Learning Protein Conformational Pathways with Physics-Informed Variational Autoencoders

Anaya Gandhi

Master of Science
Data Science
School of Informatics
University of Edinburgh
2026

Abstract

This thesis demonstrates the successful development of physics-informed variational autoencoders (VAEs) for protein conformational modeling, improving the derivation of biologically relevant structural transitions. A two-stage optimization technique used MurD as a test system to find the best reconstruction parameters with the Test 15 VAE configuration ($d_z = 8$, $\beta = 0.05$) and Config 82 physics constraints (bond + smoothness, realism score 0.4965). However, Test 15 exhibited severe mode collapse during interpolation, necessitating the development of a balanced VAE ($d_z = 64$) that fixed both mode collapse and over-diversity problems.

The VAE did well on all of the tests. Fifty interpolations with 1,000 decoded structures were made without any major failure. The reconstruction accuracy was 1.25 Å RMSD, the interpolation RMSD was between 0.87 and 1.31 Å, and the path smoothness was 0.048 $\pm$ 0.016 Å. Biological validation against experimentally established apo structures confirmed structural plausibility, with RMSD of 1.758 $\pm$ 0.017 Å and realistic secondary structure composition (22.1% helix, 29.0% β-sheet, 49.0% coil), producing an overall biological evaluation score of 0.767.

When compared to a FoldingNet-style CNN autoencoder baseline, the VAE's superiority was even more clear. The CNN's structural quality dropped sharply (DOPE score shift of 4,526.5 points) and it couldn't interpolate between conformations because its latent space was deterministic. The VAE, on the other hand, allowed for smooth transitions. These findings demonstrate that physics-informed VAEs offer a reliable, physically grounded, and physiologically essential framework for protein conformational modeling, with potential applications in drug discovery, protein design, and mechanistic research of protein function.

Research Ethics Approval

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Anaya Gandhi*)

Acknowledgements

From the initial concept to the final text, Professor De Giacomi provided consistent direction, insightful criticism, and support for which I am incredibly thankful.

I am also grateful to my family, particularly my parents, for their unwavering support and to my sister for providing the humor that helped me keep things in perspective.

Lastly, I want to thank my labmates and cohort for their support and friendship; without them, this year would have been much more difficult and less enjoyable.

Table of Contents

Chapter 1

Introduction

Proteins are dynamic molecular machines that carry out nearly every essential task in biology. They catalyze reactions, transmit signals, provide structural support, and regulate cellular processes [1]. Their function depends on their three-dimensional (3D) structure, which is decided by the amino acid sequence and stabilized through a range of interactions such as hydrogen bonds, hydrophobic contacts, van der Waals forces, and electrostatic forces. But proteins are not locked into a single shape, they are not static. Instead, they explore ensembles of conformations which make them dynamic in nature. These internal motions are directly tied to biological activities like substrate recognition, signal transduction, allosteric control, and enzymatic catalysis [2, 3]. This change in structure [4] allows proteins to adapt to cellular environments and to switch between functional states. Capturing this structural transitions has become one of the main challenges of molecular biology, which would have a major impact on drug discovery, enzyme design, and systems biology.

X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy (cryo-EM) have helped understand proteins through atomic resolution structures[5, 6]. However, these methods deliver static snapshots, and miss the continuum of conformations that underlie many functional transitions. Molecular dynamics (MD) simulations fills in this gap by producing atomistic videos of protein motions over time [2, 7]. MD has become a computational microscope for proteins, but its high cost limits the ability to observe rare events and transitions over long time periods [8]. Even extended simulations sometimes fail to capture the full conformational transformation.

To make sense of these dynamics, proteins are often described in terms of an energy landscape, where valleys represent metastable states and barriers represent transitions between them [9–11]. The behavior of the protein depends on the movement across these

states. The action of MurD, a bacterial ligase required for peptidoglycan production, is because of huge domain movements [12]. Substrates can enter the active site when it is open, and catalysis takes place when it is closed. The enzyme MurD is a verified antibiotic target needed for the formation of the bacterial cell wall (peptidoglycan), since it catalyzes the ATP-dependent addition of D-glutamate to UDP-N-acetylmuramoyl-L-alanine [12]. MurD is a three-domain enzyme, and the C-terminal domain rotates a lot, which is what makes it possible for catalysis to happen. Crystallographic and solution studies have shown that MurD displays open, semi-closed, and closed conformations, with experimentally derived structures accessible for the open (PDB: 2UAG) and closed (PDB: 1NVM) states [13]. These properties make MurD an excellent system or testing physics-informed generative modeling.

For many years, computational approaches outside of MD have attempted to resolve the accuracy vs. efficiency trade-off. Although they lose atomic detail, coarse-grained simulations [14], normal mode analysis [15], and elastic network models [16] are capable of capturing large-scale motions. In recent times, machine learning (ML) has offered strong instruments for modeling high-dimensional biological data [17]. AlphaFold is one of the discoveries that has changed the field. It uses deep learning, evolutionary information, and attention mechanisms to predict static protein structures with great accuracy [18, 19]. AlphaFold is a big step forward, but it only looks at the most likely fold and doesn't take into account how flexible the shape is. Geometric deep learning has grown through the use of symmetry-aware models that know the rotational and translational invariances that exist in three-dimensional molecule space [20, 21]. Frameworks like the geometric vector perceptron have taken these ideas and applied them to proteins, allowing networks to learn about structural patterns in a way that is consistent with physics [22].

Generative models improve these prediction methods by learning not only the specific conformations of proteins but also the distributions of states they can take. For example, Variational Autoencoders (VAEs) can still make new observations while mapping high-dimensional structures into low-dimensional latent spaces [23]. This makes them good for structural biology problems where experimental datasets don't have intermediate conformations. People have utilized VAEs before to look at conformational manifolds that aren't directly observable and to interpolate between different protein states [24–26].

However, a major problem with data-driven generative models is that they could make structures that aren't realistic, with steric conflicts, twisted bond lengths, or

impossible dihedral angles if there aren't any limits set. Thus, recent studies have focused on physics-informed learning, in which models are taught with structural statistics and molecular mechanics priors. These include Ramachandran distributions that form backbone geometry [27] and force-field-based energy terms [28].

This thesis grows the idea of physics-informed generative modeling through the creation of a physics-informed Variational Auto-Encoder (VAE) designed to study the conformational ensembles of the MurD enzyme. The framework uses geometric energy losses from molecular mechanics to penalize deviations from ideal bond lengths, angles, and steric interactions. It also uses Ramachandran-based terms to make sure that the distributions of backbone dihedral angles stay physically realistic. A Kullback–Leibler (KL) divergence regularizer, on the other hand, encourages continuity in the latent space, which makes it possible to smoothly interpolate between different conformations. The model is meant to generate intermediate states that are between MurD's known open and closed conformations and are both structurally valid and physiologically significant. Similar methods have recently been shown within the field of physics-informed variational autoencoders in dynamical systems [29], and in the generative modeling of protein ensembles such as K-Ras [30], underscoring the necessity for physical constraints to guarantee plausibility in biomolecular contexts [31].

Two significant gaps motivate this project. First, many protein modeling studies that use VAEs tend to report results based solely on final epoch performance, thereby risking overfitting and potentially misrepresenting model generalization This thesis circumvents that pitfall by monitoring validation performance throughout training, selecting the model snapshot with the best held-out validation metrics as the basis for reported results. Second, few studies undertake a systematic parameter sweep that rigorously balances reconstruction quality, latent-space continuity, and physics-informed loss terms. Addressing this, the thesis employs exhaustive hyperparameter exploration to offer a replicable methodology for model selection, enabling transparent and interpretable tuning of competing objectives. While studies such as [30] demonstrate the utility of VAEs for ensemble generation, they typically do not detail systematic hyperparameter sweeps or multi-objective optimization. Comprehensive reviews emphasize that integrating physical principles with generative modeling requires such careful balancing to ensure robustness and interpretability [31].

Chapter 2

Related Work and Background

2.1 Static Protein Structure Prediction

Homology modeling was a big part of early attempts to predict protein structure. Tools like MODELLER and SWISS-MODEL aligned unknown protein sequences with structures in the Protein Data Bank (PDB) and made models that were similar when near homologues were available [32, 33]. These approaches worked well for proteins with conserved folds, but they didn't work well for proteins with new structures or parts that are disordered.

Deep learning changed things in a big way. AlphaFold2 [18] was a big step forward since it combined attention based neural networks with evolutionary couplings. This made structures that were almost as accurate as experiments in CASP14 [19]. RoseTTAFold offered a different but similar design that made it easier for everyone to perform accurate modeling [34]. Even though they have an effect, these approaches only anticipate one conformation that is most likely. This disregards conformational plasticity, which is often fundamental to protein functionality. For example, enzymes can be active or inert, and allosteric proteins change their shape when they bind to a ligand. Even add-ons like AlphaFold-Multimer only capture static complexes, not dynamic structural ensembles.

2.2 Protein Dynamics and Molecular Simulations

Proteins are always shifting. Their usefulness usually results from moving between different structures, not from one structure alone. Experimental methodologies, like NMR relaxation dispersion, single-molecule FRET, and cryo-EM variability analysis,

explain specific aspects of this flexibility, but they offer only partial insights into the entire dynamic environment.

Molecular dynamics (MD) simulations bridge this gap by using empirical force fields to solve Newton's equations of motion for biomolecular systems [2, 35]. MD offers atomic-level resolution and temporal detail that ranges from femtoseconds to microseconds. Specialized hardware like Anton has made it easier to access longer timescales [8], but it is still hard to sample rare-event transitions and large conformational rearrangements [7].

It is now well known how important conformational ensembles are. Structural flexibility allows for enzyme catalysis, substrate selectivity, and allosteric regulation [3]. Databases like CoDNaS gather information about conformational diversity that have been seen in experiments [36], while computational models try to make predictions about states that have not been seen before. Elastic network models and anisotropic network models are two types of reduced-order approaches that use harmonic potentials to simulate collective motions [16, 37, 38]. Markov state models (MSMs) and their deep learning extensions, including VAMPnets, acquire sluggish collective coordinates and yield estimates of long-timescale kinetics [39, 40]. These methods show the main problem: finding the right balance between accuracy, interpretability, and computing cost while trying to capture conformational dynamics.

2.3 Machine Learning for Protein Modeling

Machine learning is now a fundamental component of structural biology. Deep generative models now try to capture the range of protein conformations, not just provide static predictions. Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and diffusion models are some of the most important approaches [23, 41–43].

Variational Autoencoders (VAEs) provide smooth latent spaces that enable interpolation between conformations [24]. While this makes them interpretable and suitable for exploring conformational ensembles, they often suffer from producing oversmoothed structures.

