An Iterative Framework for Generative Backmapping of Coarse Grained Proteins

Georgios Kementzidis [1] **Erin Wong** [2] **John Nicholson** [3] **Ruichen Xu** [1] **Yuefan Deng** [1]

Abstract

The techniques of data-driven backmapping from coarse-grained (CG) to fine-grained (FG) representation often struggle with accuracy, unstable training, and physical realism, especially when applied to complex systems such as proteins. In this work, we introduce a novel iterative framework by using conditional Variational Autoencoders and graph-based neural networks, specifically designed to tackle the challenges associated with such large-scale biomolecules. Our method enables stepwise refinement from CG beads to full atomistic details. We outline the theory of iterative generative backmapping and demonstrate via numerical experiments the advantages of multistep schemes by applying them to proteins of vastly different structures with very coarse representations. This multistep approach not only improves the accuracy of reconstructions but also makes the training process more computationally efficient for proteins with ultra-CG representations.

1. Introduction

Coarse-grained (CG) models simplify complex molecular systems by clustering atoms into larger particles, known as CG beads, effectively eliminating certain internal degrees of freedom (Liwo et al., 2021; Noid, 2013). This reduction leads to fewer interactions, resulting in more efficient computational models that enable spatial and temporal scales inaccessible with fine-grained (FG) simulations.By internally contracting and consolidating the intricacies of atomic interactions, CG models enable simulations that can provide more meaningful and realistic insights into phenomena at mesoscopic and macroscopic scales (Zhang et al., 2024). This trade-off, however, comes with the challenge of preserving essential molecular characteristics in the CG representation, such as the structure, dynamics, and energetics of the FG system. Consequently, CG models must be carefully designed to retain these core properties while gaining computational efficiency (Fu et al., 2022). Such models are invaluable in many fields including biomolecular research and materials science, where exploring long-timescale processes is critical but computational resources are often limited.

While CG models greatly enhance computational efficiency, significant challenges emerge when reconstructions of the original atomic details are desired. The results of CG simulations often have to be converted back to the FG scale in order for several properties and interatomic interactions to be investigated through atom-level representations. Backmapping (or de-coarsening or inverse coarsening) is a process which seeks to restore atomistic resolution from the CG representation.

This problem has been tackled in several ways in the past. One of the earliest successful attempts was given by Wassenaar et. al (Wassenaar et al., 2014): initial placement of the atoms followed by force field relaxation restores interatomic distances. However, this method is very time-consuming because of the need to run all-atom MD. Recently, several ML-based solutions have been proposed. A solution using conditional Generative Adversarial Networks (Stieffenhofer et al., 2021; 2020) was proposed, but is limited to condensed phase systems and uses voxelization. The idea of voxelization also appeared in a model trained to de-coarsen entire trajectories by conditioning the coordinate prediction on the current CG structure and the previous-timestep FG structure (Shmilovich et al., 2022). Later advancements using conditional Variational Autoencoders (Wang et al., 2022; Yang & Gómez-Bombarelli, 2023) introduced equivariant operators with the use of the appropriate graph neural networks (GNN). Lately, the most recent trends include diffusion models (Han et al., 2024; Jones et al., 2023; Liu et al., 2023), but they are usually trained and tested on C_α traces or on CG resolutions like MARTINI (Marrink et al., 2007), CABS (Koliński, 2004), UNRES (Liwo et al., 2014), meaning they

---

[1] Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11790, USA [2] Garcia Center for Polymers at Engineered Interfaces, Stony Brook University, Stony Brook, NY 11790, USA [3] Swarthmore College, Swarthmore, PA 19081, USA. Correspondence to: Yuefan Deng <yuefan.deng@stonybrook.edu>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

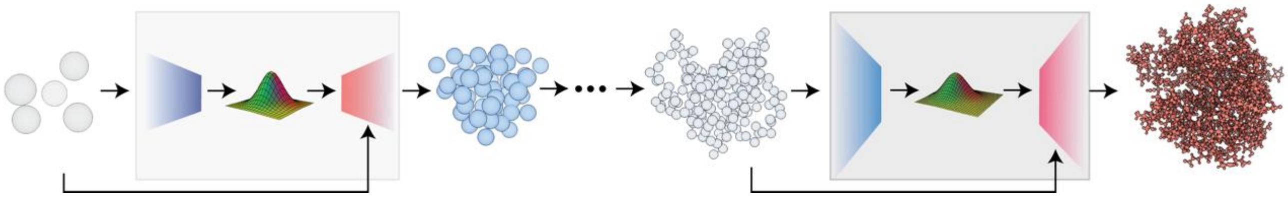

Figure 1: An illustration of the multistep generative backmapping scheme. Starting from a UCG structure, we restore atomistic resolution using independent priors (blue networks) and decoders (red networks), one resolution-step at a time.

restore the atomistic details from the residue-based coarse-graining (RBCG) resolution or a higher one – more than one bead per residue. On the other hand, there are models designed to work for arbitrary CG resolutions (Wang et al., 2022; Li et al., 2024), but they are tested only on either small molecules, or larger molecules with high-resolution CG. Furthermore, previous work emphasizes the chemical transferability of some models (Stieffenhofer et al., 2021; Yang & Gómez-Bombarelli, 2023; Liu et al., 2023).

Despite the recent advancements and their applicability to a variety of CG resolutions, little attention has been given to ultra-coarse graining (UCG), a regime in which each CG bead corresponds to tens or even hundreds of atoms (Dama et al., 2013). The speedup of UCG simulations is more pronounced (Yu et al., 2021; Yesudasan et al., 2018; Liang et al., 2023), and work on this class of applications, where the current methods underperform, manifests our focuses.

Inspired by super-resolution imaging (Saharia et al., 2022), we introduce iterative backmapping as a practical tool to alleviate the weaknesses of the previous methods on such CG representations. We develop a theoretical formulation and demonstrate the benefits of our scheme. We verify our hypothesis experimentally by applying a 2-step generative backmapping scheme to proteins with different structural and functional characteristics. An illustration of the multistep scheme is depicted in Fig. 1. When compared to the baseline CGVAE model developed by Wang and Bombarelli (Wang et al., 2022), the resulting structures are more accurate in restoring the FG structure's geometry and in conforming with the rules of basic science.

Our contributions can be summarized as follows:

- We identified the weaknesses of existing generative backmapping techniques — single-step schemes — on proteins with UCG representation.
- We developed a theoretical formulation for iterative generative backmapping of proteins using variational inference, and derive the objective's lower bound, i.e., the loss function of the model.
- We demonstrated numerically, through 2-step schemes, the advantages over the state-of-the-art 1-step baselines with a variety of metrics on two very different proteins.

2. Theory

Consider a molecule with n_0 atoms whose coordinates are $\mathbf{x}_0 \in \mathbb{R}^{n_0 \times 3}$. Let k be a positive integer, such that $\mathbf{x}_k \in \mathbb{R}^{n_k \times 3}$ is the final CG representation with n_k beads, which we aim to reconstruct. As shown in Fig. 2, we consider k coarsening operators $\Gamma_0, \Gamma_1, \ldots, \Gamma_{k-1}$ such that

$$\Gamma_i(\mathbf{x}_i) = \mathbf{x}_{i+1} \in \mathbb{R}^{n_{i+1} \times 3},$$

leading to progressively "coarser" CG representations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ with varying CG bead numbers $n_0 > n_1 > n_2 > \cdots > n_k$.

We define the ratio of the number of residues to the number of CG beads (# residues / # CG beads) as the *average CG bead size* and denote it by ρ. Notice that for higher k, ρ is higher, since each bead represents—on average—a bigger number of residues. For reference, $\rho = 1$ is equivalent to RBCG, while $\rho \gg 1$ corresponds to some UCG representations.

