JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models

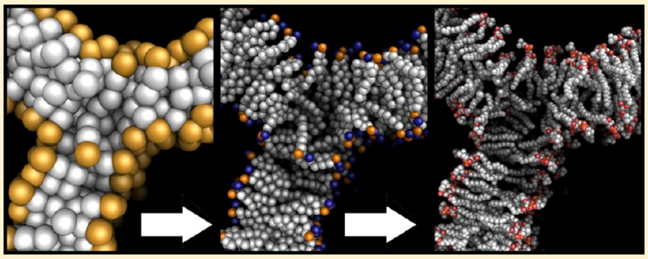
Tsjerk A. Wassenaar,*[†,‡,§,∥] Kristyna Pluhackova,[§,∥] Rainer A. Böckmann,[§] Siewert J. Marrink,[‡] and D. Peter Tieleman[†]

[†]Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

[‡]Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

[§]Computational Biology, Department of Biology, University of Erlangen-Nürnberg, Staudtstr. 5, 91058 Erlangen, Germany

S *Supporting Information*



ABSTRACT: The conversion of coarse-grained to atomistic models is an important step in obtaining insight about atomistic scale processes from coarse-grained simulations. For this process, called backmapping or reverse transformation, several tools are available, but these commonly require libraries of molecule fragments or they are linked to a specific software package. In addition, the methods are usually restricted to specific molecules and to a specific force field. Here, we present an alternative method, consisting of geometric projection and subsequent force-field based relaxation. This method is designed to be simple and flexible, and offers a generic solution for resolution transformation. For simple systems, the conversion only requires a list of particle correspondences on the two levels of resolution. For special cases, such as nondefault protonation states of amino acids and virtual sites, a target particle list can be specified. The mapping uses simple building blocks, which list the particles on the different levels of resolution. For conversion to higher resolution, the initial model is relaxed with several short cycles of energy minimization and position-restrained MD. The reconstruction of an atomistic backbone from a coarse-grained model is done using a new dedicated algorithm. The method is generic and can be used to map between any two particle based representations, provided that a mapping can be written. The focus of this work is on the coarse-grained MARTINI force field, for which mapping definitions are written to allow conversion to and from the higher-resolution force fields GROMOS, CHARMM, and AMBER, and to and from a simplified three-bead lipid model. Together, these offer the possibility to simulate mesoscopic membrane structures, to be transformed to MARTINI and subsequently to an atomistic model for investigation of detailed interactions. The method was tested on a set of systems ranging from a simple, single-component bilayer to a large protein−membrane−solvent complex. The results demonstrate the efficiency and the efficacy of the new approach.

1. INTRODUCTION

"*We need to go forward by going backward.*"—Stanley Crouch

In recent years, coarse graining methods have offered steps forward in simulation of molecular systems, complementing atomistic simulations, and allowing exploration of the behavior of larger systems over longer times. The impact of such methods on biomolecular simulations can be appreciated from recent reviews, which give a comprehensive overview of coarse-graining methods[1] and applications.[2] The processes simulated using CG models commonly range from micro- to milliseconds, which is currently beyond the time scale readily accessible with atomistic models. However, in many cases atomistic insight is required even if simulations are carried out at a coarser level. It is therefore useful to be able to convert coarse-grained models into corresponding atomistic models, either for direct inspection of the atomistic interactions or for continuation of the simulation with a higher resolution. The latter strategy is sometimes referred to as sequential multiscale simulation,[3] and is particularly interesting for studying the interaction of protein−protein and protein−lipid interactions, where the association or equilibration is first established on the coarse-grained level, as these processes are typically too slow to be simulated atomistically.

The conversion of a coarse-grained model into a corresponding atomistic model is commonly termed backmapping, inverse mapping, or reverse transformation. This process consists of two stages. The first stage is the generation of an atomistic starting structure from coarse-grained coordinates, which is followed by the second stage, consisting of the relaxation of the atomistic structure. Ideally, the quality of the atomistic structure derived initially should be such that no or little relaxation is required. This

Received: July 13, 2013

is particularly true if the objective is direct inspection of atomistic interactions, rather than subsequent atomistic simulation.

The methods that have been proposed for backmapping, usually emphasize either the first[4−7] or the second[8−11] stage. In line with this division, there are two general approaches for generating initial structures from coarse-grained coordinates. The first is aimed at near-optimal reconstruction, and uses fragments, or idealized configurations. These are selected from a database using predetermined statistical scores of the fragments with configurations of control atoms,[4] or even manually.[5] The second approach requires only an approximate structure and may use fragments,[9−11] geometrical rules,[12] or simply random placement around the corresponding coarse-grained beads.[8]

The use of fragments yields stable solutions but requires atomistic models of the fragments to be available and the correspondences between fragments and configurations of control atoms to be known. In addition, methods based on preset fragments map a CG structure to a limited set of specific atomistic configurations, while any CG structure corresponds to an ensemble of atomistic structures.

In principle, repeating the reconstruction should sample the whole region mapping to the coarse-grained configuration with probabilities reflecting the atomistic energy landscape. Therefore, it seems more appropriate to use a less determined starting structure than obtained by fitting fragments and emphasize the relaxation stage. This has been the central notion underlying the approach of Rzepiela and co-workers.[8] Their method consists of simultaneous, coupled simulation of the CG, and the derived atomistic systems. The starting structures of the latter are obtained by randomly placing the atoms on small spheres around the corresponding coarse-grained bead positions. In a number of simulated annealing (SA) runs, the atomistic system is then relaxed to a suitable atomistic structure. Drawbacks of this method are that it is restricted to a specifically modified version of GROMACS[8] and requires a redefinition of the building blocks used to construct topologies to include the correspondence of atoms and CG particles. In addition, it may require a large number of steps and considerable effort, especially with complex systems. A marked benefit over fragment-based methods is that the process can be accomplished by specifying the correspondence between atoms and CG beads. This also makes it relatively easy to specify (e.g., alternative protonation states).

If the objective is to start an atomistic simulation from a CG structure, the main requirement is that the structure resulting from backmapping is stable and sufficiently close to the equilibrium ensemble. This philosophy led Brocos and co-workers to introduce a simplified procedure to map lipid structures back to atomistic for micelle simulations, following aggregation using a CG model.[12] Their method starts with placing atoms on interpolated positions along a trace through the CG beads, and adding 'peripheral' atoms at an offset from the trace, close to the connecting atom. The resulting structure is then optimized by energy minimization to obtain a starting structure for further simulation. This approach illustrates how approximate, low-quality starting structures, combined with a simple relaxation procedure are sufficient to obtain simulation-ready atomistic models from CG lipid systems.

The merits and drawbacks of the different methods currently available lead us to formulate specific criteria for a backmapping routine. Foremost, it should yield correct structures and be simple, quick, and flexible. It should also require no or little a priori information and be generic, applicable to all types of molecules and for all force fields. Accordingly, the method should require no other information than the CG coordinates, the atomistic topology, and the correspondence mapping between atomistic and CG particles.

In our approach, we start from the simplest procedure conceivable, along the lines of the solution of Brocos et al.[12] A few elements were added to increase the quality and stability of the resulting structures, at the expense of some of the simplicity. One major refinement is a new routine for protein backbone reconstruction, which is entirely geometrical and yields interesting insights in protein secondary structure. The other improvements comprise a set of geometrical modifiers for sculpting local structure.

The first part of this paper describes the background of the problem, illustrating the development of the method and the solutions implemented for specific challenges posed by different classes of molecules. In addition, it gives an outline of the program and describes the format used for mapping definitions. The subsequent sections focus on the protocols used for testing the robustness, quality, and efficiency of our new approach. Emphasis is placed on the mapping from MARTINI CG[13−16] to GROMOS 54a7 united-aliphatic atom[17] (UA) and CHARMM36 all atom[18−20] (AA) force fields. Finally, to demonstrate the applicability to other force fields, a mesoscopic inverted hexagonal phase structure, obtained from a simulation with a simplified three-bead lipid model, is converted to MARTINI and subsequently converted to GROMOS 54a7.

The main backmapping routine is contained in a Python program called *backward.py*, which has been made available at http://cgmartini.nl. Also made available there is an automated workflow, *initram.sh*, comprising both stages of the backmapping process, tailored to the MARTINI force field and the GROMACS simulation suite. The programs and mapping definitions are also available as Supporting Information.

2. BACKGROUND AND IMPLEMENTATION

In this section, we approach the generation of a high resolution starting structure from a low resolution configuration as a largely geometrical problem.

Let $\mathbf{A}$ denote an n by 3 matrix, corresponding to an atomistic or united atom configuration of n particles in 3D Cartesian space. This configuration maps to a coarse-grained configuration, represented by an m by 3 matrix $\mathbf{C}$. In most coarse-graining schemes, each particle position, or row, in $\mathbf{C}$ corresponds to a weighted average of a set of particle positions, or rows, in $\mathbf{A}$. This means that the mapping can be written as $\mathbf{C} = \mathbf{MA}$, where $\mathbf{M}$ is the m by n mapping matrix of weights, with the jth row corresponding to the jth coarse-grained bead ($j = \{1...m\}$) and the ith column corresponding to the ith atom ($i = \{1...n\}$). Since $m < n$, it is immediately clear that every atomistic configuration maps to a single coarse-grained configuration. The reverse mapping is not defined uniquely from this, as $\mathbf{M}$ is rank deficient.

Now let $A_C = \{\mathbf{A}\text{:}\ \mathbf{MA} = \mathbf{C}\}$ denote the set of atomistic structures mapping to a given coarse-grained configuration $\mathbf{C}$. This can be written in terms of the RMSD as $A_C = \{\mathbf{A}\text{:}\ \mathrm{RMSD}(\mathbf{MA},\mathbf{C}) = (1/N\ \mathrm{tr}\ (\mathbf{MA}-\mathbf{C})^{\mathrm{T}}(\mathbf{MA}-\mathbf{C}))^{1/2} = 0\}$. Then, the backmapping challenge is to find, for any $\mathbf{C}$, a configuration $\mathbf{A} \in A_C$, with the added condition that A has a high probability according to the local energy landscape of the conformational space spanned by A_C. In other words, $\mathbf{A}$ must be a 'correct' structure, mapping to $\mathbf{C}$.