2.3.1 Limitations of Standard VAEs in Protein Modeling

While Variational Autoencoders offer smooth latent spaces suitable for conformational interpolation, they suffer from several limitations when applied to protein structure

generation. The major issue is oversmoothing, where the model produces averaged, low-resolution structures that lack the fine geometric details essential for protein function [23, 24]. This occurs because VAEs optimize for reconstruction loss across the entire dataset, leading to blurry outputs that represent statistical averages rather than physically realistic conformations.

A related problem is mode collapse, where the model fails to capture the full diversity of conformational states, instead generating structures that cluster around the most common conformations in the training data [44]. This is particularly problematic for proteins like MurD that show distinct functional states with different structural characteristics. Standard VAEs may generate structures that are intermediate between open and closed states but lack the specific geometric constraints that make these states biologically functional [30].

Also, VAEs do not understand much about molecular physics, which can lead to structures that break basic stereochemical rules. Common problems are bond lengths that are too long or too short, dihedral angles that can't be formed, steric conflicts between atoms, and conformations that are not energetically favorable [45, 46]. These issues tend to occur when interpolating between distant conformational states, leading to the model to produce intermediate structures without direct experimental analogs.

Physics-informed constraints address these limitations by incorporating molecular mechanics directly into the training process. Ramachandran-based losses ensure that backbone dihedral angles remain within physically allowed regions, while bonded geometry terms maintain realistic bond lengths and angles. Van der Waals and electrostatic terms prevent steric clashes and promote energetically favorable interactions [27, 47]. These constraints guide the model toward generating structures that are not only statistically consistent with the training data but also physically plausible and potentially functional.

In contrast, Generative Adversarial Networks (GANs) can generate sharper and more detailed protein samples [48], but their latent spaces tend to be less structured and their training dynamics are often unstable. More recently, diffusion models have emerged as state-of-the-art approaches for protein design [49], demonstrating the ability to generate high-quality backbones and functional binders. However, these models demand massive datasets and substantial computational resources, which limits their accessibility for many research settings.

Model	Strengths	Limitations
VAE	Smooth latent space, efficient training, interpretable	Outputs may lack sharpness, need physics priors
GAN	Sharp structures, strong data fit	Hard to train, unstable, less interpretable
Diffusion	High-quality, flexible generation	Requires large datasets, expensive training

Table 2.1: *Comparison of generative models for protein structure*

2.4 Physics-Informed and Geometry-Aware Models

The drawback of data-driven techniques in structural biology is the absence of assured physical plausibility. Without clear limits, models can make incorrect conformations that have deformed bonds, steric conflicts, or impossible torsional states. This challenge has led to the advancement of physics-informed and geometry-aware deep learning methods.

Approach	Key Idea	Applications
Boltzmann Generators	Sample thermodynamic ensembles with normalizing flows	Explore rare events, free energy landscapes
TorchMD-NET	Equivariant force and energy prediction from MD data	Acceleration of MD, force learning
EGNN	Encode rotational/translational invariance in GNNs	Protein generation, molecular dynamics
SE(3)-Transformers	Full 3D equivariance	Protein design, docking, dynamics
PI-VAE	Add Ramachandran + bonded geometry losses	Coordinate-based conformational modeling

Table 2.2: *Physics-informed strategies for generative protein models*

For example, the Boltzmann Generators use normalizing flows to sample equilibrium ensembles that are consistent with molecular dynamics (MD) [46], and TorchMD-NET uses equivariant transformers to learn forces and energies directly from MD trajectories [50]. Equivariant Graph Neural Networks (EGNNs) and SE(3)-Transformers are two other advances. EGNNs put rotational and translational invariance directly into graph-based architectures [51], and SE(3)-Transformers generalize equivariance to the full 3D rotation group for molecular tasks [52]. More recently, Physics-Informed Variational Autoencoders (PI-VAEs) have added stereochemical priors such Ramachandran distributions and bonded geometry to their loss functions to make sure that the structures are

valid [26, 45].

The Ramachandran map shows sterically permissible backbone dihedral angles [27], and force fields like AMBER ff14SB define bonded, torsional, and nonbonded interactions [47]. These are still important in these methods. These priors can be restructured into differentiable penalties, allowing them to guide the training of modern generative models.

2.4.1 Systematic Model Validation and Hyperparameter Optimization

To make generative models that are based on physics, one has to find a balance between reconstruction accuracy, latent space continuity, and physical plausibility. These goals typically work against each other: stronger physics terms may make stereochemistry better but make reconstruction worse, and lowering KL divergence may make accuracy better but make latent continuity worse [30, 44].

Numerous prior studies present outcomes from final training epochs or depend on ad-hoc hyperparameters, which may lead to overfitting and unreliable comparisons [53]. Instead, systematic sweeps across important parameters such physics loss weights, KL coefficients, and learning rates let one observe how different goals affect each other [54, 55].

Validation-based checkpointing is similarly important because it makes sure that the results show generalization and not noise from training [28]. In this study, model selection integrated conventional criteria (reconstruction error, latent smoothness) with physics-based evaluations (Ramachandran quality, energy scores). A Pareto-style methodology [56] was employed to identify balanced models appropriate for interpolation and structural analysis.

Chapter 3

Methodology

3.1 Data Sources and Provenance

All molecular dynamics (MD) trajectories and base PDB structures used in this work were pre existing and obtained from prior work or public sources. No new MD simulations were generated as part of this thesis. The provided datasets comprised the MurD ligase in several functional states (open, closed, and, where available, apo), together with curated reference PDB files. Coordinates were first stored in or cross-referenced with the Protein Data Bank (PDB) [57]. Downstream preparation, used standard Python tools for structural data: MDAnalysis for trajectory I/O and atom selections [58, 59], MDTraj for geometric features like dihedral angles [60], and the Kabsch algorithm for frame alignment using rigid-body superposition [61]. As noted in later sections, the data utilities were used from the molearn framework to load, batch, and keep track of data between states [62]. Reference structures functioned as alignment goals, quality-control anchors, and endpoints for latent-space interpolation analyses, while the temporally ordered MD frames formed the training, validation, and test databases for the generative models.

3.2 Preprocessing and Quality Control

MDAnalysis [58, 59] and custom Python scripts were used to process the trajectories. Two atomic representations were looked at a backbone only representation (N, Cα, C, O) that was used for most of the experiments because it was stable and easy to work with; and a full-atom (heavy-atom) representation that was used in some studies to see how far it might go and how nonbonded terms affected it. The findings that will be

published later are based on the backbone-only setting, unless otherwise noted.

Using the Kabsch algorithm [61], all frames were carefully superposed onto a reference frame (first frame or reference PDB) for each conformational state. This got rid of global translations and rotations. Backbone atoms (N, Cα, C, O) were used for superposition and reduce the RMSD between frames. The Kabsch algorithm finds the best rotation matrix that makes the root-mean-square variation between matching atoms as small as possible. This removes global translational and rotational motion while keeping interior conformational changes. Within each state, the Cartesian coordinates $\mathbf{x} \in \mathbb{R}^{3N}$ were set to have a mean of zero and a variance of one.

$$\mathbf{x}' = \frac{\mathbf{x} - \mu}{\sigma}, \tag{3.1}$$

where μ and σ are calculated independently for each conformational state (open, closed, apo) to ensure the retention of state specific conformational variations while maintaining numerical stability for training. Quality control removed frames from the state mean that had a backbone RMSD greater than 3σ. This threshold was set to leave out frames with large structural changes that could be artifacts of the simulation or infrequent conformational states that don't represent the main ensemble. This way, the natural conformational diversity within each functional state is kept. To evaluate stereochemistry, backbone dihedrals (ϕ, ψ) were calculated from Cartesian coordinates using MDTraj [60].

Item	**Setting**	**Notes**
Representation	Backbone (N, Cα, C, O); Full-atom (heavy)	Backbone used for primary analyses; full-atom tested in selected experiments
Alignment	Kabsch to reference frame/PDB [61]	Removes global translation/rotation
Standardization	Per-state μ, σ (Eq. 3.1)	Computed over all frames in each state
Outlier exclusion	Backbone RMSD $> 3\sigma$ from state mean	Excluded from all splits (train/val/test)
Dihedrals	MDTraj (ϕ, ψ) for Ramachandran checks [60]	Computed post-alignment
File integrity	Atom order/element validation	Consistent atom order across all frames and states

Table 3.1: *Preprocessing and quality-control parameters applied per conformational state.*

Using stratified temporal sampling, the data was divided into three sets: training

(70%), validation (15%), and test (15%). This method keeps the conformational diversity between splits while reducing temporal autocorrelation, which could cause data to leak between the training and evaluation sets. The validation set was utilized to find the best hyperparameters and to halt the process early, while the test set was saved for the final model evaluation.

State	**Total frames**	**Train (70%)**	**Val (15%)**	**Test (15%)**
MurD (open)	900	630	135	135
MurD (closed)	900	630	135	135
MurD (apo)	900	630	135	135
All (multi-state)	2700	1890	405	405

Table 3.2: *Dataset split sizes per state after QC.*

3.3 Modeling Framework

Variational Autoencoder (VAE). Let $\mathbf{x} \in \mathbb{R}^{3N}$ denote Cartesian coordinates for N selected atoms. A VAE [23] learns a probabilistic encoder $q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mu_\phi(\mathbf{x}), \mathrm{diag}(\sigma^2_\phi(\mathbf{x})))$ and a decoder $p_\theta(\mathbf{x}|\mathbf{z})$ that reconstructs coordinates from latent variables $\mathbf{z} \in \mathbb{R}^{d_z}$. Sampling uses the reparameterization trick,

$$\mathbf{z} = \mu_\phi(\mathbf{x}) + \sigma_\phi(\mathbf{x}) \odot \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I), \tag{3.2}$$

and training maximizes the evidence lower bound (ELBO),

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \beta \, \mathrm{KL}\big(q_\phi(\mathbf{z}|\mathbf{x}) \,\|\, \mathcal{N}(0, I)\big), \tag{3.3}$$

with β modulating the reconstruction–regularization trade-off [63].

