CG2AT2: an Enhanced Fragment-Based Approach for Serial Multiscale Molecular Dynamics Simulations

Owen N. Vickery and Phillip J. Stansfeld*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Coarse-grained molecular dynamics provides a means for simulating the assembly and interactions of macromolecular complexes at a reduced level of representation, thereby allowing both longer timescale and larger sized simulations. Here, we describe an enhanced fragment-based protocol for converting macromolecular complexes from coarse-grained to atomistic resolution, for further refinement and analysis. While the focus is upon systems that comprise an integral membrane protein embedded in a phospholipid bilayer, the technique is also suitable for membrane-anchored and soluble protein/nucleotide complexes. Overall, this provides a method for generating an accurate and well-equilibrated atomic-level description of a macromolecular complex. The approach is evaluated using a diverse test set of 11 system configurations of varying size and complexity. Simulations are assessed in terms of protein stereochemistry, conformational drift, lipid/protein interactions, and lipid dynamics.
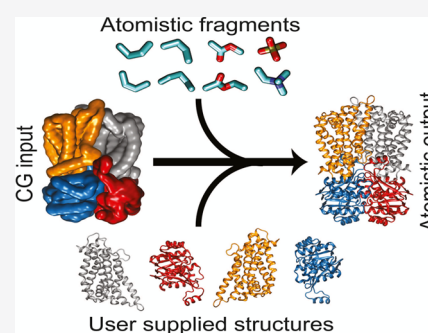
INTRODUCTION

Membrane proteins are fundamental to life. They are the gate keepers into and out of a cell and critical receptors for transmitting signals across cellular membranes. In the majority of organisms, they comprise ∼25% of genes, and from a pharmaceutical perspective, they form the targets for roughly half of all drugs.[1−3] Structural determination of these proteins has advanced considerably in the past decade, with improvements to the protocols for X-ray diffraction and the advent of atomic-resolution cryo-electron microscopy (cryo-EM) resulting in a total of over 5000 membrane protein structures, of which there are more than 1000 unique proteins (see http://memprotmd.bioch.ox.ac.uk/stats for the most recent data).[4] Despite these improvements, the lipids in the surrounding environment are frequently difficult to capture, either being lost in the purification process, too diffuse to accurately determine atomic coordinates,[5] or difficult to assign as a defined lipid type.

Molecular dynamics (MD) simulations provide an ideal tool for capturing the lipid environment around these membrane protein structures, as we have shown through our database of all membrane-embedded structures, MemProtMD.[4] The use of coarse-grained (CG) modeling allows for the observation of slow processes such as membrane bending and deformations,[6] protein−lipid interactions,[7,8] and association of soluble proteins to bilayers.[9] The simulation time required for these events is generally beyond that afforded for conventional atomistic simulations.[10] CG-MD simulations enable a well-configured springboard for studying, for example, molecular and ionic transport, ligand binding, structural dynamics, and lipid interactions of proteins.[10,11] However, first, one must translate the system from a CG representation to an atomic-level description. A number of methods have been developed for the conversion of a CG system to an atomistic one, including, for example, a simulated annealing methodology,[12] a geometric-based backward protocol,[13] machine learning algorithms,[14,15] and a previous iteration of the fragment-based CG2AT approach.[16] In all cases, the methodologies permit serial multi-scale MD simulations.[17]

Here, we describe a complete reworking and enhancement of our fragment-based approach, introducing an assortment of new features, an increased level of flexibility, enhanced scalability, and greatly improved interoperability. We apply our method to a range of molecular systems, from biological membranes to integral membrane proteins, lipid-modified lipoproteins, and DNA-bound soluble proteins.

METHODS

Implementation and Database. To enable the conversion of the CG system to an atomistic representation, the stereochemistry of each bead and molecule must be reintroduced. Various conversion methods have been previously proposed to enable conversion of a CG system in a *de novo* manner (see above). For CG2AT2, we implement a
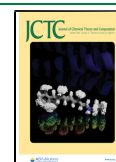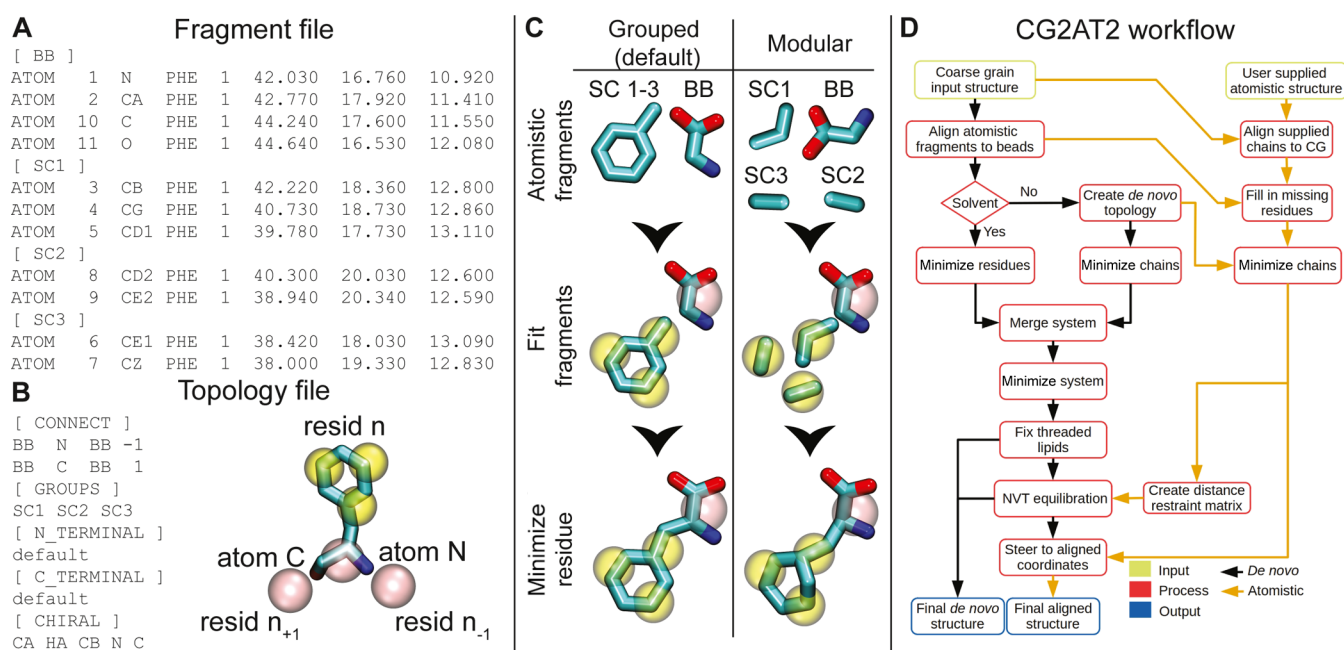
Figure 1. Workflow and database composition. (A) Fragment file for phenylalanine. Each fragment has a header containing the corresponding CG bead name, for example, backbone [BB], side chain [SC1], and so on, followed by the atoms within the fragment. If the molecule is not processed by pdb2gmx, then, the atom numbers should match the supplied topology itp file. (B) Topology file for phenylalanine. There are six sections within the topology file: CONNECT details the connectivity of the backbone and should follow a four-column convention: bead name, associated atom name, connected bead name, and numeric direction of the connected bead. GROUPS contain the fragment groupings, as shown in the (C) "Grouped" section. N_TERMINAL and C_TERMINAL contain any non-standard atomistic termini for the residue. CHIRAL contains any stereochemistry information present within the residue. This is described in the column format: central atom, atom to move, atom 1, atom 2, and atom 3. (C) Graphical depiction of the grouped (left) and modular (right) methodologies implemented within CG2AT2, where atomistic fragments are either grouped and aligned with their CG beads or aligned individually in a modular fashion. The conversion of phenylalanine is shown. Here, the grouped approach improves the conversion quality. (D) Description of the conversion of CG to AT. Initially, CG2AT2 converts the CG system in a *de novo* manner from the fragments supplied. If an atomistic protein structure is provided, this structure is aligned with the CG coordinates, with the *de novo* protein coordinates steered to the aligned protein coordinates to remove clashes with surrounding atoms, for example, lipids.

similar fragment-based methodology to the previous iteration of CG2AT, in which small atomistic fragments containing all the chemical and stereochemical information required are rigidly aligned to each CG bead.