The first stage of backmapping is the generation of an initial atomistic structure. This operation can be written as the multiplication of $\mathbf{C}$ by an n by m backmapping matrix $\mathbf{B}$,

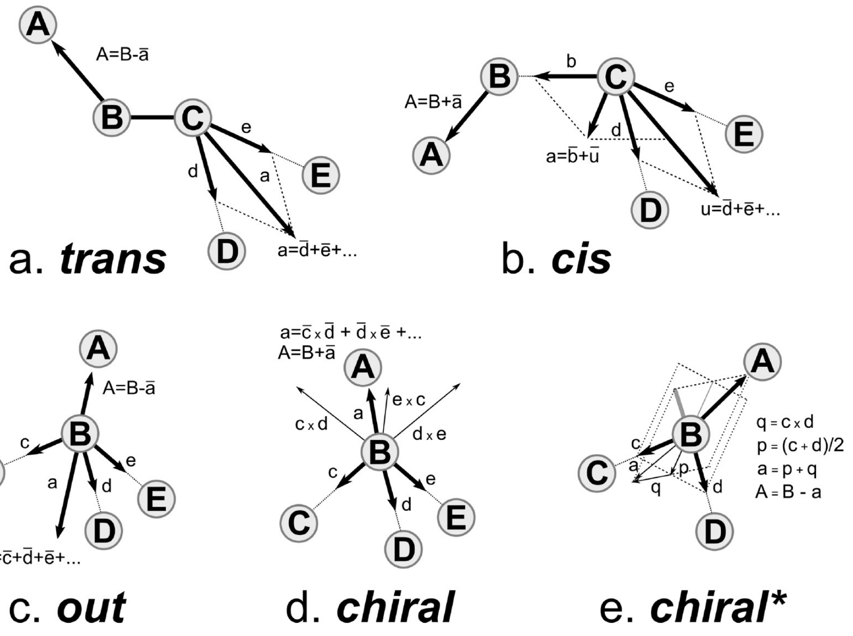


Figure 1. Modifiers for reconstruction of local geometry. Capitals denote atom position. In all images the position of particle A is to be determined from a set of known positions of control particles B, C, D. Lowercase letters denote vectors and a bar is used to denote normalization. (a) *trans*: Position of A is set as $A = B - \overline{a}$, with $a = \overline{d} + \overline{e} + ...$, and $d = D - C$, $e = E - C$. (b) *cis*: Position of A is set as $A = B + \overline{a}$, with $a = \overline{b} + \overline{u}$, where $u = \overline{d} + \overline{e} + ...$, and $b = B - C$, $d = D - C$, $e = E - C$. (c) *out*: Position of A is set as $A = B - \overline{a}$, with $a = \overline{c} + \overline{d} + \overline{e} + ...$, and $c = C - B$, $d = D - B$, $e = E - B$. (d) *chiral*: Position of A is set as $A = B + \overline{a}$, with $a = (\overline{c} \times \overline{d}) + (\overline{d} \times \overline{e}) + ...$, and $c = C - B$, $d = D - B$, $e = E - B$. (e) *chiral*: Chiral modifier with three control atoms. B, C, and D are considered the center and corners of a tetrahedron, from which the vector for placement of A is derived. Position of a is set as $A = B - \overline{a}$, with $a = p + q$, where $p = (c + d)/2$, $q = c \times d$, and $c = C - B$, $d = D - B$.

projecting the coarse-grained configuration onto the space of A_C, and yielding a projected structure $\mathbf{A_p}$: $\mathbf{A_p} = \mathbf{BC}$. In the following, this is referred to simply as projection.

If each ith row of $\mathbf{B}$ has only one nonzero entry, equal to one, on the jth position, then the position of the ith atom will be set equal to the position of the jth coarse-grained bead. In this case, the trivial solution is obtained, which is the only structure for which it is a priori certain that it meets the RMSD criterion. However, this solution will only correspond to a correct structure if the mapping is 1:1. The trivial solution is used as the starting configuration in the method of Rzepiela et al., where small, random displacements are added to it. In the approach of Brocos et al., particles are positioned at a specific bead or between two beads, which corresponds to the use of a backmapping matrix with a simple structure in which each row contains one or two entries.

In principle, the matrix B can include local and global correlations within configurations, as well as chemical information, such as bond and angle constraints, which can be written as a matrix operation.[21] However, the results from Brocos et al. have shown that a rather simple mapping may already be sufficient for generation of starting structures, and the derivation of a full backmapping matrix is therefore not necessary for our current purposes. Rather, our method starts with a simple projection in which each particle maps to the weighted average of any number of beads, provided these correspond to a single topological building block, such as a molecule or residue. The latter constraint is needed to define the scope of particles available for the reconstruction.

It is noted that a simple projection may not always allow proper reconstruction of chiral centers and double bond configurations. Therefore, the projection is followed by a stage of geometrical correction, henceforth referred to simply as correction, in which particles are repositioned using geometrical modifiers. These allow placement of a particle *cis*, *trans*, *out*, or *chiral* with respect to a set of particles, as shown in Figure 1. These modifiers are used to reconstruct, for example, the aromatic amino acid side chains, the arginine guanidinium group, and chiral centers. The protein backbone is reconstructed in a different way, explained in more detail later.

The corrected structure is subjected to force field based relaxation, consisting of energy minimization and position restrained molecular dynamics. This stage is referred to as the relaxation to distinguish it from the geometrical correction and aims to add the necessary chemical corrections to yield a structure with low energy in the conformational space spanned by A_C.

2.1. Backbone Reconstruction.

The reconstruction of a protein backbone poses a particular challenge: the peptide plane geometry and the hydrogen bonding pattern with the surroundings. Together these give rise to specific configurations, which are not reflected in the CG structures. Several methods are already available for reconstruction of proteins from low-resolution density maps or C_α only models,[6,22] but these methods invariably use fragment libraries, from which the best candidates are selected, based on the correlation with configurations of consecutive C_α atoms, while the aim of this work was a strictly geometrical approach.

Essentially, the objective is obtaining correct positions of the backbone C_i, O_i, N_{i+1}, and H_{i+1} atoms, relative to the line connecting $C_{\alpha,i}$ and $C_{\alpha,i+1}$. For helical structures, the C=O direction vector should be pointing approximately in the direction of the vector cross product $\mathbf{c}_i$ of the vectors $C_{\alpha,i+1}-C_{\alpha,i}$ and $C_{\alpha,i+2}-C_{\alpha,i}$ (Figure 2a). That this assumption is correct can be seen in Figure 2b, which shows the direction vectors of these cross products placed onto the peptide carbonyl C atoms.

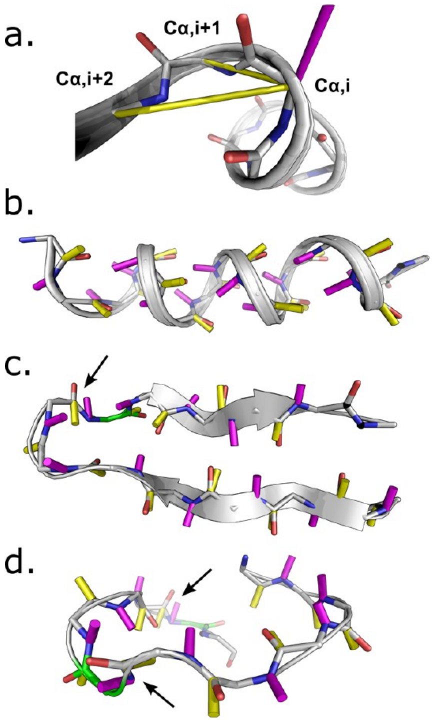


Figure 2. Peptide plane orientation predictions based on C_α triplets. (a) Cross product (purple) of vectors $C_{\alpha,i}-C_{\alpha,i+1}$ and $C_{\alpha,i}-C_{\alpha,i+2}$ (yellow) in a helix. (b) α-Helical structure with predicted carbonyl oxygen (yellow) and amide hydrogen (magenta) directions. (c) Predicted directions for β-sheet structure. Glycine, located in the turn, is marked in green. The peptide plane connecting glycine with the preceding residue, marked by an arrow, is inverted with respect to the prediction. (d) Predicted directions for a loop segment. Two glycines are marked in green, and in both cases, the peptide plane with the preceding residue is inverted with respect to the prediction.

Surprisingly, the same rule applies to extended structures, shown in Figure 2c, and even to loop regions, as can be seen in Figure 2d. In fact, the predicted carbonyl vector direction is found to be wrong only at or next to a glycine residue in a loop or turn. Here, the correct direction may be exactly opposite to the predicted direction. This can be seen in Figure 2c and d, where the glycine residues are highlighted in green and the inverted peptide planes are indicated with an arrow. For the purpose of rebuilding backbone structures from coarse-grained coordinates, this does not seem of concern, as glycine under such circumstances should be flexible enough to relax to a correct configuration.

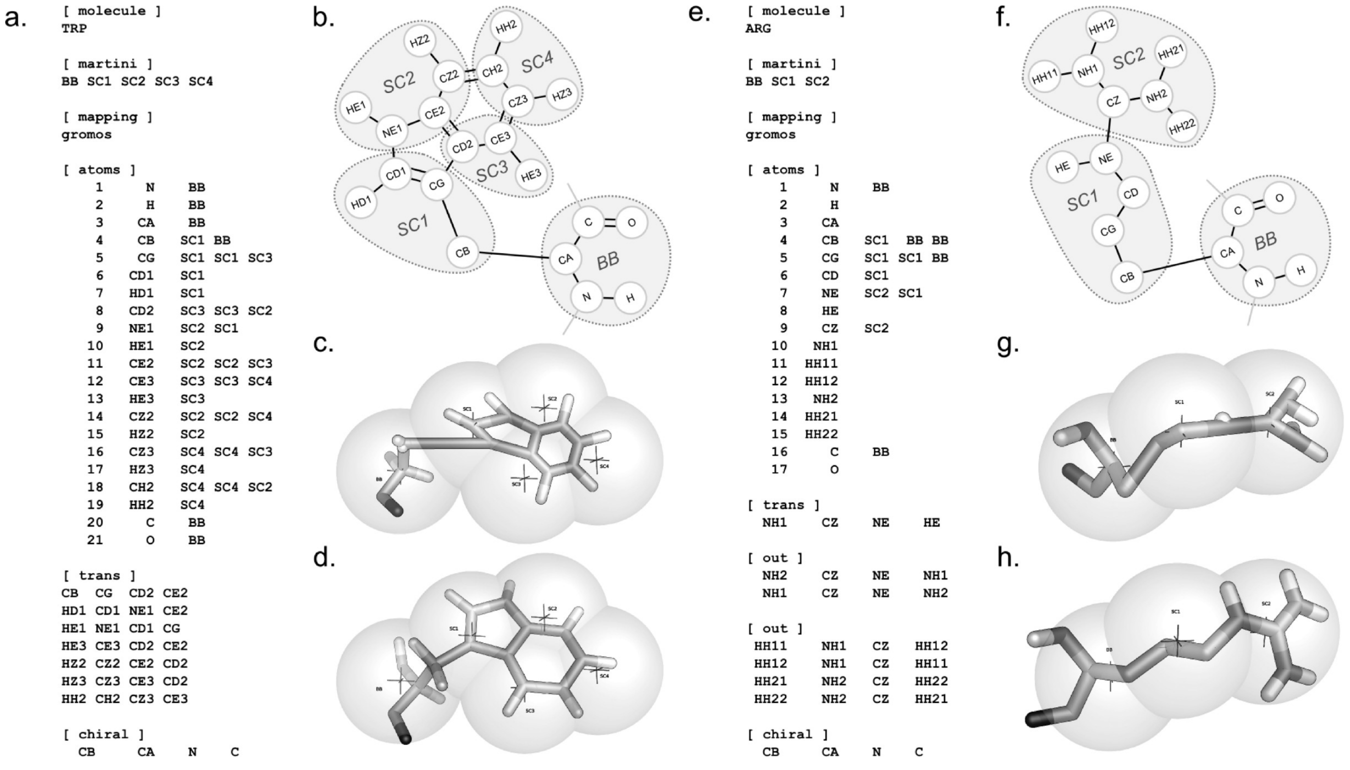
The cross product derived from consecutive C_α triplets yields a simple geometrical approximation for rebuilding amino acid backbone structures from C_α atoms or coarse-grained backbone beads: first the *i*th C_α atom is placed on the *i*th backbone node. The peptide plane direction vector is subsequently determined from the cross product $\mathbf{c}_i$ and the C_i and O_i atoms are placed at one-third from node *i* to *i*+1, with offsets in the direction of $\mathbf{c}_i$. Likewise, the atoms N_{i+1} and H_{i+1} are placed at two-thirds from node *i* to *i*+1, with offsets in the direction opposite of $\mathbf{c}_i$.