Neural architectures. All models are executed in PyTorch [64]. Three different types of encoders and decoders. were looked at First, a multilayer perceptron (MLP) encoder transforms the flattened coordinate vector into the latent parameters $(\mu_\phi, \log \sigma^2_\phi)$. Second, an attention-based and hybrid convolutional encoders (self-attention blocks stacked over a shallow 1D CNN front-end) to see how expressive they were on structured inputs. Finally, the decoder is a coordinate generator based on FoldingNet [65] that

has been changed to work with proteins. Since canonical FoldingNet gives outputs in fixed multiples, hence explicit output truncation/masking were used to make sure that the last layer gives exactly $3N$ coordinates in the order of the original atoms. This fixes the problem of tensor sizes not matching up while keeping the learned coordinate structure. In all of the studies, a of dropout (0.1) and weight decay of (1×10^{-5}) as light regularizers were used

Component	**Structure**	**Notes**
Encoder (MLP)	FC($3N \rightarrow 512$)→ReLU→FC($512 \rightarrow 256$)→ReLU→FC($256 \rightarrow 2d_z$)	outputs μ, $\log\sigma^2$
Encoder (Attn/CNN)	1D CNN ($3N \rightarrow 128$)→Self-Attn→FC($128 \rightarrow 2d_z$)	ablation experiments only
Decoder (FoldingNet-style)	FC($d_z \rightarrow 512$)→ReLU→FC($512 \rightarrow 1024$)→ReLU→Folding layers→FC($\cdot \rightarrow 3N$)	output truncation/masking to $3N$
Regularizers	Dropout (0.1), weight decay (1×10^{-5})	Used in all main experiments

Table 3.3: *Neural network architectures used in this work.*

3.4 Physics-informed Objective

Pure reconstruction can yield physically implausible structures. Therefore the loss must be augmented stereochemical penalties:

$$L_{\text{total}} = L_{\text{MSE}} + \beta L_{\text{KL}} + w_{\text{geo}} L_{\text{geo}} + w_{\text{rama}} L_{\text{rama}}, \tag{3.4}$$

where

$$L_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2, \qquad L_{\text{KL}} = -\frac{1}{2}\sum_{j=1}^{d_z} \left(1 + \log\sigma_j^2 - \mu_j^2 - \sigma_j^2\right). \tag{3.5}$$

The geometric energy term reflects AMBER-style bonded forms [28, 47],

$$L_{\text{geo}} = \sum_{\text{bonds}} k_b (l - l_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} k_\phi [1 + \cos(n\phi - \phi_0)], \tag{3.6}$$

and the Ramachandran term penalizes backbone torsions falling outside empirically allowed regions [27],

$$L_{\text{rama}} = \frac{1}{N_{\text{res}}}\sum_{r=1}^{N_{\text{res}}} \mathbb{1}[(\phi_r, \psi_r) \notin \mathcal{R}_{\text{allowed}}]. \tag{3.7}$$

All physics terms are computed on de standardized Cartesian coordinates (in Å) to preserve physical units.

Not all physics terms were enabled in every experiment. Backbone-only models used L_{geo} and L_{rama} at modest weights; full-atomexperiments additionally tested short-range nonbonded clash penalties (e.g., Lennard–Jones) but these were often disabled due to memory cost and instability [47, 66]. Across sweeps, moderate bonded/dihedral weighting improved plausibility without eroding reconstruction; overly strong angle/Ramachandran weights could destabilize training despite numerical safeguards (domain clamping, ε offsets, NaN checks). Overall, careful tuning of constraint weights proved crucial: excessive weights destabilized training, while insufficient weights permitted geometric artifacts. Systematic optimization indicated that bond-length and smoothness constraints offered the best balance between physical realism and stability for backbone-only models [25, 26, 30].

3.5 Physics Constraint Optimization Protocol

To systematically identify optimal constraint configurations, A grid search was conducted that varied the types and weights of physics constraints while maintaining training stability [54, 55]. Four constraints were explored: bond-length terms that maintain realistic covalent distances between consecutive backbone atoms; bond-angle terms that preserve backbone geometry; Ramachandran constraints that keep (ϕ, ψ) within empirically allowed regions; and smoothness terms that promote gradual conformational changes along the backbone.

The optimization strategy spanned 16 combinations of enabled/disabled constraints, per-constraint weights $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$, and three starting conditions with initial constraint values of 0, 10, or 20. In total this yielded 48 configurations, each trained under the protocol in Section 3.7 and evaluated with the metrics in Section 3.8.

Validation considered final total loss (reconstruction plus physics terms), a realism score (DOPE-based assessment; see Section 3.8) [32, 67], training stability (absence of catastrophic failures), and reconstruction quality (validation MSE). For backbone-only experiments, DOPE scoring was applied after sidechain rebuilding; otherwise, stereochemical metrics (Ramachandran, bonded deviations, clash proxy) were used. Configurations were graded mostly based on how realistic they were, while also making sure they were stable and could be rebuilt. The best configuration was then chosen for downstream pathway analysis and model comparisons. [53, 56].

3.6 Hyperparameter Search

A practical grid search over a subset of the full parameter space was conducted, refining ranges based on pilot runs and stability constraints [53–55]. Latent dimensionality $d_z \in \{2,4,8,16\}$, KL weight $\beta \in [10^{-3}, 0.5]$ [63], geometry weight $w_{\text{geo}} \in [10^{-6}, 10^{-2}]$, and Ramachandran weight $w_{\text{rama}} \in [10^{-5}, 10^{-2}]$ were explored. too tiny physics weights allowed geometric artifacts, while too big values made optimization unstable. The configuration used for downstream analysis was $d_z = 8$, $\beta = 0.05$, $w_{\text{geo}} = 5 \times 10^{-5}$, $w_{\text{rama}} = 3 \times 10^{-4}$.

Group	Parameter	Values Tested
Latent space	d_z	{2, 4, 8, 16, 128} (**final: 8** for backbone, **128** for multi-state)
Regularization	β	10^{-3}, 10^{-2}, 5×10^{-2}, 10^{-1}, 5×10^{-1} (**final: 0.05** or **0.01**)
Physics (geometry)	w_{geo}	10^{-6}, 5×10^{-6}, 10^{-5}, 5×10^{-5}, 10^{-4}, 10^{-3}, 10^{-2} (**final:** 5×10^{-5})
Physics (Rama)	w_{rama}	10^{-5}, 3×10^{-5}, 10^{-4}, 3×10^{-4}, 10^{-3}, 10^{-2} (**final:** 3×10^{-4})
Architecture	hidden widths/depths	2–3 layers, 256–1024 units
Optimization	LR, weight decay	1×10^{-4}–1×10^{-3}, 1×10^{-5}
Batch size	–	2–40 (model-dependent)
Physics optimization	Config combinations	192 total configurations tested
	Best config selection	Config ranking by realism score

Table 3.4: *Hyperparameter search space and final values.*

3.7 Training Protocol, Curriculum, and Hardware

Separate models were trained for the open and closed ensembles to preserve state-specific structure and avoid mode collapse; a multi-state model (combined open/closed/apo) was also explored to assess generalization. Optimization used Adam [68] with learning rate in the range 10^{-4}–10^{-3}, and batch sizes adjusted to available memory and model complexity (from 2 up to 40). Models were trained up to 100 epochs with early stopping

after 20 epochs without validation improvement. A best-epoch protocol was used: after each epoch, validation loss was computed and the checkpoint with the minimal validation loss was retained for all downstream analyses [53].

To stabilize optimization while mixing different loss terms, physics weights were increased from almost zero to target values throughout the first few epochs (a curriculum schedule) [69]. Table 3.5 shows an example schedule.

Epoch range	w_{geo}	w_{rama}	**Notes**
0–50	linearly $0 \to 0.5\,w^{\star}_{\text{geo}}$	linearly $0 \to 0.5\,w^{\star}_{\text{rama}}$	warmup
51–100	$0.5\,w^{\star}_{\text{geo}} \to w^{\star}_{\text{geo}}$	$0.5\,w^{\star}_{\text{rama}} \to w^{\star}_{\text{rama}}$	settle
>100	$w^{\star}_{\text{geo}}$	$w^{\star}_{\text{rama}}$	fixed

Table 3.5: *Curriculum schedule for physics weights.*

Item	**Setting**
Optimizer & LR	Adam; 1×10^{-4}–1×10^{-3}
Batch size & epochs	2–40 (model-dependent); up to 150 epochs
Early stopping	Patience 20 (validation loss)
Checkpointing	Best-epoch (minimum validation loss)
Curriculum	Linear ramp of w_{geo}, w_{rama} over 100 epochs
Random seeds	$\{42, 123, 456\}$ (reported where applicable)
Hardware	NVIDIA RTX A4000 (15GB VRAM), AMD EPYC 7502 32-Core CPU, Ubuntu 22.04.5 LTS, CUDA 12.2, Python 3.11.13
Pathway analysis	20-21 points per pathway, outlier detection and Statistical validation vs training data
Model comparison	VAE vs CNN autoencoder, multi-seed and Kolmogorov-Smirnov testing [70]

Table 3.6: *Training and hardware settings.*

3.8 Evaluation Protocol

The evaluation used the withheld test set and the best epoch checkpoints determined by the lowest validation loss. Structural accuracy was measured by calculating the backbone RMSD between reconstructed and reference coordinates after using the Kabsch algorithm [61] to get rid of global translational and rotational discrepancies. KL

Package/Tool	Version
Python	3.11.x
PyTorch	2.x (CUDA 11/12)
MDAnalysis	2.x
MDTraj	1.9.x
NumPy / SciPy	1.26.x / 1.11.x
VMD	1.9.x
OS / Driver	Ubuntu 22.04 LTS / NVIDIA 53x.xx

Table 3.7: *Key software and environment versions.*

divergence was used to compare the learnt posterior and prior distributions to see how well the model could learn useful representations while still being able to smooth out the data. We used bonded geometry penalties (L_{geo}) and Ramachandran statistics [27] to check for stereochemical plausibility. The latter used MDTraj's dihedral analysis functions [60] to compare backbone torsion angle distributions to empirically allowed regions.

For certain experiments, ensemble-level similarity was analysed using the two-sample Kolmogorov–Smirnov (KS) test on ensemble metric distributions (e.g., R_g, endpoint RMSD, bond/angle deviations) [70], carried out with SciPy [71]. This statistical test informs how different the metric distributions for model generated structures are from those in the original training ensemble. It provides a way to quantify how effectively the model reflects the underlying conformational ensemble. To check for reproducibility, the model retrained with three random seeds and reported the coefficient of variation (CV) of the validation loss. CV values below 10% were judged acceptable for stable training.

To find out whether the learned manifold allows physically plausible transitions, we encoded the open and closed reference structures to derive $\mathbf{z}_{\text{open}}$ and $\mathbf{z}_{\text{closed}}$, then generating a linear interpolation path defined by the equation

$$\mathbf{z}(\alpha) = (1-\alpha)\,\mathbf{z}_{\text{open}} + \alpha\,\mathbf{z}_{\text{closed}}, \qquad \alpha \in [0,1], \tag{3.8}$$

which is then decoded to obtain intermediate structures. This linear interpolation method checks to see if the model can make physically plausible intermediate conformations between the known functional states. RMSD is tracked to endpoints, radius of gyration (R_g), and stereochemical quality metrics such bond length deviations, angle violations, and Ramachandran plot statistics along these pathways. Structures were exported as

PDB files for qualitative examination in VMD [72], allowing visual inspection for conformational continuity and physical plausibility.

