

CGBack: Recovering protein atomistic detail using diffusion models

○Diego Ugarte^{1,2}, Yuji Sugita^{1,2,3}

¹Graduate School of Science, The University of Tokyo

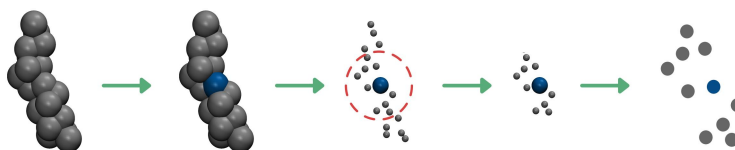
²Computational Biophysics Research Team, RIKEN R-CCS

³Theoretical Molecular Science Laboratory, RIKEN PRI

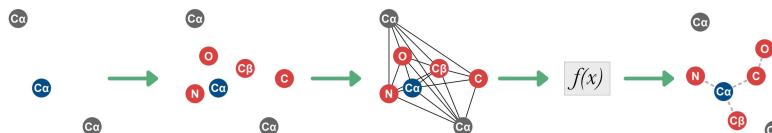
INTRODUCTION

Coarse-grained (CG) molecular dynamics enables simulations by reducing atomic detail, but this loss of resolution limits analysis. Backmapping CG structures to all-atom (AA) resolution is essential, yet existing approaches often break down for large systems. CGBack is a diffusion-based generative framework that reconstructs AA structures from CG inputs. CGBack achieves accurate, scalable backmapping for complex protein systems, enabling reliable integration into multiscale simulation workflows.

GRAPH GENERATION

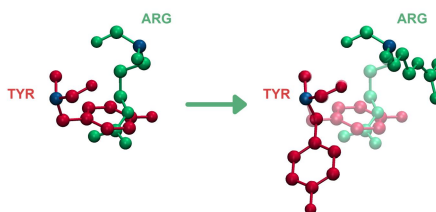
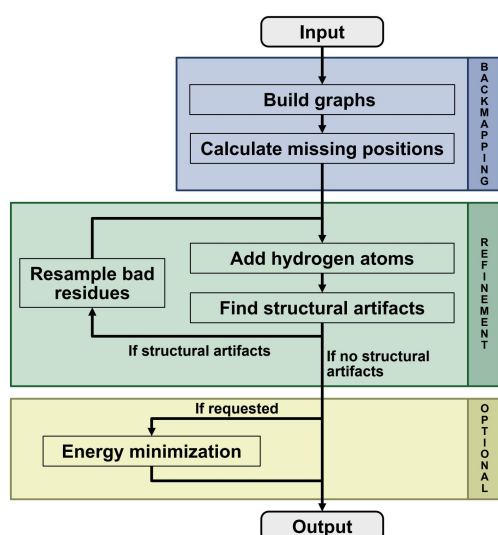


Residues within a cutoff distance of the target Cα are selected to provide local spatial context.



Missing heavy atoms of the target residue are added as nodes with random initial coordinates and connected in a fully-connected graph.

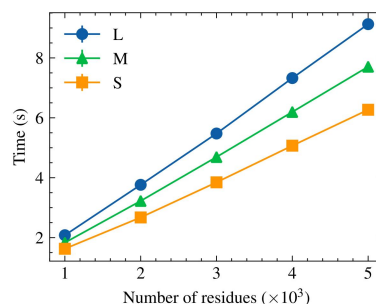
CGBack PIPELINE AND PERFORMANCE



CGBack detects ring penetrations and resamples the affected residues until the geometry is corrected.

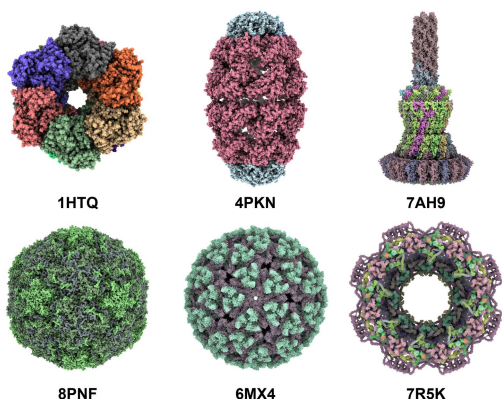


Incorrect stereochemistry at Cα or Cβ centers is fixed by resampling the residue with the correct chirality.



CGBack scales linearly with system size. The S, M, and L models offer different accuracy-speed trade-offs. For small system, the S model is enough, and for more challenging and complex systems, the L model is recommended.

BACKMAPPING LARGE COMPLEXES



CGBack backmaps large protein complexes, ranging from thousands to hundreds of thousands of residues, while maintaining stereochemical correctness and low artifact rates. Its diffusion model and refinement steps enable accurate AA reconstruction even in dense, multichain environments.

TRY CGBack!

Installation:

```
$> pip install cgback
```

Usage:

```
$> cgback input.pdb # Generates out.pdb
```



GitHub
genesis-release-r-ccs/cgback



Paper
10.1021/acs.jcim.5c01281



Slides

ACKNOWLEDGEMENTS

We thank RIKEN R-CCS for computational resources and support, and acknowledge funding from MEXT JSPS KAKENHI and the RIKEN Pioneering Project. We also thank Dr. Cheng Tan for providing simulation data and for testing CGBack together with Dr. Yangyang Zhang.