

Supplementary Material for

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, D. Peter Tieleman

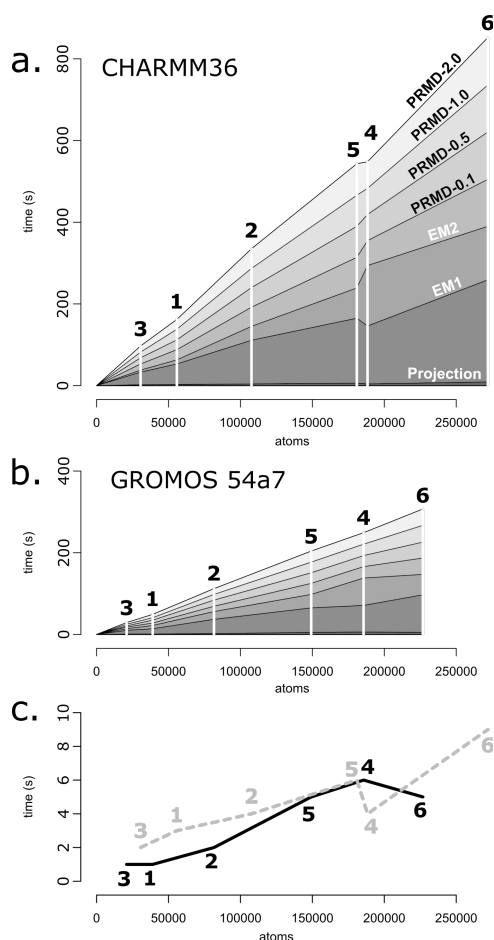


Figure S1: Timing of projection and relaxation. a. Timing of MARTINI to CHARM36 AA backmapping, showing the cumulative times versus resulting system size for each system. Systems are identified by the number as given in the methods section. The time required for each stage of the process is linear in the number of atoms. The first cycle of energy minimization is relatively most costly b. Timing of MARTINI to GROMOS 54a7 UA backmapping. c. Timing of projection only, showing a linear relation with the number of target atoms, otherwise independent of the target force field: CHARM36 (dashed line) or GROMOS 54a7 (solid line).

Using backmapping for alchemical transformations

The six test systems showed that the new method was able to reconstruct complete simulation systems quickly and accurately. The correct reconstruction of specific geometrical elements, such as chiral centers and double bonds, suggested that it should also be possible to reconstruct systems to an alternative configuration, e.g. with inverted chiral centers or double bonds or even with different chemical composition. It is noted that doing so changes the aim of the method. By default, the aim is finding a chemically correct high-resolution structure that corresponds to a coarse-grained model. However, since the method performs the mapping against a target topology, it is also possible to reconstruct to an alternative system, as long as the mapping can be resolved based on the coordinates available in the CG model. This will always be the case for inverted chirality, but is equally trivial for amino acid substitution if the number of side chain particles is the same or less in the target topology. Here, such an alchemical conversion is illustrated by backmapping cholesterol in the POPC/POPS/cholesterol bilayer to *ent*-cholesterol, in which all chiral centers are inverted (Figure S2). *ent*-Cholesterol is used experimentally to assess whether cholesterol-mediated effects are specific or aspecific. The number of chiral centers and the correspondence between the CG and UA/AA models of cholesterol make the inversion of the chiral centers rather challenging. Figure S2 shows a typical case of a CG cholesterol molecule mapped to either form of cholesterol. In all cases, all chiral centers were obtained with correct geometry, and the *ent*-cholesterol structures all relaxed to proper mirror-image structures of cholesterol.