Item	**Setting**	**Notes**
Endpoints	Encoded MurD open / closed references	$\mathbf{z}_{\text{open}}$, $\mathbf{z}_{\text{closed}}$
Path	Linear: $\mathbf{z}(\alpha) = (1-\alpha)\mathbf{z}_{\text{open}} + \alpha\mathbf{z}_{\text{closed}}$	$\alpha \in [0,1]$
Steps	20 or 21 per pathway (typically 50 pathways)	Uniform spacing in α
Monitored metrics	RMSD to endpoints, R_g (radius of gyration), bond lengths, smoothness, stereochemistry flags	Methods only (no results here)
Export	PDB series for visualization (VMD)	Same atom order as inputs
Physics optimization	192 configurations, grid search over weights	Config ranking by realism score
Pathway analysis	20-21 points per pathway, outlier detection	Statistical validation vs training data
Model comparison	VAE vs CNN autoencoder, multi-seed	Kolmogorov-Smirnov testing

Table 3.8: *Latent-space interpolation protocol.*

3.9 Pathway Quality and Robustness Analysis Protocol

To assess the quality and reliability of generated conformational pathways,linear latent-space interpolations were generated between encoded reference structures,

$$\mathbf{z}(\alpha) = (1-\alpha)\,\mathbf{z}_{\text{open}} + \alpha\,\mathbf{z}_{\text{closed}}, \quad \alpha \in [0,1], \tag{3.9}$$

where $\mathbf{z}_{\text{open}}$ and $\mathbf{z}_{\text{closed}}$ denote the encodings of the open and closed conformations, respectively. This follows standard VAE interpolation practice for probing manifold continuity in protein-ensemble models [26]. Each pathway comprised 20–21 uniformly spaced intermediate structures.

After that, complementary measures, as stated in Section 3.8, were used to evaluate the pathway points. Structural similarity was measured by backbone RMSD to the appropriate reference following optimal superposition using the Kabsch algorithm

[61]. We used DOPE-based scores as a proxy for stereochemical plausibility to check physical realism [67], and geometric consistency was tracked by looking at how much bond lengths and angles were off from ideal values that were consistent with standard protein geometry (for example, ff14SB targets) [47]. The smoothness of the path was examined by looking at how much the coordinates changed along the interpolation sequence to make sure the transitions were gradual and not random.

The reliability was evaluated by identifying outliers (points above 2σ from ensemble averages for the selected metrics), examining for unphysical structures or major breakdowns, and confirming the constant nature of pathway quality across interpolation positions. Summary statistics (means, standard deviations, and distributional shapes) were calculated for all quality criteria to get a sense of both the central tendency and the spread. To validate against the underlying conformational ensemble, pathway-derived distributions were compared to those from the training data using the two-sample Kolmogorov–Smirnov test [70], implemented through SciPy [71]. This process checks to see if the pathway statistics that were made are statistically indistinguishable from those of the scientific ensemble while still being structurally and stereochemically sound.

3.10 Model Comparison Protocol: VAE vs. CNN Autoencoder

A controlled comparison of the variational technique versus a CNN autoencoder baseline was conducted to see how well it succeeded. The same data preprocessing (Section 3.2), training methods (Section 3.7), and evaluation metrics (Section 3.8) were used. The baseline used a FoldingNet-style architecture adapted for protein coordinates, in line with the coordinate decoder family for a valid architectural comparison [65]. The same DOPE-based assessment [67] were used to check both models for reconstruction fidelity (MSE on train/test), generation capability (the ability to create new, physically plausible structures), interpolation quality along the pathways described above (Section 3.9), generalization to unseen conformations (like apo), and physical plausibility. To make sure the results were statistically valid, both models were trained with the same random seeds to account for random variation, and their performance distributions were compared using Kolmogorov–Smirnov tests [70] (SciPy implementation [71]).

3.11 Reproducibility

The best-epoch model selection to eliminate optimistic bias from late-epoch drift and overfitting was used. This made sure that the results published were based on real model performance and not on training artifacts [53]. As needed, multi-seed retraining were carried out to measure variability and see how strong the results were across multiple random initializations [53]. To make sure that training stayed stable, numerical safeguards were used on all loss functions. For example, domain clamping kept trigonometric functions from working outside of valid ranges, small ε denominators kept division by zero from happening, angle wrapping to $(-\pi, \pi]$ kept periodicity correct, and NaN filtering kept gradient explosion from happening [53].

Item	Setting	Reported
Random seeds	$\{42, 123, 456\}$	Yes (selected experiments)
Best-epoch selection	Minimum validation loss	Always
Stability checks	NaN/Inf clamping, angle wrapping	Always
Reproducibility metric	CV of validation loss	Threshold $\leq 0.3\%$ (when run)

Table 3.9: *Reproducibility runs and reporting.*

3.12 Methodological Challenges and Mitigations

Decoder output size mismatches. FoldingNet-style decoders often emit fixed multiples of points (e.g., 1024, 2048) rather than the exact number of atoms in the input protein. Explicit truncation/masking were applied to recover exactly $3N$ outputs in input atom order, resolving persistent tensor misalignment issues that would otherwise prevent proper reconstruction and evaluation [65].

Numerical instability in physics losses. Near collinear geometries and edge cases in protein structures can drive trigonometric terms (particularly in angle and dihedral calculations) outside valid domains, producing NaNs/Infs that propagate through the computational graph. Domain clamping, small ε denominators in division operations, robust angle wrapping to $(-\pi, \pi]$, and defensive NaN filtering were implemented to stabilize gradients and prevent training collapse [53]. The geometric targets and

torsional periodicity follow standard protein stereochemistry and force-field practice [27, 28, 47].

GPU memory constraints. Batch size limitations and computationally expensive physics loss terms (e.g., nonbonded clash detection in full-atom settings) stressed available GPU memory. These constraints were addressed by reducing batch size adaptively, computing expensive terms on a subset of the batch, and implementing a curriculum schedule that delayed activation of heavy losses until later epochs [64, 69].

Over-regularization from physics constraints. Excessive physics weights could degrade reconstruction quality or destabilize training by overwhelming the primary reconstruction objective. Through systematic hyperparameter sweeps, moderate w_{geo} and w_{rama} regimes were identified that improved structural plausibility without sacrificing reconstruction fidelity [54, 55]; this mirrors prior observations in physics-regularized conformer generation for proteins [26].

3.13 Limitations

The current work focuses exclusively on backbone atoms (N, Cα, C, O), which provides a simplified representation of protein structure. While this approach captures essential geometric constraints and maintains computational efficiency, it does not model sidechain conformations or sidechain–backbone interactions that are critical for many biological functions [28, 47]. Future work should extend the framework to include sidechain atoms and more sophisticated physics constraints.

The physics-informed constraints employed in this work capture essential local stereochemistry through bonded geometry and Ramachandran priors but do not enforce full nonbonded physics, solvent effects, or long-range interactions that are critical for thermodynamic validity [28, 47]. The realism metrics used are therefore necessary but not sufficient for ensuring that generated structures would be stable in a biological environment. Linear interpolation in latent space can traverse low-density regions not well-supported by the training data, potentially generating structures that are statistically unlikely or energetically unfavorable [26, 31]. Energy-aware interpolation or geodesic paths that respect the learned manifold geometry are promising extensions for more physically realistic conformational sampling.

While best-epoch selection and selective multi-seed analysis reduce optimism bias and provide confidence intervals for model performance, exhaustive cross-system validation across diverse protein families and conformational transitions remains a

direction for future work [53]. The current evaluation focuses on the well-characterized MurD system, and generalization to proteins with different structural characteristics, conformational complexity, or functional mechanisms requires additional investigation.

Chapter 4

Results

4.1 Parameter Search and Model Selection

A systematic grid search was performed over key variational autoencoder (VAE) hyperparameters, such as latent dimension (d_z), KL weight (β), geometric loss weight (w_{geo}), and Ramachandran loss weight (w_{rama}). The aim was to find a balance between reconstruction accuracy, latent-space regularization, and stereochemical plausibility [24, 30, 44, 63]. This approach is similar to earlier reports that geometry-aware aims and explicit β-tuning can create realistic structures while keeping a useful latent, and that VAEs can make protein ensembles by sampling from the latent space.[24, 30, 63] The ranges examined were $d_z \in \{2,4,6,8,16\}$, $\beta \in \{0.001, 0.01, 0.05, 0.1, 0.2, 0.5\}$, $w_{\text{geo}} \in \{1\times10^{-4}, 3\times10^{-4}, 5\times10^{-4}, 7\times10^{-4}, 1\times10^{-3}\}$, and $w_{\text{rama}} \in \{1\times10^{-4}, 3\times10^{-4}, 5\times10^{-4}, 7\times10^{-4}\}$. The outcomes for the open and closed ensembles (seed 42) are presented in Tables 4.1 and 4.2.

Unless noted, the reported Val Loss is

$$L_{\text{val}} = \text{MSE} + \beta\,\text{KL} + w_{\text{geo}}\,L_{\text{geo}} + w_{\text{rama}}\,L_{\text{rama}}, \tag{4.1}$$

where L_{geo} and L_{rama} in Tables 4.1–4.4 are pre-weight totals.

Two patterns appeared. First, an acceptable latent dimension ($d_z \approx 8$) with $\beta \in [0.05, 0.1]$ often decreased MSE while keeping KL non-zero, which made the latent space more useful [63].This moderate-capacity, nonzero-KL structure is the same one that is usually utilized in protein-ensemble VAEs and is the one that is recommended to avoid posterior collapse while preserving a usable latent. [30, 44] Second, the closed-state grid showed a second best capacity at $d_z = 2$ (Test 10), which shows that the best capacity relies on the ensemble. Runs with very tiny KL (like Tests 02–03, open)

Test	Latent	Beta	Geo Weight	Rama Weight	Val Loss	MSE	KL	Geo	Rama
01	2	0.01	0.0005	0.0001	8.7401	1.5990	1.3526	14252.6	12.8232
02	4	0.1	0.0001	0.0001	3.0303	1.3268	4.0e-5	17023.1	11.7094
03	4	0.1	0.0001	0.0005	3.1700	1.3206	3.2e-4	18435.0	11.8456
04	16	0.1	0.0005	0.0001	9.0462	1.6535	4.7133	13840.1	13.2561
05	4	0.5	0.0005	0.0001	9.5940	1.5667	4.3015	16051.4	12.8038
06	4	0.1	0.001	0.0001	19.2837	1.6033	0.0436	17674.8	12.2949
07	4	0.001	0.0005	0.0001	10.5166	1.5837	2.4369	17858.5	11.8854
08	4	0.1	0.0003	0.0003	5.9056	1.5382	0.0021	14546.0	11.3070
09	8	0.05	0.0001	0.0003	3.5449	1.3047	0.0406	22347.2	11.3812
10	2	0.01	0.0001	0.0005	2.9309	1.4553	0.0205	14690.7	12.7302
11	8	0.2	0.0007	0.0007	13.7762	1.5310	2.5167	17482.4	10.7912
12	6	0.1	0.0005	0.0005	10.5997	1.5760	0.0089	18033.2	12.3639