For CG2AT2, we have assembled a database of fragment files that are related to the commonly used force fields, for example, CHARMM,[18] AMBER,[19] GROMOS,[20] and OPLS.[21] To simplify the development and usability of the fragment database, minimal information is required beyond the atomistic fragments (Figure 1A) and an optional itp file (*e.g.*, POPC.itp).

The fragment files are stored within a single-coordinate file, where the residue is separated by bead names into fragments (Figure 1A). The variable lengths and protonation states of multimeric macromolecules such as proteins necessitate that the topologies are generated on the fly using the GROMACS[22] tool pdb2gmx. However, individual non-protein molecules, for example, palmitoyl-oleoyl-phosphatidylcholine (POPC), can use the supplied topology itp file in the database.

Additional information that cannot be retrieved from the itp and/or force field files is supplied in a topology file. This topology file contains four sections to aid the conversion. This file details inter-residue connectivity, fragment groupings, terminal residue information, and chiral groups (Figure 1B). The chiral group section enables the appropriate stereochemistry to be adhered to upon conversion. This can be aided further by grouping atomistic fragments together, so that they

are structurally aligned as a single unit to the CG beads (Figure 1C). For polymers, that is, proteins, sugars, and nucleotides, inter-residue connectivity makes sure that residues are appropriately linked, and the terminal residue state options apply appropriate charge states to the first and last residues.

An optional position restraint file for the non-protein residues can also be included. Within the supplied database, the position restraints are a harmonic potential of 50 kJ mol^{-1} nm^{-2} on the *xyz* axis. These restraints may be applied by GROMACS during the equilibration and steering steps to minimize deviation from the CG system.

Conversion Protocol. The conversion protocol of CG2AT2 can be condensed into four core steps: conversion, minimization, integration, and equilibration (Figure 1D).

Conversion. The initial conversion phase consists of atomistic fragment fitting to the CG beads. Here, the atomistic fragments are aligned by their center of mass to their respective CG bead. The aligned coordinates are then rotated to minimize bond distances with their adjacent connected beads. In this manner, each fragment can be treated individually with no knowledge of the surrounding atomistic fragments (Figure 2A). To improve upon the initial conversion, we have implemented a fragment-grouping system, detailed in an optional topology file (Figure 1B), which allows for multiple fragments to be treated as a single fragment, for example, an amino acid side chain (Figure 1C). It is possible within CG2AT2 to swap beads and residues using the flag
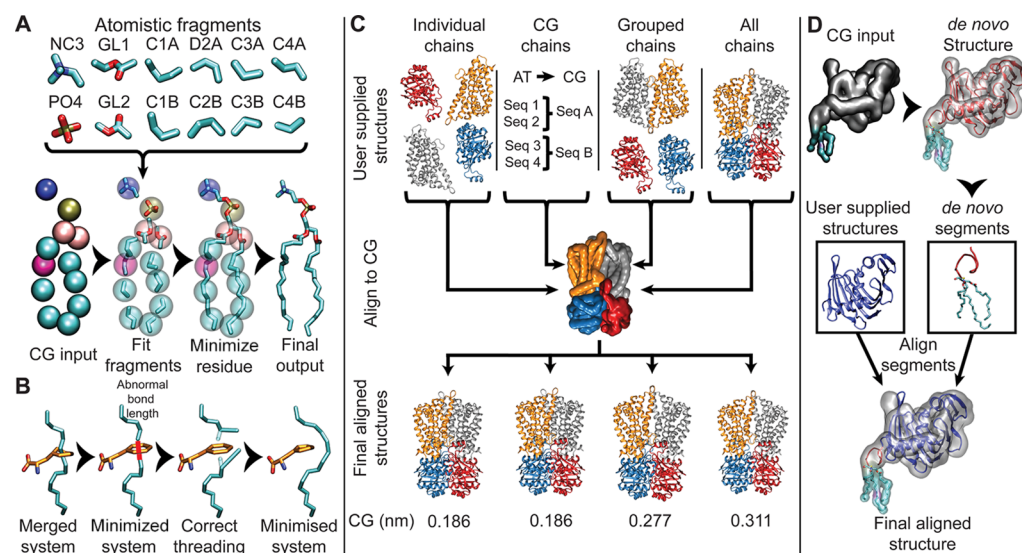
Figure 2. Graphical depiction of the conversion of Martini's POPC to CHARMM36 atomistic representation. (A) General conversion protocol implemented with CG2AT2. Here, the individual fragments are aligned within each bead and rotated to minimize the distance between connecting beads. Once aligned, the residue undergoes the first round of minimization. Hydrogens are hidden for clarity. (B) Disentanglement of a lipid tail from the side chain of phenylalanine. If an abnormally long bond is detected after the initial minimization of the merged system, the two non-hydrogen atoms and connected hydrogens are translated on a vector away from the threaded amino acid backbone. The system is then minimized to correct the bond lengths between the adjusted atoms. (C) Depiction of alignments available for the conversion of CG to AT. Here, the user is able to select whether to align the atomistic chains individually, grouped by the CG chain, in selective groups or as a single macromolecular unit. Root mean square deviation (RMSD) measurements are highlighted: Cα against CG BB. (D) Flow chart of the alignment of individually supplied atomistic chains. Here, LolB is created in the *de novo* manner depicted in the red cartoon. The supplied atomistic chains (colored blue) are aligned to the CG BB beads. The missing residues highlighted during chain alignment are copied from the *de novo* structure and mapped onto the aligned structure.

-swap. This enables the user to undertake minor mutations, for example, POPE to palmitoyl-oleoyl-phosphatidylglycerol (POPG) during the conversion process.

Minimization. After a successful conversion, each molecule type undergoes a single round of minimization by the steepest descent algorithm. This corrects the majority of bond lengths and angles and any steric clashes.

Integration. Following a successful build of each molecule type, the system components are reintegrated and minimized. In rare cases upon reintegration, it is possible for lipid tails to be threaded through aromatic structures such as phenylalanine. To detect these instances, abnormal bond lengths are used as an indicator of a threaded molecule. Once identified, the two atom groups connected via the abnormal bond are translated by three times the normalized vector between the backbone of the residue and the center of the two atom groups. This translation is sufficient to circumvent the threaded residue. The system is subsequently minimized to correct any distortions and reassessed (Figure 2B).

Equilibration. There are a number of levels of equilibration implemented by CG2AT2, depending on the required output. In the simplest and fastest instance, a non-equilibrated atomistic system may be produced by using the flag -o none. This means that the system is only energy-minimized as part of the conversion, following the reintegration of the complete system. Beyond this, the default level of equilibration for CG2AT2 is to relax the system with a 5 ps NVT simulation with position restraints applied to the backbone Cα atoms of the protein, if included.

Protein Conversion Types. To establish the atomistic protein coordinates, CG2AT2 enables both a *de novo* protocol, where the atoms are constructed from the CG beads, and/or

an aligned approach, where a supplied atomistic protein structure is fitted with the CG protein coordinates. The decision of which approach to use depends upon whether one wishes to retain any protein conformational changes that have arisen from the CG simulation. The *de novo* method is also well suited to retain protein interactions with non-protein groups such as lipids and minor conformational changes. Conversely, if particular atomic interactions are important, for example, the orientation of the backbone within the selectivity filter of a potassium channel, or if the CG simulation is being used to establish an optimal bilayer as a starting point for an atomistic simulation, the aligned method should be used. CG2AT2 allows for a range of conversion types for the user; the choice of which to use depends upon the hypotheses being addressed by the simulation. Upon successful conversion, a RMSD value comparing the backbone of each CG monomer to the converted atomistic protein is provided along with a summary of the atomistic system.