2.2. Implementation. The implementation follows a modular software design, introduced previously to streamline development of GRID-computing based workflows.[23] An important aspect of the software model is the separation of streams for input and output and for static data and provenance. The model also promotes modular process-oriented implementation, where each process is designed as a stand-alone program, which can be coupled to form more complex workflows. Although the backmapping protocol consists of three stages: projection, correction, and relaxation, they are organized as two processes. The projection and correction are geometrical operations on a set of positions, which have been implemented in the Python program *backward.py*. The second process, contained in the bash wrapper *initram.sh*, is a workflow combining the projection/correction with relaxation.

The objective of *backward.py* is converting a coarse-grained structure into an atomistic one. Because there is no fundamental difference between this inverse mapping and forward mapping, the scope has been broadened to conversion of a set of coordinates at one resolution to a set of coordinates at another resolution. From this objective, it follows that the system coordinates are the only mandatory input. The conversion requires information about mapping between the resolutions, but this information does not depend on the system, and should therefore be defined within the scope of the program, not in the input, and be available as a library of mapping definitions.

Given a structure and a target force field, *backward.py* performs the projection according to the mapping information contained in molecule (or residue) based mapping definitions of which examples are shown in Figure 3. These definitions are derived from the (back)mapping matrices introduced in the previous section, written in a sparse notation, using particle names to signify rows and columns. Each mapping definition starts with a name, corresponding to the name of the molecule, followed by a tag indicating the low-resolution force-field (e.g., [martini] or [cooke]), and a listing of the beads of that residue in the order of the topology for that force field. This is then followed by the name of the higher resolution force field for which the mapping is defined. The next section, [atoms], lists the atoms in the order corresponding to the higher-resolution forcefield. Each line in the atom list consists of a number, the target atom name and a list of source particles of which the average is taken to determine the target particle position. If a low-resolution particle is listed twice, then it gets a relative weight of 2, allowing more precise positioning of target particles. Note how this corresponds to a line of the backmapping matrix, where the entries corresponding to the low-resolution particles are set to the relative weights, with the sum of each row equal to 1. If an atom has no listing of source particles, then it is positioned at a small random offset from the preceding atom, allowing more efficient writing of mapping definitions. After the atom list, any number of geometric modifiers may be added. This simple approach makes it trivial to generate starting structures with molecules for which no fragments are available but for which it is possible to generate a corresponding atom list.

A specific aim of this project was to provide a method that would be flexible enough to generate starting structures for molecules with alternative protonation states or with virtual sites,[24] without requiring separate fragments for each possible case. For this reason, the target atom list can be read from a topology file, which then takes precedence over the order in the mapping definitions. Atoms present in the definition, but not in the target list, are ignored, while particles in the target list, but not in the definition, will be given the coordinates of the preceding particle, with a small random offset. This ensures that the starting structure obtained and the topology provided match, and can be used directly for the relaxation stage. The topology parser included in *backward.py* is specific to the GROMACS topology format but could easily be extended to work with other packages by adding parsers for the corresponding formats.



```
[ molecule ]
TRP

[ martini ]
BB SC1 SC2 SC3 SC4

[ mapping ]
gromos

[ atoms ]
    1     N    BB
    2     H    BB
    3    CA    BB
    4    CB   SC1 BB
    5    CG   SC1 SC1 SC3
    6   CD1   SC1
    7   HD1   SC1
    8   CD2   SC3 SC3 SC2
    9   NE1   SC2 SC1
   10   HE1   SC2
   11   CE2   SC2 SC2 SC3
   12   CE3   SC3 SC3 SC4
   13   HE3   SC3
   14   CZ2   SC2 SC2 SC4
   15   HZ2   SC2
   16   CZ3   SC4 SC4 SC3
   17   HZ3   SC4
   18   CH2   SC4 SC4 SC2
   19   HH2   SC4
   20     C    BB
   21     O    BB

[ trans ]
CB  CG  CD2 CE2
HD1 CD1 NE1 CE2
HE1 NE1 CD1 CG
HE3 CE3 CD2 CE2
HZ2 CZ2 CE2 CD2
HZ3 CZ3 CE3 CD2
HH2 CH2 CZ3 CE3

[ chiral ]
  CB     CA    N    C
```

```
[ molecule ]
ARG

[ martini ]
BB SC1 SC2

[ mapping ]
gromos

[ atoms ]
    1     N    BB
    2     H
    3    CA
    4    CB   SC1  BB BB
    5    CG   SC1 SC1 BB
    6    CD   SC1
    7    NE   SC2 SC1
    8    HE
    9    CZ   SC2
   10   NH1
   11  HH11
   12  HH12
   13   NH2
   14  HH21
   15  HH22
   16     C    BB
   17     O

[ trans ]
  NH1     CZ     NE     HE

[ out ]
  NH2     CZ     NE     NH1
  NH1     CZ     NE     NH2

[ out ]
 HH11     NH1    CZ    HH12
 HH12     NH1    CZ    HH11
 HH21     NH2    CZ    HH22
 HH22     NH2    CZ    HH21

[ chiral ]
  CB      CA     N     C
```

Figure 3. Mapping and reconstruction of tryptophan and arginine for the GROMOS 54a7 force field from MARTINI CG. (a) Mapping definition for tryptophan. The mapping file starts with a header [molecule], followed by the name of the molecule or residue. The next header, [martini], indicates that the mapping applies to the low-resolution force field MARTINI, and is followed by a listing of the coarse grained particles, in the order dictated by that force field. Under [mapping] the higher-resolution (target) force field is listed, followed by the actual mapping section [atoms]. This section lists the atoms in the order of the higher-resolution force field, with each line containing a number, the name of the target force field and a number of source particles. If no source particles are given, the definition of the preceding particle is reused. After the atom list any number of geometric modifiers can be given. (b) Schematic view of tryptophan CG/UA mapping. (c) Overlay of projected structure and original CG positions of tryptophan, based on the mapping definition. (d) Overlay of final, relaxed structure and original CG positions of tryptophan. (e) Mapping definition for arginine. (f) Schematic view of arginine CG/UA mapping. (g) Overlay of projected structure and original CG positions of arginine, based on the mapping definition. (h) Overlay of final, relaxed structure and original CG positions of arginine.

The mapping definitions may contain a number of geometrical modifiers, which are processed in order, after the initial projection. Modifiers are preceded by a header, indicating the type. Each modification is written on a single line, giving the particle to be repositioned, followed by the control atoms (see Figure 1). If the target particle is not in the atom list, it is added as control particle, available for subsequent modifications. This allows writing sequences of modifications to realize complex reconstructions, which may be necessary in some cases. An example of this is the arginine side chain shown in Figure 3e−h.

Unlike *backward.py*, the wrapper *initram.sh* is tailored to GROMACS, which is used to perform the relaxation simulations. The wrapper starts with a call to *backward.py*, and then performs two cycles of energy minimization (EM), followed by a series of position restrained molecular dynamics simulations. For *initram.sh*, the target topology is mandatory input, as it is required for running the simulations.

The first cycle of EM is performed with nonbonded interactions turned off between particles within certain groups of molecules, such as proteins and lipids. This is done to avoid high forces due to overlapping atoms. This step is followed by a cycle of EM with all interactions turned on. After EM, a series of short molecular dynamics (MD) runs is performed, with increasing integration time steps. The default protocol was developed to be robust enough to backmap complex systems and comprises two cycles of 500 EM steps and four cycles of 500 steps of MD, with time steps of 0.2, 0.5, 1.0, and 2.0 fs. For simple systems, it is possible to use a more efficient scheme, consisting of fewer steps. To facilitate such alternative schemes, the number of steps of each cycle and the series of time steps can be specified on the command line. To avoid large deviations from the coarse-grained configuration, the simulations are run with harmonic position restraints to the coordinates obtained by the projection from CG to UA/AA. These restraints are taken from the target topology as provided by the user.

3. METHODS

3.1. Backmapping and Simulation. Backmapping from MARTINI to GROMOS united atom representation or to CHARMM36 all atom representation was performed with *backward.py* and *initram.sh*. Backmapping from the Cooke 3-bead lipid model to MARTINI was performed with *backward.py* alone. For *initram.sh*, default settings were used for the number of integration steps (500) and the series of time steps (0.2, 0.5, 1.0, and 2.0 fs). Position restraints were applied to protein and lipid heavy atoms. For the timings reported, the backmapping procedure was run on an Intel Xeon 5160 "Woodcrest" chip, using four cores, running at 3.0 GHz. For all systems, except for the mesoscopic inverted hexagonal phase, a 20 ns simulation was run after backmapping to assess the stability of the resulting models. All simulations were performed using GROMACS version 4.5.5,[25−27] unless stated otherwise.

Simulations with the GROMOS 54a7 force field were run in the NpT ensemble, with temperature coupled weakly to a heat bath at 310 K, unless stated otherwise, using the v-rescale (Bussi) thermostat[28] with a coupling time of 0.1 ps, and coupled to a reference pressure of 1.0 bar, using a Parrinello−Rahman barostat[29] with a coupling time of 2.0 ps and a compressibility of 4.5 × 10^{-5} bar^{-1}. Electrostatic interactions were calculated using particle-mesh Ewald (PME) summation,[30] with a real-space cutoff of 1.5 nm. Van der Waals interactions were switched to zero between 1.2 and 1.4 nm to avoid cutoff artifacts.[31] Neighborlists were updated every 10 steps. All bonds were constrained using the LINCS algorithm.[21] Water was modeled explicitly using the simple point charge (SPC) model.[32] The integration time step used was 2 fs, and every 10 steps, the center of mass motion was removed for two groups (protein and or membrane and water with or without ions) separately. Lipid topologies for GROMOS 54a7 were taken from Poger and Mark.[33]

Simulations with the CHARMM36 force field used the GROMACS port from Piggot et al.[20] These simulations were performed in the NpT ensemble, with the temperature coupled weakly to a heat bath at 310 K, unless stated otherwise, using the v-rescale thermostat[28] with a coupling time of 0.1 ps, and coupled to a reference pressure of 1.0 bar, using a Parrinello−Rahman manostat[29] with a coupling time of 4.0 ps and a compressibility of 4.5 × 10^{-5} bar^{-1}. Electrostatic interactions were calculated using PME, with a real-space cutoff of 1.2 nm. Van der Waals interactions were switched to zero between 1.2 and 1.4 nm. Neighbor lists were updated every 5 steps. Bonds involving hydrogens were constrained using the LINCS algorithm.[21] Water was modeled explicitly using the CHARMM three point transferable intermolecular potential (TIP3P) model.[34,35] The integration time step used was 2 fs and the overall center of mass motion was removed every 10 steps. Topologies for CHARMM36 were generated according to the implementation in GROMACS.[18,20,36]

CG simulations from which starting structures were derived were performed using MARTINI.[13−16] This force field has proven suitable for simulation of many biomolecular processes. A comprehensive overview of the application of MARTINI coarse-grained simulations was provided recently.[37] CG structures and topologies were constructed according to the Martini 2.1 polarizable force field using the script *martinize.py*,[38] available from http://cgmartini.nl. The proteins were simulated using a MARTINI based elastic network, called RubberBands (manuscript in preparation), which is similar to ElNeDyn[39] and is implemented in *martinize.py*. Membranes were built using a recently developed program for generating custom coarse-grained membranes, called *insane.py* (INSert membrANE, manuscript in preparation). CG simulations were performed under NpT conditions, with temperature coupled weakly to a heat bath at a temperature depending on the model (see below), using the Berendsen thermostat[40] with a coupling time of 1.0 ps, and coupled to a reference pressure of 1.0 bar, using a Berendsen barostat[40] with a coupling time of 5.0 ps and a compressibility of 4.5 × 10^{-5} bar^{-1}. Long range electrostatic interactions beyond 1.2 nm were calculated using particle-mesh Ewald (PME) summation.[30] Van der Waals interactions were switched to zero between 0.9 and 1.2 nm to avoid cutoff artifacts. Neighborlists were updated every 10 steps. Water was modeled using the MARTINI polarizable water model (PW)[16] with a relative dielectric permittivity of 2.5. The integration time step used was 20 fs and every step the center of mass motion was removed for protein and membrane independently.

