Dscam isoform specificity. Nature *449*, 487–491.

8. Pinelo, J.E.E., Manandhar, P., Popovic, G., Ray, K., Tasdelen, M.F., Nguyen, Q., Iavarone, A.T., Offenbacher, A.R., Hudson, N.E., and Sen, M. (2023). Systematic mapping of the conformational landscape and dynamism of soluble fibrinogen. J. Thromb. Haemost. *21*, 1529–1543.

9. Labasque, M., Hivert, B., Nogales-Gadea, G., Querol, L., Illa, I., and Faivre-Sarrailh, C. (2014). Specific contactin N-glycans are implicated in neurofascin binding and autoimmune targeting in peripheral neuropathies. J. Biol. Chem. *289*, 7907–7918.

10. Li, J., Su, Y., Xia, W., Qin, Y., Humphries, M.J., Vestweber, D., Cabañas, C., Lu, C., and Springer, T.A. (2017). Conformational equilibria and intrinsic affinities define integrin activation. EMBO J. *36*, 629–645.

From simple to complex: Reconstructing all-atom structures from coarse-grained models using cg2all

Yui Tik Pang,[1] Lixinhao Yang,[2] and James C. Gumbart[1,2,*]
[1]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA
*Correspondence: gumbart@physics.gatech.edu
https://doi.org/10.1016/j.str.2023.12.004

In this issue of *Structure*, Heo and Feig present cg2all, a novel deep-learning model capable of efficiently predicting all-atom protein structures from coarse-grained (CG) representations. The model maintains high accuracy, even when the CG model is simplified to a single bead per residue, and has a number of promising applications.

Deep-learning methodologies have undergone swift advances since the introduction of the transformer architecture in 2017.[1] Transformers leverage a deep-learning technique known as "attention," which allows neural-network models to focus on interrelated segments within a data sequence when generating outputs. This approach has demonstrated exceptional efficacy in a wide range of generative deep-learning applications, from natural language processing to computer vision, yielding highly realistic and convincing results. In the realm of biology, AlphaFold2 employs transformers to discern patterns in protein sequence data, achieving unparalleled accuracy in protein structure prediction.[2] In many cases, these predictions are precise matches to experimental structures.

However, a machine learning method that can predict protein conformations beyond just one (or a few) structures is still missing, and molecular dynamics (MD) simulation remains the predominant tool to study protein dynamics in silico. Traditionally, MD simulations represent proteins and their surrounding biological environments atom by atom, which contributes to their accuracy but also incurs a high computational cost. One approach to mitigate the computational cost of MD simulations is the use of coarse-grained (CG) models, which reduce the resolution of the system from atomistic to several beads per residue. Over the years, CG models have been utilized to address numerous questions across fields such as biology and materials science.[3]

Coarse graining enables MD simulations to explore protein conformations more rapidly; however, the reduced resolution limits the accuracy and insights that can be derived. For example, hydrogen bonds and salt bridges play important roles in protein structure and dynamics; yet, without hydrogen atoms, CG models fail to capture these interactions explicitly. Prior attempts using library-based methods to reconstruct all-atom models from CG models[4,5] have achieved modest accuracy but frequently require extensive optimization to eliminate steric clashes and other physical imperfections.

Here, Heo and Feig introduce cg2all, a deep-learning approach that converts CG models back into their corresponding all-atom representations (Figure 1).[6] Cg2all utilizes the SE(3) transformer,[7] a variant of the self-attention mechanism that remains equivariant under 3D roto-translations, as well as a rigid-body block representation of the protein inspired by AlphaFold2, to construct all-atom structures from CG models at various resolutions. The model is trained using high-resolution X-ray crystal structures from the Protein Data Bank (PDB), with a loss function that incorporates both data-dependent terms as well as physics-based terms, such as torsion energies.

The authors trained individual instances of cg2all for a variety of CG models, with resolutions ranging from one to eight beads per residue, including some well-established models such as MARTINI[8] and PRIMO.[9] All models exhibited excellent reconstruction accuracy for the validation set, with the average heavy-atom RMSD being 0.31 Å and 0.18 Å for reconstructions from MARTINI and PRIMO, respectively. Remarkably, cg2all also achieved an average heavy-atom RMSD of 0.46 Å when reconstructing from the center-of-mass one-bead-per-residue model, underscoring cg2all's capability to predict and construct side-chain-atom coordinates using only