Given this formulation, $\mathbf{x}_k$ is the result of applying successive coarsening operators to the FG structure:

$$\mathbf{x}_k = (\Gamma_{k-1} \circ \Gamma_{k-2} \circ \ldots \circ \Gamma_1 \circ \Gamma_0)(\mathbf{x}_0).$$

We define and decompose the conditional distribution $p(\mathbf{x}_0|\mathbf{x}_k)$ as:

$$p(\mathbf{x}_0|\mathbf{x}_k) = \prod_{i=0}^{k-1} p_i(\mathbf{x}_i|\mathbf{x}_{i+1}).$$

The distribution is factorized into multiple conditional distributions, considering the intermediate CG representations. The Markov property, which dictates conditional independence between a representation $\mathbf{x}_i$ and its non-immediate coarser versions $\mathbf{x}_j, j > i+1$, is required for this factorization to hold. In backmapping, given a representation $\mathbf{x}_i$ of some resolution, the property implies that any additional knowledge of non-adjacent coarse representations $\mathbf{x}_j, j > i+1$ does not provide more information about $\mathbf{x}_i$ than the adjacent coarse representation $\mathbf{x}_{i+1}$. For instance, in Fig. 2, $\mathbf{x}_{k-1}$ is more informative than $\mathbf{x}_k$. In

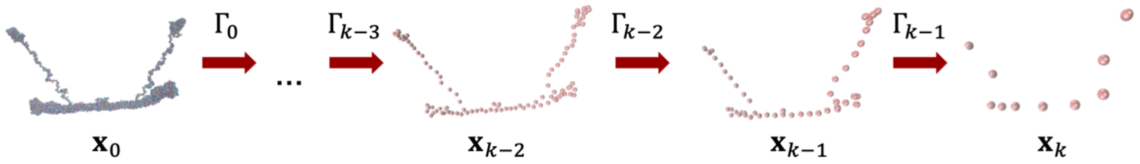

Figure 2: An FG conformation $\mathbf{x}_0$ with k progressively coarser and less informative CG conformations $\mathbf{x}_i$. The average CG bead size ρ is increasing.

principle, this property is not guaranteed, which can be rectified by defining appropriate coarsening operators Γ_i through topology-conserving algorithms to preserve the FG molecule's geometry (Wang et al., 2022; Freddolino et al., 2009; Arkhipov et al., 2006; Wang & Gómez-Bombarelli, 2019).

First, we focus on $p_i(\mathbf{x}_i|\mathbf{x}_{i+1})$, the probability of reconstructing a true, high-resolution $\mathbf{x}_i$ given an input CG $\mathbf{x}_{i+1}$. Employing variational inference (Blei et al., 2017) and following (Wang et al., 2022), we parametrize the conditional distribution by introducing the approximate posterior distribution $q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i, \mathbf{x}_{i+1})$ with parameters ϕ_i, where $\mathbf{z}_i$ denotes a latent space variable. Also, we describe the original conditional distribution as an integral over the latent space with

$$p_i(\mathbf{x}_i|\mathbf{x}_{i+1}) = \int p_{\theta_i}(\mathbf{x}_i|\mathbf{x}_{i+1}, \mathbf{z}_i)\, p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1})\, d\mathbf{z}_i,$$

where the integrand distributions have parameters $\{\theta_i, \psi_i\}$. Due to the complexity and high dimensionality of the FG conformation space, $p_i(\mathbf{x}_i|\mathbf{x}_{i+1})$ is intractable. So, when designing the de-coarsening operator, instead of maximizing the log-likelihood $\log p_i(\mathbf{x}_i|\mathbf{x}_{i+1})$, we maximize the evidence lower bound (ELBO):

$$\log p_i(\mathbf{x}_i|\mathbf{x}_{i+1}) \geq \mathbb{E}_{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\, \mathbf{x}_{i+1})}\left[\log p_{\theta_i}(\mathbf{x}_i|\mathbf{x}_{i+1},\, \mathbf{z}_i)\right] + D_{\mathrm{KL}}\Big(p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1}) \;\|\; q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\, \mathbf{x}_{i+1})\Big), \tag{1}$$

where we used Jensen's inequality, and D_{KL} is the Kullback–Leibler divergence, a measure of the statistical distance between two distributions. The derived ELBO, consisting of a reconstruction term and a regularization term, can be maximized with the relevant $p_{\theta_i}(\mathbf{x}_i|\mathbf{x}_{i+1}, \mathbf{z}_i)$, $p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1})$, and $q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i, \mathbf{x}_{i+1})$ (by using neural networks) to be illustrated in Section 3.

Following the above inequality, we derive the k-step backmapping ELBO:

$$\log p(\mathbf{x}_0|\mathbf{x}_k) \geq \sum_{i=0}^{k-1}\Big(\mathbb{E}_{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\, \mathbf{x}_{i+1})}\left[\log p_{\theta_i}\big(\mathbf{x}_i|\mathbf{x}_{i+1},\, \mathbf{z}_i\big)\right] + D_{\mathrm{KL}}\big(p_{\psi_i}\big(\mathbf{z}_i|\mathbf{x}_{i+1}\big) \,\big\|\, q_{\phi_i}\big(\mathbf{z}_i|\mathbf{x}_i,\, \mathbf{x}_{i+1}\big)\big)\Big). \tag{2}$$

where the new ELBO is split into k independent components, each consisting of a reconstruction term and a regularization term. Due to the terms' independence, we can design k models, each trying to reconstruct $\mathbf{x}_i$ from $\mathbf{x}_{i+1}$. We can thus set up k independent optimization processes to approximate the ELBO of $\log p(\mathbf{x}_0|\mathbf{x}_k)$ for a k-step scheme with the following apparent advantages:

- **Optimized utilization of processing power and RAM:** Even with high-performance computing and modern GPUs, the demands for RAM are high in the case of a single model restoring the FG resolution from a CG structure with very few beads, particularly due to expensive operations on graphs.
- **Varied reconstruction strategies:** We get to look at the k tasks separately and introduce variations to the network architectures, the inductive bias, or even the objective functions. This allows adjusting the model to properties aligned with each scale independently and, likely, more efficiently.
- **Enhanced detection of errors:** A lumped single-step model lacks the ability to reveal model or implementation errors, unlike the k-step model that progressively advances to show all details including errors.
- **Reduced and smoother search spaces:** It is less likely for each task to be trapped in local minima during training because the search space for each problem is reduced, and it can be explored more thoroughly and differently.

Overall, k-step backmapping, essentially a "divide and conquer" strategy, divides a challenging task into k relatively easier sub-tasks and then "conquers" them. However, the cost of the extra human and computing resources in RAM and floating-point operations in executing the k optimization processes must be well balanced with the gain of accuracy of the k-step backmapping scheme. Choosing a proper k for the "most gain and least pain" requires a given task and, in our work, schemes with $k = 2$, i.e., 2-step backmapping schemes, allow demonstration of the schemes' features.

3. Method

The theoretical framework, especially the ELBO derivation, has indicated the choice of architecture and form of the objective function. After elaborating on the design of 1-step schemes, including successful implementations by other groups, we describe our multistep method.

3.1. Relevant architectures

The variational inference and the derivation of the ELBO evoke the use of a conditional Variational Autoencoder (c-VAE) (Sohn et al., 2015), an extension of the original VAE (Kingma & Welling, 2013). The c-VAEs are used in semi-supervised learning for approximating distributions conditioned on some auxiliary variables. In this case, the low-resolution CG representation we aim to reconstruct is considered the auxiliary variable. Following the examples of recent backmapping solutions with c-VAEs (Wang et al., 2022; Yang & Gómez-Bombarelli, 2023) we employ neural networks $p_{\theta_i}(\mathbf{x}_i|\mathbf{x}_{i+1}, \mathbf{z}_i)$, $p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1})$, $q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i, \mathbf{x}_{i+1})$ with parameters $\{\theta_i, \psi_i, \phi_i\}$ to approximate the corresponding conditional distributions. They are the decoder, prior, and encoder, respectively, following the conventional terminology. The encoder approximates a lower dimensional posterior distribution $(\mathbf{x}_i, \mathbf{x}_{i+1}) \rightarrow \mathbf{z}_i$, shaping it as a multivariate Gaussian with learnable mean vector and covariance matrix. Similarly, during the inference stage—when the higher resolution information is unknown—the prior approximates the posterior distribution $\mathbf{x}_{i+1} \rightarrow \mathbf{z}_i$. With the KL divergence in Eq. 1, we make sure that the encoder and the prior give rise to similar latent representations. Finally, a sample of the latent space passes through the decoder to complete the reconstruction $(\mathbf{x}_{i+1}, \mathbf{z}_i) \rightarrow \mathbf{x}_i$.

