arginines and lysines. Work is underway in our group to address this problem and will be the subject of a forthcoming publication.

To facilitate the implementation of the SIRAH force field, a tarball containing a complete set of interaction topology files in Gromacs format (version 4.5.5) is freely available at <http://www.sirahff.com>. The documentation contained therein includes step-by-step tutorials, prestabilized WT4 solvation boxes, and mapping scripts to convert from PDB to CG and from CG to PSF formats, facilitating the visualization and analysis of molecular systems.

## AUTHOR INFORMATION

### Corresponding Author

\*Tel/Fax: +598-2522 0910. E-mail: spantano@pasteur.edu.uy.

### Author Contributions

L.D. and M.R.M. equally contributed to this work

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was partially funded by FOCEM (MERCOSUR Structural Convergence Fund), COF 03/11. A.F.B. is beneficiary of a National Fellowship provided by ANII. M.R.M. and S.P. are researchers from the National Scientific Program of ANII (SNI). We thank Paolo Carloni and Anna Tramontano for very useful discussions.

## REFERENCES

- (1) Klein, M. L.; Shinoda, W. Large-scale molecular dynamics simulations of self-assembling systems. *Science* **2008**, *321*, 798–800.
- (2) Saiz, L.; Klein, M. L. Computer simulation studies of model biological membranes. *Acc. Chem. Res.* **2002**, *35*, 482–489.
- (3) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*, 1 ed.; Taylor & Francis Group: New York, 2009; pp 1–455.
- (4) Sansom, M. S.; Scott, K. A.; Bond, P. J. Coarse-grained simulation: a high-throughput computational approach to membrane proteins. *Biochem. Soc. Trans.* **2008**, *36*, 27–32.
- (5) Orsi, M.; Essex, J. W. The ELBA force field for coarse-grain modeling of lipid membranes. *PLoS One* **2011**, *6*, e28637.
- (6) Trylska, J.; Tozzini, V.; Chang, C. E.; McCammon, J. A. HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophys. J.* **2007**, *92*, 4179–4187.
- (7) Sieradzan, A. K.; Niadzvedtski, A.; Scheraga, H. A.; Liwo, A. Revised backbone-virtual-bond-angle potentials to treat the L- and D-amino acid residues in the coarse-grained united residue (UNRES) force field. *J. Chem. Theory Comput.* **2014**, *10*, 2194–2203.
- (8) Basdevant, N.; Borgis, D.; Ha-Duong, T. Modeling protein–protein recognition in solution using the coarse-grained force field SCORPION. *J. Chem. Theory Comput.* **2013**, *9*, 803–813.
- (9) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
- (10) Pasi, M.; Lavery, R.; Ceres, N. PaLaCe: A coarse-grain protein model for studying mechanical properties. *J. Chem. Theory Comput.* **2013**, *9*, 785–793.
- (11) Han, W.; Schulten, K. Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: Improved backbone hydration and interactions between charged side chains. *J. Chem. Theory Comput.* **2012**, *8*, 4413–4424.
- (12) Arkhipov, A.; Yin, Y.; Schulten, K. Four-scale description of membrane sculpting by BAR domains. *Biophys. J.* **2008**, *95*, 2806–2821.
- (13) Spiga, E.; Alemani, D.; Degiacomini, M. T.; Cascella, M.; Dal Peraro, M. Electrostatic-consistent coarse-grained potentials for molecular simulations of proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3515–3526.
- (14) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. Coarse-grained model of proteins incorporating atomistic detail of the active site. *Phys. Rev. Lett.* **2005**, *95*, 218102.
- (15) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, P. H.; Hünenberger, S. J.; Marrink, S. J. Martini extension to carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.
- (16) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. Multiscale modeling of emergent materials: biological and soft matter. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869–1892.
- (17) Savelyev, A.; Papoian, G. A. Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. U. S. A* **2010**, *107*, 20340–20345.
- (18) Bereau, T.; Deserno, M. Generic coarse-grained model for protein folding and aggregation. *J. Chem. Phys.* **2009**, *130*, 235106.
- (19) Tschöp, W.; Kremer, K.; Han, O.; Batoulis, J.; Bürger, T. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.* **1998**, *46*, 61–74.
- (20) Faller, R. Systematic Coarse Graining of Polymers and Biomolecules. In *Multiscale Modelling Methods for Applications in Materials Science*; Kondov, I., Sutmann, G., Eds.; Julich, Germany, September 16–20, 2013; pp 135–150.
- (21) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.
- (22) Ingólfsson, H. I.; López, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The power of coarse graining in biomolecular simulations. *WIREs Comput. Mol. Sci.* **2013**, *4*, 225–248.
- (23) Brini, E.; Algaer, E. A.; Ganguly, P.; Rodriguez-Ropero, F.; van der Vegt, N. F. A. Systematic coarse-graining methods for soft matter simulations—a review. *Soft Matter* **2013**, *9*, 2108–2119.

