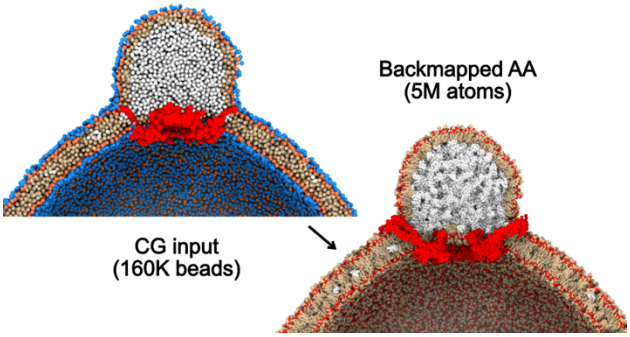
Backmapping with Mapping and Isomeric Information

Siyoung Kim[1]
[1]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637 USA.

ABSTRACT

I present a powerful and flexible backmapping tool named Multiscale Simulation Tool (`mstool`) that converts a coarse-grained (CG) system into all-atom (AA) resolution and only requires AA to CG mapping and isomeric information (*cis/trans/chiral*). The backmapping procedure includes two simple steps: a) the AA atoms are randomly placed near the corresponding CG beads according to the provided mapping scheme. b) the energy minimization is performed with two modifications in the forcefield. First, the nonbonded interactions are replaced with a cosine function to ensure numerical stability. Second, additional dihedrals are imposed to maintain the molecule's isomeric properties. To test the tool's simplicity and robustness, I have demonstrated a backmapping of the four-bead CG lipid model into the corresponding AA resolution (resolution increased by a factor of 34) without using an intermediate resolution. The tool is freely available at github.com/ksy141/mstool.

TABLE OF CONTENTS (TOC)



INTRODUCTION

The CG molecular dynamics (MD) simulations have become a popular method in studying biological processes with time and length scales beyond the reach of the AA-MD simulations,[1–5] such as large-scale membrane deformation, organelle biogenesis, and the complete virion or capsid of a virus.[6–9] In this context, resolution transformation from CG to AA, or backmapping, is a necessary step to obtain detailed atomistic interactions between the molecules, a level of detail not available in the CG resolution. The CG MD and backmapping can complement each other: for example, one can perform a preliminary CG MD to navigate the system's equilibrium distribution followed by additional AA MD, using initial structure obtained from backmapping, to understand the system at a finer resolution.

While it is straightforward to map AA models into CG models, the reverse requires recovery of the degree of freedom that has been integrated away. The fragment-based

reconstruction, a class of backmapping method, replaces each CG bead with the corresponding atomistic fragment or arrangement.[10,11] While this is a convenient and powerful method, it requires an AA structure and/or predefined database for fragments, which should be updated once a new type of CG is introduced. Furthermore, if an input CG system is highly CG, the initial AA structure before relaxation may have many unphysical contacts, leading to energy minimization failure.

Wassenaar et al. have devised the geometrical approach in which the atoms are positioned using the local geometrical information, provided by the user, before the relaxation step.[12] Such approach gives structures that are consistent with the user-provided bonded and isomeric properties before relaxation. However, this approach requires a user to carefully define multiple sets of local geometrical arrangements for a molecule, which makes it challenging to use. Furthermore, even though the initial positions of atoms may preserve the local geometrical information, the initial energy and force values may still be large, which can result flipped isomeric properties for some molecules after energy minimization.

The above approaches (fragment-based reconstruction or geometrical projection) have focused on the construction of the initial AA structures from CG structures, followed by standard energy minimization. In this paper, I present a backmapping approach that uses a modified forcefield for energy minimization. It is powerful enough that the initial positions of atoms can be random without a need for fragment alignment or geometrical projection, greatly simplifying the user input into a minimal set of information: mapping scheme and isomer properties. The tool replaces the nonbonded interactions with a cosine function to ensure the numerical stability of energy minimization while keeping the bonded interactions intact. In other words, the system is relaxed with higher priority given on the bonded interactions while the cosine nonbonded interactions ensure that no atoms are overlapping. In addition, a set of dihedral terms is imposed during relaxation to maintain the provided isomeric properties of the molecules (*cis/trans/chiral*). While the idea is simple, the tool is powerful and capable of constructing a complete AA structure from a highly CG system of a large size.

The rest of the article is structured as follows. I will first describe the modifications on the forcefield used during energy minimization. Then, I will present the backmapping of lipid and/or protein systems. All the examples described in this paper are available at the Github link with a step-by-step guideline.

METHODS

For each type of molecule (residue) present in the CG structure, the mapping scheme describes which AA atoms belong to which CG beads. It also should include the molecule's isomeric properties such as *cis*, *trans*, and/or *chiral*. The predefined mapping files based on the Martini forcefield are read if no user mapping is provided,[13–16] located at `$mstool/mapping/martini.protein.c36m.dat` and `$mstool/mapping/martini.lipid.c36.dat`.

The backmapping procedure consists of three steps: Ungrouping, energy minimization, and checking a backmapped structure. The first step places the AA atoms near the corresponding CG bead according to the mapping scheme, implemented as `mstool.Ungroup` in the `mstool` package. The output of this step is an AA structure, but not at the energy minima as the AA atoms are clustered at the locations of the CG beads.