Additionally, molecular representations necessitate graph-based algorithms when designing the encoder, prior, and decoder. In order to accurately predict interatomic forces and other physical and geometric properties at the quantum, atomic, and molecular levels, researchers (Gilmer et al., 2017; Gasteiger et al., 2020; Schütt et al., 2021; Satorras et al., 2021; Schütt et al., 2017; Unke & Meuwly, 2019) have developed several neural models for graph-structured data, and their architectures proved versatile across diverse tasks (Reiser et al., 2022). For instance, Message Passing Neural Networks (MPNNs) (Gilmer et al., 2017) predicted the local minima in the energy landscape of molecules (Mansimov et al., 2019) as well as the primary structures of conformations of alpha carbon traces (Dauparas et al., 2022), while the continuous-filter convolution network SchNet enabled the correct derivation of CG force fields (Husic et al., 2020). These same architectures can function as the backbone for the various components of the c-VAE.

3.2. Backmapping solutions and 1-step baseline

Our reference 1-step scheme is CGVAE (Wang et al., 2022), which uses a c-VAE and the variational inference framework with $k = 1$, and such a choice has a compelling rationale. First, this model, equivariant under translations and rotations, is probabilistic, and thus the same input CG can generate diverse—while accurate and natural—FG outputs. Unlike some of its predecessors (Stieffenhofer et al., 2021; 2020), it is not limited to condensed phase systems. Second, among the models handling CG representations of varying resolutions, this model is much coarser than the typical MARTINI or RBCG, and is among the most accurate. The encoder and prior consist of continuous convolution steps with SchNet-like models (Schütt et al., 2017). In particular, they use MPNNs (Gilmer et al., 2017) with radial basis transformations (Gasteiger et al., 2020) to process information at the FG and CG level, as well as pooling operations to map numerical representations from the FG to the CG space. Similarly, the decoder is designed to process the CG conformation and samples of the latent space through convolutions and lifting operations that ensure equivariance. The model's objective function is based on the ELBO of Eq. 1, in which the reconstruction term $\mathbb{E}_{q_{\phi_0}(\mathbf{z}_0|\mathbf{x}_0, \mathbf{x}_1)} \log p_{\theta_0}(\mathbf{x}_0|\mathbf{x}_1, \mathbf{z}_0)$ of the right-hand side is computed as a weighted sum of the MSD between the true $\mathbf{x}_0$ and the generated $\hat{\mathbf{x}}_0$, and a term penalizing incorrect bond lengths of adjacent atoms in the generated graph.

Another model of interest is GENZPROT (Yang & Gómez-Bombarelli, 2023), which, resembling CGVAE with subtle differences, is uniquely suited for specific backmapping tasks. Like CGVAE, GENZPROT employs a c-VAE framework and leverages GNNs to encode structural information, incorporating node embeddings with both atomic and residue-type information. However, it differs by adopting an internal coordinate-based representation, predicting bond lengths, bond angles, and torsion angles instead of Cartesian coordinates. This approach preserves physical plausibility and chemical connectivity, a crucial feature for proteins. Additionally, GENZPROT introduces a hierarchical message-passing scheme that operates on three levels: atom–atom, atom–residue, and residue–residue interactions, capturing both short and long-ranged dependencies. Its key strength lies in its transferability, achieved through training on diverse protein structures, enabling it to generalize across a wide range of protein structures and environments. While GENZPROT excels in reconstructing FG structures from

Algorithm 1 Backmapping a protein in k steps (inference phase)

```
Input: $\Gamma_0, \Gamma_1, \ldots, \Gamma_{k-1}$; trained networks $p_{\theta_i}(\mathbf{x}_i|\mathbf{x}_{i+1}, \mathbf{z}_i)$ and $p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1})$
$\mathbf{x}_k \leftarrow (\Gamma_{k-1} \circ \cdots \circ \Gamma_0)(\mathbf{x}_0)$
for $i = k-1$ down to $0$ do
    Sample $\mathbf{z}_i \sim p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1})$
    $\widehat{\mathbf{x}}_i \sim p_{\theta_i}(\mathbf{x}_i|\mathbf{x}_{i+1}, \mathbf{z}_i)$
    $\mathbf{x}_i \leftarrow \widehat{\mathbf{x}}_i$
end for
```

RBCG mappings, it is not compatible with very coarse mappings, limiting the scope of its applications. Nevertheless, its ability to provide chemically accurate and reliable FG structures makes it an excellent tool we see in parallel to our proposed multistep schemes.

3.3. k-step schemes

We describe how to build the 2-step scheme, which can be scaled to any multistep scheme. Starting from a very coarse representation $\mathbf{x}_2$, we use a c-VAE model to predict the coordinates of the alpha carbon (Cα) traces of the molecule $\mathbf{x}_1$, instead of the atomistic representation. In the absence of a real chemical structure with covalent bonds between atoms, but artificial bond–graph edges between beads or Cα atoms, we use a CGVAE, because it is designed to work with arbitrary CG mapping protocols. During the second step, we restore the atomistic resolution $\mathbf{x}_0$ given the generated RBCG representation $\mathbf{x}_1$. To take advantage of the uniformity of amino acids as protein building blocks, as well as the chemical transferability of the RBCG → FG reconstruction (Yang & Gómez-Bombarelli, 2023), we choose GENZPROT for this task. Because proteins can be in multiple protonation states, we ignore hydrogens and focus on reconstructing heavier atoms. In total, we combine CGVAE's flexibility with GENZPROT's chemical specificity. Both models could be replaced by similar c-VAE models if they are more accurate or suitable for each of the two tasks.

Given $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2$, we separately and simultaneously design two operators, one for each step, namely the $\mathbf{x}_2 \rightarrow \mathbf{x}_1$ and $\mathbf{x}_1 \rightarrow \mathbf{x}_0$ reconstructions. Thus, the search space of each optimization task is simplified and can be explored diversely. For example, during RBCG → FG, we penalize unphysical bond, angle, or dihedral magnitudes, incorporating physics and chemistry insights into the training, while using amino acid types as node embeddings. Moreover, we relax the computational and memory limitations; each model can be deeper, allowing it to focus on more details for better optimization, as escapes from local minima are unrealistic structures. Their synergistic impact yields a more precise operator. Figure 1 illustrates the more general k-step reconstruction process, while Algorithm 1 outlines the steps.

3.4. Evaluation Metrics

We adapt the following metrics to assess the performance of a given scheme:

- **Reconstruction Accuracy:** The accuracy is quantified with the RMSD between the true $\mathbf{x}_0$ and the generated $\widehat{\mathbf{x}}_0$ FG structures computed using only the n_0 heavy (non-hydrogen) atom coordinates:

$$\mathrm{RMSD}(\mathbf{x}_0, \widehat{\mathbf{x}}_0) = \sqrt{\frac{\|\mathbf{x}_0 - \widehat{\mathbf{x}}_0\|_2^2}{n_0}},$$

  averaged over multiple samples. We ignore hydrogens to account for the multiple protonation states of proteins.

- **Sample Quality:** We quantify how accurately the bond-graph is restored during backmapping. We compute the Graph Edit Distance (Sanfeliu & Fu, 1983) of the true G and generated $\widehat{G}$ bond graph, and divide it by the number of edges, to normalize the value:

$$\lambda(G, \widehat{G}) = \frac{\mathrm{GED}(G, \widehat{G})}{|E|},$$

  given the bond graphs $G = (V, E)$ and $\widehat{G} = (\widehat{V}, \widehat{E})$. A lower GED means that it takes fewer edit operations, i.e., vertex/edge insertions/deletions/substitutions, to transform one graph into another, accounting for the total number of edges $|E|$ for normalization purposes.