The *de novo* method is reliant on the coordinates of the CG beads. To improve the secondary structure, an additional correction is applied to the amide bond, after the fragment alignment. In this correction, the amide bond is aligned to the cross-vector of the two proceeding backbone beads, as originally outlined by Wassenaar *et al.*[13] This enables the majority of the backbone hydrogen bond network to be recovered. Conversely, if a higher degree of similarity to the atomistic input structure is required, the backbone hydrogen bond network within the supplied atomistic structure is mapped. This network is used as distance restraints to guide the *de novo* NVT equilibration and thereby aid in the recovery of the secondary structure lost during the CG simulation. This allows the production of either an independent *de novo* or

guided *de novo* structure. The latter may be implemented using the flag -disre and an input -a atomistic protein coordinate file (Figure S1).

The aligned method requires a supplied atomistic structure, using the -a flag, and builds upon the *de novo* conversion. With the *de novo* method, the user is able to select the degree of information taken from the CG system. CG2AT2 contains four alignment methods for the supplied atomistic chains: individual atomistic chains (default), user-defined grouped atomistic chains, whole atomistic protein, or grouped based on the CG subunits (Figure 2C). Initially, the supplied structure is separated into individual chains, identified by a greater than 4 Å amide bond length. The amino acid sequence of each atomistic chain is then indexed to the CG protein sequence. Depending upon the choice of grouping, the atomistic backbone is fitted to the CG model. If only a partial structure is detected during the sequence alignment, CG2AT2 implements a hybrid build approach, that is, combining *de novo* and aligned protein structures (Figure 2D). In this instance, any residues missing from the input atomistic structure but within the CG protein are incorporated from the *de novo* method. This is particularly useful for adding flexible, unfolded residues to a protein, for example, growing residues onto a core CG protein to add lipoprotein tethers.[23]

The energy-minimized *de novo* non-hydrogen atoms are steered onto the aligned protein coordinates generated by the aforementioned structural alignment. The steering is undertaken in a multi-step process of increasing harmonic restraints for 2 ps each ($100-10,000$ kJ mol^{-1} nm^{-2}). This methodology preserves interactions between protein and non-protein residues with minimal perturbation and compares favorably with other embedding methodologies, such as g_membed, alchembed, inflategro, and the original CG2AT,[16,24−26] in retaining protein−lipid interactions from CG. To minimize non-protein molecule displacement during the steered MD, an additional positional restraint is applied to the non-protein and non-hydrogen atoms, as described by the supplied position restraint file.

Macromolecular Systems. We tested CG2AT2 upon a range of increasingly complex systems from plain lipid bilayers to multi-membrane-spanning protein complexes. The systems were all taken from ongoing projects within our research group using a mixture of Martini versions 2.0,[27] 2.2,[28,29] and 3.0 (beta).

For all CG Martini simulations, an elastic network of 1000 kJ mol^{-1} nm^{-2} was applied between all backbone beads between 0.5 and 1 nm. Electrostatics were described using the reaction field method, with a cutoff of 1.1 nm using the potential shift modifier, and the van der Waals interactions were shifted between 0.9 and 1.1 nm. The temperature and pressure were kept constant throughout the simulation at 310 K and 1 bar, respectively, with protein, lipids, and water/ions coupled separately to temperature baths by the V-rescale method[30] and a semi-isotropic Parrinello−Rahman barostat.[31]

In the atomistic simulations, electrostatics were described using PME, with a cutoff of 1.2 nm, and the van der Waals interactions were shifted between 1 and 1.2 nm. The TIP3P water model was used; the water bond angles and distances were constrained using SETTLE.[32] All other bonds were constrained using the LINCS algorithm.[33] The production simulations were run in an NPT ensemble with temperature V-rescale coupling at 310 K with protein, lipids, and water/ions coupled individually and semi-isotropic Parrinello−Rahman

barostat at 1 bar. All CGMD systems contain a salt concentration of 0.15 M NaCl. After 1 μs of CGMD simulation, the CG systems were converted to a CHARMM36 atomistic form. To access the stability of the CG2AT conversions, the systems were equilibrated for 1 ns with 1000 kJ mol^{-1} nm^2 position restraints applied to the Cα atoms after their conversion. A further 100 ns unrestrained MD simulation was then performed on each atomistic system. The following systems are ordered by system size from small ($\sim$37,000 atoms) to large (>2,000,000 atoms).

System 1—a POPC Membrane. A CG bilayer of 128 POPC lipids was solvated with 1682 Martini water beads with a total system size of 3256 beads. By comparison, the converted atomistic system has 37,830 atoms.

System 2—a Mixed-Lipid Bilayer. A bilayer consisting of 80 POPC, 10 palmitoyl-oleoyl-phosphatidylserine (POPS), and 32 cholesterol (CHOL) molecules was solvated with 1778 Martini water beads, creating a total system size of 3170 beads. The converted atomistic system has 36,362 atoms.

System 3—SRY. The sex-determining region Y (SRY) protein recognizes and binds to double-stranded nucleotides. We used the *Homo sapiens* structure (PDB ID: 1J46[34]) to create the CG representation and for CG2AT2 conversion. The system consists of a CG tetrameric SRY with 13 DNA base pairs, solvated with 18,661 Martini water beads, yielding a system size of 4640 beads. The atomistic system contains 53,425 atoms.

System 4—δ-Opioid. The class A G-protein-coupled receptor (δ-opioid) activates the β-arrestin and G-protein pathways in the central nervous system. We used the *H. sapiens* structure (PDB ID: 4N6H[35]) to create the CG representation and for CG2AT2 conversion. The system consists of a CG δ-opioid and 237 POPC molecules solvated with 3743 Martini water beads, yielding a system size of 7338 beads. The converted atomistic system contains 82,569 atoms.

System 5—Lgt. The lipoprotein diacylglyceryl transferase (Lgt) transfers the diacylglyceryl group from a phospholipid to the thiol group of a conserved cysteine residue of the prolipoprotein. We used the *Escherichia coli* model (PDB ID: 5AZC[36]) to create the CG representation and for CG2AT2 conversion. The system consists of a CG Lgt, 74 POPG, and 290 palmitoyl-oleoyl-phosphatidylethanolamine (POPE) molecules solvated with 5926 Martini water beads, yielding a system size of 11,674 beads. The converted atomistic system contains 126,028 atoms.

System 6—LspA. The lipoprotein signal peptidase (LspA) catalyzes the removal of signal peptides from prolipoproteins. We used the *Staphylococcus aureus* model (PDB ID: 6RYO[37]) to create the CG representation and for CG2AT2 conversion. The system consists of a CG LspA, 74 POPG, and 290 POPE molecules solvated with 7399 Martini water beads, producing a system size of 12,928 beads. The system was converted to an atomistic system of 142,690 atoms.

System 7—Lnt. The lipoprotein N-acyl transferase (Lnt) catalyzes the phospholipid-dependent N-acylation of the N-terminal cysteine of apolipoproteins. We used the *E. coli* model (PDB ID: 5N6H[38]) to create the CG representation and for CG2AT2 conversion. The system consists of a CG Lnt with 75 POPG and 288 POPE molecules, solvated with 7840 Martini water beads and a total system size of 14,091 beads. The converted atomistic system has 153,299 atoms.