The next step is called Reduced Nonbonded Energy Minimization (REM) and implemented as `mstool.REM` in the `mstool` package. During REM, the energy of the system is minimized with a modified forcefield. First, CHARMM36m protein[17] and CHARMM36 lipid[18] forcefield are applied to the output of `mstool.Ungroup`. Then, the tool replaces the positive part of the LJ potential (U > 0) with the cosine repulsion interactions, expressed as

$$U_{LJ}(r) = \min\left[Acos^2\left(\frac{\pi}{2\sigma}r\right), 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)\right] \quad (1)$$

The Coulomb interaction is described by

$$U_c(r) = Cq_iq_j\cos^2\left(\frac{\pi}{2r_c}r\right) \quad (2)$$

By default, $A = 100\ kJ/mol$, $C = 50\ kJ/mol$. Nonbonded interactions are cut-offed at $r_c = 1.2\ nm$. Eqs 1 and 2 are illustrated in Fig. 1A. These soft nonbonded interactions ensure the energy and force values do not diverge even if atoms are at a very close distance and smoothes the potential energy surface, making energy minimization easier.

In addition to using modified nonbonded interactions, additional dihedral terms are applied during REM to maintain the provided isomeric properties of molecules, which should be written in the mapping files. A set of improper dihedral terms are applied for *chiral* (Fig. 1B). A single improper dihedral term is applied for each *cis* or *trans* (Fig. 1C). The force constant, K, is set to $300\ kJ/mol/rad^2$, and $\theta_0 = 0^\circ\ (cis)\ or\ 180^\circ\ (trans)$.

Every amino acid except for glycine has a chiral center at the carbon alpha (CA) atom, and their chirality should be denoted in a mapping scheme. However, a peptide bond, which is dominantly in *trans* conformation, cannot be specified in a mapping scheme because it is a cross-residue property, therefore cannot be defined within a single residue. To prevent any *cis* peptide bonds, `mstool` applies a *trans* dihedral (Fig. 1D) in every peptide bond that it detects during REM. With the modified forcefield, `mstool` uses OpenMM for energy minimization.[19,20]

The last step in the backmapping procedure is checking the resulting AA structures, implemented `mstool.CheckStructure` in the `mstool` package. It reports which isomeric properties are reviewed (written in the mapping schemes) and flipped (inconsistent with initial mapping scheme) isomeric properties.

All the backmapping was performed on a MacBook Pro (16-inch, 2021) with an Apple M1 Pro chip. Ramachandran plots were obtained using MDAnalysis.[21,22] The C36-c parameters were used for triolein.[23] The structures were visualized with ChimeraX.[24,25]

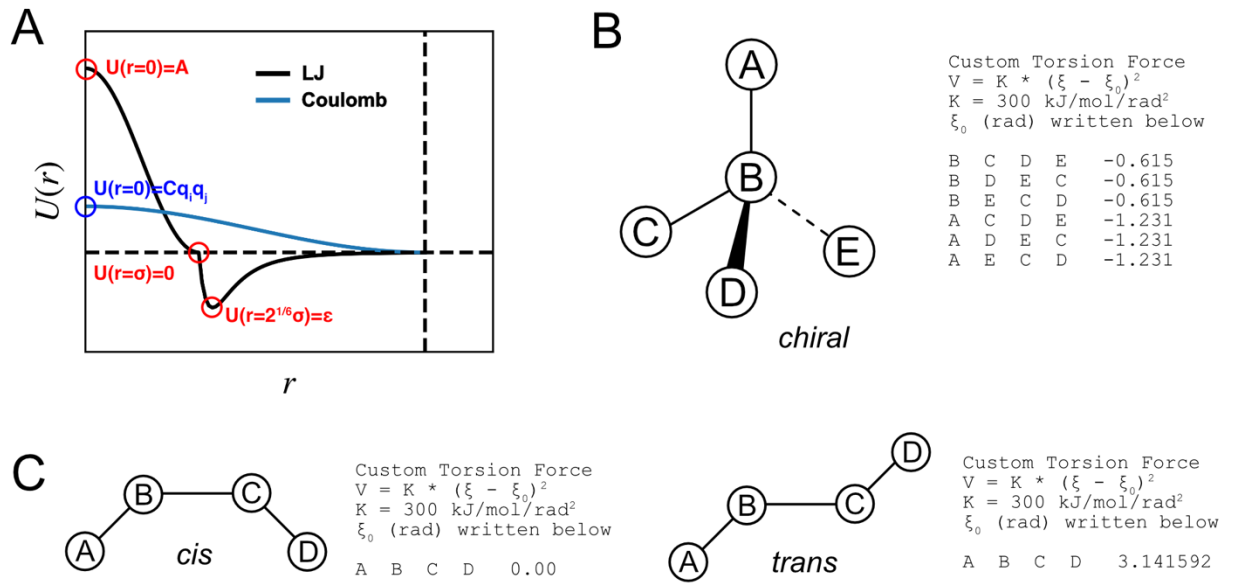


Figure 1. Illustration of the modified forcefield. (A) Nonbonded interactions. (B) Dihedrals for *chiral*. (C) Dihedrals for *cis* or *trans*.

RESULTS

Lipid

Triolein is a neutral lipid that nucleates inside a bilayer when its concentration is above a critical concentration.[26] Recently, I have developed a highly CG (four CG beads), triolein model based on the Grime-Madsen implicit lipid model that is able to nucleate at the critical concentration.[27,28] To test the performance of `mstool`, I have backmapped a CG membrane structure in which triolein was clustered inside a phospholipid bilayer into a AA lipid bilayer (Fig. 2A). The first step was performed by `mstool.Ungroup`, which is simply ungrouping atoms from their CG beads, resulting in an intermediate AA structure. The second step was done by `mstool.REM`, leading to an energy minimized structure. Finally, the isomeric properties of the molecules were reviewed by `mstool.CheckStructure`, and no flipped isomers were reported. The conversion increased the resolution of the structure 34-fold without any bridging resolutions, demonstrating the robustness of `mstool`.