- (24) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J. Chem. Theory Comput.* **2010**, *6*, 1711–1725.
- (25) Zeida, A.; Machado, M. R.; Dans, P. D.; Pantano, S. Breathing, bubbling, and bending: DNA flexibility from multimicrosecond simulations. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2012**, *86*, 021903.
- (26) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. Another coarse grain model for aqueous solvation: WAT FOUR? *J. Chem. Theory Comput.* **2010**, *6*, 3793–3807.
- (27) Darre, L.; Tek, A.; Baaden, M.; Pantano, S. Mixing atomistic and coarse grain solvation models for MD simulations: Let WT4 handle the bulk. *J. Chem. Theory Comput.* **2012**, *8*, 3880–3894.
- (28) Darre, L.; Machado, M. R.; Pantano, S. Coarse-grained models of water. *WIREs Comput. Mol. Sci.* **2012**, *2*, 921–930.
- (29) Gonzalez, H. C.; Darre, L.; Pantano, S. Transferable mixing of atomistic and coarse-grained water models. *J. Phys. Chem. B* **2013**, *117*, 14438–14448.
- (30) Machado, M. R.; Dans, P. D.; Pantano, S. A hybrid all-atom/coarse grain model for multiscale simulations of DNA. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18134–18144.
- (31) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725.
- (32) Dans, P. D.; Darre, L.; Machado, M. R.; Zeida, A.; Brandner, A. F.; Pantano, S. Assessing the Accuracy of the SIRAH Force Field to Model DNA at Coarse Grain Level. In *Advances In Bioinformatics and Computational Biology*; Setubal, J. C., Almeida, N. F., Eds.; Springer International Publishing: Gwerbestrasse, Switzerland, 2013; pp 71–81.
- (33) Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragnolini, T.; Chebaro, Y.; St-Pierre, J. F.; Kalimeri, M.; Barducci, A.; Laurin, Y.; Tek, A.; Baaden, M.; Nguyen, P. H.; Derreumaux, P. The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* **2014**, *43*, 4871–4893.
- (34) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theor. Comp.* **2008**, *4*, 819–834.
- (35) Maritan, A.; Micheletti, C.; Trovato, A.; Banavar, J. R. Optimal shapes of compact strings. *Nature* **2000**, *406*, 287–290.
- (36) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037–10041.
- (37) 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.
- (38) Darden, T. A.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (40) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (41) Simmerman, H. K.; Jones, L. R. Phospholamban: Protein structure, mechanism of action, and role in cardiac function. *Physiol Rev.* **1998**, *78*, 921–947.
- (42) Mortishire-Smith, R. J.; Pitzenberger, S. M.; Burke, C. J.; Middaugh, C. R.; Garsky, V. M.; Johnson, R. G. Solution structure of the cytoplasmic domain of phospholamban: Phosphorylation leads to a local perturbation in secondary structure. *Biochemistry* **1995**, *34*, 7603–7613.
- (43) Mascioni, A.; Karim, C.; Zamoon, J.; Thomas, D. D.; Veglia, G. Solid-state NMR and rigid body molecular dynamics to determine domain orientations of monomeric phospholamban. *J. Am. Chem. Soc.* **2002**, *124*, 9392–9393.
- (44) Pantano, S.; Carafoli, E. The role of phosphorylation on the structure and dynamics of phospholamban: a model from molecular simulations. *Proteins* **2007**, *66*, 930–940.
- (45) Haenni, D.; Zosel, F.; Reymond, L.; Nettels, D.; Schuler, B. Intramolecular distances and dynamics from the combined photon statistics of single-molecule FRET and photoinduced electron transfer. *J. Phys. Chem. B* **2013**, *117*, 13015–13028.