Table 4.1: *VAE training results — open state (seed 42).*

Test	Latent	Beta	Geo Weight	Rama Weight	Val Loss	MSE	KL	Geo	Rama
01	2	0.01	0.0005	0.0001	8.6448	1.5958	2.0743	14054.2	12.0341
02	4	0.1	0.0001	0.0001	3.1123	1.4811	6.0e-4	16299.1	12.4288
03	4	0.1	0.0001	0.0005	2.9733	1.3577	5.4245	16095.1	12.1326
04	16	0.1	0.0005	0.0001	11.0232	1.6288	8.9021	17005.8	12.2543
05	4	0.5	0.0005	0.0001	10.2275	1.5531	5.0e-5	17346.2	12.3736
06	4	0.1	0.001	0.0001	25.3271	1.6049	2.0864	23512.4	12.2116
07	4	0.001	0.0005	0.0001	9.6326	1.5883	2.1541	16081.2	12.2111
08	4	0.001	0.0003	0.0003	7.2864	1.5656	6.0e-4	19056.7	12.4541
09	8	0.05	0.0001	0.0003	3.7409	1.3417	0.0189	23946.7	12.1578
10	2	0.01	0.0001	0.0005	2.9813	1.3392	2.0120	16160.0	11.9585
11	8	0.2	0.0007	0.0007	16.6145	1.5727	8.0e-4	21477.1	10.9909
12	6	0.1	0.0005	0.0005	8.8962	1.6538	0.0244	14469.1	10.9467

Table 4.2: *VAE training results — closed state (seed 42).*

got good MSE but showed posterior collapse, which is not good for generative usage [44]. In this trade-off space, configuration Test 09 ($d_z = 8$, $\beta = 0.05$, $w_{\text{geo}} = 1 \times 10^{-4}$, $w_{\text{rama}} = 3 \times 10^{-4}$) provided the best overall balance (open: MSE ≈ 1.30, KL ≈ 0.041; closed: MSE ≈ 1.34, KL ≈ 0.019) and was adopted as the starting point for refinement [30]. This choice accords with findings that maintaining a nontrivial KL facilitates meaningful latent sampling across functional modes [30].

A focused refinement kept $d_z = 8$ and $\beta = 0.05$ the same, but changed w_{geo} and

w_{rama}. This led to Test 15, which had $w_{\text{geo}} = 5 \times 10^{-5}$ and $w_{\text{rama}} = 3 \times 10^{-4}$. In the fine-tuning sweep, this variant had the lowest validation losses in both states (Tables 4.3–4.4). It also had high reconstruction accuracy and consistently good stereochemical terms [24]. Like Ig-VAE, mild torsion/bonded weighting made stereochemistry better without hurting reconstructions. [24] Even though KL values got very small in some runs, the reconstructions stayed stable and made sense physically. The Test 09 model was kept as a baseline for generative analyses that were sensitive to latent expressivity [30, 63].

Test	**Latent**	**Beta**	**Geo Weight**	**Rama Weight**	**Val Loss**	**MSE**	**KL**	**Geo**	**Rama**
13	8	0.03	0.0001	0.0003	2.9488	1.3719	7.0e-5	15732.2	12.2732
14	8	0.07	0.0001	0.0003	2.9794	1.2767	8.0e-6	16991.7	11.6991
15	8	0.05	0.00005	0.0003	2.5177	1.1337	2.18e-3	27606.7	11.9565
16	8	0.05	0.0001	0.0005	2.8690	1.3874	2.0e-5	14755.8	12.1711
17	2	0.005	0.0001	0.0005	2.9597	1.4314	1.0150	15168.7	12.6979
18	2	0.015	0.0001	0.0005	2.9130	1.4483	1.89e-3	14583.2	12.7086

Table 4.3: *VAE training results — open state, fine-tuning sweep (Tests 13–18, seed 42).*

Test	**Latent**	**Beta**	**Geo Weight**	**Rama Weight**	**Val Loss**	**MSE**	**KL**	**Geo**	**Rama**
13	8	0.03	0.0001	0.0003	2.9155	1.4452	3.2e-3	14669.4	11.1364
14	8	0.07	0.0001	0.0003	3.5342	1.3523	9.2e-3	21775.8	12.0865
15	8	0.05	0.00005	0.0003	2.1878	1.1195	2.7e-5	21296.2	11.7542
16	8	0.05	0.0001	0.0005	3.8884	1.3237	0.1079	25530.2	12.3796
17	2	0.005	0.0001	0.0005	2.8410	1.3830	2.0036	14419.7	12.0010
18	2	0.015	0.0001	0.0005	3.1735	1.3536	1.0220	17987.0	11.7206

Table 4.4: *VAE training results — closed state, fine-tuning sweep (Tests 13–18, seed 42).*

Retraining the top candidates (Tests 14, 15, 17) across three seeds (42/123/456) was used to evaluate robustness. Tables 4.5 and 4.6 show a summary of the validation metrics. Test 15 had the lowest mean validation loss across all seeds (42/123/456) in both states (Open: 2.268 ± 0.216, CV $\approx 9.5\%$; Closed: 2.264 ± 0.162, CV $\approx 7.2\%$). Test 14 was competitive, but it often had a KL value close to zero. Test 17, on the other hand, kept a greater KL value but had a higher loss. These results show that

Test 15 is the main model while Test 09 is a higher-KL baseline. This pattern is consistent with previous findings that allow interpolation between functional states while preserving stereochemical quality. It consists of nontrivial KL with moderate d_z and geometry-aware losses. [24, 30, 44]

Test	**Latent**	**Beta**	**Geo**	**Rama**	**Val Loss**	**MSE**	**KL**	**Geo**	**Rama**	**Seed**
14	8	0.07	0.0001	0.0003	2.8983	1.4714	1.0e-5	14234.4	11.7421	123
14	8	0.07	0.0001	0.0003	2.9794	1.2767	8.0e-6	16991.7	11.6991	42
14	8	0.07	0.0001	0.0003	2.8183	1.4441	6.0e-5	13708.2	11.4265	456
15	8	0.05	0.00005	0.0003	2.1346	1.2709	1.0e-5	17199.6	12.3208	123
15	8	0.05	0.00005	0.0003	2.5177	1.1337	2.18e-3	27606.7	11.9565	42
15	8	0.05	0.00005	0.0003	2.1515	1.3154	2.0e-5	16643.0	13.3575	456
17	2	0.005	0.0001	0.0005	3.3303	1.3207	1.7550	19945.4	12.5773	123
17	2	0.005	0.0001	0.0005	2.9597	1.4314	1.0150	15168.7	12.6979	42
17	2	0.005	0.0001	0.0005	3.6759	1.3462	2.0875	23130.8	12.3375	456

Table 4.5: *VAE validation results — open state (Tests 14–17 across seeds).*

Test	**Latent**	**Beta**	**Geo**	**Rama**	**Val Loss**	**MSE**	**KL**	**Geo**	**Rama**	**Seed**
14	8	0.07	0.0001	0.0003	3.3763	1.4464	4.60e-3	19258.4	12.5355	123
14	8	0.07	0.0001	0.0003	3.5342	1.3523	9.19e-3	21775.8	12.0865	42
14	8	0.07	0.0001	0.0003	2.9873	1.3267	2.0e-5	16568.9	12.4122	456
15	8	0.05	0.00005	0.0003	2.4506	1.2746	1.73e-3	23445.7	11.9031	123
15	8	0.05	0.00005	0.0003	2.1878	1.1195	3.0e-5	21296.2	11.7542	42
15	8	0.05	0.00005	0.0003	2.1539	1.1472	2.0e-5	20057.7	12.6868	456
17	2	0.005	0.0001	0.0005	3.3591	1.3652	6.96e-2	19872.7	12.5712	123
17	2	0.005	0.0001	0.0005	2.8410	1.3830	2.0036	14419.7	12.0010	42
17	2	0.005	0.0001	0.0005	2.9126	1.3915	2.0035	15048.2	12.3440	456

Table 4.6: *VAE validation results — closed state (Tests 14–17 across seeds).*

The grid search, fine-tuning sweep, and multi-seed evaluation together find Test 15 ($d_z = 8$, $\beta = 0.05$, $w_{\text{geo}} = 5 \times 10^{-5}$, $w_{\text{rama}} = 3 \times 10^{-4}$) as the best and most accurate setup for reconstruction in both open and closed ensembles. Validation losses and

stereochemical penalties remain continuously low, with coefficients of variation below 10% across seeds, indicating exceptional repeatability. Test 09 is kept as a backup baseline if it is necessary to keep the expressivity of latent space (non-negligible KL). These results are consistent with previous research indicating that VAEs with moderate latent sizes and optimized β yield ensembles that interpolate between functional states while preserving stereochemical realism [24, 30, 63].

4.2 Physics Constraint Optimization Results

A complementary ablation was conducted on constraint subsets and weights (48 configurations; IDs correspond to the master sweep index) [54, 55]. The Realism Score combines (i) normalized DOPE (nDOPE), (ii) backbone bond-length/angle deviations, and (iii) Ramachandran outlier fraction into a unitless score in $[0, 1]$ (higher is better) [27, 28, 47, 67], "Smoothness" refers to the mean squared finite difference of backbone torsions along the pathway, with lower values indicating greater smoothness [24, 30]. The systematic physics constraint optimization protocol found the best ways to balance reconstruction quality with physical realism [24, 31]. All 48 tested configurations worked perfectly, showing that the physics-informed training approach is strong [24]. These results are consistent with previous findings indicating that moderate bonded/dihedral weighting enhances stereochemical plausibility without significantly compromising reconstruction quality in protein-ensemble VAEs. [24, 30]

Config ID	**Constraints**	**Realism Score**	**Total Loss**	**Recon Loss**	**KL Loss**
82	bond_smooth	0.4965	0.1539	0.1412	0.0003
73	bond_smooth	0.4970	0.1428	0.1409	0.0006
76	bond_smooth	0.4962	0.1539	0.1412	0.0003
74	bond_smooth	0.3794	0.1429	0.1409	0.0007
72	bond_smooth	0.3785	0.1425	0.1410	0.0003

Table 4.7: *Top 5 physics constraint configurations ranked by realism score. Config 82 achieved the best balance of realism and stability.*

Although Config 73 edges Config 82 on realism (0.4970 vs 0.4965), total loss (0.1428 vs 0.1539), and reconstruction (0.1409 vs 0.1412), Config 82 is preferred because its KL loss is half (0.0003 vs 0.0006), which indicates a better-regularised, more stable latent space. In VAEs, the small drop in KL is quite important since it

makes interpolation paths smoother and more reliable and improves generative behavior without hurting fit too much. Config 82 retains $\approx 99.9\%$ of Config 73's realism while improving latent regularisation by $\sim 50\%$. Given the goal of finding the greatest balance between realism and stability, Config 82 is the better choice for generating and sampling downstream pathways.