Martini lipids are one of the most popular CG models.[13–16] It has been shown that they accurately capture the physics and chemistry of many biological processes and reproduce the bilayer properties such as thickness and area per lipid.[29,30] To test the efficacy of `mstool`, I have backmapped a spherical bilayer at Martini resolution with 20 different Martini lipids into an AA resolution system. The backmapped structure of the CG bilayer had no flipped isomers (Fig. 2B). This shows the robustness of isomeric dihedrals

applied during REM (Fig.1B and Fig. 1C) as each cholesterol molecule (residue name CHL1) has 8 chiral centers, and each PIP3 molecule (residue name SAPI33, renamed as PI3A in `mstool`) has 7 chiral centers and 4 cis bonds.

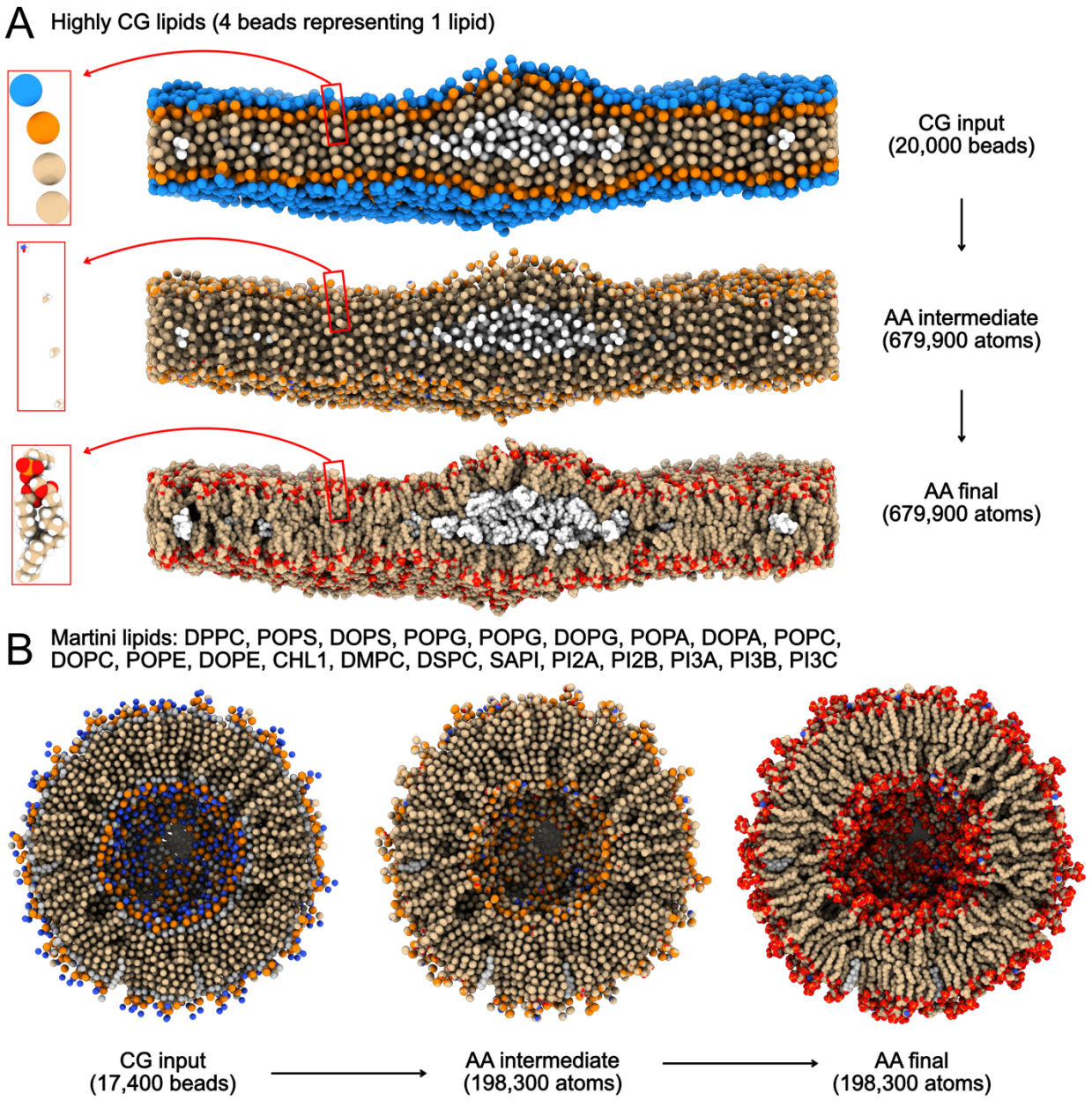


Figure 2. Lipid backmapping of (A) highly CG lipids and (B) Martini lipids.