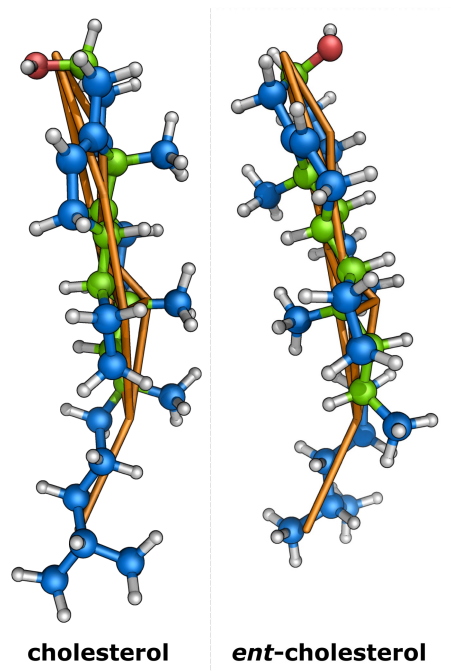


Figure S2: Mapping of CG cholesterol to AA cholesterol and *ent*-cholesterol. Side views of MARTINI CG cholesterol structure (orange) with the corresponding CHARMM36 cholesterol (left panel) and *ent*-cholesterol (right panel) structures obtained through projection-relaxation. The eight chiral centers are highlighted in green.

Tutorial 1: Reverse transformation of aquaporin 1 water channel in POPC bilayer

In this tutorial a MARTINI coarse grained system containing aquaporin-1 embedded in a bilayer and surrounded by explicit solvent is converted to GROMOS united atom representation. Aquaporins are water transporting proteins commonly found in cell membranes. They have been the subject of numerous experimental and simulation studies.

The possibility to convert a system from low to high resolution is of particular interest for systems like aquaporin. Several methods exist for merging a protein and a membrane to build an atomistic system, but the equilibration of such complex systems typically takes hundreds of nanoseconds, or even microseconds if the bilayer has a mixed composition. This is impractical in terms of atomistic simulations. However, on the coarse-grained level these time scales are easily reached. In addition, the setup of a complex protein-membrane-solvent system is much easier for coarse-grained systems. Yet coarse-grained models do not allow simulation of atomic scale processes, such as permeation of single water molecules through aquaporins. Using a method for resolution conversion, the best of both can be combined. The system can be built and equilibrated on the coarse-grained level and subsequently converted, followed by investigation of the process of interest at atomistic resolution. This tutorial focuses on the resolution conversion of a previously built and equilibrated system.

The MARTINI coarse-grained coordinates and the GROMOS 54a7 united atom target topology can be downloaded from <http://www.biotechnik.nat.uni-erlangen.de/research/boeckmann/downloads/MPS/>. A tutorial explaining how this system was built is given elsewhere¹.

Before starting the conversion, it is important to assert that molecules in the structure file are not split over the periodic boundaries. It is also important that all dependencies, like the force field parameters and molecular topologies, are available. In addition, the target topology should not list solvent molecules and ions, as these are determined and added based on the coarse-grained system. If the conversion program `backward.py` and the wrapper `initram.sh` are not available, they can be downloaded as an archive (`backward.zip`) from <http://md.chem.rug.nl/cgmartini/index.php/downloads/tools/240-backward>.

This archive also contains the mapping definitions. The file can be unzipped, e.g., in the working directory:

```
unzip backward.zip
```

The directory should now contain the following files/directories:

```
aquaporin_CG.gro
aquaporin_UA.top
backward.py
initram.sh
Mapping/
popc_UA.itp
```

If the program is located elsewhere, the paths in the following commands should be modified accordingly. The conversion is performed in a single step, using `initram.sh`, which in turn calls `backward.py` and the GROMACS programs `grompp` and `mdrun` for relaxation runs. It is important that the latter are available, but it is assumed that this is the case. If not, the program will fail, indicating that one of the dependencies cannot be found.

The mandatory arguments to `initram.sh` are the coarse-grained input structure (`-f aquaporin_CG.gro`), the target topology file (`-p aquaporin_UA.top`) and the name of the target force field (`-to gromos54a7`). We here add the option `-keep`, which causes all intermediate files generated to be retained rather than deleted. This allows tracking the process.

```
./initram.sh -f aquaporin_CG.gro -p aquaporin_UA.top -to gromos54a7 -keep
```

After checking the dependencies, the process starts with a call to `backward.py`. This program sets up the conversion, reading in the mapping definitions for the force field chosen. It will report the residues available

for the conversion for that force field, which should look like:

Residues defined for transformation from martini to gromos54a7:

```
['POPC', 'GLN', 'ILE', 'POPE', 'POPS', 'GLY', 'GLU', 'CYS', 'HIS', 'SER', 'PRO',  
'CHOL', 'CL4', 'ASN', 'HEP', 'VAL', 'DOPE', 'DPPC', 'DOPC', 'THR', 'ASP', 'TRP',  
'LYS', 'PHE', 'ALA', 'MET', 'LEU', 'ARG', 'TYR']
```