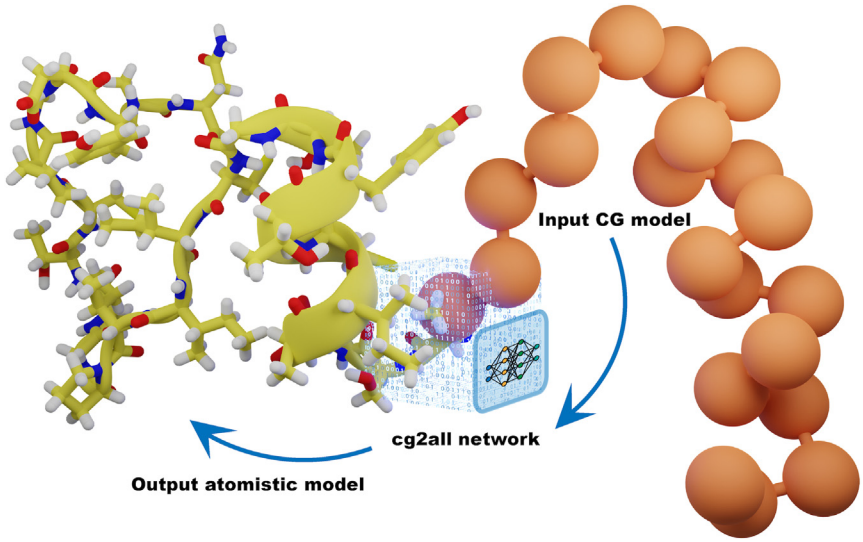

Figure 1. Schematic representation of the transformation of a CG model (orange spheres) into an all-atom structure (yellow cartoon and sticks representation) using the cg2all network

residue-level information. Interestingly, when the C_α position was used instead of the residue center of mass for the CG model, the heavy-atom RMSD increased to 0.96 Å, suggesting that center-of-mass positions are more informative than C_α positions. The authors further showcased cg2all's utility in converting MD trajectories run at CG resolution to atomistic resolution. Furthermore, the converted atomistic structures are stable enough to directly initiate all-atom MD simulations without additional optimization, and they remained folded after 10 ns.

One important implication of this study is that it offers a computationally efficient and deterministic function to map any CG model to its all-atom equivalent. This opens up possibilities for applications that can leverage the benefits of both the speed of CG models and the accuracy of all-atom models. One such application demonstrated by the authors is cryo-EM structure refinement. One effective approach, molecular dynamics flexible fitting (MDFF), employs restrained MD simulations to gradually drive the all-atom structure into the electron density from cryo-EM, a process that can take hours or even days to complete.[10] This is largely due to the numerous local minima in the energy landscape generated by all-atom force fields, causing simulation time to be spent repeatedly sampling these minima. However, with the use of cg2all, structure optimization can be performed using a simple potential function that operates at the one-bead-per-residue CG level while still maintaining the density restraint at atomistic resolution to fully utilize the information provided by the density. Using the proposed mechanism, a level of accuracy similar to MDFF (C_α-RMSD 0.36 vs. 0.35 Å) was achieved in an order-of-magnitude lower computational time.

The aforementioned example is just one instance of how cg2all can be employed for multi-scale sampling. An intriguing potential application could involve running MD simulations at CG resolution but with forces calculated at the all-atom level. While such a setup would eliminate the speed advantage gained from having fewer particles in CG models, the acceleration from a CG-enabled larger simulation timestep and a smoother energy landscape would still apply. However, additional research is required to determine how to accurately reproduce the dynamics of the all-atom system while using a CG model. Conversely, cg2all can be used to systematically parameterize bespoke CG force fields from general all-atom force fields through direct force matching, which is often challenging but essential for accuracy in simpler CG models such as single-bead-per-residue models. Alternatively, simpler and faster versions of cg2all can be created through "distillation" to serve as a system-specific neural-network potential, replacing traditional analytical CG force fields.

While further validation of cg2all's performance is needed, this work presents an attractive and promising approach for converting CG models to their all-atom counterparts, with significant potential for speeding up many applications that typically require the accuracy of atomistic resolution. Additionally, this work serves as a demonstration of the application of the SE(3) transformer in protein-related tasks, which could influence future developments in areas such as neural-network-potential design and other protein-structure-prediction applications.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (R01-GM148586).

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. Adv. Neural Inf. Process *30*, 5998–6008.

2. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

3. Joshi, S.Y., and Deshmukh, S.A. (2021). A review of advancements in coarse-grained molecular dynamics simulations. Mol. Simul. *47*, 786–803.

4. Alford, R.F., Leaver-Fay, A., Jeliazkov, J.R., O'Meara, M.J., Dimaio, F.P., Park, H., Shapovalov, M.V., Renfrew, P.D., Mulligan, V.K., Kappel, K., et al. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J. Chem. Theory Comput. *13*, 3031–3048.

5. Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins *77*, 778–795.

6. Heo, L., and Feig, M. (2024). One bead per residue can describe all-atom protein structures. Structure *32*, 97–111.

7. Fuchs, F., Worrall, D., Fischer, V., and Welling, M. (2020). SE (3)-transformers: 3D roto-translation equivariant attention networks. Adv. Neural Inf. Process *33*, 1970–1981.

8. Souza, P.C.T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., Grünewald, F., Patmanidis, I., Abdizadeh, H., Bruininks, B.M.H., Wassenaar, T.A., et al. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nat. Methods *18*, 382–388.

9. Gopal, S.M., Mukherjee, S., Cheng, Y.M., and Feig, M. (2010). A coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. Proteins *78*, 1266–1281.

10. Trabuco, L.G., Schreiner, E., Gumbart, J., Hsin, J., Villa, E., and Schulten, K. (2011). Applications of the molecular dynamics flexible fitting method. J. Struct. Biol. *173*, 420–427.