Protein

`mstool` supports multiple backmapping options for protein. The equilibration timescale of protein is much longer than that of lipid and usually beyond timescales attainable by AA-MD. Therefore, it is possible that a backmapped protein structure is physically plausible (low potential energy) but is not biologically feasible that may not be equilibrated even after performing AA-MD simulations. One should choose a protein backmapping scheme that best fits the problem at hand. If there is an initial or reference AA protein structure, I highly recommend using it, combined with the loop modeler provided in `mstool`, introduced as the last option in this session. However, if there is no initial AA protein structure, or the final CG structure is very different from the AA structure, the first two options can serve as an alternative. Backmapping of T-cell intracellular antigen 1 from Martini resolution to AA resolution will be used as a test case.[31] A short CG-MD simulation was performed to stray away from the initial AA structure (Fig. 3A).

The first option is to randomly place atoms of each amino acid at their corresponding CG beads, just like lipid backmapping, through `mstool.Ungroup`. For instance, the Martini resolution has one backbone bead (BB) per residue; Therefore, the backbone atoms at the AA level (N, HN, CA, HA, C, O) are initially placed where the BB atom locates for each amino acid. The intermediate AA structure is then relaxed through REM and then reviewed.

The second option, which is the default option in `mstool`, is only slightly different from the first option in ungrouping. Instead of randomly placing AA atoms from CG beads, it predicts the positions of the backbone atoms using the previously published geometrical algorithm.[12,32] In short, it uses the positions of three consecutive backbone beads to obtain the positions of backbone atoms. In both options, the side chain atoms are placed randomly near their corresponding CG beads during `mstool.Ungroup`. The rest of the procedure (`mstool.REM` and `mstool.CheckStructure`), is the same.

The backmapped structures using the first and second options are shown in Fig. 3B and Fig. 3C, respectively, along with their Ramachandran plots. Although both options provide a reasonable structure, there are some Ramachandra outliers. Therefore, I highly recommend the last option, which leverages the initial AA structure.

Most of the cases, protein AA structures are present before running CG simulations. The last option of backmapping CG protein structures into biologically sound AA ones is to use these reference AA structures. While CG-MD simulations allow global changes in protein tertiary structures, they do not usually change the secondary structures. For instance, the Martini protein forcefield by design is not expected to change any secondary structures. If CG-MD simulations are run with an elastic network model (ENM), it is more obvious that protein structure changes little. Therefore, one can copy a rigid domain of an atomistic structure and align it into its corresponding CG structure. The loops that connect the rigid domains can be built using `mstool.LoopModeler` (details in Discussion) or other loop modelers. For instance, T-cell intracellular antigen 1 has two rigid domains, connected by a loop. Therefore, it is reasonable to use the reference AA structure of the

two domains in the backmapping process. The backmapped structure using this approach has fewer Ramachandran outliers (Fig. 3D).