If the target force field is not available, e.g., due to an error in the name, the list will be empty. The whole process takes a few minutes, generating a number of files. The structures obtained after each stage of relaxation are numbered. The first stage, the generation of the united atom starting structure (0-backward.gro) and the complete united atom topology (backmapped.top) takes about a second. The first cycle of energy minimization is started subsequently. This process is run in verbose mode, printing the potential energy to the screen. Due to the strain in the starting structure, the potential energy may be very high initially, but usually drops quickly, reaching negative values in a few hundreds steps.

At the end of the process, the final, relaxed structure (6-prmd-0.002.gro) is copied to backmapped.gro. This file can be used together with the matching topology file backmapped.top to start subsequent high resolution simulations.

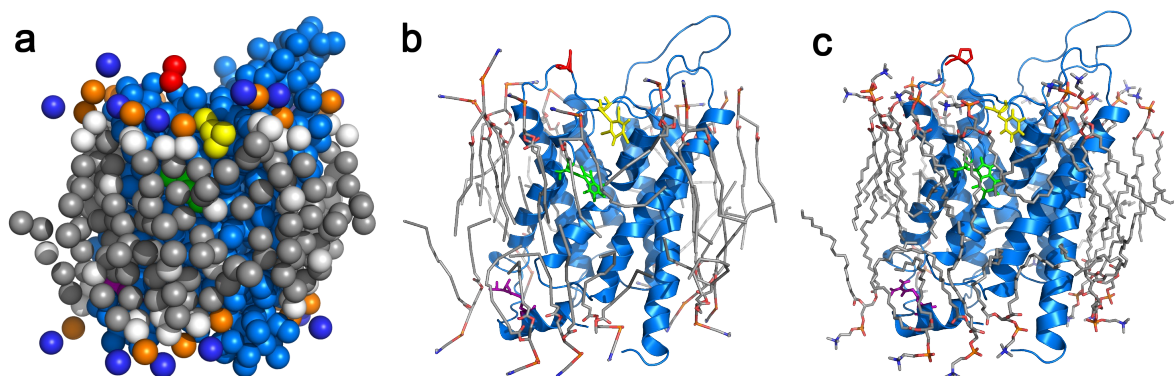


Figure S3: Monomer of aquaporin 1 (marine) in a POPC bilayer. Only the lipid molecules closest to the protein are visualized. Water and ions were omitted for clarity. Some residues are highlighted, and shown as sticks in UA representation: P133 (red), F208 (yellow), R243 (magenta) and W215 (green). a. CG representation (aquaporin_CG.gro) b. Backward structure (0-backward.gro) c. final geometry (backmapped.gro).

Tutorial 2: Adding a mapping definition for reverse transformation of hydroxyproline

In this tutorial, the process of adding a residue mapping definition is demonstrated, using hydroxyproline (Figure S4) as example. Hydroxyproline is a modified amino acid, which is a natural constituent of collagen. Parameters for collagen and hydroxyproline are available for MARTINI and for GROMOS 54a7.

Adding a mapping definition for resolution conversion is straightforward if the correspondence between particles is known. The MARTINI structure for hydroxyproline, like proline, consists of a backbone bead (BB) and a single side-chain bead (SC1), with default particle correspondence for amino acids: BB maps to $\text{N-C}_\alpha\text{-C=O}$, while SC1 maps to the atoms $\text{C}_\beta\text{-C}_\gamma\text{-C}_\delta\text{-O}_\delta\text{-H}_\delta$. The mapping definition is written to a new file in the directory Mapping/. The name of the file is not important, but should be descriptive, e.g., hyp.gromos.map. This file should be opened with an editor. The first lines written to the file are

```
[ molecule ]  
; Hydroxyproline  
HYP
```

The mapping is molecule or residue based, and the tag [molecule] starts a new entry. The line following the tag is a comment, indicated by the semicolon, which lists the actual name of the molecule. Comments are optional and are ignored by the parser. The first non-comment line following the [molecule] tag should contain the name of the molecule. This name should correspond to a molecule type as defined in the target topology, or to a residue name in the structure file.