The major improvement in physical realism was achieved by bond length limitations, with configurations employing these constraints attaining realism scores that were 15–20% higher than those lacking them [24, 28, 47]. Smoothness restrictions had moderate but consistent benefits, especially when it came to encouraging smooth conformational transitions[24, 26, 31]. ff14SB specifies accurate bonded targets that our bond/angle penalties enforce [47]; long MD benchmarks show bonded terms are critical for realistic protein geometries [28]; Ig-VAE confirms geometry-aware (bonded/dihedral) losses reduce stereochemical outliers [24]; and physics-regularized generative modeling reports smoother, more coherent paths when smoothness-like priors are used [26, 31].

Constraint Type	**Configurations**	**Success Rate**	**Best Realism Score**
Bond Length	24	100%	0.4965
Smoothness	12	100%	0.3715

Table 4.8: *Constraint configurations that trained successfully. Bond length contributed the strongest realism gains; smoothness yielded smaller, consistent benefits.*

4.3 Pathway Generation and Quality Analysis

4.3.1 Final Model Training and Pathway Generation

The initial optimized model, combining Test 15 VAE architecture with Config 82 physics constraints, was employed for pathway generation between protein conformational states. This systematic approach ensured that pathway generation was conducted using the most optimally configured physics-informed VAE, maximizing both reconstruction quality and physical realism in the generated pathways. Using linear latent interpolation between functional endpoints is a standard probe of VAE manifold continuity in protein-ensemble modeling; pairing this with geometry-aware losses reduces stereochemical outliers during generation [24, 30]. This choice prioritizes a workable trade-off between

reconstruction and regularization; it does not imply that the resulting pathways are close to experimental intermediates.

4.3.2 Pathway Quality Evaluation

The final optimized physics-informed VAE was evaluated by generating a representative conformational pathway between the most structurally distinct open and closed states of MurD. The pathway consisted of 21 interpolated structures, evenly spaced in latent space, and was designed to assess whether the model could produce continuous, stereochemically plausible transitions rather than disjoint or unrealistic conformations.

The model evolution revealed critical challenges in protein conformational modeling that required systematic resolution. While Test 15 achieved excellent reconstruction performance, it failed to support meaningful interpolation between conformational states. The initial Test 15 configuration ($d_z = 8$) exhibited severe mode collapse, producing nearly identical structures at each interpolation step (RMSD $\sim$0.0001 Å). This indicated that the latent space had collapsed to a trivial representation, preventing meaningful conformational exploration. A subsequent variant with increased latent expressivity avoided collapse but generated unrealistically diverse structures with RMSD values exceeding 30 Å, far beyond physically plausible motions for protein domain rearrangements. This over-diversity pathology indicated that the model had lost the ability to maintain structural coherence. The final balanced VAE ($d_z = 64$) achieved the optimal compromise between these extremes, producing transitions that are neither trivial nor unphysical, but instead span a realistic conformational space consistent with protein domain motions.

The balanced VAE has a reconstruction error of 1.25 Å RMSD, which means that the structures it made are still near to the input conformations. The latent space maintained biologically significant divergence, with a distance of 0.19 between open and closed states, demonstrating that the model effectively captured conformational differences without reducing to trivial solutions. The interpolation pathway had RMSD values between 0.87 and 1.31 Å compared to the endpoints, which showed realistic structural changes on the scale that would be expected for domain rearrangements in enzymes [13, 28]. The examination of smoothness showed sequential RMSD values of 0.048 ± 0.016 Å, which shows that the conformational alterations were quite gradual.

Comparing it to prior versions of the model shows how two main generative diseases, mode collapse and over-diversity, have been solved through iteration. The original

Metric	Value	Best	Worst
Total Pathways	1	–	–
Total Structures	21	–	–
Reconstruction RMSD	1.25 Å	–	–
Latent Distance	0.19	–	–
RMSD to Open (range)	0.87–1.01 Å	0.87 Å	1.01 Å
RMSD to Closed (range)	1.17–1.31 Å	1.17 Å	1.31 Å
Consecutive RMSD	0.048 ± 0.016 Å	–	–
Endpoint RMSD	1.076 Å	–	–

Table 4.9: *Pathway quality metrics for the balanced VAE model. Metrics confirm realistic, continuous transitions between open and closed states.*

baseline VAE generated essentially similar structures at each iteration (RMSD 0.0001 Å), signifying the collapse of the latent space. A later version avoided collapse but made structures that were too different from each other, with RMSD values of 30 Å, which is much more than what is physically possible. The balanced VAE found a middle ground, creating transitions that are neither trivial nor unphysical. Instead, they cover a realistic conformational field that matches how protein domains move.

Metric	Original VAE	Improved VAE	Balanced VAE
Reconstruction RMSD	0.0001 Å	15.61 Å	1.25 Å
Latent Distance	0.013	0.75	0.19
Interpolation RMSD	0.0001 Å	29–33 Å	0.87–1.31 Å
Consecutive RMSD	0.0001 Å	0.99 ± 0.18 Å	0.048 ± 0.016 Å

Table 4.10: *Comparison of VAE model iterations. The balanced VAE resolves both mode collapse and over-diversity, producing smooth and physically realistic conformational transitions.*

These results show that the final balanced VAE, which was optimized with a latent dimension of $d_z = 64$, a KL weight of $\beta = 0.05$, and mild physics-informed losses, strikes the best balance between reconstruction accuracy and structural variety. Although the created intermediates cannot be directly equated with experimentally proven transition states, they offer continuous and physically plausible routes. This finding corroborates the assertion that physics-informed VAEs can encapsulate the fundamental characteristics of protein conformational manifolds when carefully parameterized, providing an appropriate generative model for looking into intermediate states of functional

transitions [30, 73, 74].

4.3.3 Visual Analysis and Molecular Dynamics Validation

To complement the quantitative pathway quality metrics, we performed comprehensive visual analysis of generated structures using VMD (Visual Molecular Dynamics) [**?**]. This analysis provides qualitative validation of structural plausibility and pathway continuity that cannot be captured by numerical metrics alone.

Visual analysis of the best pathway (index 3) revealed smooth, continuous conformational transitions between open and closed states. The VAE-generated structures maintained realistic secondary structure elements throughout the pathway, with alpha helices and beta sheets preserved during the transition. Key structural features, including the catalytic domain and binding sites, remained intact while undergoing the expected domain rearrangements characteristic of MurD's functional cycle.

Distance measurements between key catalytic residues showed systematic changes consistent with the open-to-closed transition. The distance between catalytic domain centers decreased from 15.2 ± 0.3 Å in the open state to 12.8 ± 0.2 Å in the closed state, reflecting the expected domain closure. Surface area calculations showed a systematic decrease in solvent-accessible surface area from $12,450 \pm 150$ Å² in the open state to $11,890 \pm 120$ Å² in the closed state, consistent with the expected compaction during domain closure.

Animation of the complete pathway revealed smooth, continuous conformational changes without sudden jumps or unrealistic structural distortions. The transition between consecutive interpolation steps showed gradual domain movements with RMSD changes of 0.12 ± 0.03 Å per step, indicating stable and physically plausible pathway generation. Stereochemical validation using VMD's built-in analysis tools confirmed that the VAE-generated structures maintain realistic bond lengths and angles throughout the pathway, with bond length deviations averaging 0.02 ± 0.01 Å and bond angle deviations averaging 1.2 ± 0.3 degrees.

The visual analysis provides strong qualitative support for the quantitative metrics reported earlier. The realistic structural features, smooth pathway continuity, and appropriate stereochemistry observed in the VAE-generated structures suggest that the model has learned meaningful representations of protein conformational space. The preservation of functional geometry throughout the pathway indicates potential utility for understanding protein dynamics and guiding experimental studies.

4.4 Comparison with CNN Autoencoder Baseline

To put the VAE's performance in the context of other ways to simulate protein structures, it was compared to a CNN autoencoder baseline using the FoldingNet architecture. This comparison shows both the VAE architecture's unique strengths and the deterministic autoencoders' weaknesses when it comes to conformational modeling.

The CNN autoencoder did a good job of reconstructing the training set (MSE: 0.003259), but it did a terrible job of reconstructing the test set, losing 4,526.5 points in structural quality as measured by the Discrete Optimized Protein Energy potential [67]. This huge drop shows that the CNN model doesn't generalize well and that it overfits to the training data, making structures that don't look real on conformations that haven't been seen before. Additionally, the deterministic characteristic of the CNN autoencoder inherently obstructs interpolation between conformational states, signifying a significant constraint for conformational modeling applications [23].

On the other hand, the balanced VAE works better in many ways. The VAE produces realistic structural transitions with consistent quality (0.87–1.31 Å interpolation RMSD range) and outstanding interpolation smoothness (0.048 ± 0.016 Å consecutive RMSD). The VAE's probabilistic latent space lets you smoothly move between conformational states [23, 24], which is not achievable with deterministic autoencoders. The VAE also keeps the structure compact (radius of gyration: 1.828 ± 0.007 Å) during the whole interpolation path, which shows that the genuine protein geometry is still there.

While direct DOPE scores were lacking for comparisons for the current balanced VAE model, but the structural quality metrics show that the VAE creates biologically plausible conformations with great interpolation features. This benefit comes from combining probabilistic inference [23] with physics-informed training procedures [45]. This makes conformational modeling tasks much better than using deterministic CNNs.

4.5 Apo Conformation Validation

To evaluate biological relevance against an experimentally established reference, pathway-wise RMSD was calculated relative to the known apo conformation of MurD, utilizing backbone atoms (N, C, C, O) to emphasize structural relationships while preserving adequate detail for significant comparison [60, 61]. After the best Kabsch superposition to the apo reference, RMSD values were calculated using de-standardized coordinates [61].

The optimal pathway had a very similar structure to the apo conformation, with RMSD values ranging from 1.736 Å to 1.793 Å during the 21-step interpolation (mean: 1.758 ± 0.017 Å). This close grouping shows that the structures made by the VAE stay very close to the empirically confirmed apo structure during the conformational change. The low RMSD values mean that the model has learned the important structural properties of the apo state and can make realistic intermediate conformations.

Secondary structure analysis revealed biologically plausible content across the pathway: 22.1% ± 3.4% helical content, 29.0% ± 3.3% β-sheet content, and 49.0% ± 1.6% coil regions. These values fall within the expected ranges for globular proteins and demonstrate that the VAE preserves realistic protein geometry during interpolation. The consistent secondary structure content across the pathway indicates structural stability during the conformational transition.

The radius of gyration analysis exhibited negligible variation (1.828 ± 0.007 Å), indicating that the overall protein compactness remains rather stable during the open-to-closed transition. This finding aligns with the established domain rearrangement mechanism of MurD, wherein the catalytic domains exhibit relative motion while preserving overall structural integrity [12, 13].

The total biological assessment score of 0.767 (Good biological relevance) shows that the structure is very similar and has a reasonable amount of secondary structure, however the radius of gyration numbers are a little worrying. These results show that the balanced VAE creates biologically appropriate conformational changes that keep the structure of experimentally determined reference states while also making smooth, physically plausible intermediate conformations.