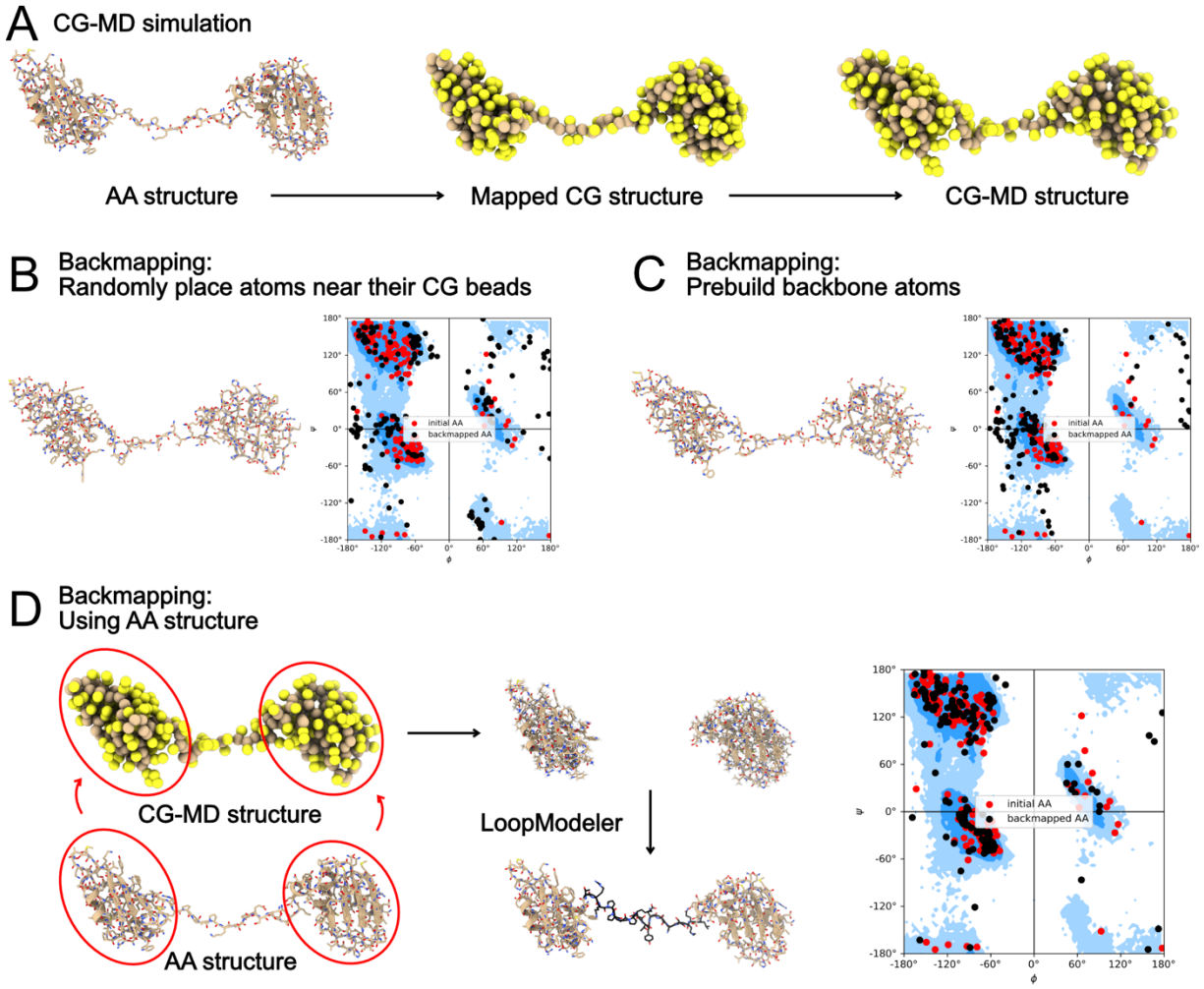


Figure 3. Protein backmapping. (A) CG-MD simulation of the target protein. (B) First option. (C) Second option (default in `mstool.Ungroup`). (D) The third option using AA structure and `mstool.LoopModeler`.

Membrane protein and lipid

Backmapping of membrane protein and lipid is not different from backmapping of either lipid (Fig. 2) or protein (Fig. 3). However, it's important to note that the protein and the lipid should not be backmapped separately and then combined because the resulting structure will likely have unphysical contact between protein and lipid. All the molecules should be present during REM to get the final backmapped structure with no bad contacts. In this section, I will use the structure of outer membrane protein F (PDB: 2OMF),[33] integral in facilitating diffusion of molecules in the outer membrane of *E. coli*, to demonstrate backmapping of membrane protein and lipid. From the initial CG structure, a short CG-MD simulation was performed to stray away from the initial structure (Fig. 4A).

The first approach is to randomly place the protein and lipid AA atoms near their corresponding CG beads. The second option, which is the default, is the same with the first option except for protein backbone atoms, prebuilt based on three consecutive protein backbone beads in the ungrouping step. Both approaches gave a reasonable structure for the protein; However, as seen in the previous session, there are non-negligible Ramachandran outliers with these two options (Fig. 4B and Fig. 4C).

Considering the minimal conformational changes of the protein during the CG-MD simulation, it is reasonable to align the initial protein AA structure to the final CG structure and ungroup lipid AA atoms from lipid CG beads. However, because the initial AA protein structure is slightly different from the final CG protein structure, there are inevitably bad contacts between the initial AA protein atoms aligned to the final CG protein structure and lipid atoms ungrouped from the final CG lipid beads. Fortunately, REM is capable of relaxing such systems with unphysical contacts. This third option applies the AA protein forcefield (CHARMM36m) into the initial AA protein atoms aligned to the final CG protein structure during REM and relaxes a system with lipid atoms. The Ramachandran plot of the backmapped structure using the third option (Fig. 4D) shows smaller number of outliers compared to the first two options (Fig. 4B and Fig. 4C).

If the input protein structure is not complete, has a modified protein residue such as glycosylation or phosphorylation, or has a ligand, the conventional protein forcefield will not be applicable. To deal with such cases, `mstool` has an option to treat the input protein structure as a space-filler. In this case, the tool does not apply the AA protein forcefield to the provided protein structure but creates ENM for the input structure with moderate LJ parameters assigned to the input protein atoms. This way, regardless of the input protein structure, it will not move during REM, and the lipid atoms are repulsed from the space-filler. Then at the last step of REM, the space-filler is replaced with the input protein structure. Therefore, the Ramachandran plot of the final structure is identical to the initial AA protein structure (Fig. 4E); However, there are no bad contacts between protein and lipid atoms in the backmapped structure. The procedure of this option is illustrated in Fig. 4F.

Although this work uses simple examples to showcase the backmapping, `mstool` can work for larger and more complicated systems. The heroic backmapping of a CG model of lipid droplet biogenesis is shown in Fig. 4G. It is a gigantic spherical bilayer with a diameter of 60 nm to show the budding of LD for the first time in simulation.[34] Protein shown in red is a seipin oligomer (PDB: 6DS5),[35] which facilitates triolein (shown in white) clustering and is required for proper lipid droplet maturation. This CG system is a great example of protein backmapping using `mstool.LoopModeler`. The luminal domains of seipin are rigid and change little during the CG-MD simulation. This suggests that one can cut the luminal part of the AA protein structure and copy it into the CG protein structure. In contrast, the transmembrane segments were flexible and became open up during CG-MD.[34] Therefore, each transmembrane segment of the AA structure was copied into its corresponding CG structure. A total of 22 loop structures that connect the luminal domain and the transmembrane segments was made using `mstool.LoopModeler` all at once. This way, while the AA structure of the luminal

domain and transmembrane segments are preserved, the backmapped model also can capture the opening of seipin transmembrane segments during lipid droplet maturation.