The next few lines to add to the file set the lower resolution definition. It consists of a tag with the name of the force field, followed by a list of particle names in the same order as in that force field. The mapping written here is for MARTINI, where hydroxyproline consists of two beads, BB and SC1:

```
[ martini ]  
BB SC1
```

After the definition of the lower resolution like this, one or several mappings to a higher resolution can be specified. Each of these starts with a tag [mapping], followed by a list of force field names, to which the mapping applies. Here, we write a mapping for GROMOS 54a7, but the same mapping applies to other versions of the GROMOS force field, and all are added:

```
[ mapping ]  
gromos43a1 gromos43a2 gromos45a3 gromos53a5 gromos53a6 gromos54a7
```

The next section, tagged [atoms], lists the particles to construct from the lower resolution. These particles are given in the order in which they need to be written, according to the definition in the target force field. Each particle definition consists of a number, the particle name, and a list of corresponding beads in the lower resolution representation. The number is added to make it easier to write mapping definitions for complex molecules, but is not used otherwise. The position of the target particle is set to the average of the positions of the beads listed. If no beads are listed, then the position of the previous particle is used. The resulting initial mapping for hydroxyproline is relatively simple:

```
[ atoms ]  
1 N BB  
2 CA  
2  
3 CB BB SC1  
4 CG SC1  
5 OD1  
6 HD1  
7 CD2  
8 C BB  
9 O
```

Because MARTINI hydroxyproline is a two-bead model, the mapping ends up ill-defined. However, since this is an amino-acid, the backbone (N,H,C $_{\alpha}$,C,O) will be reconstructed using a specialized routine, rather than from the mapping defined above. This will yield well-defined positions for the backbone, in turn allowing reconstructing the side chain using geometrical modifiers. These modifiers are processed in the order given. We start with repositioning the C $_{\beta}$, setting the chirality of the C $_{\alpha}$:

```
[ chiral ]
CB CA N C
```

This places C $_{\beta}$ onto C $_{\alpha}$, using N and C as control atoms to define the chirality. Note that reversing N and C would invert the chirality, yielding a D-aminoacid. Now, the side chain atom C $_{\delta}$ is placed onto N, on the same side as C $_{\beta}$ to reconstruct the ring. This can be easily achieved using the cis operator:

```
[ cis ]
CD2 N CA CB
```

Finally, the hydroxyl group is reset to be connected to C $_{\gamma}$. The latter is a chiral center, but because GROMOS is a united atom force field, the chirality is maintained using an improper dihedral, and it is sufficient to place the oxygen pointing outward from the ring. For this purpose we here use the trans operator. The same operation is used to place the hydrogen atom. This causes two particles to get the same coordinates, but in such cases the last atom is given a small random displacement to avoid overlaps.

```
[ trans ]
OD1 CG CB CA
HD1 CG CB CA
```

This finishes the mapping definition of hydroxyproline, and the file can be saved and closed. If the aim was adding a normal molecule, then the procedure described thus far would be sufficient to add the mapping definition. However, we here add an amino acid, which needs to be added to the lists of residues for backbone reconstruction and for handling of terminal protonation states. These lists are defined in two places. In the directory Mapping/, where the mapping definitions are located, there is a file `__init__.py`, which contains the code for reading the files and doing the conversion. To list an additional amino acid, the file can be opened in an editor. The variables that are most likely to need updating are located near the top of the file. Among these is a variable called `_aminoacids`, which contains a listing of residue names of protein building blocks. The residuename "HYP" should be added to this list:

```
# Listing of aminoacids
_aminoacids = [ "ALA", "CYS", "ASP", "GLU", "PHE", "GLY", "HIS", "ILE", "LYS",
"LEU", "MET", "ASN", "PRO", "GLN", "ARG", "SER", "THR", "VAL", "TRP", "TYR",
"ACE", "NH2", "HYP", ]
```

After adding the residue name, the file can be saved and closed. The residue name is subsequently added in a similar manner to the main backmapping program "backward.py", which is located one directory higher than `__init__.py`. `backward.py` should also be opened in an editor. Again, the variables most likely to require updating are located near the top of the file. In this case, the variable is called `AminoAcids`. This list is written as a single string, rather than a list of strings, and the residue name HYP can simply be added at the end, preceded by a space:

```
AminoAcids = "ALA CYS ASP GLU PHE GLY HIS ILE LYS LEU MET ASN PRO GLN ARG SER
THR VAL TRP TYR ACE NH2 HYP".split()
```