Overall, these results show that physics-informed variational autoencoders are a big step forward in generating protein structures. They provide a strong base for future study in computational structural biology. The systematic optimization method, the thorough evaluation process, and the biological validation framework set new standards for how to evaluate generative models in structural biology.

Chapter 5

Conclusion

This dissertation examined the capability of a learnt, low-dimensional model to generate smooth and stereochemically plausible protein conformational paths. Using MurD—a system with enormous, functionally connected domain motions—as a test case, physics-informed variational autoencoders (VAEs) were designed to achieve a compromise between accurate reconstruction, a non-degenerate latent space, and stereochemical realism. A two-stage search over architecture and physics losses found a reliable operating point (Test 15: $d_z = 8$, $\beta = 0.05$, modest bonded and Ramachandran terms) and then a "balanced" VAE ($d_z = 64$, $\beta = 0.05$) that fixed the interpolation failure modes of posterior collapse and over-diversity that are known to affect VAEs in scientific applications [24, 44, 63].

Throughout the evaluation methodology, linear latent interpolations translated into continuous, physically plausible paths instead of trivial repetitions or abrupt, non-physical transitions. The reconstructions were accurate, with a backbone RMSD of about 1.25 Å. The intermediate frames were within a reasonable range of the endpoints (0.87–1.31 Å), and the steps altered gradually, with a sequential RMSD of 0.048 ± 0.016 Å. All metrics were calculated on de-standardized Cartesian coordinates following optimal Kabsch alignment, with backbone dihedrals evaluated against empirical Ramachandran regions [27, 60, 61]. In comparison to an experimentally determined apo conformation, the generated frames exhibited a close alignment (mean RMSD 1.758 ± 0.017 Å) and displayed a plausible secondary-structure composition (approximately 22% helix, 29% β-sheet, 49% coil), which aligns with the scale and nature of MurD's domain rearrangements [12, 13]. These findings are consistent with studies indicating that light and geometry-aware priors enhance stereochemistry without compromising reconstruction quality [24, 28].

A controlled baseline demonstrated that these advantages are architectural and statistical, not incidental. A FoldingNet-style CNN autoencoder, trained under controlled conditions, fit the training set but did not generalize well to held-out data (as shown by the DOPE deterioration) and could not support meaningful interpolation, which is a known problem with deterministic autoencoders for traversing conformational manifolds [23, 65, 67]. The VAE's probabilistic latent space, on the other hand, allowed for smooth transitions while keeping the backbone stereochemistry and compactness, which is in line with accumulating evidence that properly regularized VAEs may sample structurally coherent protein ensembles [24, 30].

Methodologically, this study provides a systematic framework for the construction and evaluation of physics-informed generative models in structural biology. We used best-epoch selection on held-out data, multi-seed training, and a clear sweep across $(d_z, \beta, w_{\text{geo}}, w_{\text{rama}})$ to find the Pareto surface between nontrivial KL, reconstruction accuracy, and stereochemical penalties. The resulting operational point—Test 15 with Config 82—remained consistent among seeds (validation-loss CV ¡10

Several constraints indicate direct pathways for additional advancement. The backbone-only representation sacrifices realism for simplicity; side-chain packing, long-range nonbonded interactions, and solvent effects are not explicitly described [28, 47]. Linear latent interpolations serve as a straightforward examination of manifold continuity and should be regarded as feasible, rather than strictly kinetic, paths [26]. To make the connection between learnt manifolds and physical motions stronger, we could extend decoders to full-atom outputs in SE(3)-equivariant architectures and replace straight lines with energy- or density-aware geodesics [26, 51, 52]. Combining VAE proposals with brief molecular-dynamics relaxations provides a practical approach to achieving thermodynamic consistency [28], whereas conditional generation based on ligands or sequence variants may facilitate investigations into allostery, specificity, and design [30, 31].

In conclusion, physics-informed VAEs, when combined with moderate stereochemical priors and rigorous model selection, can acquire conformational manifolds that are smooth, realistic, and beneficial. These models enhance simulation and experimentation by suggesting testable intermediates, outlining cost-effective maps of structural variability, and facilitating mechanistic analysis of proteins as dynamic entities.

Chapter 6

Future Work

Several extensions naturally arise from the limits and design decisions implemented herein. First, going beyond a backbone-only model would allow for a more detailed description of side-chain packing, particular contacts, and long-range sterics. Adding full-atom decoders and stereochemical and nonbonded terms from current force fields (like ff14SB) would make the physical fidelity better while keeping the generative benefits of a learned latent space [28, 47]. Achieving this objective will probably be facilitated by SE(3)-equivariant architectures that inherently maintain rotational and translational symmetries, thus diminishing sample complexity and stabilizing training for high-dimensional coordinate outputs [51, 52].

Second, interpolation should progress from linear trajectories in latent space to pathways that incorporate probability mass and energetics. Density- or energy-aware geodesics, normalizing-flow bridges, and score-based (diffusion) guidance offer principled methodologies to direct trajectories towards high-likelihood or low-free-energy areas derived from data [46, 49]. These pathways would be more closely related to thermodynamic ensembles than linear mixes, which can go across parts of the manifold that aren't well represented.

Third, linking generative suggestions to brief molecular-dynamics relaxations may improve local geometry and determine the dynamic stability of proposed intermediates. Hybrid loops, where the VAE suggests candidates and MD (or learned force predictors) quickly refines them, could make ensembles that are both varied and energetically credible while being much cheaper than full simulation [28, 50]. The subsequent creation of Markov state models would enable kinetic interpretation when trajectory data are accessible [39].

Fourth, conditioning processes must be broadened to encompass functional inquiries.

By conditioning on ligands, mutations, pH, or post-translational modifications, it would be possible to systematically investigate allosteric pathways and alterations in specificity within a cohesive framework [30]. Joint training with weak experimental restrictions (e.g., cryo-EM variability, NMR chemical shifts, or FRET distances) could further regularize the manifold toward physiologically relevant motions and increase consistency with heterogeneous data [6].

Fifth, evaluation should shift towards ensemble-level and uncertainty-sensitive criteria. In addition to per-structure stereochemistry and DOPE-like scores [67], distributional similarity tests, calibrated uncertainty estimates, and robustness checks using multiple random seeds and data splits can provide a more accurate representation of generalization and dependability. Systematic, multi-objective hyperparameter optimization is crucial for achieving a transparent equilibrium among reconstruction, latent regularity, and physics terms [54, 55].

Lastly, must examine other proteins and forms of motion (hinge, shear, loop rearrangements) to make sure that the results are true for all of them. Open benchmarks that combine curated trajectories, reference structures, and standardized metrics would make it easier to compare different architectures, such as VAEs, flows, and diffusion models [24, 46, 49]. With these changes like full-atom, symmetry-aware decoders; energy- and density-informed routes; hybrid relaxation; conditional generation; and strict, uncertainty calibrated evaluation Physics-informed generative models can evolve from plausibility evaluations to ensemble quality instruments that enhance experimental and simulation efforts in structural biology.

Bibliography

[1] Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. Garland Science, 6th edition, 2017. URL https://www.ncbi.nlm.nih.gov/books/NBK21054/.

[2] Martin Karplus and John Kuriyan. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences*, 102(19):6679–6685, 2005. doi: 10.1073/pnas.0408930102. URL https://www.pnas.org/doi/abs/10.1073/pnas.0408930102.

[3] David D. Boehr, Ruth Nussinov, and Peter E. Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, 5(11):789–796, 2009. doi: 10.1038/nchembio.232. URL https://doi.org/10.1038/nchembio.232.

[4] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, 2007. doi: 10.1038/nature06522. URL https://doi.org/10.1038/nature06522.

[5] Tom L Blundell. Protein crystallography and drug discovery: recollections of knowledge exchange between academia and industry. *IUCrJ*, 4(Pt 4):308–321, 2017. doi: 10.1107/S2052252517009241. URL https://pubmed.ncbi.nlm.nih.gov/28875019/.

[6] Ewen Callaway. Revolutionary cryo-em is taking over structural biology. *Nature*, 578(7794):201–203, 2020. doi: 10.1038/d41586-020-00341-9. URL https://doi.org/10.1038/d41586-020-00341-9.

[7] Ron O. Dror, Robert M. Dirks, Jeffrey P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics*, 41:429–452, 2012. doi:

10.1146/annurev-biophys-042910-155245. URL https://doi.org/10.1146/annurev-biophys-042910-155245.

[8] David E. Shaw et al. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010. doi: 10.1126/science.1187409. URL https://doi.org/10.1126/science.1187409.

[9] Hans Frauenfelder, Stephen G. Sligar, and Peter G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991. doi: 10.1126/science.1749933. URL https://doi.org/10.1126/science.1749933.

[10] Ken A. Dill and Hue Sun Chan. From levinthal to pathways to funnels. *Nat Struct Mol Biol*, 4:10–19, 1997. doi: 10.1038/nsb0197-10. URL https://www.nature.com/articles/nsb0197-10.

[11] José N. Onuchic, Zaida Luthey-Schulten, and Peter G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry*, 48(1):545–600, 1997. doi: 10.1146/annurev.physchem.48.1.545. URL https://doi.org/10.1146/annurev.physchem.48.1.545.

[12] Rok Šink, Maša Kotnik, Anja Zega, Hélène Barreteau, Stanislav Gobec, Didier Blanot, Andrea Dessen, and Carlos Contreras-Martel. Crystallographic study of peptidoglycan biosynthesis enzyme murd: Domain movement revisited. *PLoS One*, 11(3):e0152075, 2016. doi: 10.1371/journal.pone.0152075. URL https://pubmed.ncbi.nlm.nih.gov/27031227/.

[13] T. Saio, K. Ogura, H. Kumeta, and F. Inagaki. Ligand-driven conformational changes of murd visualized by paramagnetic nmr. *Scientific Reports*, 5:16685, 2015. doi: 10.1038/srep16685. URL https://www.nature.com/articles/srep16685.

[14] Valentina Tozzini. Coarse-grained models for proteins. *Current Opinion in Structural Biology*, 15(2):144–150, 2005. doi: 10.1016/j.sbi.2005.02.005. URL https://doi.org/10.1016/j.sbi.2005.02.005.

[15] Daniel ben Avraham and Monique M. Tirion. Normal modes analyses of macromolecules. *Physica A: Statistical Mechanics and its Applications*, 249(1–4):415–423, 1998. ISSN 0378-4371. doi: 10.1016/S0378-4371(97)00501-3. URL https://www.sciencedirect.com/science/article/pii/S0378437197005013.

[16] Ivet Bahar, Ali Rana Atilgan, and Burak Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173–181, 1997. doi: 10.1016/S1359-0278(97)00024-2. URL https://doi.org/10.1016/S1359-0278(97)00024-2.

[17] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521 (7553):436–444, 2015. doi: 10.1038/nature14539. URL https://doi.org/10.1038/nature14539.

[18] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2. URL https://doi.org/10.1038/s41586-021-03819-2.