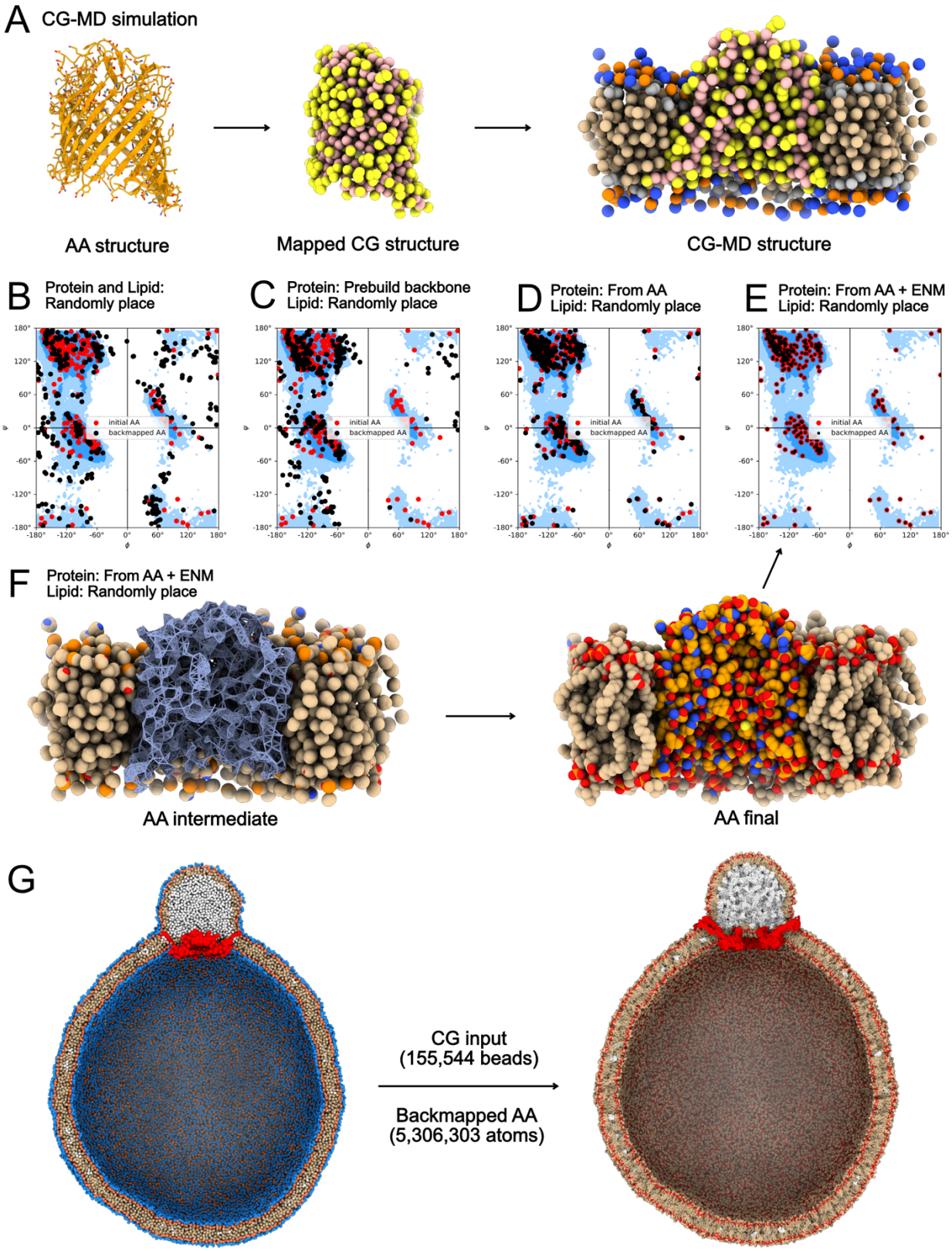


Figure 4. Lipid and protein backmapping. (A) CG-MD simulation of the target protein. (B-E) Ramachandran plots of the backmapped protein structures using multiple options. (F) Workflow that treats an input structure as a space-filling rock structure. (G) Heroic backmapping of lipid droplet biogenesis model.

DISCUSSION

CG-MD simulations are helpful in understanding complex biological processes by allowing access to time and length scales beyond the scales typical of AA-MD simulations. However, CG models lack atomistic details that are critical in understanding protein structures, protein-lipid interactions, and protein-ligand interactions. Therefore, converting CG structures into AA ones in a biologically and physically sound manner is an important tool in the field of CG modeling of biophysical systems.

In this work, I have developed a backmapping tool that converts CG structures into AA ones and implemented into a python package called `mstool`. The procedure consists of three steps for resolution conversion. First, `mstool.Ungroup` simply places atoms near their corresponding beads. Second, `mstool.REM` applies a modified forcefield to a system and then performs energy minimization. The modified forcefield replaces the positive part of LJ with a cosine function and adds additional dihedrals to isomers. The former modification ensures numerical stability during energy minimization, and the latter ensures consistency in isomeric properties. Finally, the backmapped structure is reviewed by `mstool.CheckStructure`.

These three steps are decoupled. Therefore, each step can be combined with other backmapping tools. For instance, one can consider using approaches to create a better intermediate AA structure before energy minimization such as fragment alignment[10,11] or geometrical projection.[12] Once an intermediate structure is obtained using either of these approaches or other methods, `mstool.REM` can be performed on the structure. Obviously, one can backmap CG structures into AA structures with other backmapping tools and then review flipped isomers with `mstool.CheckStructure`.

It should be noted that `mstool.REM` can be used for other purposes other than backmapping. For example, if an AA structure has multiple bad contacts, the standard energy minimization algorithm will crash because of the high potential energy. In such cases, `mstool.REM` can help resolve all the bad contacts so that standard energy minimization can be run afterward.