The backmapping procedure and results were compared with the Simulated Annealing (SA) based protocol,[8] which was obtained from http://cgmartini.nl/ and which matches our new method in scope. Other methods were not included, because of lack of generality or because they could not be retrieved.

For the SA based method, the modified version of GROMACS 3.3.1 developed by Rzepiela et al.[8] was downloaded from http://www.cgmartini.nl, compiled, and used to set up a backmapping topology. A matching atomistic starting structure was generated by placing atoms randomly around the corresponding CG particles, using the tool *g_fg2cg*. This structure was subsequently used to run the coupled reverse-transformation simulated-annealing simulation.[8] The backmapping was run with the parameters mentioned in the original paper: An initial capping force of 15 000 kJ mol^{-1} nm^{-1} was used, with a capping increase rate of 100 kJ mol^{-1} nm^{-1} ps^{-1}. The restraining force constant was set to 12 000 kJ mol^{-1} nm^{-2}, the water restraining force constant was 400 kJ mol^{-1} nm^{-2}, and the number of steps at which the decoupling of the UA system from the CG configuration was commenced was set to 5000. The initial temperature was 400 K for water and 1300 K for the other components, which was brought down to 300 K in 60 ps by simulated annealing. One series was performed with the temperature coupled weakly to a heat bath, using the Nosé−Hoover thermostat[41,42] with a coupling time of 0.1 ps, corresponding to the original protocol,[8] while in a second series stochastic coupling was used, according to suggestion of Schäfer et al.,[43] with an inverse friction constant of 0.1 ps. The stochastic coupling should suppress local heating, which may cause distortions.

3.2. Models.

We tested the method on lipid bilayers, proteins in solution, and membrane proteins. The protein containing systems were taken from ongoing research projects, except for WALP20, which was chosen because it was used previously as a test system for development of a backmapping method.[8]

Protein AA and UA topologies were produced using the GROMACS' tool *pdb2gmx*. Except for YvoA, which was kindly provided by Simon Fillenberg (University of Erlangen), structures were obtained from the Protein DataBank: 2LCN[44] (WALP), 1J4N[45] (bovine aquaporin 1), 3IJ4,[46] and 2QTS[47] (ASIC1a).

The systems are described in more detail below.

3.2.1. DOPC Bilayer (1).

A rectangular dioleoyl phosphatidylcholine (DOPC) bilayer consisting of 200 lipids was built and solvated with 2351 MARTINI polarizable water (PW), yielding a system consisting of 9853 beads. The corresponding GROMOS 54a7 UA representation comprised 39 012 united atoms, and the CHARMM36 system measured 55 812 atoms. After energy minimization the CG bilayer was equilibrated at 303 K and 1 bar (NpT ensemble) for 100 ns.

3.2.2. Mixed Bilayer (2).

A rectangular bilayer consisting of 236 palmitoyl oleyl phosphatidylcholine (POPC), 32 palmitoyl oleyol phosphatidylserine (POPS), and 66 cholesterol (CHOL) molecules, divided equally over both leaflets, was built and solvated with 5464 PW. Then, 76 sodium and 44 chloride beads were added, corresponding to a concentration of 150 mM NaCl and compensating the net charge of the membrane. The total system comprised 20 524 beads. The corresponding GROMOS 54a7 UA representation measured 81 634 united atoms, and the CHARMM36 system consisted of 107 700 atoms. After energy minimization the CG bilayer was equilibrated at 310 K and 1 bar (NpT ensemble) for 100 ns.

*3.2.3. WALP20 in DPPC Bilayer (**3**).* A model of the WALP20 transmembrane helix,[48] sequence WWALA LALAL ALALA LALWW, was built from the WALP19-P10 crystal structure (2LCN)[44] by mutating the individual residues in PyMOL. The structure obtained was energy minimized in vacuum, and coarse-grained using *martinize.py*. The peptide was inserted in a membrane consisting of 120 dipalmitoyl phosphatidylcholine (DPPC) molecules and solvated with 1223 PW molecules using *insane.py*. The resulting model contained 5153 beads and is comparable to the one Rzepiela and co-workers used to test their reverse transformation.[8] The corresponding GROMOS 54a7 UA representation consisted of 20 883 particles and the system in CHARMM36 representation measured 30 607 atoms. The CG system was energy minimized and equilibrated (NpT, 310 K, 1 bar) for 10 ns, and the final structure was used for reverse transformation using the method from Rzepiela and using the method developed here.

*3.2.4. YvoA (**4**).* The HTH-type transcriptional repressor YvoA is a soluble protein from *Bacillus subtilis*. A crystal structure of this protein was recently solved and kindly provided by Simon Fillenberg (University of Erlangen) for simulation studies. The structure was coarse-grained and solvated in a rhombic dodecahedron box with 14 706 PW, and 163 sodium ions and 161 chloride ions to compensate the net charge of the protein and mimic isotonic salinity. The total size of the resulting system was 45 500 beads. The corresponding GROMOS 54a7 UA representation measured 185 590 particles, and the CHARMM36 system contained 188 358 atoms. The CG system was equilibrated and run for 1 μs NpT simulation (310 K, 1 bar). The final structure was used for reverse transformation.

*3.2.5. Aquaporin (**5**).* The tetrameric biological unit of bovine aquaporin 1 (PDB ID 1J4N) was coarse-grained and set up with rectangular periodic boundary conditions, inserted in a membrane consisting of 318 POPC lipids, and solvated with 10 030 PW, 104 sodium ions, and 116 chloride ions. The total size of the system thus obtained was 36 520 beads. The system in GROMOS 54a7 representation consisted of 149 188 particles and with CHARMM36 resulted in 180 904 atoms. After energy minimization and relaxation of the CG system, a production NpT simulation of 100 ns was run (310 K, 1 bar), of which the final structure was used for reverse transformation.

*3.2.6. Acid-Sensing Ion Channel 1a (**6**).* The chicken Acid-Sensing Ion Channel (ASIC) 1a trimeric biological unit was built by combining crystal structures 3IJ4[46] and 2QTS,[47] extending the intracellular N- and C-termini of the former, based on the configuration of the latter. The resulting structure was coarse-grained and set up with rectangular periodic boundary conditions, inserted in a membrane consisting of 466 POPC lipids and solvated with 15 259 PW, 244 sodium ions, and 223 chloride ions. The total size of the coarse-grained system was 55 173 beads, which yielded 226 796 united atoms in GROMOS 54a7 representation, and in CHARMM36 representation measured a total of 272 014 atoms. The final structure of a 1 μs NpT CG simulation (310 K, 1 bar) was used for reverse transformation.

*3.2.7. DOPC Inverted Hexagonal Phase (**7**).* An inverted hexagonal phase structure formed by 2006 three-bead model[49] lipid molecules (6018 particles) was generously provided by Clement Arnarez (University of Groningen). This structure was used for multistage backmapping from a very low resolution model to atomistic detail. The system was first converted to MARTINI representation, using *backward.py*, yielding a system containing 28 084 particles. After conversion, the structure was solvated with 144 791 MARTINI water beads and then energy minimized. The minimized solvated structure was subsequently converted to GROMOS54a7 representation using *backward/initram*, yielding a final system with a total of 1 845 816 atoms.

3.3. Remapping. To assess the stability and robustness of the approach, relaxed structures obtained after the reverse transformation of systems **1**−**6** were remapped (i.e., coarse-grained again) also using *backward.py*, and then converted to high resolution once more. As it is problematic to convert atomistic to coarse-grained water, due to the undefined mapping from multiple solvent molecules to single particles, the original coarse-grained solvent configuration was used for both stages of reverse transformation. Pairs of corresponding high-resolution representations, before and after remapping, were compared to assess the similarity.

3.4. Analysis. To assess the accuracy of the backmapping procedure, the RMSD was calculated for each pair of high-resolution structures before and after remapping. In principle, remapping should retain the structure and yield low remapping RMSDs. This makes the RMSD an indicator of the quality of a backmapping procedure.[50] The RMSD for proteins was determined for backbone atoms, heavy atoms and all atoms, after fitting on C_α atoms. Lipid RMSD was calculated on a per-molecule basis, over all heavy atoms, after a least-squares fit on heavy atoms. RMSD values for lipids are provided as the average over all molecules, with standard deviations.

For bilayer and solvent, the mapping quality was further assessed by investigation of the radial distribution function (RDF) and the bilayer density profiles. The RDFs and densities were determined for single frames, taken from the end of each cycle of simulation.

Protein structures obtained from backmapping were also processed using the protein quality checker *WhatCheck*[51] to assess the quality of the structures.

To investigate the extent to which secondary structure of proteins is retained during CG simulation and backmapping, for all protein structures, the secondary structure was determined using the method of Kabsch and Sanders.[52] The secondary structure similarity between starting structures and final structures was quantitated using the Jaccard index,[53] defined as $J(A,B) = |A\cap B|/|A\cup B|$. For the total secondary structure, the numerator is the number of corresponding secondary structure elements, which is divided by the number of residues, whereas for α-helical and extended regions, the numerator is the number of residues that are both of the given type and the denominator the number of residues in which at least one is of the given type. The index is 1 for identity and 0 for complete dissimilarity.

4. RESULTS AND DISCUSSION

The results are summarized in Tables 1 and 2 and in Figures 4−9. In addition, Figure S1 in the Supporting Information shows the cumulative times required for the different stages of backmapping for each system, as a function of the number of target particles. For both CHARMM36 and GROMOS 54a7, the time required for each stage scales linearly with the number of particles in the system. The time for geometric reconstruction is in the order of seconds (Supporting Information Figure S1c). The systems containing protein (**3**−**6**) were also reconstructed using the SA based method proposed earlier by Rzepiela et al.[8] This protocol, excluding the setup of the systems, took 11:53 min for WALP, 2:24 h for YvoA, 2:11 h for aquaporin 1, and 3:23 h for

Table 1. Timing of Backmapping and Remapping RMSD

system	time[a]	molecule	D[b]	D_{BB}	$D_{SIM,BB}$
		G54a7 Projection/Relaxation			
1	0:51	DOPC	0.122 (0.012)		
2	1:43	POPC	0.108 (0.012)		
		POPS	0.108 (0.011)		
		CHOL	0.062 (0.018)		
3	0:31	WALP20	0.062	0.021	0.131
		DPPC	0.114 (0.013)		
4	4:11	YvoA	0.094	0.045	0.581
5	3:06	Aquaporin 1	0.092	0.042	0.366
6	5:19	ASIC1a	0.089	0.040	0.842
		G54a7 Simulated Annealing			
3	11:53	WALP20	0.093	0.038	0.204
		DPPC	0.142 (0.070)		
4	2:30:27	YvoA	0.108	0.055	0.674
5	2:11:32	Aquaporin 1	0.112	0.066	0.416
6	3:23:02	ASIC1a	NA	NA	NA
		CHARMM36 Projection/Relaxation			
1	2:42	DOPC	0.119 (0.010)		
2	5:27	POPC	0.121 (0.011)		
		POPS	0.163 (0.023)		
		CHOL	0.082 (0.028)		
3	1:46	WALP20	0.060	0.031	0.125
		DPPC	0.123 (0.012)		
4	9:06	YvoA	0.094	0.049	0.547
5	8:58	Aquaporin 1	0.083	0.048	0.427
6	14:22	ASIC1a	0.089	0.048	0.450

[a] Time for reverse transformation using *initram* or the simulated annealing protocol. [b] RMSD is given in nanometers for all heavy atoms (D), for backbone heavy atoms (D_{BB}) and for backbone heavy atoms after 20 ns of simulation ($D_{SIM,BB}$). The RMSD for lipids is given as average value with the standard deviation in parentheses.