System 8—LolB. This protein is an outer membrane lipoprotein receptor and receives a mature lipoprotein from

Table 1. Similarities between the Martini Backbone and Cα Atom of Each Residue after End 1 and 2^a

system	protein monomers	de novo vs CG		de novo NVT + 100 ns sim	aligned vs CG	aligned vs input AT	aligned + 100 ns sim
		MIN	NVT				
SRY (3)	1	0.072	0.093	0.358	0.352	0.012	0.196
δ-opioid (4)	1	0.089	0.102	0.266	0.162	0.009	0.212
Lgt (5)	1	0.090	0.164	0.323	0.192	0.010	0.255
LspA (6)	1	0.088	0.150	0.293	0.159	0.011	0.230
Lnt (7)	1	0.088	0.127	0.183	0.098	0.010	0.198
LolB (8)	1	0.088	0.154	0.209	0.181	0.011	0.369
Kir2.2 (9)	4	0.087	0.105	0.386	0.152	0.010	0.261
WzmWzt (10)	4	0.087	0.147	0.329	0.187	0.013	0.225
Lpt (11)	11	0.090	0.217	n/a	0.305	0.011	n/a

^{a}The final RMSDs of the Cα atoms after 100 ns of atomistic simulation are reported in nm.
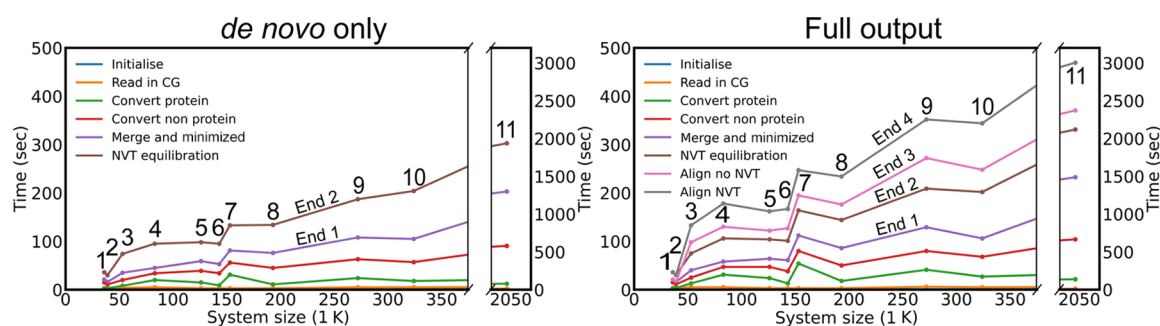

Figure 3. Timings for the conversion from CG to the CHARMM36 atomistic representation. The data show the cumulative time plotted against the final atomistic size of each system. End 1 corresponds to the minimized de novo system, end 2 corresponds to the de novo after which a short NVT simulation has been undertaken, end 3 corresponds to the production of the aligned structure without NVT equilibration, and end 4 corresponds to the production of the aligned structure with an NVT equilibration. The conversion was run on an i7-6800K Intel processor, GeForce GTX 1080 GPU, and 64 Gb of RAM using GROMACS 2020 with the parallelization of CG2AT2 limited to 8 threads.

LolA for delivery to the outer membrane of Gram-negative bacteria. We used the *E. coli* model (PDB ID: 1IWM[39]) to create the CG representation and for CG2AT2 conversion. The system consists of a CG LolB with 157 POPG and 363 POPE molecules solvated with 9736 Martini water beads with a salt concentration of 0.15 M NaCl, yielding a system size of 17,529 beads. The atomistic system consists of 193,138 atoms.

System 9—Kir2.2. Kir2.2 is an inward-rectifying potassium channel. We used the *Gallus gallus* model (PDB ID: 3SPI[40]) to create the CG representation and for CG2AT2 conversion. The system consists of a CG tetrameric Kir2.2 with 4 phosphatidylinositol-4,5-bisphosphate (PIP$_2$) and 465 POPC molecules solvated with 14,960 Martini water beads with a total system size of 24,272 beads. The atomistic system contains 272,230 atoms.

System 10—Wzm-Wzt. Wzm-Wzt transports fully elongated, lipid-linked lipopolysaccharide (LPS) O-antigens across the inner membrane, using an undecaprenyl-linked carrier. We used the *Aquifex aeolicus* structure (PDB ID: 6M96[41]) to create the CG representation and for CG2AT2 conversion. The system consists of a CG tetrameric Wzm-Wzt with 125 POPG and 504 POPE molecules solvated with 18,661 Martini water beads with therefore a system size of 28,938 beads. The atomistic system contains 324,468 atoms.

System 11—Lpt Complex. The LPS transport (Lpt) protein complex spans the periplasm of Gram-negative bacteria and shuttles LPS molecules from the inner to outer membrane. We used *E. coli* structures and homology models (LptA: 2R19,[42] LptC: 3MY2,[43] LptB2FG: Swiss-model[44] based on 6MIT,[45] and LptDE: Swiss-model based on 4Q35[46]) to create the CG representation and for CG2AT2 conversion. The

system consists of a CG Lpt complex composed of 4 LptA, 2 LptB, 1 LptC, 1 LptD, 1 LptE, 1 LptG, and 1 LptF with 147 LPSs, 284 POPG, and 1150 POPE molecules solvated with 133,459 Martini water beads with a total CG system size of 181,263 beads. The atomistic system has 2,039,940 atoms.

RESULTS

To assess the quality of the CG to atomistic conversion, CG2AT2 was evaluated against 11 macromolecular systems of varying size and complexity. To demonstrate the initial accuracy, RMSDs were calculated for each de novo endpoint against the input CG structure; this yielded excellent conversion RMSDs of less than 0.09 nm. The RMSD between the aligned structures to the CG demonstrates the fluctuations of the CG structure over time. Therefore, the RMSD between the aligned and supplied atomistic structures was measured. This yielded RMSDs of less than 0.013 nm, within the range of a standard minimized system. These results are summarized within Table 1.

For CG2AT2 to be included in a workflow, the conversion must ideally be undertaken in a high-throughput manner. While the conversion time is highly dependent upon the specifications of the machine used to run CG2AT2, we provide a set of benchmarks as a guide for the user. Figure 3 shows the cumulative time required for each step in the conversion with the possible endpoints for the conversion shown. CG2AT2 takes roughly 4 min for a full conversion of a 100 K atom system, producing an equilibrated de novo and aligned system. As expected, the majority of the time to convert the system is taken by the MD equilibration steps, namely, the NVT and the multi-step steering.
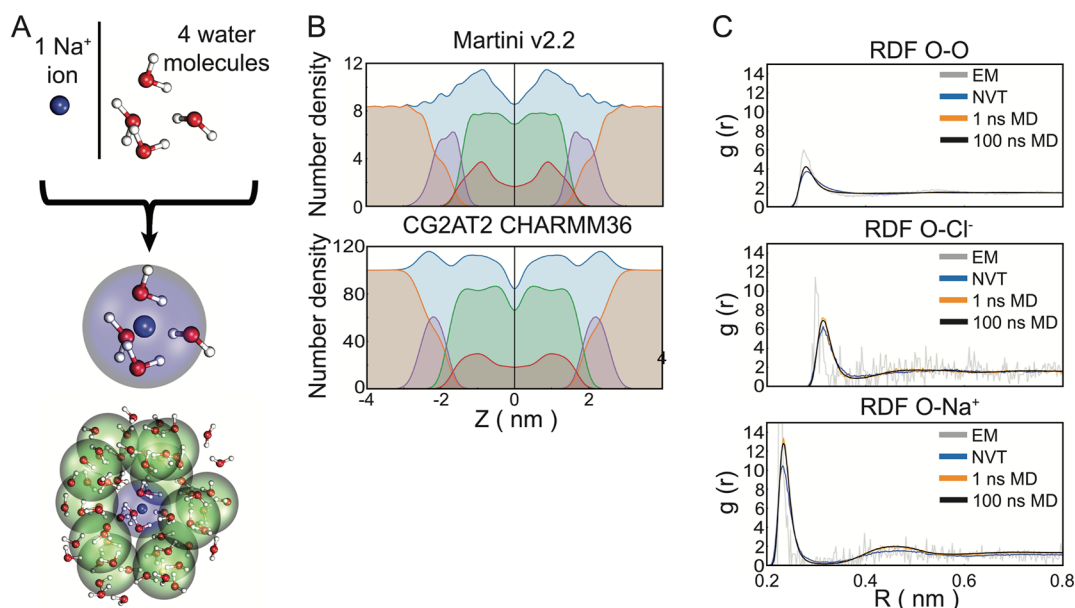
Figure 4. Solvent conversion. (A) Ion conversion; a tetrahedral arrangement of water is placed over a single ion located at the center of the CG ion bead. A zoomed-out view of a Na^+ ion surrounded by a cluster of CG water molecules. (B) Number density profiles of the complex membrane system. Colors: system (blue), water and ions (orange), sterols (red), phospholipid tails (green), and the phospholipid headgroup (purple). (C) O, Cl^-, and Na^+ atom RDFs referenced to the TIP3P oxygen atom. EM: complete system energy minimization by the steepest descent algorithm. *NVT*: optional unrestrained 5 ps *NVT* equilibration. 1 ns MD: an initial 1 ns position-restrained *NPT* equilibration. 100 ns MD: a 100 ns *NPT* production run.