- **Steric Clash Score:** A steric clash occurs when two non-bonded atoms of the protein are unnaturally close to each other — such a positioning would normally lead to a strong repulsion force. When it comes to backmapping, the fewer the steric clashes, the more realistic the resulting conformation. If the distance between any two heavy (non-hydrogen) sidechain atoms from neighboring residues falls within a threshold distance, then we report a steric clash. The threshold distance is set to 1.2 Å, as in (Yang & Gómez-Bombarelli, 2023). We compute the percentage of residues with a steric clash.

For each setup, we carry out statistical analysis, and report the mean and standard deviation of each metric by performing independent experiments. We also conduct visual assessment of the secondary structure of reconstructed conformations by examining the Ramachandran plot (Ramachandran et al., 1963) to illustrate the distributions of the main backbone dihedrals. A strong match between the Ramachandran plots of the true and generated FG structures signifies a more accurate backmapping scheme.

4. Implementation

We apply the 2-step iterative scheme to two proteins and compare it with CGVAE, our 1-step baseline. We choose CGVAE because it is probabilistic, equivariant under translations and rotations, and the most accurate model applied to UCG protein representations.

4.1. Data

The first molecule we apply the models to is the eukaryotic translation initiation factor 4E, namely the protein eIF4E (PDB: 4TPW (Papadopoulos et al., 2014)). The structure we use has 181 residues, and long trajectories of the molecule can be found in (Hasegawa et al., 2024). Also, we choose mappings with very few beads, specifically $n_2 = 10, 19, 50$. Considering the number of residues in the molecule, this corresponds to average CG bead sizes $\rho = 18.10, 9.53, 3.62$, respectively. The second molecule is a structural ensemble of a Nuclear Localization Signal (NLS 99–140) peptide (Fuertes et al., 2017), labeled as PED000151 in the Protein Ensemble Database (Ghafouri et al., 2024). It has 48 residues and was reduced to $n_2 = 5, 8$ beads, leading to average CG bead sizes of $\rho = 9.60, 6.00$, respectively. In contrast, previous works (Wang et al., 2022; Yang & Gómez-Bombarelli, 2023; Li et al., 2024) have reported CG bead sizes ranging from 0.33 to 1.67.

We choose two proteins with different structural and dynamical features. eIF4E is globular, compact, with a more limited range of conformations. On the other hand, PED00151 is an Intrinsically Disordered Protein (IDP), meaning its three-dimensional structure is not stable under normal conditions. The radius of gyration R_g (Lobanov et al., 2008) of the two proteins over the course of 1 μs equilibrium simulations shows their flexibility or lack thereof. More specifically, with a mean R_g of 1.627 nm and a standard deviation of 0.027, as opposed to a mean of 2.152 nm and a larger standard deviation of 0.285, eIF4E is clearly more compact than the flexible PED00151.

4.2. Training

Prior to model training, we collect three independent 1 μs simulations of eIF4E from (Hasegawa et al., 2024), leading to a total of 3000 FG conformations. Similarly, we collect 9 746 simulation frames of the PED000151e000 ensemble from (Yang & Gómez-Bombarelli, 2023). We use AUTOGRAIN (Wang et al., 2022) to form various CG operators:

$$\mathbf{x}_0 \to \mathbf{x}_2,\ \mathbf{x}_0 \to \mathbf{x}_3,\ \mathbf{x}_1 \to \mathbf{x}_2,\ \mathbf{x}_1 \to \mathbf{x}_3,\ \text{and}\ \mathbf{x}_2 \to \mathbf{x}_3.$$

AUTOGRAIN's objective function prevents distant beads in a high-resolution representation from being assigned to the same low-resolution CG bead. This way, we create geometry-conserving CG mapping operators with the specified number of beads. Moreover, we form the corresponding $\mathbf{x}_0 \to \mathbf{x}_1$ mappings and representation $\mathbf{x}_1$ using MDTraj (McGibbon et al., 2015) by selecting the alpha carbon Cα of each residue. By visual inspection, we claim that knowledge of the coarser representation does not provide more information than the RBCG representation. This way we can invoke the Markov property and take advantage of conditional independence, as described in Section 2. For a more detailed look at the CG representations, plots are provided in Appendix G.

We train multiple CGVAE-like models with a train–validation–test split (80-10-10). In a similar way, GENZPROT, trained on a variety of proteins (Yang & Gómez-Bombarelli, 2023), is fine-tuned on eIF4E and PED000151 structures with an additional split on the corresponding datasets. We set a number of epochs with early stopping and an adaptive learning-rate scheduler, in case the validation-loss improvement stagnates. To search the hyperparameter space efficiently and extensively we use the Weights and Biases software (Biewald, 2020) with the hyperband algorithm (Li et al., 2018). For the 2-step scheme, the batch size was set to 2. Due to memory overload, the batch size for the 1-step scheme was set to 1 with gradient accumulation (Hermans et al., 2017). You can find detailed tables with the hyperparameters corresponding to each condition in Appendix D.

We train our models on a high-performance computing cluster. For 1-step backmapping, we use two NVIDIA A100 Tensor Core 80 GB GPUs with an Intel Haswell CPU due to the heavy memory and computational load, mostly originating from operations on graphs. We incorporate data parallelization (Li et al., 2020) in the original PyTorch (Paszke et al., 2019) code. For 2-step backmapping, the memory and computational load from the UCG → RBCG reconstruction is much lower, so we only use one NVIDIA Tesla K80 24 GB GPU with an Intel Ice Lake CPU for each step.

4.3. Results

To evaluate the performance of our proposed iterative backmapping schemes, we conduct a comparative analysis against the 1-step CGVAE baseline using two proteins of different structural characteristics: the globular protein eIF4E and the intrinsically disordered protein PED00151. We apply the 2-step backmapping scheme to both proteins. The results are reported in Tables 1 and 2 and are illustrated in Fig. 3 — note that the vertical axis for sample quality and steric-clash scores is in logarithmic scale. We record the mean and standard deviation from three experiments with different random seeds. We observe consistent improvements across multiple reconstruction metrics when employing multistep schemes.

Table 1: Results on the protein eIF4E.

METRIC	SCHEME	18.10 (10)	9.53 (19)	3.62 (50)
RECONSTRUCTION ACCURACY RMSD (Å)	1-STEP	7.563 ± 0.027	2.718 ± 0.018	1.364 ± 0.024
	2-STEP	1.503 ± 0.045	1.337 ± 0.123	0.997 ± 0.111
SAMPLE QUALITY (NORMALIZED GED)	1-STEP	49.562 ± 0.437	1.476 ± 0.009	0.301 ± 0.032
	2-STEP	0.016 ± 0.002	0.012 ± 0.001	0.009 ± 0.001
STERIC CLASH SCORE (%)	1-STEP	99.999 ± 0.001	23.030 ± 1.287	3.144 ± 0.057
	2-STEP	4.264 ± 0.728	3.733 ± 0.809	1.716 ± 0.315

Table 2: Results on the protein PED00151.

METRIC	SCHEME	9.60 (5)	6.00 (8)
RECONSTRUCTION ACCURACY RMSD (Å)	1-STEP	7.488 ± 0.129	5.707 ± 0.244
	2-STEP	5.368 ± 0.059	4.118 ± 0.066
SAMPLE QUALITY (NORMALIZED GED)	1-STEP	29.665 ± 0.753	15.000 ± 2.799
	2-STEP	0.330 ± 0.015	0.197 ± 0.010
STERIC CLASH SCORE (%)	1-STEP	99.970 ± 0.032	99.099 ± 0.799
	2-STEP	25.611 ± 0.258	13.631 ± 0.474

Quantitative Evaluation We assess reconstruction accuracy using heavy-atom RMSD, sample quality via normalized GED, and structural plausibility through the steric clash score. For eIF4E, the 2-step scheme significantly reduces the RMSD, especially at the highest CG bead size (10 beads, $\rho = 18.10$), where the RMSD drops from 7.56 Å (1-step) to 1.50 Å (2-step). Improvements are also consistent at finer resolutions ($\rho = 9.53$ and 3.62), with RMSD values falling below 1.00 Å for the finest CG. This performance gain confirms that intermediate CG representations help guide the reconstruction towards more accurate FG configurations. The sample quality improves by more than an order of magnitude across all resolutions; for instance, at $\rho = 18.10$ it decreases from 49.56 to 0.016. The steric-clash score demonstrates a pattern similar to the other metrics, indicating that intermediate resolutions lead to more chemically plausible reconstructions.