ASIC1a (Table 1), which is approximately 40−60 times slower than backmapping to GROMOS 54a7 with the new method. The quality of the resulting structures, and a more detailed comparison between the two methods, is given in the following sections, which focus on the different classes of molecules.

4.1. Lipids.

Six of the seven test systems include lipids, of which **1**, **2**, and **7** are pure bilayers. Systems **1** and **2** were used for development and initial tests of the method. The first of these is a solvated DOPC bilayer and the second contains a ternary mixture of POPC, POPS, and cholesterol, solvated with water and ions.

The basis of this work was the assumption that a united atom or all-atom force field might be robust enough to allow starting from the trivial mapping solution, and this assumption was first tested on the DOPC system. The initial results showed that the basic assumption was correct, except that the correct chirality and *cis*/*trans* isomerism could not be guaranteed. To solve this, the mapping scheme was extended and the geometric modifiers were introduced. These allowed a more controlled reconstruction, which in turn increased the stability and decreased the length of the required relaxation cycles.

Figure 4 summarizes the backmapping stages and results for lipids. Oleoyl double bonds were reconstructed correctly to *cis* configuration, using a combination of modifiers. The acyl ester C−O−CO-C dihedral appeared more problematic, especially in UA representation. This torsion angle has a preference for trans configurations, with a small local energy minimum around 0 degrees. In CHARMM36, the distribution is properly reflected in the torsion angle potential and nonoptimal projections can be easily corrected during relaxation. However, in GROMOS UA representation, the torsion angle has two equal energy minima with a high barrier, preventing rotation. As a consequence, the geometric reconstruction must yield only correct configurations, avoiding trapping *cis* structures. To achieve this, a series of modifiers was used to reconstruct the local geometry.

Table 2. Stability of Secondary Structure During (CG) Simulation and Backmapping[a]

system	force field	total	extended	α-helix
		YvoA (**3**)		
100 ns AA MD	Charmm	0.80	0.87	0.77
Projection-relaxation	Charmm	0.52	0.42	0.66
- after 20 ns MD	Charmm	0.60	0.49	0.80
Projection-relaxation	Gromos	0.55	0.40	0.73
- after 20 ns MD	Gromos	0.67	0.67	0.82
Simulated annealing	Gromos	0.44	0.30	0.39
- after 20 ns MD	Gromos	0.58	0.63	0.57
		Aquaporin (**4**)		
100 ns AA MD	Gromos	0.66	n/a	0.76
Projection-relaxation	Charmm	0.52	n/a	0.63
- after 20 ns MD	Charmm	0.59	n/a	0.71
Projection-relaxation	Gromos	0.58	n/a	0.73
- after 20 ns MD	Gromos	0.57	n/a	0.69
Simulated annealing	Gromos	0.32	n/a	0.30
- after 20 ns MD	Gromos	0.43	n/a	0.48
		ASIC-1a (**5**)		
100 ns AA MD	Charmm	0.79	0.91	0.73
Projection-relaxation	Charmm	0.50	0.31	0.68
- after 20 ns MD	Charmm	0.54	0.50	0.67
Projection-relaxation	Gromos	0.66	0.45	0.83
- after 20 ns MD	Gromos	0.68	0.68	0.76
Simulated annealing	Gromos	0.19	0.03	0.05

[a] Secondary structure similarity between the original model and the final structure is given as Jaccard index, $J(A,B) = |A \cap B|/|A \cup B|$, where the numerator is the number of residues that are both of a given type and the denominator is the number of residues in which at least one is of the given type. For the total secondary structure, the numerator is the number of corresponding secondary structure elements, which is divided by the number of residues.

Glycerol-based lipids also contain a chiral center, which, in UA models, is imposed by an improper dihedral potential. In AA models, the tetrahedral configuration is maintained by the repulsive interactions between the substituents, which prohibit inversion. POPS has a second chiral center in the serinyl headgroup, and cholesterol has three chiral centers in UA representation and eight in AA. In all cases, the chirality was set to the correct stereoisomer in the geometric reconstruction by *backward.py*.

The coarse-grained model of **1** has an area per lipid of 0.69 nm^2 at 303 K, which corresponds well to the experimental value of 0.674 nm^2.[54] This area per lipid is retained during backmapping. In the subsequent unrestrained MD simulation, the area per lipid drops within 5 ns of UA simulation to 0.64 nm^2, and drops within 8 ns of AA simulation to 0.67 nm^2.

To assess the robustness of the backmapping procedure, the high resolution structures obtained were coarse-grained and subsequently converted again. For each lipid the RMSD between the initial and final high-resolution structures was determined (see Table 1). The RMSD values thus obtained for the diacyl glycerols were around 0.12 nm, while cholesterol yielded an average remapping RMSD of 0.06 and 0.08 nm in UA and AA

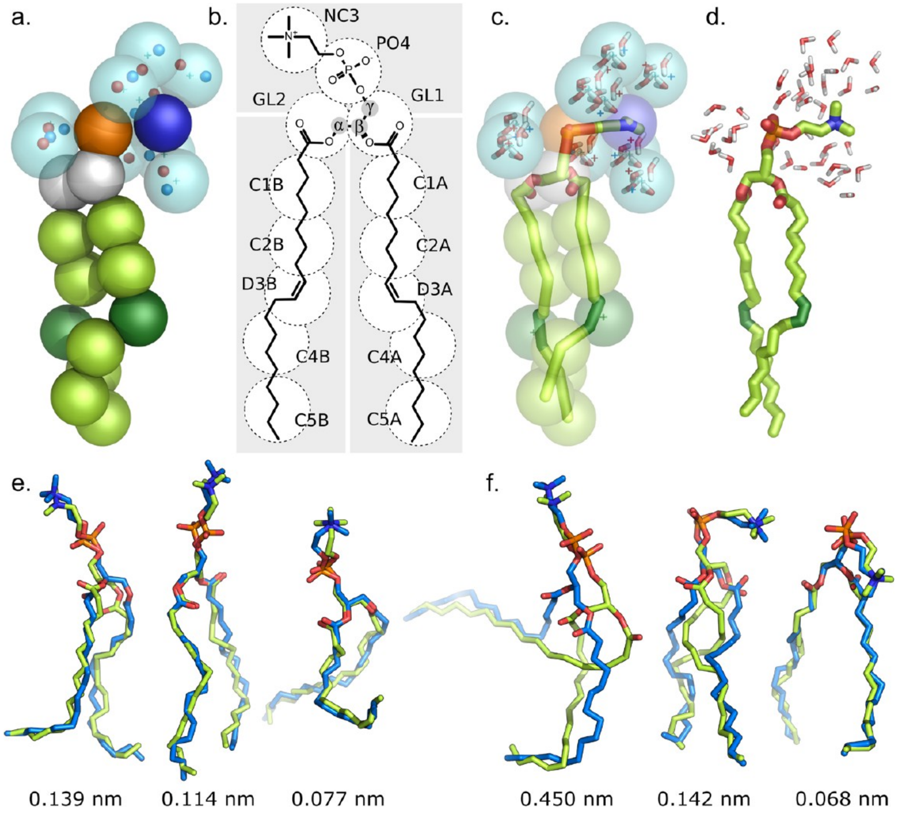


Figure 4. Backmapping of lipid molecules. (a) MARTINI DOPC molecule with headgroup solvated by polarizable water. (b) Mapping between CG beads and non-hydrogen atoms. (c) Projected structure. The local geometry around the acyl ester groups and the double bonds was constructed applying geometrical modifiers after mapping. (d) Atomistic DOPC molecule with solvated headgroup after relaxation. Only non-hydrogen atoms are shown. (e) Overlays of GROMOS 54a7 DPPC lipid structures before and after remapping with initram, corresponding to maximal (0.139 nm), average (0.114 nm), and minimal (0.077 nm) lipid RMSD. (f) Overlays of lipid structures before and after remapping using CG/UA coupled simulated annealing, corresponding to maximal (0.450 nm), average (0.142 nm), and minimal (0.068 nm) RMSD. In the maximum and average RMSD structures the tails appeared to have crossed close to the glycerol group.

representation respectively, due to the rigidity of the sterol moiety. The DPPC RMSD in the WALP containing system 3 was 0.114 ± 0.014 nm, which is slightly lower than the corresponding value of 0.141 ± 0.069 nm obtained after remapping with the SA protocol. The cause of this can be understood from Figure 5e and f, which show overlays illustrating the degree of structural difference associated with low, average and high DPPC RMSD for both methods. The lipids in Figure 4f corresponding to maximal and average RMSD appear to have the acyl tails intertwined, near the ester groups, suggesting a possible problem inherent to the SA protocol. This was found to be caused by the random placement of the atoms around the corresponding CG beads, which occasionally caused crossing of atoms belonging to different tails.

4.2. Protein. Unfortunately, the simple projection-relaxation approach resulted in undefined peptide planes in proteins. This was solved with a new algorithm for reconstruction of the protein backbone from consecutive CG backbone beads, as explained above. With this algorithm, the protein structures in models **3−6** were reconstructed to UA and AA representation. The systems **3−6** were also backmapped using the SA protocol for comparison. An example of the projection-relaxation process is given in Figure 5, which shows the WALP20 structures resulting from different stages of backmapping, illustrating how the helical structure builds up. The figure also shows the structure obtained with the SA protocol, which restores most of the helix, except for one turn.

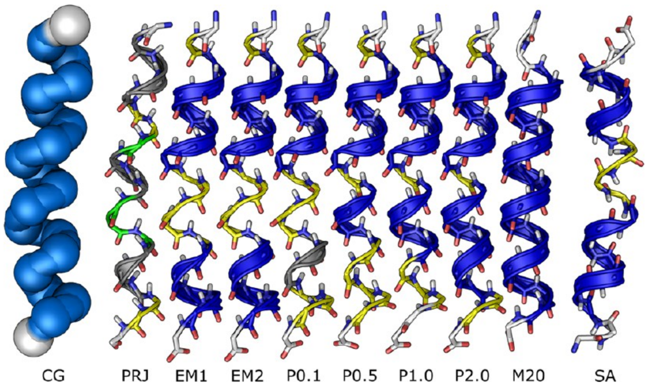


Figure 5. Build up of WALP20 secondary structure during backmapping. From left to right: MARTINI CG structure of WALP20 (CG), projected structure obtained with *backward.py* (PRJ), structure after bonded-only energy minimization (EM1), structure after energy minimization with full interactions (EM2), structure after position-restrained MD simulation with 0.1 fs time step (P0.1), with 0.5 fs time step (P0.5), with 1.0 fs time step (P1.0), and with 2.0 fs time step (P2.0), structure after 20 ns of unrestrained MD simulation (M20), and structure obtained through backmapping using the simulated annealing protocol from Rzepiela et al. (SA). Colors correspond to DSSP[52] determined secondary structure: α-helix (blue), 3^{10}-helix (gray), π-helix (purple), turn (yellow), bend (green), β-bridge (black), and unstructured (white).