Solvent. A significant update to CG2AT now allows for the full conversion of the CG systems, including water molecules and ions. Within the new implementation of solvent conversion, CG2AT2 places a randomly rotated arrangement of solvent molecules over each CG solvent bead. The solvent fragment can contain any number of molecules in any configuration. Here, we utilize a fragment containing four water molecules in a roughly tetrahedral arrangement taken from an equilibrated atomistic MD system. Additional solvent molecules/models can easily be added as additional solvent fragments. As CG ions are treated as solvated, the solvent fragment is overlaid over the CG ion coordinate. This process is demonstrated within Figure 4. The similarity of the CG to AT solvent bulk properties is demonstrated in Figure 4B. Here, we see similar density profiles in the complex membrane before and after conversion. To assess the quality of the solvent conversion, we analyzed the radial distribution function (RDF) of the water oxygen atom against Na^+, Cl^-, and itself (Figure 4C). The quality of the initial solvent conversion is demonstrated by the minimal shift in the peak RDF values of less than 0.014 nm between initial minimization and equilibrium values, demonstrating the need for minimal equilibration after conversion.

Lipids. An advantage of converting CG simulations to atomistic representation is the retention of the equilibrated bilayer properties, such as the correctly mixed fluid-phase lipid bilayer. We tested a conversion of a complex bilayer composed of POPC, POPS, and CHOL at a ratio of 8:1:3 solvated with 0.15 M NaCl. An identical system was created by the webserver CHARMM-GUI as a direct comparison.[47] Both systems were equilibrated for 1 ns in the *NPT* ensemble before production runs. The area per lipid (APL) was compared between the CG, CG2AT2, and CHARMM-GUI. Between the CG2AT2 conversion and CHARMM-GUI membrane, the APLs of the lipid components are in excellent agreement

(Figure S2A). We also calculated the lipid tail order parameter profiles for the AT systems for both bilayers produced by CG2AT2 and CHARMM-GUI (Figure S2B). After 100 ns, the order parameters of both sets of simulations closely match. Interestingly, however, the rate of convergence to the final profile varies between the two methods, with CG2AT2 converging to a mean Δ-S_{CH} at less than 0.005 roughly ~25 ns faster than CHARMM-GUI (Figure S2C).

Assessing the Quality of the Conversion by Evaluating Protein Dynamics. We employed a set of nine protein test systems, spanning a range of secondary structure complexities and architectures. These systems were used to evaluate the conversion procedure of CG2AT2 in terms of the stability and quality of the dynamics over a subsequent 100 ns of atomistic simulation after conversion. In all cases, the majority of the secondary structure elements were recovered from the CG simulations (Figure 5).

LolB was used to demonstrate the hybrid conversion approach (Figure 2D). The β-barrel core of the LolB protein was solved by X-ray crystallography; however, this lacks the N-terminal region, which is anchored to the membrane by a triacyl-cysteine. The N-terminal domain and the lipoprotein tether were then grown onto the β-barrel in the CG representation using the method outlined in Rao et al.[23]

To measure the similarity of the CG and the converted systems, structures were compared using $C\alpha$ and BB RMSDs. A summary of conversion RMSDs is shown in Table 1. Depending on the level of equilibration chosen, there is an increasing degree of conformational drift. The *de novo* conversion supplies a system closest to the initial CG representation, with a post-minimization RMSD of 0.087 ± 0.005 nm compared to 0.140 ± 0.036 nm after *NVT* equilibration. While a comparison between the aligned protein structure and the CG model can be made, this mostly highlights the deviation of the CG conformation from the
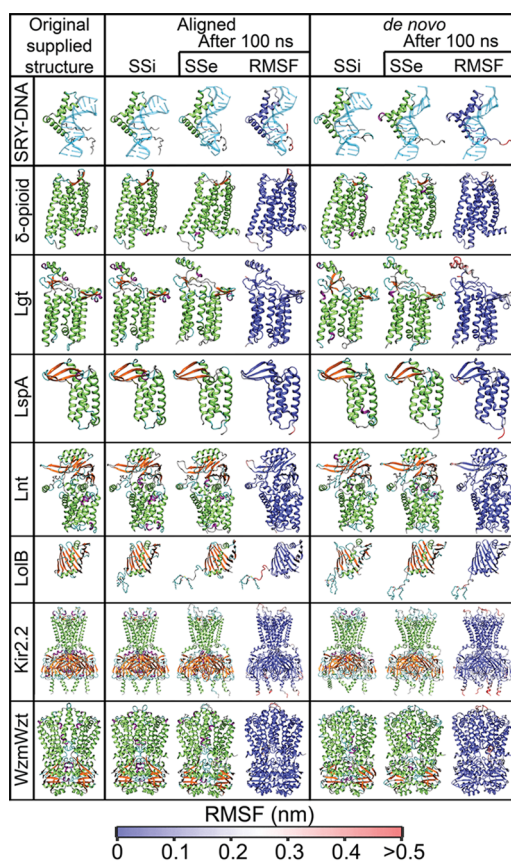
Figure 5. Protein secondary structure and stability. Original supplied structures from the RCSB database (SSi). The protein structure after initial conversion color by the secondary structure (SSe). The protein secondary structure after 100 ns of atomistic simulation, colored by the end secondary structure (RMSF). The protein structure after 100 ns colored by the RMSF, on a blue to red scale, with each monomer aligned individually. The color scheme used for the secondary structure is as follows. Turn: cyan, extended sheet: orange, bridge: tan, α-helix: lime, 3_{10}-helix: purple, π-helix: blue, and coil: silver.

input atomistic structure. A more useful measurement is between the aligned structure generated by the steered MD step and the input atomistic structure. This typically provides an RMSD of 0.011 ± 0.002 nm, highlighting that through using the steered MD, the supplied structure can be accurately reproduced. By applying variations in the aligned RMSD, any clashes between aligned monomers and/or deviations arising due to the hybridization method can be easily detected and highlighted by CG2AT2 (0.015 nm cutoff).

The backbone provides much of the structural stability of the protein. Therefore, we calculated a Ramachandran plot to assess the stereochemical quality of the *de novo* and aligned output of CG2AT2 (Figure S3). As a guide, a good atomic-resolution protein model will generally provide over 90% of residues within the allowed regions.[48] Here, the *de novo* conversion shows an excellent level of quality with 93−99% within the allowed regions with the majority of the test systems above 96%. This is also in excellent agreement to the aligned structures, which were resolved at the atomic resolution.