For PED00151, the more flexible IDP, results show a clear but more moderate improvement with the 2-step scheme. The RMSD decreases from 7.49 Å (1-step) to 5.37 Å (2-step) at $\rho = 9.60$ and from 5.71 Å to 4.12 Å at $\rho = 6.00$. These gains are less dramatic than in eIF4E, highlighting the inherent challenges in backmapping disordered proteins. In terms of steric clashes, PED00151 shows extremely high scores with the 1-step baseline (nearly 100 %), indicating severe structural artifacts. The 2-step scheme dramatically reduces these values to 25.61 % and 13.63 %, respectively, showing that iterative refinement significantly enhances physical realism.

Notice that, unlike sample quality and steric-clash score, reconstruction accuracy had a more modest improvement. This is because normalized GED and steric-clash score percentages are structurally binary, whereas RMSD is continuous. Even if atom positions are slightly off in a way that still contributes to RMSD, correcting just enough to form correct or remove incorrect bonds dramatically reduces GED. That is why GED can improve significantly even when RMSD does not shift much. An example is illustrated in Fig. 4, where the modest improvement of RMSD$(\mathbf{x}_0, \widehat{\mathbf{x}}_0)$ from 5.64 Å for 1-step to 3.65 Å for 2-step actually leads to a much more realistic structure. This example also raises questions about the reliability of RMSD as a metric often used in the literature to quantify structural deviation.

Secondary Structure Recovery We further validate backmapping quality by analyzing the Ramachandran plots of reconstructed FG structures, shown in Appendix C for more clarity. For eIF4E, the 2-step scheme yields dihedral-angle distributions that closely match the native FG dataset, particularly at $\rho = 3.62$. In contrast, the 1-step scheme fails to reproduce realistic backbone angles at high ρ, indicating distorted or unnatural secondary structures. This trend highlights the necessity of using intermediate resolutions to preserve features such as alpha helices and beta sheets during reconstruction.

For PED00151, even the 2-step scheme produces broader and noisier dihedral angle distributions than those of the reference FG dataset. This aligns with the nature of IDPs, where the high entropy of conformational states imposes a hard limit on reconstruction precision.

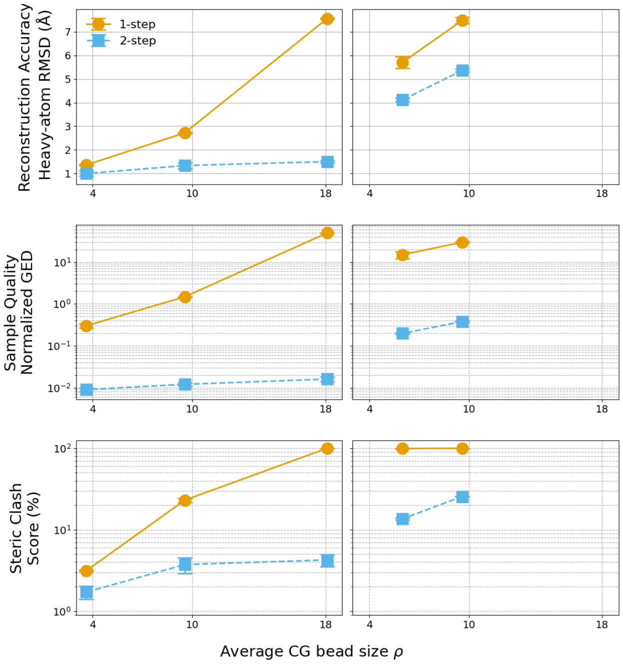

Figure 3: The results for eIF4E (left) and PED00151 (right). The vertical axis in the bottom two rows is shown on a logarithmic scale.

Importance of Bead Size and Molecule Properties Our experiments also reveal that the benefits of multistep schemes are more pronounced at higher average CG bead sizes. At lower ρ and for more compact molecules, 1-step schemes are capable of achieving decent reconstructions because the task is less ambiguous. For instance, for eIF4E at $\rho = 3.62$ all three metrics approach their ideal values, and the additional step in the 2-step scheme yields only modest improvements. At coarser resolutions, however, single-step models struggle owing to their limited capacity and the computational burden of restoring high-resolution detail from minimal input. Hardware constraints further emphasize this issue: because of GPU-memory limits we could train only 1-step models with shallow architectures (e.g., batch size = 1, encoder-convolution depth = 1), resulting in a restricted search space. By contrast, our 2-step scheme enables deeper networks for each stage, leading to more thorough optimization and a lower risk of convergence to sub-optimal minima.

5. Conclusion, Limitations, and Future Work

We present a practical framework for improving backmapping from ultra-coarse-grained (UCG) protein representations using multistep generative models. By decomposing the reconstruction into smaller, intermediate steps, our method enables deeper, more specialized networks at each resolution level—reducing memory bottlenecks and improving accuracy, especially for high CG bead sizes. Compared to single-step methods, multistep schemes offer not only better reconstruction metrics but also practical benefits such as modular design, improved convergence, and easier debugging.

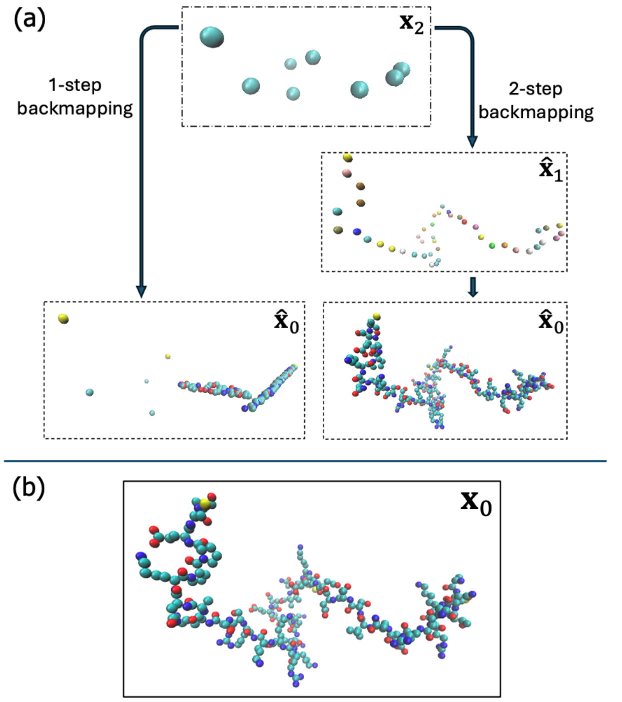

Figure 4: (a) Starting from $\mathbf{x}_2$ with $n_2 = 8$, we restore the FG representation $\mathbf{x}_0$ of PED00151 using 1-step and 2-step schemes. (b) The ground truth $\mathbf{x}_0$.

While 2-step backmapping improves structural realism over direct baselines, challenges remain, especially for IDPs, where coarse-grained representations carry less recoverable information. In future work, we aim to implement 3-step schemes, investigate more thoroughly the effect of inherent flexibility on backmapping success, apply multistep schemes to more proteins or even other macromolecules, as well as explore different strategies to accurately reconstruct disordered proteins or embed additional information into the reconstruction process when targeting flexible residues.

References

Arkhipov, A., Freddolino, P. L., and Schulten, K. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure*, 14(12):1767–1777, 2006.

Biewald, L. Experiment tracking with weights and biases. Software available from `https://www.wandb.com/`, 2020.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

Dama, J. F. et al. The theory of ultra-coarse-graining. 1. general principles. *Journal of Chemical Theory and Computation*, 9(5):2466–2480, 2013.

Dauparas, J. et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615): 49–56, 2022.