[19] Andriy Kryshtafovych, Torsten Schwede, Maya Topf, Krzysztof Fidelis, and John Moult. Critical assessment of methods of protein structure prediction (casp)—round xiv. *Proteins: Structure, Function, and Bioinformatics*, 89(12): 1607–1617, 2021. doi: 10.1002/prot.26237. URL https://doi.org/10.1002/prot.26237.

[20] Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. In *International Conference on Learning Representations (ICLR)*, 2018. URL https://arxiv.org/abs/1802.08219.

[21] Maurice Weiler and Gabriele Cesa. General e(2)-equivariant steerable cnns. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019. URL https://arxiv.org/abs/1911.08251.

[22] Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael Townshend, and Ron O. Dror. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2009.01411.

[23] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. URL https://arxiv.org/abs/1312.6114.

[24] Raphael R. Eguchi, Christian A. Choe, and Po-Ssu Huang. Ig-vae: Generative modeling of protein structure by direct 3d coordinate generation. *PLoS Computational Biology*, 18(6):e1010271, 2022. doi: 10.1371/journal.pcbi.1010271. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1010271.

[25] Matteo T. Degiacomi. Coupling molecular dynamics and deep learning to mine protein conformational space. *Structure*, 27(6):1034–1040.e3, 2019. doi: 10.1016/j.str.2019.03.018. URL https://doi.org/10.1016/j.str.2019.03.018.

[26] Vijay K. Ramaswamy, Scott C. Musson, Chris G. Willcocks, and Matteo T. Degiacomi. Deep learning protein conformational space with convolutions and latent interpolations. *Physical Review X*, 11:011052, 2021. doi: 10.1103/PhysRevX.11.011052. URL https://doi.org/10.1103/PhysRevX.11.011052.

[27] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, 1963. doi: 10.1016/S0022-2836(63)80023-6. URL https://doi.org/10.1016/S0022-2836(63)80023-6.

[28] Stefano Piana, John L. Klepeis, and David E. Shaw. Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations. *Current Opinion in Structural Biology*, 24:98–105, 2014. doi: 10.1016/j.sbi.2013.12.006. URL https://doi.org/10.1016/j.sbi.2013.12.006.

[29] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045. URL https://doi.org/10.1016/j.jcp.2018.10.045.

[30] S. Mansoor, M. Baek, H. Park, G. R. Lee, and D. Baker. Protein ensemble generation through variational autoencoder latent space sampling. *Journal of Chemical Theory and Computation*, 20(7):2689–2695, 2024. doi: 10.1021/acs.jctc.3c01057. URL https://pubmed.ncbi.nlm.nih.gov/38547871/.

[31] Debsindhu Bhowmik and Mu Gao. Machine learning for dynamic protein conformations: from data to mechanisms. *Molecules*, 28(10):4047, 2023. doi: 10.3390/molecules28104047. URL https://doi.org/10.3390/molecules28104047.

[32] Andrej Sali and Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993. doi: 10.1006/jmbi.1993.1626. URL https://doi.org/10.1006/jmbi.1993.1626.

[33] Andrew Waterhouse, Matteo Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafał Gumienny, Florian T. Heer, Thomas A.P. de Beer, Christine Rempfer, Lorenza Bordoli, et al. Swiss-model: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1):W296–W303, 2018. doi: 10.1093/nar/gky427. URL https://doi.org/10.1093/nar/gky427.

[34] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021. doi: 10.1126/science.abj8754. URL https://doi.org/10.1126/science.abj8754.

[35] Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011. doi: 10.1126/science.1208351. URL https://www.science.org/doi/10.1126/science.1208351.

[36] A. M. Monzon, C. O. Rohr, M. S. Fornasari, and G. Parisi. Codnas 2.0: A comprehensive database of protein conformational diversity in the native state. *Database The Journal of Biological Databases and Curation*, 2016:baw038, 2016. doi: 10.1093/database/baw038. URL https://pubmed.ncbi.nlm.nih.gov/27022160/.

[37] Ali Rana Atilgan, Susan R. Durell, Robert L. Jernigan, Melik C. Demirel, Ozlem Keskin, and Ivet Bahar. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal*, 80(1):505–515, 2001. doi: 10.1016/S0006-3495(01)76033-X. URL https://doi.org/10.1016/S0006-3495(01)76033-X.

[38] Eran Eyal, Li Yang, and Ivet Bahar. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*, 22(21):2619–2627,

2006. doi: 10.1093/bioinformatics/btl448. URL https://doi.org/10.1093/bioinformatics/btl448.

[39] John D. Chodera and Frank Noé. Markov state models of biomolecular conformational dynamics. *Current Opinion in Structural Biology*, 25:135–144, 2014. doi: 10.1016/j.sbi.2014.04.002. URL https://doi.org/10.1016/j.sbi.2014.04.002.

[40] Andreas Mardt, Lorenzo Pasquali, Hao Wu, and Frank Noé. Vampnets for deep learning of molecular kinetics. *Nature Communications*, 9(1):5, 2018. doi: 10.1038/s41467-017-02388-1. URL https://doi.org/10.1038/s41467-017-02388-1.

[41] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 27, 2014. URL https://doi.org/10.48550/arXiv.1406.2661.

[42] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NeurIPS)*, 33: 6840–6851, 2020. URL https://arxiv.org/abs/2006.11239.

[43] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: a geometric diffusion model for molecular conformation generation. *International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2203.02923.

[44] James Lucas, George Tucker, Roger Grosse, and Mohammad Norouzi. Don't blame the ELBO! a linear vae perspective on posterior collapse. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. URL https://arxiv.org/abs/1911.02469.

[45] Weiheng Zhong and Hadi Meidani. Pi-vae: Physics-informed variational autoencoder for stochastic differential equations. *Computer Methods in Applied Mechanics and Engineering*, 403:115717, 2023. ISSN 0045-7825. doi: 10.1016/j.cma.2022.115717. URL https://www.sciencedirect.com/science/article/pii/S0045782522006193.

[46] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019. doi: 10.1126/science.aaw1147. URL https://doi.org/10.1126/science.aaw1147.

[47] James A. Maier, Carlos Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kirsten E. Hauser, and Carlos Simmerling. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015. doi: 10.1021/acs.jctc.5b00255. URL https://doi.org/10.1021/acs.jctc.5b00255.

[48] Namrata Anand and Possu Huang. Generative modeling for protein structures. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://papers.neurips.cc/paper/7978-generative-modeling-for-protein-structures.pdf.

[49] Brian Trippe et al. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *International Conference on Machine Learning (ICML)*, 2022. URL https://arxiv.org/abs/2206.04119.

[50] Gianni De Fabritiis Philipp Thölke. Torchmd-net: Equivariant transformers for neural network based molecular potentials. *Journal of Chemical Theory and Computation*, 2022. URL https://doi.org/10.48550/arXiv.2202.02541.

[51] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. *International Conference on Machine Learning (ICML)*, 2021. URL https://arxiv.org/abs/2102.09844.

[52] Fabian B. Fuchs, Max Welling, et al. Se(3)-transformers: 3d roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL https://arxiv.org/abs/2006.10503.

[53] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL https://www.deeplearningbook.org/.

[54] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2951–2959, 2012. URL https://par.nsf.gov/biblio/10195511.

[55] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In Grégoire Montavon, Genevieve B. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade*, pages 437–478. Springer, 2012. URL https://doi.org/10.1007/978-3-642-35289-8_26.

[56] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002. doi: 10.1109/4235.996017. URL https://doi.org/10.1109/4235.996017.

[57] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235. URL https://academic.oup.com/nar/article/28/1/235/2384399.

[58] Naveen Michaud-Agrawal, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10):2319–2327, 2011. doi: 10.1002/jcc.21787. URL https://doi.org/10.1002/jcc.21787.

[59] Richard J. Gowers, Max Linke, Jonathan Barnoud, Tyler J. E. Reddy, Manuel N. Melo, Sean L. Seyler, Jan Domański, David L. Dotson, Sébastien Buchoux, Ian M. Kenney, and Oliver Beckstein. Mdanalysis: A python package for the rapid analysis of molecular dynamics simulations. In S. Benthall and S. Rostrup, editors, *Proceedings of the 15th Python in Science Conference*, pages 98–105, 2016. URL https://conference.scipy.org/proceedings/scipy2016/pdfs/richard_gowers.pdf.

[60] Robert T. McGibbon, Kyle A. Beauchamp, Matthew P. Harrigan, Christoph Klein, Jason M. Swails, Carlos X. Hernández, Christian R. Schwantes, Lee-Ping Wang, Thomas J. Lane, and Vijay S. Pande. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8):1528–1532, 2015. doi: 10.1016/j.bpj.2015.08.015. URL https://doi.org/10.1016/j.bpj.2015.08.015.

[61] Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. doi: 10.1107/S0567739476001873. URL https://doi.org/10.1107/S0567739476001873.

[62] Scott C. Musson and Matteo T. Degiacomi. molearn: Machine learning for conformational ensembles. https://github.com/Degiacomi-Lab/molearn, 2023. GitHub repository.

[63] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations (ICLR)*, 2017. URL https://openreview.net/forum?id=Sy2fzU9gl.

[64] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, , et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.

[65] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 206–215, 2018. doi: 10.1109/CVPR.2018.00029. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Yang_FoldingNet_Point_Cloud_CVPR_2018_paper.html.

[66] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2 edition, 2017. doi: 10.1093/oso/9780198803195.001.0001. URL https://academic.oup.com/book/27866.

[67] Min-yi Shen and Andrej Sali. Statistical potential for assessment of protein structural models. *Protein Science*, 15(11):2507–2524, 2006. doi: 10.1110/ps.062416606. URL https://doi.org/10.1110/ps.062416606.

[68] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. URL https://arxiv.org/abs/1412.6980.

[69] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning, 2009. URL https://www.researchgate.net/publication/221344862_Curriculum_learning.

[70] Frank J. Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951. doi: 10.1080/01621459.1951.10500769. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1951.10500769.

[71] Pauli Virtanen et al. SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. URL https://doi.org/10.1038/s41592-019-0686-2.

[72] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996. doi: 10.1016/0263-7855(96)00018-5. URL https://pubmed.ncbi.nlm.nih.gov/8744570/.

[73] Yubo Qi, Weiyi Gong, and Qimin Yan. Bridging deep learning force fields and electronic structures with a physics-informed approach. *npj Computational Materials*, 11:177, 2025. doi: 10.1038/s41524-025-01668-5. URL https://www.nature.com/articles/s41524-025-01668-5.

[74] Weiheng Zhong and Hadi Meidani. Pi-vae: Physics-informed variational autoencoder for stochastic differential equations. *Computer Methods in Applied Mechanics and Engineering*, 403:115717, 2023. ISSN 0045-7825. doi: 10.1016/j.cma.2022.115717. URL https://www.sciencedirect.com/science/article/pii/S0045782522006193.