While the RMSD and Ramachandran plots demonstrate the accuracy and quality of the conversion, they provide little information on the stability of the systems. Therefore, the converted systems were simulated for 100 ns atomistic MD simulations using the standard CHARMM36 force field

settings. The subsequent conformational drift of the protein was analyzed by RMSD, root mean square fluctuation (RMSF), and secondary structure retention (Table 1, Figure 5, Table S1 and Figure S4). In all cases, the structures remained stable throughout the 100 ns simulation. The global deviation of the simulations from their initial configuration is highlighted by the RMSD. The *de novo* simulations demonstrated a slightly larger degree of conformational drift over the course of the simulation with an average RMSD of 0.293 ± 0.066 nm compared to 0.243 ± 0.052 nm for the aligned systems. To pinpoint areas of instability and dynamic regions of the converted system, we evaluated the RMSF calculations. The nine test systems shown here demonstrate comparable RMSFs to the aligned systems, with only the extended loops and short helices showing values greater than 0.2 nm. This is most noticeable within the periplasmic region of Lgt and the N- and C-terminal regions of Kir2.2.

While the tertiary and secondary structures within the Martini simulations are retained *via* the use of elastic networks, minor changes can still occur within the protein structure. We therefore expect to see the majority of the secondary structure to be recovered during the *de novo* conversion, as detailed in the original atomistic protein coordinate file. A qualitative comparison of the secondary structures in Figure 5 demonstrates that a large degree of the secondary structure is recovered and retained over the 100 ns simulation. There are also only minor variations in the secondary structure retention, when using the *de novo* approach, and this is in excellent agreement with the aligned systems. This is further highlighted by the Jaccard indices, which measure the similarity between two sets of data ranging from uncorrelated (0) to correlated (1). Representing the secondary structure correlation to a reference structure, the numbers presented here represent the mean of the simulations. The aligned structures were used as a guide to estimate the secondary structure variability over 100 ns of atomistic simulation, with a mean index of 0.81 ± 0.07. A Jaccard index was calculated for the *de novo* conversion in relation to the input structure to test the ability of CG2AT2 to recover the original secondary structure, with a mean index of 0.69 ± 0.09. A Jaccard index of 0.72 ± 0.07 was also calculated to compare the *de novo* simulation to the *de novo* starting structure. These values demonstrate that the dynamics of the *de novo* conversion are marginally higher than those of the aligned systems. This appears to arise from drift in the β-strands within the CG system. Recovery of short helices can be seen in Figure 5, in particular the LspA and Lgt *de novo* conversions.

Protein−Ligand Interactions. Two of the core advantages of CG simulations are straightforward system configuration, for example, insane and quick equilibration. A number of papers have demonstrated the use of CG simulations to find difficult small-molecule ligand coordination sites, lipid binding sites, and protein−protein interactions.[49] Here, we demonstrate the conversion of the tetrameric Kir2.2 protein coordinated with PIP_2 lipids. To generate a Kir2.2:PIP_2 binding pose, the system was simulated for 1 μs. The final frame from the CG simulation captured two bound PIP_2 molecules, as described previously (Corey et al.[7]). We show qualitatively that the steering process to generate the aligned systems minimally perturbs the binding pattern between the lipid and the protein (Figure 6). The RMSDs between the *de novo* and aligned PIP_2 non-hydrogen atoms shown in binding sites 1 and 2 are 0.25 and 0.29 nm, respectively.
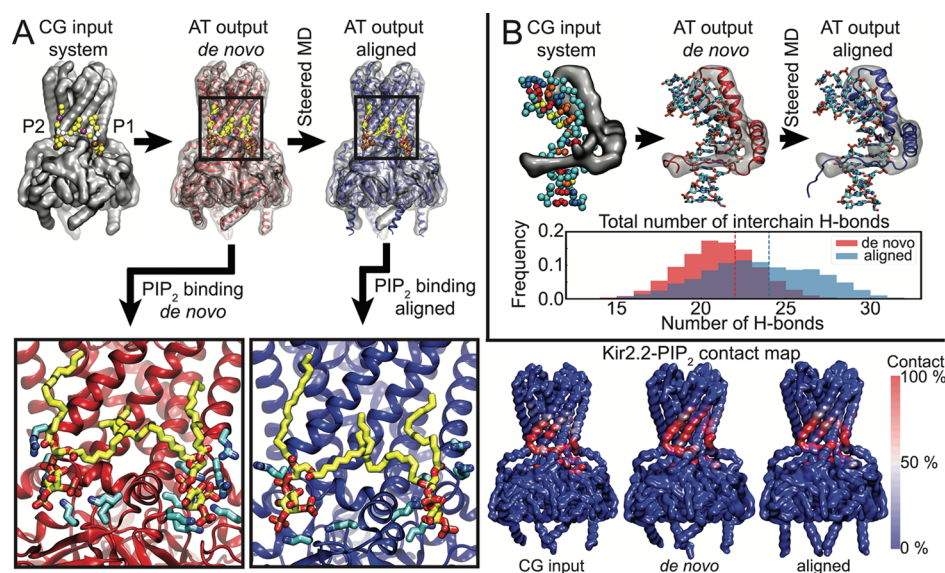
Figure 6. Protein/ligand conversion and retention. (A) Conversion of the Kir2.2 bound to four PIP$_2$ lipids. Shown here are two PIP$_2$ lipids bound in pose 1 and 2. The PIP$_2$ binding sites found in the CG simulations are retained upon both *de novo* and aligned conversions. (Bottom right) The contact maps of PIP$_2$ binding to Kir2.2. The last 100 ns of the CG simulation is shown as a comparison with the 100 ns atomistic simulations. (B) Conversion of the SRY protein in complex with double-stranded DNA. The number of DNA interchain hydrogen bonds over the 100 ns is shown; the dotted line highlights the starting number of hydrogen bonds.

The increased RMSD arises from the inevitable rearrangement of amino acids, leading to small steric clashes and rearrangement of electrostatic interactions between the aligned protein and PIP$_2$. To test the stability of the PIP$_2$ binding pose, the lipid contact maps of the converted systems were compared to those of the CG system. To provide a comparable map, the contacts within the last 100 ns of the CG simulation were compared to those of the 100 ns atomistic simulation. As shown by Figure 6, the contact maps between the three simulations are near-identical, demonstrating the preservation of the protein−ligand interface. The Pearson correlation coefficient between the converted atomistic simulations and the CG simulation provides values of 0.71 and 0.70 for the *de novo* and aligned systems, respectively. The high correlation of 0.77 between the *de novo* and aligned systems demonstrates that little modification has occurred to the protein−PIP$_2$ interactions as a result of the steered MD alignment.

Another complex system conversion problem is protein complexes that contain DNA. Here, we demonstrate the variability of CG2AT2 with the conversion of the SRY protein in a complex with 14 base pairs of double-stranded DNA (Figure 6B). The multimeric non-protein residues are converted in a similar manner through the protein *de novo* conversion with the exception of the further processing of the protein backbone. To analyze the stability of the protein−DNA complex over the course of the simulations, the number of DNA interchain hydrogen bonds was calculated. We show in Figure 6B that the hydrogen bonding pattern remains stable in both *de novo* and aligned atomistic simulations with the total number of H-bonds being 29.0 ± 3.0 and 26.3 ± 2.5, respectively.

■ DISCUSSION

We have described an updated fragment-based approach of CG2AT for the accurate conversion of complex CG representations to atomistic details. Here, we have demonstrated against a variety of systems that CG2AT2 can provide a gradient of conversion approaches from completely *de novo* to aligned, depending upon the degree of protein conformational changes in the CG system the user wishes to retain. A graphical representation of the workflow of CG2AT is demonstrated by the multi-component CG Lpt system, composed of 11 protein monomers and 2 lipid bilayers (1434 lipids and 147 LPS) (Figure 7).