Freddolino, P. L. et al. Application of residue-based and shape-based coarse-graining to biomolecular simulations. In *Coarse-graining of Condensed Phase and Biomolecular Systems*, pp. 299–315. CRC Press, 2009.

Fu, X. et al. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *arXiv preprint arXiv:2210.07237*, 2022.

Fuertes, G. et al. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in saxs vs. fret measurements. *Proceedings of the National Academy of Sciences*, 114(31):E6342–E6351, 2017.

Gasteiger, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. arXiv preprint arXiv:2003.03123, 2020.

Ghafouri, H. et al. Ped in 2024: improving the community deposition of structural ensembles for intrinsically disordered proteins. *Nucleic Acids Research*, 52(D1): D536–D544, 2024.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pp. 1263–1272. PMLR, 2017.

Han, X. et al. The latent road to atoms: Backmapping coarse-grained protein structures with latent diffusion. *arXiv preprint arXiv:2410.13264*, 2024.

Hasegawa, K. et al. Structural and dynamical analyses of apo and cap-binding eif4e: An in silico study. *bioRxiv*, pp. 2024.05.18.594816, 2024.

Hermans, J. R., Spanakis, G., and Möckel, R. Accumulated gradient normalization. In *Asian Conference on Machine Learning*, pp. 342–355. PMLR, 2017.

Husic, B. E. et al. Coarse graining molecular dynamics with graph neural networks. *The Journal of Chemical Physics*, 153(19):194101, 2020.

Jones, M. S., Shmilovich, K., and Ferguson, A. L. DiAMoNDBack: Diffusion-denoising autoregressive model for non-deterministic backmapping of cα protein traces. *Journal of Chemical Theory and Computation*, 19(21): 7908–7923, 2023.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114, 2013.

Koliński, A. Protein modeling and structure prediction with a reduced representation. *Acta Biochimica Polonica*, 51: 349–371, 2004.

Li, J., Meng, Z., and Liang, S. Towards deep generative backmapping of coarse-grained molecular systems. In *Proceedings of the 2024 2nd Asia Conference on Computer Vision, Image Processing and Pattern Recognition*, 2024.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., and Talwalkar, A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(185):1–52, 2018.

Li, S. et al. PyTorch Distributed: Experiences on accelerating data parallel training. arXiv preprint arXiv:2006.15704, 2020.

Liang, D. et al. Coarse-grained modeling of the sars-cov-2 spike glycoprotein by physics-informed machine learning. *Computation*, 11(2):24, 2023.

Liu, Y., Lin, G., and Chen, M. Backdiff: a diffusion model for generalized transferable protein backmapping. *arXiv preprint arXiv:2301.xxxxx*, 2023.

Liwo, A. et al. A unified coarse-grained model of biological macromolecules based on mean-field multipole–multipole interactions. *Journal of Molecular Modeling*, 20:2306, 2014.

Liwo, A. et al. Theory and practice of coarse-grained molecular dynamics of biologically important systems. *Biomolecules*, 11(9):1347, 2021.

Lobanov, M. Y., Bogatyreva, N. S., and Galzitskaya, O. V. Radius of gyration as an indicator of protein structure compactness. *Molecular Biology*, 42(4):623–628, 2008.

Mansimov, E., Mahmood, O., Cho, K., and Chao, D. Molecular geometry prediction using a deep generative graph neural network. *Scientific Reports*, 9(1):20381, 2019.

Marrink, S. J. et al. The martini force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.

McGibbon, R. T. et al. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8):1528–1532, 2015.

Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics*, 139(9), 2013.

Papadopoulos, E. et al. Structure of the eukaryotic translation initiation factor eif4e in complex with 4egi-1 reveals an allosteric mechanism for dissociating eif4g. *Proceedings of the National Academy of Sciences*, 111(31): E3187–E3195, 2014.

Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, pp. 8024–8035, 2019.

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, 1963.

Reiser, P. et al. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93, 2022.

Saharia, C. et al. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022.

Sanfeliu, A. and Fu, K. S. A distance measure between attributed relational graphs for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13 (3):353–362, 1983.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, pp. 9323–9332. PMLR, 2021.

Schütt, K., , et al. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Schütt, K. T., Unke, O. T., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Shmilovich, K. et al. Temporally coherent backmapping of molecular trajectories from coarse-grained to atomistic resolution. *The Journal of Physical Chemistry A*, 126 (48):9124–9139, 2022.

Sohn, K., Lee, H., and Yan, X. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Stieffenhofer, M., Wand, M., and Bereau, T. Adversarial reverse mapping of equilibrated condensed-phase molecular structures. *Machine Learning: Science and Technology*, 1(4):045014, 2020.

Stieffenhofer, M., Bereau, T., and Wand, M. Adversarial reverse mapping of condensed-phase molecular structures: Chemical transferability. *APL Materials*, 9(3):031103, 2021.

Unke, O. T. and Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of Chemical Theory and Computation*, 15(6):3678–3693, 2019.

Wang, W. and Gómez-Bombarelli, R. Coarse-graining autoencoders for molecular dynamics. *npj Computational Materials*, 5(1):125, 2019.

Wang, W. et al. Generative coarse-graining of molecular conformations. In *International Conference on Machine Learning*, pp. to appear. PMLR, 2022.

Wassenaar, T. A. et al. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of Chemical Theory and Computation*, 10(2):676–690, 2014.

Yang, S. and Gómez-Bombarelli, R. Chemically transferable generative backmapping of coarse-grained proteins. In *Proceedings of the 40th International Conference on Machine Learning*. PMLR, 2023.

Yesudasan, S., Wang, X., and Averett, R. D. Coarse-grained molecular dynamics simulations of fibrin polymerization: effects of thrombin concentration on fibrin clot structure. *Journal of Molecular Modeling*, 24:1–14, 2018.

Yu, A. et al. A multiscale coarse-grained model of the sars-cov-2 virion. *Biophysical Journal*, 120(6):1097–1104, 2021.

Zhang, Z. et al. Learning coarse-grained force fields for fibrogenesis modeling. *Computer Physics Communications*, 295:108964, 2024.

A. Table of Notations

Table 3: Explanation of symbols and notation in the paper.

Symbol	Meaning
$\mathbf{x}_i$	Coordinates at step i
n_i	Number of particles at step i
$\mathbf{z}_i$	Latent space variable at step i
Γ_i	Coarsening operator at step i
$p_{\theta_i}(\mathbf{x})$	Distribution of $\mathbf{x}$ with parameters θ_i
$\widehat{\mathbf{x}}_i$	Predicted coordinates at step i
$p(\mathbf{x}\|\mathbf{y})$	Conditional probability of $\mathbf{x}$ given $\mathbf{y}$
$D_{\mathrm{KL}}(p\\|q)$	Kullback–Leibler Divergence of p and q
ρ	Average CG bead size in unit of residue count

B. Detailed Derivations

Given two representations $\mathbf{x}_i, \mathbf{x}_{i+1}$, we derive the Evidence Lower Bound (ELBO):

$$\begin{aligned}
\log p_i(\mathbf{x}_i|\mathbf{x}_{i+1}) &= \log \mathbb{E}_{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\Big[\tfrac{p_i(\mathbf{x}_i|\mathbf{x}_{i+1})}{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\Big] \\
&\geq \mathbb{E}_{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\Big[\log \tfrac{p_i(\mathbf{x}_i|\mathbf{x}_{i+1},\,\mathbf{z}_i)\;p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1})}{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\Big] \\
&= \mathbb{E}_{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\big[\log p_{\theta_i}(\mathbf{x}_i|\mathbf{x}_{i+1},\,\mathbf{z}_i)\big] \;+\; \mathbb{E}_{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\Big[\log \tfrac{p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1})}{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\Big] \\
&= \mathbb{E}_{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\big[\log p_{\theta_i}(\mathbf{x}_i|\mathbf{x}_{i+1},\,\mathbf{z}_i)\big] \;+\; D_{\mathrm{KL}}\Big(p_{\psi_i}(\mathbf{z}_i|\mathbf{x}_{i+1}) \;\|\; q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})\Big),
\end{aligned} \tag{1}$$

In the second row, we have applied Jensen's inequality.