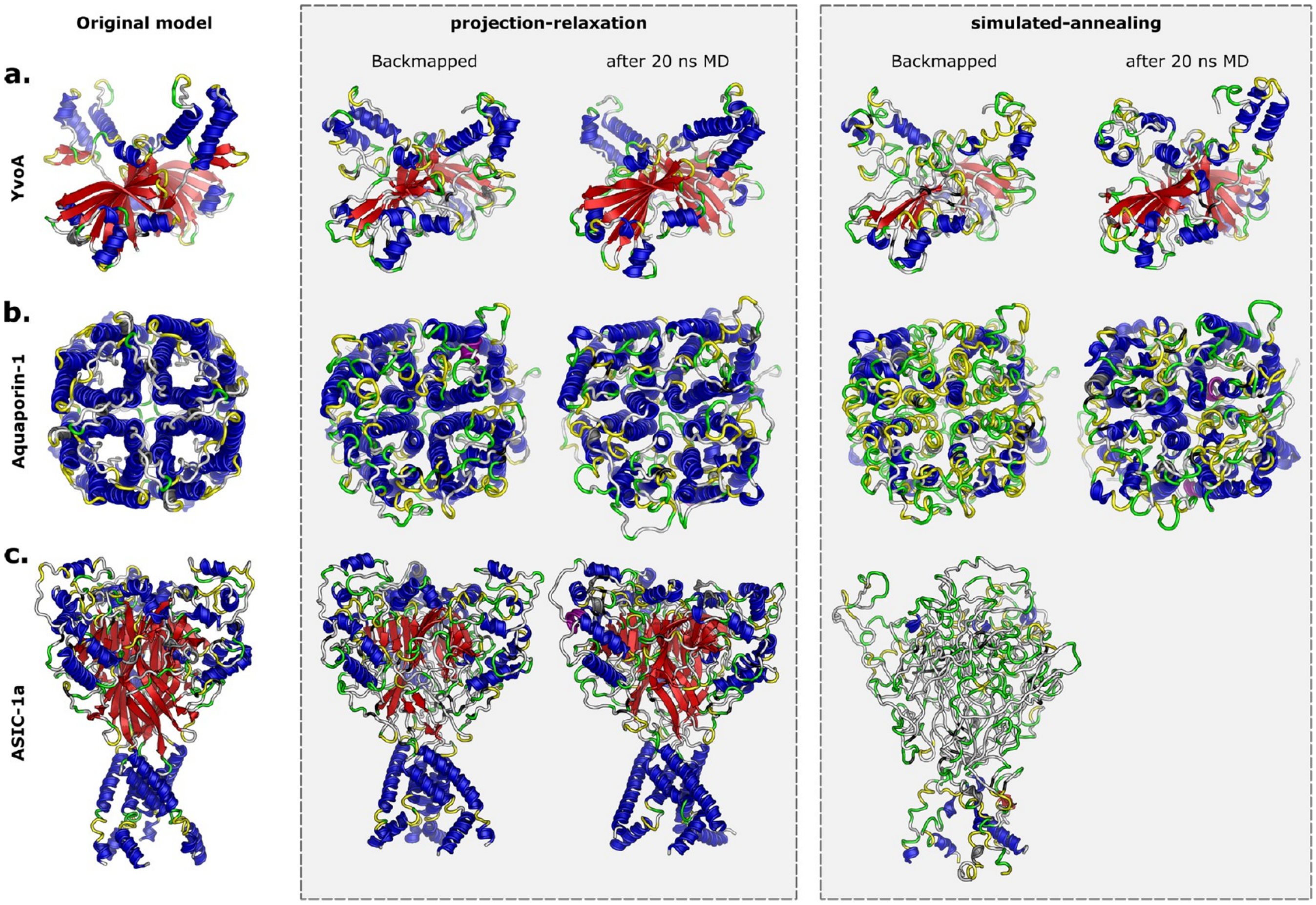


Figure 6. Conservation of protein secondary and tertiary structure with backmapping. On the left are the original models, derived from the crystal structures. The central box contains the structures obtained by backmapping the last frame from a 100 ns coarse grained simulation using the new method and the same structure after 20 ns unrestrained MD simulation. The box on the right containes the structures obtained by backmapping the last frame from a 100 ns coarse grained simulation using the simulated-annealing based method from Rzepiela et al. and the same structure after 20 ns of unrestrained MD simulation. Secondary structure assignment was done with DSSP.[52] The color scheme used is explained in the caption of 5. The proteins are (a) HTH-type transcriptional repressor YvoA. (b) Bovine aquaporin 1. (c) Acid sensing ion channel ASIC1a.

The secondary structure in MARTINI is maintained by specific bonded terms and the original structural elements should therefore be retained in the structures obtained by backmapping. This is especially true for helical elements, while MARTINI is known to have problems with maintaining extended regions. A qualitative comparison of secondary structure in the original models, used to set up the simulations, and the structures obtained by backmapping is given in Figure 6. A quantitative comparison, by means of structure content correlation using the Jaccard index,[53] is presented in Table 2.

From Figure 6, it is clear that with the new method more secondary structure is restored during backmapping than with the simulated annealing approach. This can also be seen from the Jaccard indices given in Table 2, which represent the secondary structure correlation of the given structure with the original model. The Jaccard index is given for all secondary structure, as well as for α-helical and extended regions. For reference, the same indices are calculated for corresponding structures obtained from 100 ns AA or UA MD simulations of the original models. The high-resolution reference simulations demonstrate the stability of extended regions ($J \approx 0.9$), while helical regions show some variability ($J \approx 0.75$). As compared to these reference simulations, the projection-relaxation does not restore extended regions well. The helical regions are restored better. The simulated annealing approach gives worse recovery of secondary structure, as reflected both in Figure 6 and in the Jaccard indices.

The problem with restoring extended structures is probably mainly due to the placement of backbone beads in MARTINI. These are positioned at the center of mass of each residues' backbone atoms (N,C_α,C,O). As a consequence, extended regions yield near-linear strands of beads, which do not contain the information required for correct rebuilding of the peptide planes using the backbone reconstruction routine. Indeed, a preliminary check with the ElNeDyn mapping,[39] which places backbone beads on C_α positions, suggests improved reconstruction of extended regions (data not shown).

An example illustrating the reconstruction of secondary and tertiary structure is YvoA (Figure 6a). This protein has a characteristic structure with two chains, each consisting of two domains. The C-terminal dimerization domain is characterized by a central β-sheet structure, whereas the N-terminal DNA binding domain has a so-called winged helix-turn-helix fold. During 1 μs of coarse grained simulation the characteristic features of the crystal structure are lost, and the resulting structure is more compact than the crystal structure. The shape of the structure obtained from backmapping reflects the structure obtained in the CG simulation. However, most of the secondary structure elements can already be identified, although some

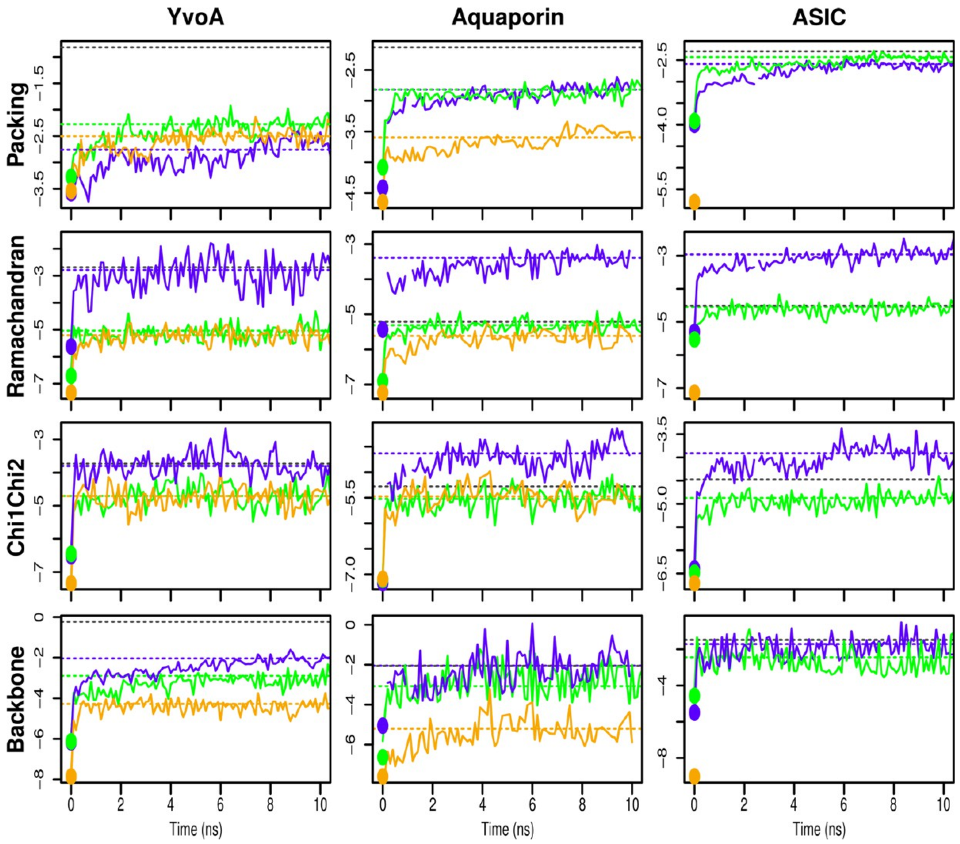


Figure 7. Protein quality after backmapping. Z-scores are shown for four protein quality markers for YvoA (**4**, left), aquaporin (**5**, center), and ASIC (**6**, right), as determined by WhatCheck.[51] Protein quality markers are, from top to bottom, new packing Z-score, Ramachandran Z-score, Chi1/Chi2 Z-score, and Backbone Z-score. Each panel show the results for projection-relaxation to CHARMM36 (blue) and to GROMOS 54a7 (green) and for the simulated annealing protocol to GROMOS 54a7 (orange). The initial backmapped structure is marked with a dot, while the line shows the score during the subsequent MD simulation. The score obtained from a corresponding 100 ns CHARMM36 (YvoA/ASIC) or GROMOS 54a7 (Aquaporin) simulation is shown as a black dashed line.

extended strands are missing or broken, resulting in a relatively low structure correlation ($J_{\mathrm{Extended}} \approx 0.4$). After 20 ns equilibration MD, the characteristic winged topology observed in the crystal structure is recovered. The correlation of extended structure elements increases more using GROMOS 54a7 (J = 0.67) than with CHARMM36 (J = 0.60). The SA protocol restores fewer structural elements, with low correlations for extended (J = 0.30) and helical (J = 0.39) regions. This is attributed to the constant application of biasing potentials, which are aimed at retaining the CG structure, rather than at regaining a native fold. It is noted that in the subsequent MD simulation the correlation increases, but stays below that observed using the projection-relaxation approach.

Similar results are obtained for aquaporin 1 (Figure 6b) and ASIC1a (Figure 6c). It is noted that the SA protocol failed to reconstruct the secondary structure in ASIC1a, reflected in negligible α-helical and extended region correlations (J = 0.03 and J = 0.05). The resulting structure is shown in Figure 6c, right panel, and could not be used for further MD simulation. The results did not change significantly using stochastic coupling to avoid local heating, as suggested by Schäfer.[43] The results suggest that the SA protocol has difficulty restoring protein structure from relaxed MARTINI CG models, especially for complex systems.