CG2AT2 now performs complete reconstruction of atomistic systems from CG, including water. This is achieved almost entirely within the python-based code, with GROMACS called to generate topologies and to run the equilibration simulations. CG2AT2 can convert complete systems with no pre-processing of the input files, via a supplied comprehensive database of over 107 unique residue types, at the time of publication. The current individual fragment libraries include standard CHARMM36, CHARMM36 with virtual sites, and Stockholm lipids (Slipids)[50] for conversion from CG Martini versions 2 and 3. To enable the variability of the residue types, we have included a range of updated force fields (AMBER99SB-ILDN with Slipids and CHARMM36) containing additional parameters and the base-level force fields. However, it should be noted that all GROMACS-readable force fields are compatible with CG2AT2.

The protocol outlined here generates the key files required for a production run simulation including an equilibrated atomistic system and topology files. This enables the user to implement high-throughput robust and accurate conversion of CG systems with minimal further processing. In combination with tools such as TS2CG,[51] CG2AT2 may enable the conversion of triangulated surfaces all the way down to the full atomic-level detail and thereby permit flexible approaches for converting between molecular systems at multiple levels of granularity.

■ AVAILABILITY

CG2AT2 is available from github, https://github.com/owenvickery/cg2at, and may be added to a Conda installation
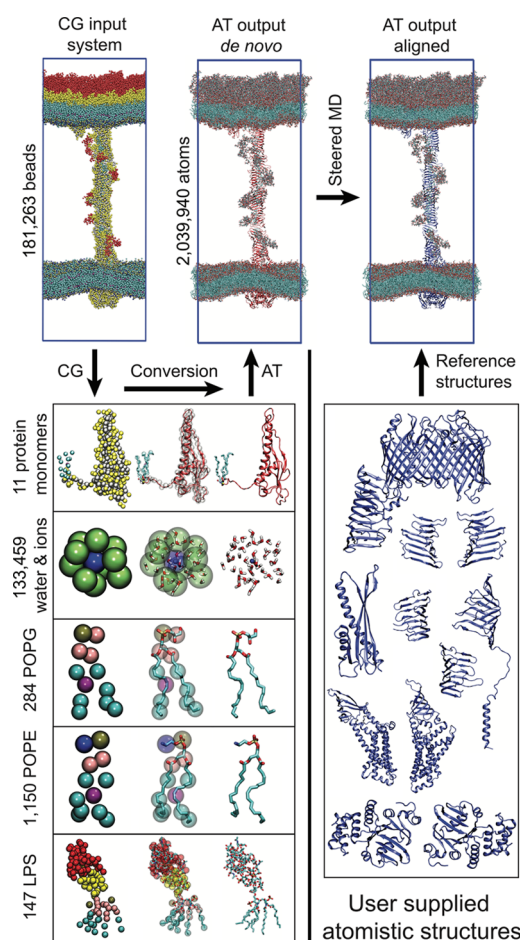
Figure 7. CG2AT2 conversion of the Lpt system to CHARMM36. The Martini CG system (top left) is separated into individual residue types and then converted into their atomistic representation (bottom left). The individual residue types are recombined, providing the *de novo* system before being steered to the reference structures provided (bottom right). For clarity, the solvent has been omitted from the images of the system.

using conda install -c stansfeld_rg cg2at. For full usage instructions, please see https://github.com/owenvickery/cg2at.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00295.

> Plots from atomistic MD simulation for both protein and lipids after conversion detailing the structural stability of the converted protein and the dynamic properties of the lipids (PDF)

AUTHOR INFORMATION

Corresponding Author

Phillip J. Stansfeld − *School of Life Sciences & Department of Chemistry, University of Warwick, Coventry CV4 7AL, U.K.;* orcid.org/0000-0001-8800-7669; Phone: +44 (0) 2476523864; Email: phillip.stansfeld@warwick.ac.uk

Author

Owen N. Vickery − *School of Life Sciences & Department of Chemistry, University of Warwick, Coventry CV4 7AL, U.K.*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.1c00295

REFERENCES

(1) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E. H.; Britto, R.; Bursteinas, B.; Bye-A-Jee, H.; Coetzee, R.; Cukura, A.; Silva, A. D.; Denny, P.; Dogan, T.; Ebenezer, T. G.; Fan, J.; Castro, L. G.; Garmiri, P.; Georghiou, G.; Gonzales, L.; Hatton-Ellis, E.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Jokinen, P.; Joshi, V.; Jyothi, D.; Lock, A.; Lopez, R.; Luciani, A.; Luo, J.; Lussi, Y.; MacDougall, A.; Madeira, F.; Mahmoudy, M.; Menchi, M.; Mishra, A.; Moulang, K.; Nightingale, A.; Oliveira, C. S.; Pundir, S.; Qi, G.; Raj, S.; Rice, D.; Lopez, M. R.; Saidi, R.; Sampson, J.; Sawford, T.; Speretta, E.; Turner, E.; Tyagi, N.; Vasudev, P.; Volynkin, V.; Warner, K.; Watkins, X.; Zaru, R.; Zellner, H.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Blatter, M. C.; Bolleman, J.; Boutet, E.; Breuza, L.; Casals-Casas, C.; de Castro, E.; Echioukh, K. C.; Coudert, E.; Cuche, B.; Doche, M.; Dornevil, D.; Estreicher, A.; Famiglietti, M. L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Keller, G.; Kerhornou, A.; Lara, V.; Le Mercier, P.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T. B.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pourcel, L.; Pozzato, M.; Pruess, M.; Rivoire, C.; Sigrist, C.; Sonesson, K.; Stutz, A.; Sundaram, S.; Tognolli, M.; Verbregue, L.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Garavelli, J. S.; Huang, H.; Laiho, K.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Q.; Wang, Y.; Yeh, L. S.; Zhang, J.; Ruch, P.; Teodoro, D. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480−D489.

(2) Wallin, E.; Heijne, G. V. Genome-Wide Analysis of Integral Membrane Proteins from Eubacterial, Archaean, and Eukaryotic Organisms. *Protein Sci.* **1998**, *7*, 1029−1038.

(3) Koehler Leman, J.; Ulmschneider, M. B.; Gray, J. J. Computational Modeling of Membrane Proteins. *Proteins: Struct., Funct., Bioinf.* **2015**, *83*, 1−24.

(4) Newport, T. D.; Sansom, M. S. P.; Stansfeld, P. J. The MemProtMD Database: A Resource for Membrane-Embedded Protein Structures and Their Lipid Interactions. *Nucleic Acids Res.* **2019**, *47*, D390−D397.

(5) Gahbauer, S.; Böckmann, R. A. Comprehensive Characterization of Lipid-Guided G Protein-Coupled Receptor Dimerization. *J. Phys. Chem. B* **2020**, *124*, 2823–2834.

(6) Esteban-Martín, S.; Risselada, H. J.; Salgado, J.; Marrink, S. J. Stability of Asymmetric Lipid Bilayers Assessed by Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2009**, *131*, 15194–15202.

(7) Corey, R. A.; Vickery, O. N.; Sansom, M. S. P.; Stansfeld, P. J. Insights into Membrane Protein-Lipid Interactions from Free Energy Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 5727–5736.

(8) Corradi, V.; Mendez-Villuendas, E.; Ingólfsson, H. I.; Gu, R.-X.; Siuda, I.; Melo, M. N.; Moussatova, A.; DeGagné, L. J.; Sejdiu, B. I.; Singh, G.; Wassenaar, T. A.; Delgado Magnero, K.; Marrink, S. J.; Tieleman, D. P. Lipid-Protein Interactions Are Unique Fingerprints for Membrane Proteins. *ACS Cent. Sci.* **2018**, *4*, 709–717.