Moreover, we can combine multiple ELBO's to compute a lower bound for $\log p(\mathbf{x}_0|\mathbf{x}_k)$:

$$\begin{aligned}
\log p(\mathbf{x}_0|\mathbf{x}_k) &= \log \prod_{i=0}^{k-1} p_i\big(\mathbf{x}_i|\mathbf{x}_{i+1}\big) \\
&= \sum_{i=0}^{k-1} \log p_i\big(\mathbf{x}_i|\mathbf{x}_{i+1}\big) \\
&\geq \sum_{i=0}^{k-1} \Big(\mathbb{E}_{q_{\phi_i}(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1})}\big[\log p_{\theta_i}\big(\mathbf{x}_i|\mathbf{x}_{i+1},\,\mathbf{z}_i\big)\big] \\
&\quad + D_{\mathrm{KL}}\big(p_{\psi_i}\big(\mathbf{z}_i|\mathbf{x}_{i+1}\big) \,\|\, q_{\phi_i}\big(\mathbf{z}_i|\mathbf{x}_i,\,\mathbf{x}_{i+1}\big)\big)\Big).
\end{aligned} \tag{2}$$

C. Secondary Structure

The Ramachandran plots illustrating the dihedral angle distributions across the true FG and generated/backmapped structures.

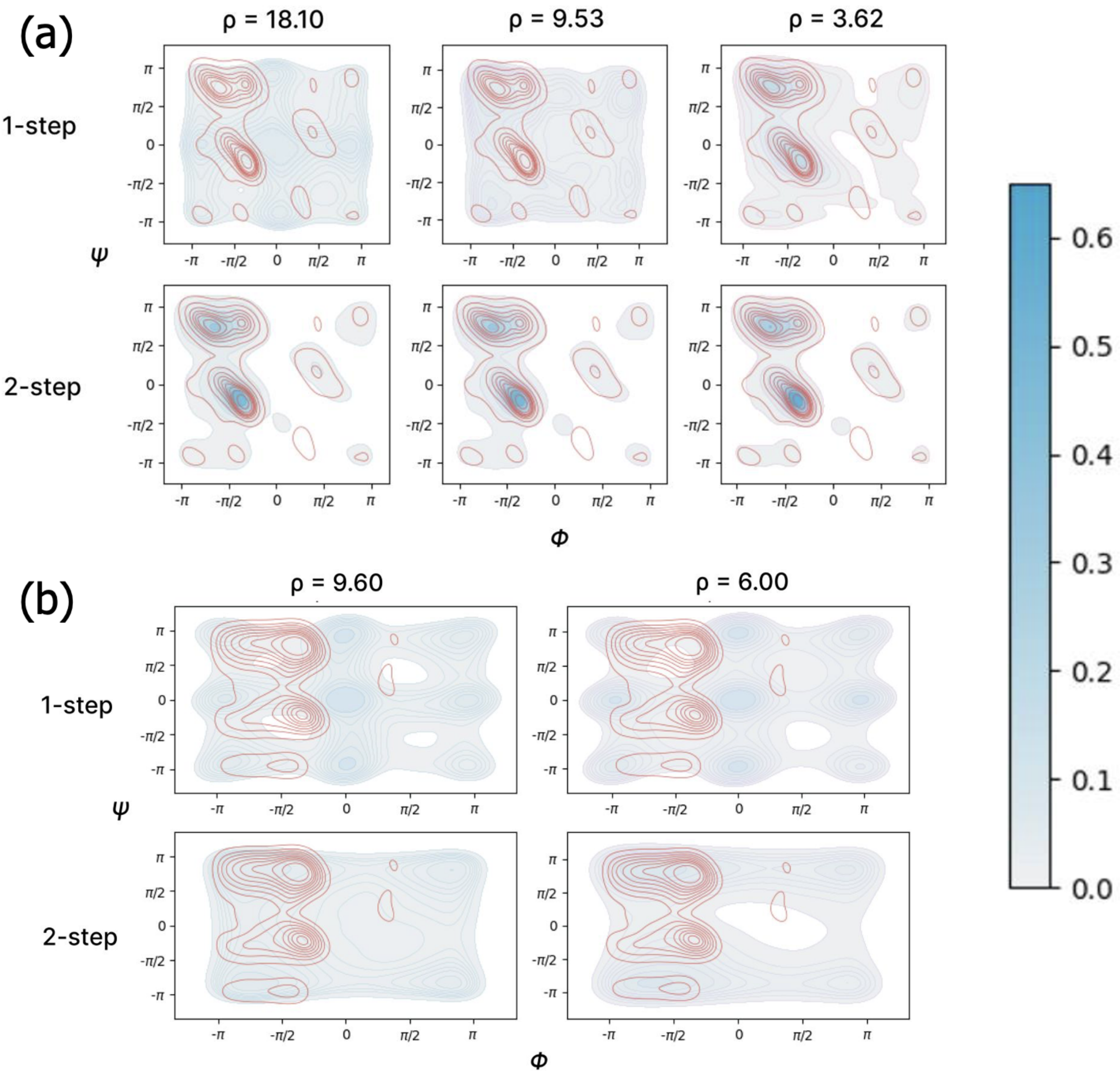

Figure 5: Ramachandran plots for different schemes and CG bead sizes: (a) eIF4E, (b) PED00151. The contours correspond to the true distribution while the color corresponds to the probability density of the (ϕ, ψ) angle combinations across the backmapped structures.

D. Hyperparameter Tables

We performed hyperparameter optimization using Weights & Biases (Biewald, 2020). For the UCG→RBCG or UCG→FG reconstructions, we employed CGVAE (Wang et al., 2022), tuned with the hyperband algorithm (Li et al., 2018) to minimize validation loss. For downstream backmapping (RBCG→FG), we fine-tuned a pretrained GenZProt, keeping its original hyperparameters (Yang & Gómez-Bombarelli, 2023).

Some comments related to the implications of memory overload in 1-step models:

- In 1-step schemes, we used a batch size of 1, with gradient accumulation every 2 steps, leading to an effective batch size of 2.
- For very coarse CG mappings, convolution depths had to be low.
- The node embedding dimension matches or exceeds the maximum number of mapped FG particles per bead.

Table 4: Hyperparameters for the 1-step backmapping (UCG → FG) of eIF4E using CGVAE.

# of atoms (FG)		$n_0 = 1488$	
Dataset size		3000	
# of CG beads	$n_1 = 10$	$n_1 = 19$	$n_1 = 50$
Hyperparameter		**Value**	
Edge feature dimension K	6	6	6
Graph loss weight γ	1.303	29.000	1.188
Encoder Convolution Depth	1	1	1
Prior Convolution Depth	1	1	1
Decoder Convolution Depth	1	4	4
FG cutoff d_{cut}	13.262	14.947	12.865
CG cutoff D_{cut}	31.738	33.064	23.657
Node embedding dimension F	400	400	400
Batch size	2	2	2
Learning rate	0.000078	0.000064	0.000054
Activation functions	swish	swish	swish
Training epochs	100	100	100
Order for multi-hop graph	3	1	1
Regularization strength β	0.002	0.040	0.002
Factor	0.210	0.131	0.106

Table 5: Hyperparameters for the first step of 2-step backmapping (UCG → RBCG) of eIF4E using CGVAE.

# of atoms (CG)		$n_1 = 177$	
Dataset size		3000	
# of CG beads	$n_2 = 10$	$n_2 = 19$	$n_2 = 50$
Hyperparameter		**Value**	
Edge feature dimension K	6	6	6
Graph loss weight γ	0.8014	2.6083	4.3745
Encoder Convolution Depth	4	6	3
Prior Convolution Depth	4	6	3
Decoder Convolution Depth	6	5	7
FG cutoff d_{cut}	40.296	47.771	43.03657
CG cutoff D_{cut}	81.525	69.453	115.64641
Node embedding dimension F	389	421	563
Batch size	2	2	2
Learning rate	0.0001814	0.0001719	0.00010339
Activation functions	swish	swish	swish
Training epochs	30	30	30
Order for multi-hop graph	1	1	6
Regularization strength β	0.000199	0.00541	0.022151
Factor	0.53875	0.72021	0.76064

Table 6: Hyperparameters for the 1-step backmapping (UCG → FG) of PED00151 using CGVAE.