There are potential applications of this backmapping procedure that leverages the efficiency and fast equilibration of CG models. For example, `mstool.LoopModeler` can predict a structure of missing loops without the use of a structure template. It is powerful to model any lengths of missing loops and multiple loops all at once. It is also simple to use, taking only a fasta file of protein chains and a protein structure file. The loop modeler automatically detects which parts are not present in a provided AA structure (except for the N-termini and C-termini) by comparing the sequence and residue number of the AA

structure with the sequence of the provided fasta and build loop structures. The details of the workflow and quality of the loops will be discussed in a separate paper.

Another application is a membrane builder. CG membranes are easier to construct and equilibrate than AA membranes. One can build a CG membrane with a desired lipid composition at the Martini resolution, equilibrate the membrane at CG level, and then backmap into the AA one. A prototype of a membrane builder is included in the repository.

Finally, the backmapping tool can be used in studying lipid docking. Lipid-synthesizing or lipid-metabolizing proteins modify lipids inside their reaction chambers. However, due to their flexibility, lipids sometimes do not appear with a solid electron density in crystallography or cryo-EM of the protein and lipid complex. Therefore, it is a challenge to identify a lipid binding pose inside the protein even though this structural model is critical in understanding the mechanism of protein. Exhaustive conformation search, which rotates all the rotatable bonds of small ligands and finds the best conformation, is usually not suitable for lipids because lipids are extremely flexible compared to small ligands. REM can be an alternative to this. For instance, one can place lipid atoms at one point inside the reaction chamber and then run `mstool.REM` multiple times to obtain the most statistically significant pose or the lowest potential energy pose. This approach was used to identify a docking pose of lyso-phosphatidylinositol in membrane-bound O-acyltransferase 7.[36]

CONCLUSIONS

I present a robust and flexible backmapping tool with a minimal input set from users: mapping and isomeric information. Leveraging the efficacy of the modified forcefield, the tool is capable of converting CG systems of any resolution and of large scale to the AA one.

ACKNOWLEDGMENTS

I thank my Ph.D. advisor, Gregory A. Voth, for his mentorship and support. I am incredibly appreciative of my coworkers at D. E. Shaw Research for helpful discussion, Avi Robinson-Mosher, Peter Skopp, Robert McGibbon, Stefano Piana-Agostinetti, Qi Wang, Virginia Jiang, Dan Kozuch, Jim Valcourt, Parker de Waal, Jacob Pessin, and Brent Gregersen. Finally, I appreciate Won Hee Ryu and anonymous reviewers for their time to review this paper.

REFERENCES

(1) Marrink, S. J.; Corradi, V.; Souza, P. C. T.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. P. Computational Modeling of Realistic Cell Membranes. *Chem. Rev.* **2019**, *119* (9), 6184–6226. https://doi.org/10.1021/acs.chemrev.8b00460.

(2) Corradi, V.; Sejdiu, B. I.; Mesa-Galloso, H.; Abdizadeh, H.; Noskov, S. Yu.; Marrink, S. J.; Tieleman, D. P. Emerging Diversity in Lipid–Protein Interactions. *Chem. Rev.* **2019**, *119* (9), 5775–5848. https://doi.org/10.1021/acs.chemrev.8b00451.

(3) Jin, J.; Pak, A. J.; Durumeric, A. E. P.; Loose, T. D.; Voth, G. A. Bottom-up Coarse-Graining: Principles and Perspectives. *J. Chem. Theory Comput.* **2022**, *18* (10), 5759–5791. https://doi.org/10.1021/acs.jctc.2c00643.

(4) Kim, S.; Swanson, J. M. J.; Voth, G. A. Computational Studies of Lipid Droplets. *J. Phys. Chem. B* **2022**, *126* (11), 2145–2154. https://doi.org/10.1021/acs.jpcb.2c00292.

(5) Álvarez, D.; Sapia, J.; Vanni, S. Computational Modeling of Membrane Trafficking Processes: From Large Molecular Assemblies to Chemical Specificity. *Current Opinion in Cell Biology* **2023**, *83*, 102205. https://doi.org/10.1016/j.ceb.2023.102205.

(6) Simunovic, M.; Srivastava, A.; Voth, G. A. Linear Aggregation of Proteins on the Membrane as a Prelude to Membrane Remodeling. *Proceedings of the National Academy of Sciences* **2013**, *110* (51), 20396–20401. https://doi.org/10.1073/pnas.1309819110.

(7) Jarin, Z.; Pak, A. J.; Bassereau, P.; Voth, G. A. Lipid-Composition-Mediated Forces Can Stabilize Tubular Assemblies of I-BAR Proteins. *Biophysical Journal* **2021**, *120* (1), 46–54. https://doi.org/10.1016/j.bpj.2020.11.019.

(8) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A Multiscale Coarse-Grained Model of the SARS-CoV-2 Virion. *Biophysical Journal* **2021**, *120* (6), 1097–1104. https://doi.org/10.1016/j.bpj.2020.10.048.

(9) Yu, A.; Lee, E. M. Y.; Briggs, J. A. G.; Ganser-Pornillos, B. K.; Pornillos, O.; Voth, G. A. Strain and Rupture of HIV-1 Capsids during Uncoating. *Proceedings of the National Academy of Sciences* **2022**, *119* (10), e2117781119. https://doi.org/10.1073/pnas.2117781119.

(10) 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* (4), 1157–1166. https://doi.org/10.1021/ct100569y.

(11) Vickery, O. N.; Stansfeld, P. J. CG2AT2: An Enhanced Fragment-Based Approach for Serial Multi-Scale Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2021**, *17* (10), 6472–6482. https://doi.org/10.1021/acs.jctc.1c00295.