The file can now be saved and closed. The amino acid hydroxyproline is now available for backmapping from MARTINI to GROMOS 54a7. To illustrate this, the collagen structure 1CAG can be downloaded from

the PDB and coarsegrained using martinize.py:

```
martinize.py -f 1CAG.pdb -o CG.top -x CG.pdb
```

In addition, a united atom description is generated, using the tool pdb2gmx from GROMACS. This does require that the atom CD from hydroxyproline is first renamed CD2 and that the acetic acid molecules are removed from the PDB file. This can be done use the stream editor sed, which will modify a file in-place if the option -i is given:

```
sed -i -e 's/CD HYP/CD2 HYP/' -e '/ACY/d' 1CAG.pdb
pdb2gmx -f 1CAG.pdb -o AA.pdb -p AA.top
```

At this point we have a coarse grained structure and a united atom topology, which we use as input to backward.py to obtain a united atom structure.

```
backward.py -f CG.pdb -p AA.top -o backmapped.gro -po backmapped.top -from
martini -to gromos54a7
```

The resulting structure is not relaxed. In this example, we only perform one cycle of energy minimization, with full interactions, to relax the structure. We first generate the parameter file for 2500 steps of steepest descent energy minimization without periodic boundary conditions:

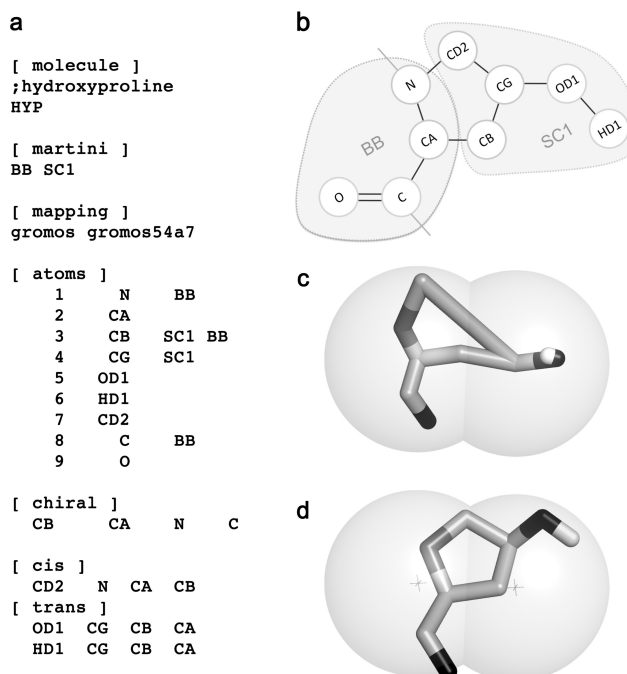
```
echo -e "integrator=steep \n nsteps=2500 \n pbc=no"> em.mdp
```

Then we generate the run input file using the GROMACS tool grompp and subsequently run the energy minimization using mdrun:

```
grompp -f em.mdp -c backmapped.gro -p backmapped.top -o relax.tpr
mdrun -v -deffnm relax
```

As a final step, the structures before and after coarse graining and backmapping can be superposed and compared. This should show that all hydroxyl groups of hydroxyproline have inverted chirality. This is caused by an inverted improper dihedral in the GROMOS force field definition of that residue.

Figure S4: Mapping and reconstruction of hydroxyproline for the GROMOS 54a7 force field from MARTINI CG. a. Mapping definition for 4-hydroxyproline. b. Schematic view of tryptophan CG/UA mapping. c. Overlay of projected structure and original CG positions of hydroxyproline, based on the mapping definition. d. Overlay of final, relaxed structure and original CG positions.



References:

(1) Pluhackova, K., Wassenaar, T. A. and Böckmann, R. A. In Membrane Biogenesis: Methods and Protocols; Doron, R., M. H. J., Eds.; Springer Science+Business Media, 2013; Vol. 1033; Chapter Molecular Dynamics Simulations of Membrane Proteins.