# of atoms (FG)	$n_0 = 376$	
Dataset size	10000	
# of CG beads	$n_1 = 5$	$n_1 = 8$
Hyperparameter	**Value**	
Edge feature dimension K	6	6
Graph loss weight γ	8.0601	1.6931
Encoder Convolution Depth	1	2
Prior Convolution Depth	1	2
Decoder Convolution Depth	4	4
FG cutoff d_{cut}	14.8905	11.4050
CG cutoff D_{cut}	29.2027	22.0893
Node embedding dimension F	450	400
Batch size	1	1
Learning rate	0.000165	0.0000986
Activation functions	swish	swish
Training epochs	100	100
Order for multi-hop graph	1	1
Regularization strength β	0.0214	0.0311
Factor	0.2294	0.7787

Table 7: Hyperparameters for the 2-step backmapping (UCG → RBCG) of PED00151 using CGVAE.

# of atoms (FG)	$n_1 = 46$	
Dataset size	10000	
# of CG beads	$n_2 = 5$	$n_2 = 8$
Hyperparameter	**Value**	
Edge feature dimension K	6	6
Graph loss weight γ	2.4581	0.7310
Encoder Convolution Depth	4	5
Prior Convolution Depth	4	5
Decoder Convolution Depth	10	10
FG cutoff d_{cut}	47.9878	57.6470
CG cutoff D_{cut}	50.1431	100.343
Node embedding dimension F	569	507
Batch size	2	2
Learning rate	0.0000975	0.000116
Activation functions	swish	swish
Training epochs	30	30
Order for multi-hop graph	1	3
Regularization strength β	0.00057	0.00207
Factor	0.639	0.916

E. Comparison of CG Bead Sizes

As indicated in previous sections, this work explores reconstruction of UCG representations, unlike

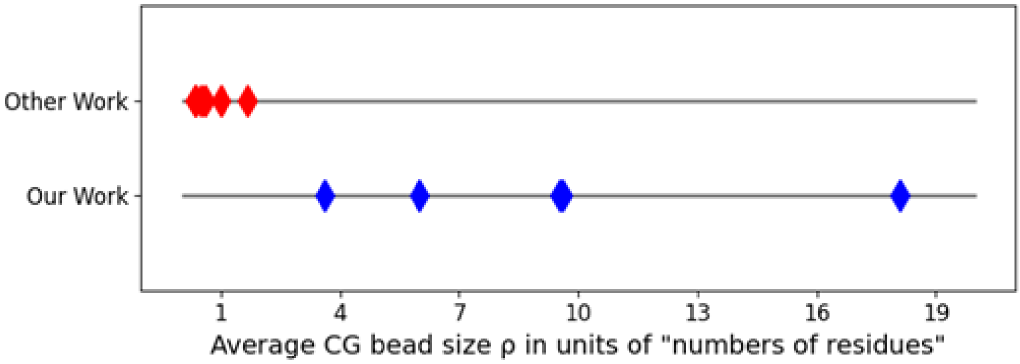

Figure 6: The CG bead sizes ρ in previous papers are much smaller— and closer to $\rho = 1.00$—than the ones we try in our work.

F. Radius of Gyration

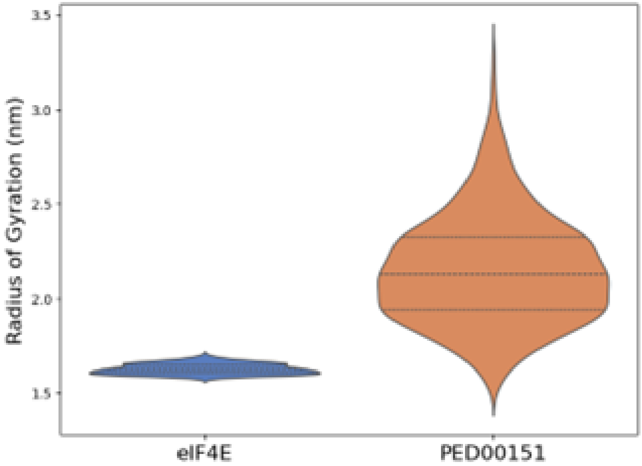

Figure 7: Radius of gyration distribution of the two proteins we test the schemes on. The difference in variance indicate eIF4E is more compact, while PED00151 is more flexible.

G. CG Representations

Using AUTOGRAIN (Wang et al., 2022) and MDTraj (McGibbon et al., 2015) we develop the various CG mapping operators and apply them to the FG datasets used in the experiments.

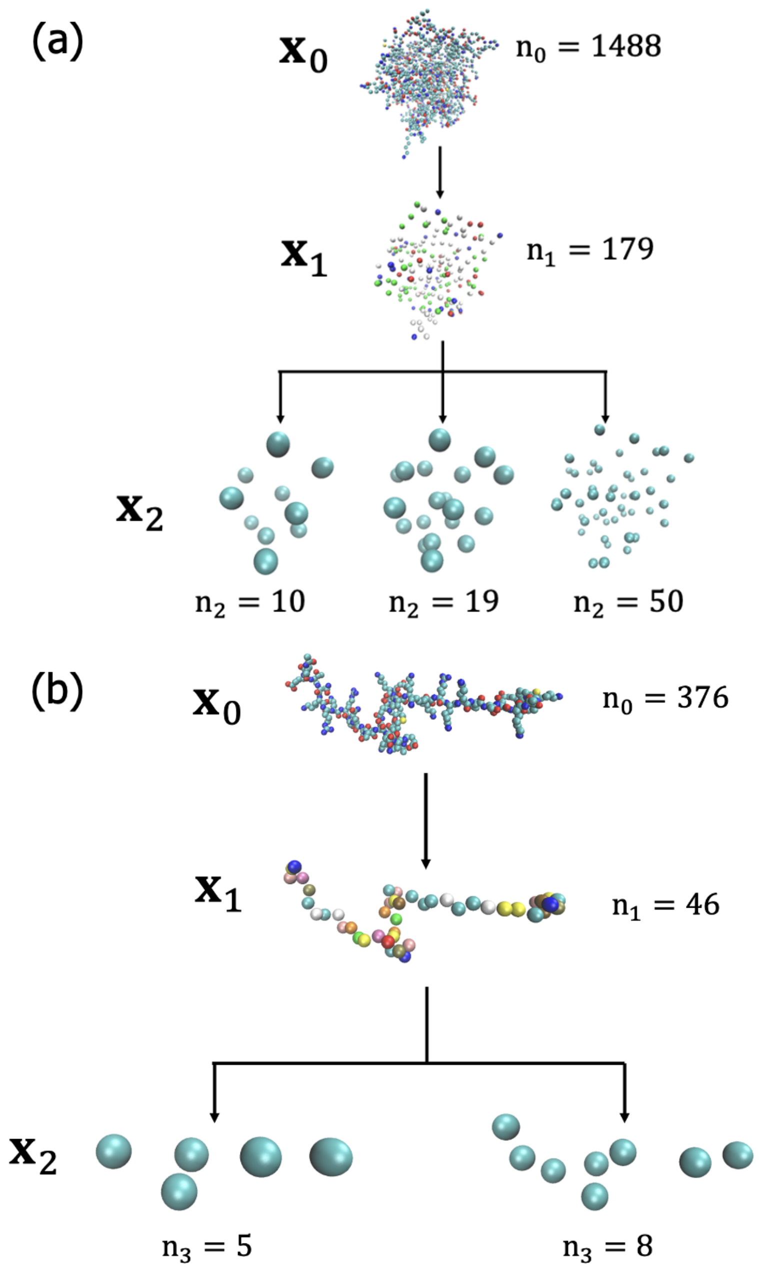

Figure 8: The different FG and CG representations of (a) eIF4E and (b) PED00151 used in our experiments.