(12) 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* (2), 676–690. https://doi.org/10.1021/ct400617g.

(13) Marrink, S. J.; De Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108* (2), 750–760. https://doi.org/10.1021/jp036508g.

(14) 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* (27), 7812–7824. https://doi.org/10.1021/jp071097f.
(15) 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* (5), 819–834. https://doi.org/10.1021/ct700324x.
(16) 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* (1), 687–697. https://doi.org/10.1021/ct300646g.
(17) 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* **2017**, *14* (1), 71–73. https://doi.org/10.1038/nmeth.4067.
(18) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D. Jr.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114* (23), 7830–7843. https://doi.org/10.1021/jp101759q.
(19) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9* (1), 461–469. https://doi.org/10.1021/ct300857j.
(20) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLOS Computational Biology* **2017**, *13* (7), e1005659. https://doi.org/10.1371/journal.pcbi.1005659.
(21) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* **2011**, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.
(22) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations; Austin, Texas, 2016; pp 98–105. https://doi.org/10.25080/Majora-629e541a-00e.
(23) Campomanes, P.; Prabhu, J.; Zoni, V.; Vanni, S. Recharging Your Fats: CHARMM36 Parameters for Neutral Lipids Triacylglycerol and Diacylglycerol. *Biophysical Reports* **2021**, *1* (2), 100034. https://doi.org/10.1016/j.bpr.2021.100034.
(24) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis: UCSF ChimeraX Visualization System. *Protein Science* **2018**, *27* (1), 14–25. https://doi.org/10.1002/pro.3235.

(25) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF CHIMERAX : Structure Visualization for Researchers, Educators, and Developers. *Protein Science* **2021**, *30* (1), 70–82. https://doi.org/10.1002/pro.3943.
(26) Hamilton, J. A.; Small, D. M. Solubilization and Localization of Triolein in Phosphatidylcholine Bilayers: A 13C NMR Study. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78* (11), 6878–6882. https://doi.org/10.1073/pnas.78.11.6878.
(27) Grime, J. M. A.; Madsen, J. J. Efficient Simulation of Tunable Lipid Assemblies Across Scales and Resolutions. arXiv October 11, 2019. http://arxiv.org/abs/1910.05362 (accessed 2023-08-14).
(28) Kim, S.; Li, C.; Farese, R. V.; Walther, T. C.; Voth, G. A. Key Factors Governing Initial Stages of Lipid Droplet Formation. *J. Phys. Chem. B* **2022**, *126* (2), 453–462. https://doi.org/10.1021/acs.jpcb.1c09683.
(29) Marrink, S. J.; Risselada, J.; Mark, A. E. Simulation of Gel Phase Formation and Melting in Lipid Bilayers Using a Coarse Grained Model. *Chemistry and Physics of Lipids* **2005**, *135* (2), 223–244. https://doi.org/10.1016/j.chemphyslip.2005.03.001.
(30) Ingólfsson, H. I.; Melo, M. N.; van Eerden, F. J.; Arnarez, C.; Lopez, C. A.; Wassenaar, T. A.; Periole, X.; de Vries, A. H.; Tieleman, D. P.; Marrink, S. J. Lipid Organization of the Plasma Membrane. *J. Am. Chem. Soc.* **2014**, *136* (41), 14554–14559. https://doi.org/10.1021/ja507832e.
(31) Wang, I.; Hennig, J.; Jagtap, P. K. A.; Sonntag, M.; Valcárcel, J.; Sattler, M. Structure, Dynamics and RNA Binding of the Multi-Domain Splicing Factor TIA-1. *Nucleic Acids Research* **2014**, *42* (9), 5949–5966. https://doi.org/10.1093/nar/gku193.
(32) Lombardi, L. E.; Martí, M. A.; Capece, L. CG2AA: Backmapping Protein Coarse-Grained Structures. *Bioinformatics* **2016**, *32* (8), 1235–1237. https://doi.org/10.1093/bioinformatics/btv740.
(33) Cowan, S. W.; Schirmer, T.; Rummel, G.; Steierf, M.; Ghosh, R.; Pauptitt, R. A.; Jansonius, J. N.; Rosenbusch, J. P. Crystal Structures Explain Functional Properties of Two E. Coli Porins. *Nature* **1992**, *358*, 727–733.
(34) Kim, S.; Chung, J.; Arlt, H.; Pak, A. J.; Farese, R. V.; Walther, T. C.; Voth, G. A. Seipin Transmembrane Segments Critically Function in Triglyceride Nucleation and Lipid Droplet Budding from the Membrane. *eLife* **2022**, *11*, e75808. https://doi.org/10.7554/eLife.75808.
(35) Yan, R.; Qian, H.; Lukmantara, I.; Gao, M.; Du, X.; Yan, N.; Yang, H. Human SEIPIN Binds Anionic Phospholipids. *Developmental Cell* **2018**, *47* (2), 248-256.e4. https://doi.org/10.1016/j.devcel.2018.09.010.
(36) Wang, K.; Lee, C.-W.; Sui, X.; Kim, S.; Wang, S.; Higgs, A. B.; Baublis, A. J.; Voth, G. A.; Liao, M.; Walther, T. C.; Farese, R. V. The Structure of Phosphatidylinositol Remodeling MBOAT7 Reveals Its Catalytic Mechanism and Enables Inhibitor Identification. *Nat Commun* **2023**, *14* (1), 3533. https://doi.org/10.1038/s41467-023-38932-5.