Although more secondary structure was recovered with the new method, the remapping RMSDs obtained from both methods were approximately equal for the larger proteins (Table 1), with values around 0.09 nm for the projection-relaxation approach and just over 0.1 nm for the other method. A recent comparison of several reconstruction methods suggested that remapping RMSDs around 0.1 nm should be considered accurate for CG models such as MARTINI,[50] which means that both methods qualify in terms of accuracy.

It is noted that neither the conservation of secondary structure nor the remapping RMSD provides information about the overall quality of the resulting structures. Therefore, the structures obtained from backmapping and the trajectories from the subsequent 20 ns MD simulations were processed with the protein quality checker WhatCheck.[51] This program provides Z-scores for a set of parameters, indicating the deviation from the average values as derived from a database of structures. It is known that for MD simulations and homology models, the Z-scores are usually lower than for high-resolution crystal structures.[55] Therefore, the Z-scores were compared to those obtained from 100 ns AA/UA MD simulations, starting from the original models, shown in Figure 6. The results are presented in Figure 7. In this figure, the reference values from the AA/UA simulations are indicated by a black dashed line. The structures obtained from backmapping are marked with a dot, and it is clear that backmapping with either method yields low Z-scores. During equilibration, the score increases quickly, reaching values corresponding to those observed in AA/UA simulation, except for the packing in YvoA and aquaporin, and the backbone configuration in YvoA. The SA approach yields lower scores, and

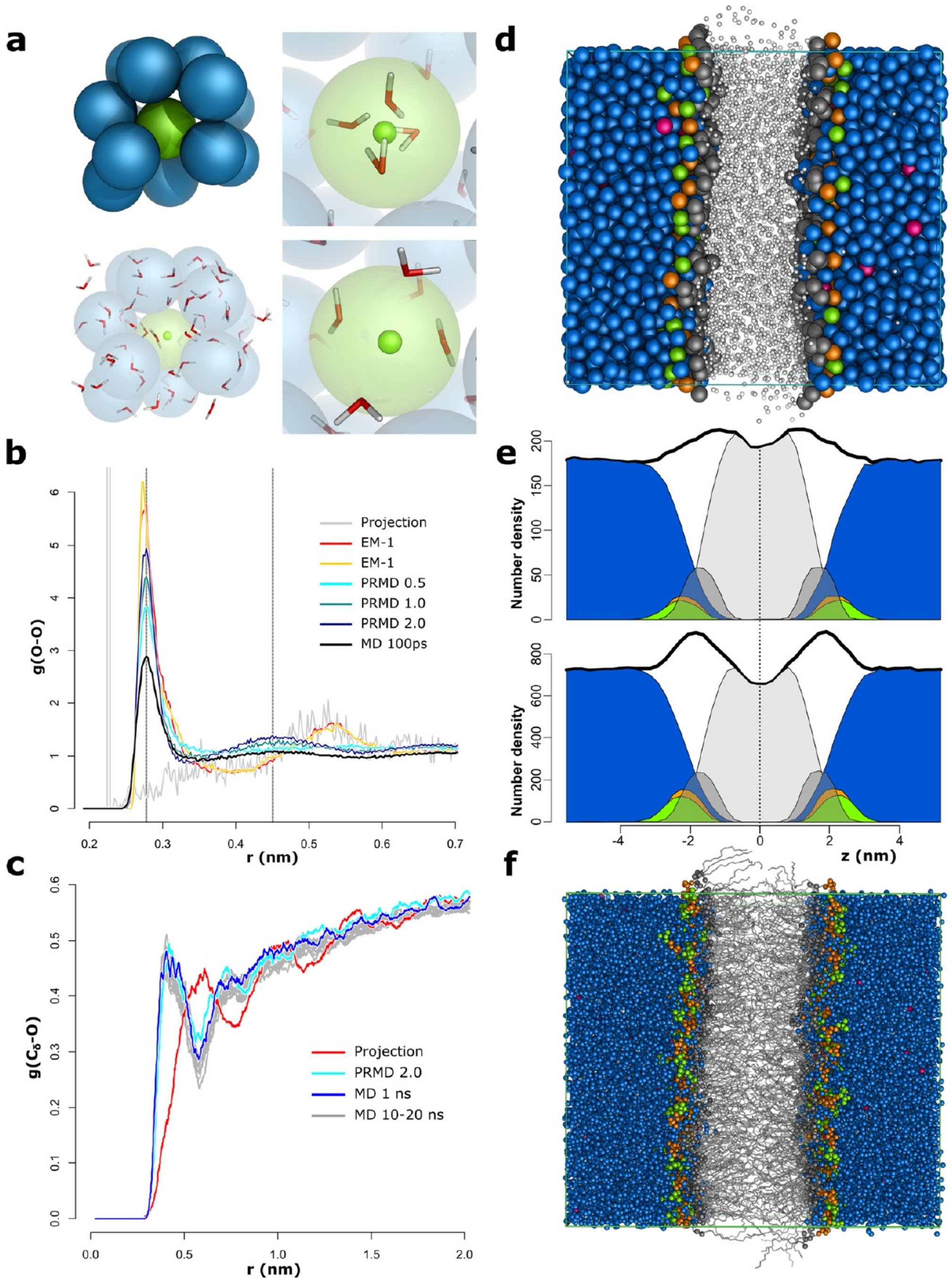


Figure 8. Solvent backmapping. (a) Backmapping of water and ions. Clockwise: CG ion surrounded by CG water, placement of tetrahedral four-water cluster over center of CG bead and placement of Na^+ at center, hydrated Na^+ configuration after energy minimization, relaxed Na^+/H_2O configuration. (b) SPC oxygen RDF in YvoA containing system (4), showing the relaxation of the water structure with the stages of the backmapping process (EM1: Energy minimization with intramolecular nonbonded interactions turned off. EM2: All interactions included. P0.5: Position-restrained MD with 0.5 fs time step. P1.0: 1.0 fs time step. P2.0: 2.0 fs time step). (c) Glycerol-C_β -SPC oxygen RDFs, determined for the membrane ASIC1a containing system (6), showing the relaxation of the water structure in the lipid linker/ester region during backmapping. The gray lines are the reference RDFs determined every nanosecond from the last 10 ns of the 20 ns unrestrained MD simulation. (d) Coarse-grained POPC/POPS/CHOL system (2). Colors indicate groups: water (blue), terminal head groups (chartreuse), phosphate groups (orange), glycerol linkers (dark gray), tail (light gray). (e) Number density profiles of selected particles in POPC/POPS/CHOL system (2). Top panel corresponds to CG, bottom panel to UA. Colors are in accordance with the particle groups in the CG representation. (f) United atom POPC/POPS/CHOL system (2). Colors indicate groups corresponding to those in CG representation.

this effect is not alleviated during the simulation. Furthermore, CHARMM36 usually yields higher scores than GROMOS54a7.

Of course it is unrealistic to expect conversion from a lower resolution model to atomistic representation to yield the same quality as a high-resolution structure obtained by crystallography or atomistic simulation. Coarse grained models simply lack a certain amount of information, but the coarse-grained force field may also cause a systematic deviation from the atomistic behavior. The problem of MARTINI to maintain the structural integrity of proteins is a good example, and this is reflected in the

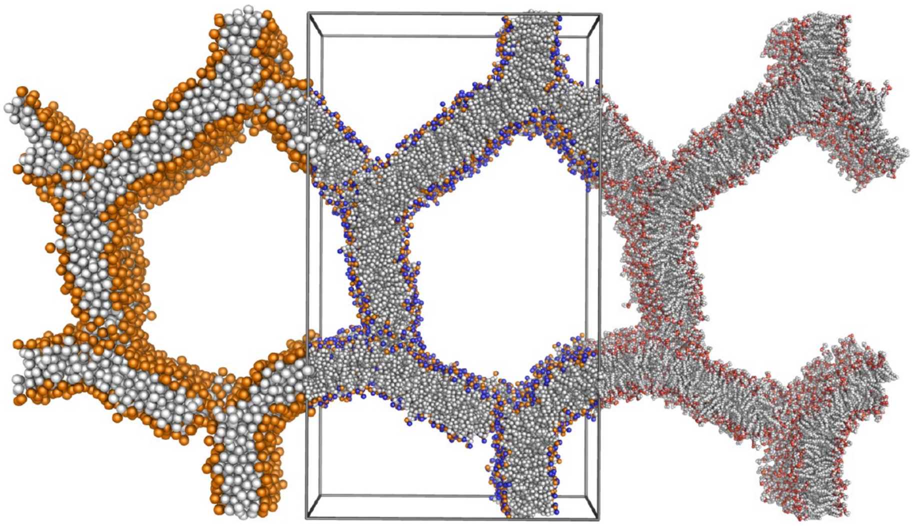


Figure 9. Backmapping of a three-bead lipid model inverted hexagonal phase to GROMOS 54a7. Combined view of a hexagonal phase bilayer as original three-bead lipid model structure (left), converted to MARTINI representation (middle, 14 beads per lipid), and converted to GROMOS 54a7 representation 54 particles per lipid). For clarity, solvent has been omitted from the images for MARTINI and GROMOS.

quality of the structure obtained by backmapping. Therefore, if the objective is, for example, investigation of interactions between molecules, it may be a good choice to restrict intraprotein dynamics by applying a stiff elastic network. Furthermore, if CG simulations are used to equilibrate a membrane around a protein, a better high-resolution representation of the whole system can be obtained by replacing the protein with the original structure after the initial stage of backmapping. This can be a suitable strategy to build atomistic protein–membrane systems with multiple lipid types. For such systems conventional methods, such as merging an equilibrating bilayer with a protein, are not optimal, as the presence of the protein may have an effect on the distribution of the lipids, potentially requiring microseconds to re-equilibrate.

4.3. Solvent. Different aspects from reverse transformation of solvent are illustrated in Figure 8. From the start, the aim was providing a method for backmapping of complete systems, including solvent, similar in scope to the method of Rzepiela et al. Their solution uses a bundled water model,[56] which is coupled to the coarse-grained water particle during the simulated annealing. Rather than using an alternative water model, backward simply places four water molecules in a tetrahedral configuration over each CG water particle (Figure 8a). The water molecules in the cluster are placed close together to avoid uncontrolled, potentially catastrophic overlaps between neighboring clusters. During energy minimization, the water molecules push each other outward and the interactions with neighboring clusters and other molecules, including lipid and protein, optimize the packing. This process is illustrated in Figure 8b, which shows the water O–O RDF in the YvoA based system (**4**) for the end of each stage. The initial RDF is characterized by an intense, narrow peak at 0.225 nm, due to the placement of water molecules close around each CG bead. In addition, there is a broad peak around 0.53 nm, corresponding to the first shell in the MARTINI W–W RDF. During the first cycle of energy minimization, the main peak shifted to the equilibrium value of 0.278 nm, while the second maximum remained largely unchanged. The latter shifted toward 0.450 nm during the cycles of position-restrained MD, which corresponds to the second maximum of the SPC O–O RDF.[57] After relaxation the water still appeared more ordered, but this effect dissipated within 100 ps during the subsequent unrestrained simulation.

A specific reason for including solvent in reverse transformation is to have an explicit solvent environment during relaxation and to retain the hydration of proteins and bilayers. Figure 8c shows the RDF between the glycerol C_δ and water OW atoms in the mixed bilayer system, illustrating the bilayer hydration before and after relaxation. Initially, there is long-range structuring, corresponding to the packing of CG water beads around the lipid linker regions. During relaxation, the water molecules reposition, and the profile converges to the equilibrium profile.