(9) Naughton, F. B.; Kalli, A. C.; Sansom, M. S. P. Association of Peripheral Membrane Proteins with Membranes: Free Energy of Binding of GRP1 PH Domain with Phosphatidylinositol Phosphate-Containing Model Bilayers. *J. Phys. Chem. Lett.* **2016**, *7*, 1219–1224.

(10) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898–7936.

(11) Sengupta, D.; Prasanna, X.; Mohole, M.; Chattopadhyay, A. Exploring GPCR-Lipid Interactions by Molecular Dynamics Simulations: Excitements, Challenges, and the Way Forward. *J. Phys. Chem. B* **2018**, *122*, 5727–5737.

(12) Rzepiela, A. J.; Schäfer, L. V.; Goga, N.; Risselada, H. J.; De Vries, A. H.; Marrink, S. J. Reconstruction of Atomistic Details from Coarse-Grained Structures. *J. Comput. Chem.* **2010**, *31*, 1333–1343.

(13) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **2014**, *10*, 676–690.

(14) Stieffenhofer, M.; Bereau, T.; Wand, M. Adversarial Reverse Mapping of Condensed-Phase Molecular Structures: Chemical Transferability. *APL Mater.* **2021**, *9*, 031107.

(15) An, Y.; Deshmukh, S. A. Machine Learning Approach for Accurate Backmapping of Coarse-Grained Models to All-Atom Models. *Chem. Commun.* **2020**, *56*, 9312–9315.

(16) Stansfeld, P. J.; Sansom, M. S. P. From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 1157–1166.

(17) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.

(18) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2016**, *14*, 71–73.

(19) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins* **2010**, *78*, 1950–1958.

(20) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(21) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(22) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous Parallelization and Acceleration of Molecular Dynamics Simulations in GROMACS. *J. Chem. Phys.* **2020**, *153*, 134110.

(23) Rao, S.; Bates, G. T.; Matthews, C. R.; Newport, T. D.; Vickery, O. N.; Stansfeld, P. J. Characterizing Membrane Association and Periplasmic Transfer of Bacterial Lipoproteins through Molecular Dynamics Simulations. *Structure* **2020**, *28*, 475–487.

(24) Wolf, M. G.; Hoefling, M.; Aponte-Santamaría, C.; Grubmüller, H.; Groenhof, G. G-Membed: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. *J. Comput. Chem.* **2010**, *31*, 2169–2174.

(25) Kandt, C.; Ash, W. L.; Peter Tieleman, D. Setting up and Running Molecular Dynamics Simulations of Membrane Proteins. *Methods* **2007**, *41*, 475–488.

(26) Jefferys, E.; Sands, Z. A.; Shi, J.; Sansom, M. S. P.; Fowler, P. W. Alchembed: A Computational Method for Incorporating Multiple Proteins into Complex Lipid Geometries. *J. Chem. Theory Comput.* **2015**, *11*, 2743–2754.

(27) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(28) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.

(29) 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. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687–697.

(30) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(31) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(32) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(33) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(34) Murphy, E. C.; Zhurkin, V. B.; Louis, J. M.; Cornilescu, G.; Clore, G. M. Structural Basis for SRY-Dependent 46-X,Y Sex Reversal: Modulation of DNA Bending by a Naturally Occurring Point Mutation. *J. Mol. Biol.* **2001**, *312*, 481–499.

(35) Fenalti, G.; Giguere, P. M.; Katritch, V.; Huang, X.-P.; Thompson, A. A.; Cherezov, V.; Roth, B. L.; Stevens, R. C. Molecular Control of δ-Opioid Receptor Signalling. *Nature* **2014**, *506*, 191–196.

(36) Mao, G.; Zhao, Y.; Kang, X.; Li, Z.; Zhang, Y.; Wang, X.; Sun, F.; Sankaran, K.; Zhang, X. C. Crystal Structure of E. Coli Lipoprotein Diacylglyceryl Transferase. *Nat. Commun.* **2016**, *7*, 1–12.

(37) Olatunji, S.; Yu, X.; Bailey, J.; Huang, C.-Y.; Zapotoczna, M.; Bowen, K.; Remškar, M.; Müller, R.; Scanlan, E. M.; Geoghegan, J. A.; Olieric, V.; Caffrey, M. Structures of Lipoprotein Signal Peptidase II from Staphylococcus Aureus Complexed with Antibiotics Globomycin and Myxovirescin. *Nat. Commun.* **2020**, *11*, 140.

(38) Wiktor, M.; Weichert, D.; Howe, N.; Huang, C.-Y.; Olieric, V.; Boland, C.; Bailey, J.; Vogeley, L.; Stansfeld, P. J.; Buddelmeijer, N.; Wang, M.; Caffrey, M. Structural Insights into the Mechanism of the Membrane Integral N-Acyltransferase Step in Bacterial Lipoprotein Synthesis. *Nat. Commun.* **2017**, *8*, 25–28.

(39) Takeda, K.; Miyatake, H.; Yokota, N.; Matsuyama, S. I.; Tokuda, H.; Miki, K. Crystal Structures of Bacterial Lipoprotein Localization Factors, LolA and LolB. *EMBO J.* **2003**, *22*, 3199–3209.

(40) Hansen, S. B.; Tao, X.; MacKinnon, R. Structural Basis of PIP2 Activation of the Classical Inward Rectifier K+ Channel Kir2.2. *Nature* **2011**, *477*, 495–498.

(41) Caffalette, C. A.; Corey, R. A.; Sansom, M. S. P.; Stansfeld, P. J.; Zimmer, J. A Lipid Gating Mechanism for the Channel-Forming O Antigen ABC Transporter. *Nat. Commun.* **2019**, *10*, 1–11.

(42) Suits, M. D. L.; Sperandeo, P.; Dehò, G.; Polissi, A.; Jia, Z. Novel Structure of the Conserved Gram-Negative Lipopolysaccharide Transport Protein A and Mutagenesis Analysis. *J. Mol. Biol.* **2008**, *380*, 476–488.

(43) Tran, A. X.; Dong, C.; Whitfield, C. Structure and Functional Analysis of LptC, a Conserved Membrane Protein Involved in the Lipopolysaccharide Export Pathway in Escherichia Coli. *J. Biol. Chem.* **2010**, *285*, 33529–33539.

(44) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46*, W296−W303.

(45) Owens, T. W.; Taylor, R. J.; Pahil, K. S.; Bertani, B. R.; Ruiz, N.; Kruse, A. C.; Kahne, D. Structural Basis of Unidirectional Export of Lipopolysaccharide to the Cell Surface. *Nature* **2019**, *567*, 550−553.

(46) Qiao, S.; Luo, Q.; Zhao, Y.; Zhang, X. C.; Huang, Y. Structural Basis for Lipopolysaccharide Insertion in the Bacterial Outer Membrane. *Nature* **2014**, *511*, 108−111.

(47) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.

(48) Laskowski, R. A.; MacArthur, M. W.; Thornton, J. M.; Bank, P. D. Validation of protein models derived from experiment. *Curr. Opin. Struct. Biol.* **1998**, *8*, 631−639.

(49) Souza, P. C. T.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein−Ligand Binding with the Coarse-Grained Martini Model. *Nat. Commun.* **2020**, *11*, 1−11.

(50) Grote, F.; Lyubartsev, A. P. Optimization of Slipids Force Field Parameters Describing Headgroups of Phospholipids. *J. Phys. Chem. B* **2020**, *124*, 8784−8793.

(51) Pezeshkian, W.; König, M.; Wassenaar, T. A.; Marrink, S. J. Backmapping Triangulated Surfaces to Coarse-Grained Membrane Models. *Nat. Commun.* **2020**, *11*, 2296.