- (46) Glenner, G. G.; Keiser, H. R.; Bladen, H. A.; Cuatrecasas, P.; Eanes, E. D.; Ram, J. S.; Kanfer, J. N.; DeLellis, R. A. Amyloid. VI. A comparison of two morphologic components of human amyloid deposits. *J. Histochem. Cytochem.* **1968**, *16*, 633–644.
- (47) Geddes, A. J.; Parker, K. D.; Atkins, E. D.; Beighton, E. “Cross-beta” conformation in proteins. *J. Mol. Biol.* **1968**, *32*, 343–358.
- (48) Astbury, W. T.; Dickinson, S.; Bailey, K. The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem. J.* **1935**, *29*, 2351–2360.
- (49) Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 315–320.
- (50) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003**, *302*, 1364–1368.
- (51) Viana, I. F. T.; Dhalia, R.; Marques; Marques, E. T. A.; Lins, R. D. Influence of Scaffold Stability and Electrostatics on Top7-Based Engineered Helical HIV-1 Epitopes. In *Advances In Bioinformatics and Computational Biology*; Setubal, J. C., Almeida, N. F., Eds.; Springer International Publishing: Gwerbestrasse, Switzerland, 2013; pp 94–103.
- (52) Berman, H. M.; Ten Eyck, L. F.; Goodsell, D. S.; Haste, N. M.; Kornev, A.; Taylor, S. S. The cAMP binding domain: An ancient signaling module. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 45–50.
- (53) Akimoto, M.; Zhang, Z.; Boulton, S.; Selvaratnam, R.; VanSchouwen, B.; Gloyd, M.; Accili, E. A.; Lange, O. F.; Melacini, G. A mechanism for the auto-inhibition of hyperpolarization-activated cyclic nucleotide-gated (HCN) channel opening and its relief by cAMP. *J. Biol. Chem.* **2014**, *289*, 22205–22220.
- (54) Canaves, J. M.; Taylor, S. S. Classification and phylogenetic analysis of the cAMP-dependent protein kinase regulatory subunit family. *J. Mol. Evol.* **2002**, *54*, 17–29.
- (55) Berrera, M.; Pantano, S.; Carloni, P. Catabolite activator protein in aqueous solution: A molecular simulation study. *J. Phys. Chem. B* **2007**, *111*, 1496–1501.
- (56) Berrera, M.; Pantano, S.; Carloni, P. cAMP Modulation of the cytoplasmic domain in the HCN2 channel investigated by molecular simulations. *Biophys. J.* **2006**, *90*, 3428–3433.
- (57) Thiru, A.; Nielispach, D.; Mott, H. R.; Okuwaki, M.; Lyon, D.; Nielsen, P. R.; Hirshberg, M.; Verreault, A.; Murzina, N. V.; Laue, E. D. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *EMBO J.* **2004**, *23*, 489–499.
- (58) Aasland, R.; Stewart, A. F. The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. *Nucleic Acids Res.* **1995**, *23*, 3168–3173.
- (59) Mendez, D. L.; Mandt, R. E.; Elgin, S. C. Heterochromatin protein 1a (HP1a) partner specificity is determined by critical amino acids in the chromo shadow domain and C-terminal extension. *J. Biol. Chem.* **2013**, *288*, 22315–22323.
- (60) Jahn, R.; Scheller, R. H. SNAREs—Engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 631–643.
- (61) Sudhof, T. C.; Rothman, J. E. Membrane fusion: Grappling with SNARE and SM proteins. *Science* **2009**, *323*, 474–477.
- (62) Pantano, S.; Montecucco, C. The blockade of the neurotransmitter release apparatus by botulinum neurotoxins. *Cell. Mol. Life Sci.* **2014**, *71*, 793–811.
- (63) Chen, X.; Tomchick, D. R.; Kovrigin, E.; Arac, D.; Machius, M.; Sudhof, T. C.; Rizo, J. Three-dimensional structure of the complexin/SNARE complex. *Neuron* **2002**, *33*, 397–409.
- (64) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK<sub>a</sub> values. *Proteins* **2005**, *61*, 704–721.