The actual hydration of the bilayer before and after backmapping is illustrated in Figure 9d–f. Parts d and f of Figure 9 show the fully hydrated POPC/POPS/CHOL bilayer in MARTINI and GROMOS 54a7 UA representation, respectively. Figure 8e shows the corresponding density profiles. The CG and UA profiles both show the hydration extending to the acyl ester region, corresponding to the tail of the CG linker region, depicted in gray. The densities are in accordance with current models of bilayer hydration,[58] and show that the reverse mapping with solvent retains the correct profiles.

4.4. Multistage Backmapping. The foregoing sections demonstrate the efficacy of the projection-relaxation approach to convert MARTINI based models to UA/AA representation. To investigate how generic the method is, we decided to attempt to take an inverted hexagonal phase structure obtained from a simulation of 2006 three-bead model lipids[49] and convert it to GROMOS 54a7 representation. The three-bead model has a minimal amount of information, which makes it difficult, if not practically impossible, to directly convert to a high-resolution model. For this reason, a multistage approach was taken, converting to MARTINI representation first, solvating, and then converting the resulting structure to high resolution. The inverted hexagonal phase is shown at the different resolutions in Figure 9. It is evident that the final structure is not fully relaxed, as there is too little disorder in the tail region. This is a consequence of the lack of information in the three-bead model, forming a single rod from which two acyl chains need to be generated. It indicates that for such conversions the intermediate stage should be simulated sufficiently long to allow relaxation.

Yet the results do suggest how advantage can be taken from different levels of resolution. A mesoscopic model allows self-assembly and/or phase transitions at a scale that is not feasible with intermediate and higher resolutions. The intermediate resolution can subsequently be used to relax the local features after which conversion to a high-resolution model can be used to investigate specific interactions with atomistic detail.

5. CONCLUSIONS

We presented a new method for inverse mapping of coarse-grained structures to atomistic representations. It primarily differs from other available methods in the geometric approach taken for the reconstruction of the initial structure, prior to relaxation, and in the flexibility to make the resulting structure match a given target topology. The latter allows specifying protonation states and virtual site definitions, and could, in principle, render backmapping a suitable method for effecting alchemical modifications. The stable conversion of CG cholesterol to either cholesterol or ent-cholesterol, shown in the Supporting Information, illustrates this latter point. The method is fast and robust, and should allow high-throughput processing of coarse-grained structures.

The new method performs integral backmapping, reconstructing complete systems, including the solvent. The resolution conversion is performed based on simple definitions of particle correspondences, making it easy to add new molecules and new force fields. Currently, the MARTINI CG force field is implemented, together with the GROMOS UA force fields and the atomistic force fields CHARMM36 and AMBER[59] in combination with Slipids.[60] In addition, the mapping between MARTINI and a three-bead lipid model[49] is available. Alternative atomistic force fields, for example, OPLS,[61] are currently being implemented.

Overall, the backmapping is accurate, reflected in low RMSDs between start and end structures after coarse-graining and backmapping. The accuracy of the protein backmapping is for an important part due to a new algorithm for backbone building, which is combined with geometric reconstruction of side chains, using the information available in the CG model. The backbone reconstruction is based on a very simple, yet strikingly accurate approximation of the peptide plane orientation, using only three control points, C_α or backbone bead, per residue.

High-resolution protein structures obtained by conversion from MARTINI may initially have a low quality score, as determined with WhatCheck,[51] and a short MD simulation can be added to increase the quality. Yet the quality of structures obtained with the new method is higher than those obtained with the earlier SA based approach. There is still room for improvement, but a limiting factor may be the quality of protein models in MARTINI.

Taken together, the projection-relaxation approach presented here offers a simple and effective way to obtain a high-resolution view from CG configurations, which opens the route to obtaining detailed insights in underlying atomistic mechanisms. The software and latest list of mapping definition files have been made available at http://cgmartini.nl.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The supporting infomation contains *backward.py* together with the wrapper *initram.sh* and mapping topologies for the CHARMM, GROMOS, AMBER, and Slipids force fields. In addition, it describes the use of the backmapping method for alchemical transformations and contains two tutorials explaining the backmapping of aquaporin 1 from MARTINI to GROMOS 54a7 using *initram.sh*, and how to add a new mapping definition (hydroxyproline). This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tsjerkw@gmail.com.

Author Contributions

[‖]T.A.W. and K.P. contributed equally to this work

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the Natural Science and Engineering Research Council (DPT) and by the German Science Foundation (DFG BO 2963/2-1). D.P.T. is an Alberta Innovates Health Solutions Scientist and Alberta Innovates Technology Futures Strategic Chair (Bio)Molecular Simulation. The authors thank Andrzej Rzepiela from the Swiss Institute of Bioinformatics for comments and suggestions.

■ REFERENCES

(1) Noid, W. G. *J. Chem. Phys.* **2013**, *139*, 090901.

(2) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. *WIRE Comp. Mol. Sci.* **2013**, DOI: 10.1002/wcms.1169.

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

(4) Stansfeld, P. J.; Sansom, M. S. *J. Chem. Theory Comput.* **2011**, *7*, 1157−1166.

(5) Shih, A. Y.; Freddolino, P. L.; Sligar, S. G.; Schulten, K. *Nano Lett.* **2007**, *7*, 1692−1696.

(6) Feig, M.; Rotkiewicz, P.; Kolinski, A.; Skolnick, J.; C., L. B., III *Proteins* **2000**, *41*, 86−97.

(7) Heath, A. P.; Kavraki, L. E.; Clementi, C. *Proteins: Struc., Funct. Bioinf.* **2007**, *68*, 646−661.

(8) Rzepiela, A. J.; Schäfer, L. V.; Goga, N.; Risselada, H. J.; De Vries, A. H.; Marrink, S. J. *J. Comput. Chem.* **2010**, *31*, 1333−1343.

(9) Hess, B.; Leon, S.; van der Vegt, N.; Kremer, K. *Soft Matter* **2006**, *2*, 409−414.

(10) Peter, C.; Kremer, K. *Soft Matter* **2009**, *5*, 4357−4366.

(11) Thøgersen, L.; Schiøtt, B.; Vosegaard, T.; Nielsen, N. C.; Tajkhorshid, E. *Biophys. J.* **2008**, *95*, 4337−4347.

(12) Brocos, P.; Mendoza-Espinosa, P.; Castillo, R.; Mas-Oliva, J.; Piñeiro, A. *Soft Matter* **2012**, *8*, 9005−9014.

(13) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.

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

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

(16) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput. Biol.* **2010**, *6*, e1000810.

(17) Schmid, N.; Eichenberger, A.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A.; Gunsteren, W. *Europ. Biophys. J.* **2011**, *40*, 843−856.

(18) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.

(19) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *J. Phys. Chem. B* **2010**, *114*, 7830−7843.

(20) Piggot, T. J.; Piñeiro, A.; Khalid, S. *J. Chem. Theory Comput.* **2012**, *8*, 4593−4609.

(21) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(22) Rotkiewicz, P.; Skolnick, J. *J. Comput. Chem.* **2008**, *29*, 1460−1465.

(23) Wassenaar, T. A.; Dijk, M.; Loureiro-Ferreira, N.; Schot, G.; Vries, S. J.; Schmitz, C.; Zwan, J.; Boelens, R.; Giachetti, A.; Ferella, L.; Rosato, A.; Bertini, I.; Herrmann, T.; Jonker, H. R. A.; Bagaria, A.; Jaravine, V.; Güntert, P.; Schwalbe, H.; Vranken, W. F.; Doreleijers, J. F.; Vriend, G.; Vuister, G. W.; Franke, D.; Kikhney, A.; Svergun, D. I.; Fogh, R. H.; Ionides, J.; Laue, E. D.; Spronk, C.; Jurkša, S.; Verlato, M.; Badoer, S.; Dal Pra, S.; Mazzucato, M.; Frizziero, E.; Bonvin, A. M. J. J. *J. Grid. Comput* **2012**, *10*, 743−767.

(24) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20*, 786−798.

(25) Berendsen, H.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(26) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(27) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(28) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(29) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

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

(31) van der Spoel, D.; van Maaren, P. J. *J. Chem. Theory Comput.* **2006**, *2*, 1−11.

(32) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermol. Forces* **1981**, 331−342.

(33) Poger, D.; Mark, A. E. *J. Chem. Theory Comput.* **2010**, *6*, 325−336.

(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926−935.

(35) MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(36) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. *J. Chem. Theory Comput.* **2010**, *6*, 459−466.

(37) Marrink, S. J.; Tieleman, D. P. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.

(38) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.

(39) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. *J. Chem. Theory Comput.* **2009**, *5*, 2531−2543.

(40) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(41) Nosé, S. *Mol. Phys.* **1984**, *100*, 191−198.

(42) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.

(43) Schäfer, L. V.; de Jong, D. H.; Holt, A.; Rzepiela, A. J.; de Vries, A. H.; Poolman, B.; Killian, J. A.; Marrink, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 1343−1348.

(44) Courtney, J. M.; Vostrikov, V. V.; Hinton, J. F.; Koeppe, R. E., II *Biophys. J.* **2011**, *100*, 636.

(45) Sui, H.; Han, B.-G.; Lee, J. K.; Peter, W.; Jap, B. K. *Nature* **2001**, *414*, 872.

(46) Gonzales, E. B.; Kawate, T.; Gouaux, E. *Nature* **2009**, *460*, 599−604.

(47) Jasti, J.; Furukawa, H.; Gonzales, E. B.; Gouaux, E. *Nature* **2007**, 316−323.

(48) de Planque, M. R. R.; Killian, J. A. *Mol. Membr. Biol.* **2003**, *20*, 271−284.

(49) Cooke, I. R.; Kremer, K.; Deserno, M. *Phys. Rev. E* **2005**, *72*, 011506.

(50) Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. *Proteins: Struc., Funct. Bioinf.* **2010**, *78*, 1266−1281.

(51) Vriend, G.; Sander, C. *J. Appl. Crystallogr.* **1993**, *26*, 47−60.

(52) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

(53) Jaccard, P. *New Phytol.* **1912**, *11*, 37−50.

(54) Kučerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencer, J.; Jackson, A.; Katsaras, J. *Biophys. J.* **2008**, *95*, 2356−2367.

(55) Law, R. J.; Capener, C.; Baaden, M.; Bond, P. J.; Campbell, J.; Patargias, G.; Arinaminpathy, Y.; Sansom, M. S. *J. Mol. Graph. Model.* **2005**, *24*, 157−165.

(56) Fuhrmans, M.; Sanders, B. P.; Marrink, S.-J.; Vries, A. H. *Theor. Chem. Acc.* **2010**, *125*, 335−344.

(57) Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954−9960.

(58) White, S. H.; Ladokhin, A. S.; Jayasinghe, S.; Hristova, K. *J. Biol. Chem.* **2001**, *276*, 32395−32398.

(59) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struc., Funct. Bioinf.* **2006**, *65*, 712−725.

(60) Jämbeck, J. P. M.; Lyubartsev, A. P. *J. Phys. Chem. B* **2012**, *116*, 3164−3179.

(61) Kaminski, G.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

■ NOTE ADDED IN PROOF

It was brought to our attention that our routine for backbone reconstruction is similar to the method used for identification of secondary structure in C-alpha only models by Levitt and Greer (*J. Mol. Biol.* **1977**, *114*, 181−293).