Novelty Assessment Report

Paper: Flow Autoencoders are Effective Protein Tokenizers
PDF URL: https://openreview.net/pdf?id=5p9uled7JM
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

Protein structure tokenizers enable the creation of multimodal models of protein structure, sequence, and function. Current approaches to protein structure tokenization rely on bespoke components that are invariant to spatial symmetries, but that are challenging to optimize and scale. We present Kanzi, a flow-based tokenizer for tokenization and generation of protein structures. Kanzi consists of a diffusion autoencoder trained with a flow matching loss. We show that this approach simplifies several aspects of protein structure tokenizers: frame-based representations can be replaced with global coordinates, complex losses are replaced with a single flow matching loss, and SE(3)-invariant attention operations can be replaced with standard attention. We find that these changes stabilize the training of parameter-efficient models that outperform existing tokenizers on reconstruction metrics at a fraction of the model size and training cost. An autoregressive model trained with Kanzi outperforms similar generative models that operate over tokens, although it does not yet match the performance of state-of-the-art continuous diffusion models.

Core Task Landscape

This paper addresses: **Protein Structure Tokenization and Generation**

A total of **39 papers** were analyzed and organized into a taxonomy with **17 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Discrete Structure Tokenization Methods**
- **Continuous Structure Representation and Embedding**
- **Multimodal Protein Language Models**
- **Evaluation and Benchmarking of Tokenization**
- **Structure Prediction and Reconstruction**
- **Protein Language Model Architectures and Training**
- **Computational Methods for Structure Discretization**

Complete Taxonomy Tree

- Protein Structure Tokenization and Generation Survey Taxonomy
- Discrete Structure Tokenization Methods
  - Vector-Quantized Autoencoder Approaches (5 papers)
  - [1] Learning the language of protein structure (Gaujac, 2024) View paper
  - [3] Balancing locality and reconstruction in protein structure tokenizer (Jiayou Zhang, 2024) View paper
  - [9] FoldToken2: Learning compact, invariant and generative protein structure language (Zhangyang Gao, 2024) View paper
  - [32] FoldToken3: Fold Structures Worth 256 Words or Less (Zhangyang Gao, 2024) View paper
  - [35] FoldToken4: Consistent & Hierarchical Fold Language (Zhangyang Gao, 2024) View paper
  - Geometry-Constrained Tokenization (2 papers)
  - [2] Protein Structure Tokenization via Geometric Byte Pair Encoding (Michael Sun, 2025) View paper
  - [19] GCP-VQVAE: A Geometry-Complete Language for Protein 3D Structure (Mahdi Pourmirzaei, 2025) View paper
  - Structure Alphabet and Symbolic Encoding (3 papers)
  - [4] Bilingual language model for protein sequence and structure (Michael Heinzinger, 2024) View paper
  - [11] Tokenizing foldable protein structures with machine-learned artificial amino-acid vocabulary (Xiaohan Lin, 2023) View paper
  - [29] SaProt: Protein Language Modeling with Structure-aware Vocabulary (Jin Su, 2023) View paper
  - Domain-Specific Structural Tokenization (1 papers)
  - [17] Tokenizing Loops of Antibodies (Fang, 2025) View paper
- Continuous Structure Representation and Embedding
  - Flow-Based and Diffusion Autoencoders ★ (2 papers)
  - [0] Flow Autoencoders are Effective Protein Tokenizers (Anon et al., 2026) View paper
  - [22] ProteinAE: Protein Diffusion Autoencoders for Structure Encoding (Li Shaoning, 2025) View paper
  - Continuous Latent Space Models (2 papers)
  - [16] The Continuous Language of Protein Structure (Lukas Billera, 2024) View paper
  - [21] HD-Prot: A Protein Language Model for Joint Sequence-Structure Modeling with Continuous Structure Tokens (Yi Zhou, 2025) View paper
  - All-Atom and Fine-Grained Encoding (2 papers)
  - [7] Bio2Token: All-atom tokenization of any biomolecular structure with Mamba (Liu, 2024) View paper
  - [33] P(all-atom) Is Unlocking New Path For Protein Design (Wei Qu, 2024) View paper

- Multimodal Protein Language Models
  - Joint Sequence-Structure Language Models (3 papers)
  - [12] Elucidating the Design Space of Multimodal Protein Language Models (Hsieh, 2025) View paper
  - [13] DPLM-2: A Multimodal Diffusion Protein Language Model (Wang Xin-you, 2024) View paper
  - [14] LM2Protein: A Structure-to-Token Protein Large Language Model (Chang Zhou, 2025) View paper
  - Structure-Aware Sequence Models (3 papers)
  - [5] ProtTeX: Structure-In-Context Reasoning and Editing of Proteins with Large Language Models (Ma Zicheng, 2025) View paper
  - [20] Integrating Functional Knowledge into Protein Design: A Novel Approach to Tokenization and Noise Injection for Function-Aware Protein Language Models (Tang, 2025) View paper
  - [24] Distilling Structural Representations into Protein Sequence Models (Jeffrey Ouyang-Zhang, 2024) View paper
  - Cross-Scale and Multi-Omics Integration (2 papers)
  - [6] Unified Cross-Scale 3D Generation and Understanding via Autoregressive Modeling (Lu, 2025) View paper
  - [25] Life-Code: Central Dogma Modeling with Multi-Omics Sequence Unification (Liu, 2025) View paper
- Evaluation and Benchmarking of Tokenization (3 papers)
  - [15] From Static Structures to Ensembles: Studying and Harnessing Protein Structure Tokenization (Zijing Liu, 2025) View paper
  - [27] Protein Structure Tokenization: Benchmarking and New Recipe (Yuan Xinyu, 2025) View paper
  - [37] Tokenized and continuous embedding compressions of protein sequence and structure. (Amy X. Lu, n.d.) View paper
- Structure Prediction and Reconstruction
  - Single-Sequence and Limited Homology Prediction (2 papers)
  - [8] Single-sequence protein structure prediction using supervised transformer protein language models (Wenkai Wang, 2022) View paper
  - [18] GhostFold: Accurate protein structure prediction using structure-constrained synthetic coevolutionary signals (Nitesh Mishra, 2025) View paper
  - Conformational Ensemble Generation (2 papers)
  - [26] Conformational ensembles for protein structure prediction (Jiaan Yang, 2025) View paper
  - [38] ProteinConformers: Benchmark Dataset for Simulating Protein Conformational Landscape Diversity and Plausibility (Y Zhou, n.d.) View paper
- Protein Language Model Architectures and Training
  - Long-Context and Interaction Modeling (2 papers)
  - [28] Pairing interacting protein sequences using masked language modeling (Umberto Lupo, 2023) View paper
  - [31] Long-context Protein Language Modeling Using Bidirectional Mamba with Shared Projection Layers (Wang, 2024) View paper
  - Structural Attention and Transformer Mechanisms (1 papers)
  - [23] Transformers trained on proteins can learn to attend to Euclidean distance (I. Ellmen, 2025) View paper
  - Controllable Generation and Fine-Tuning (1 papers)
  - [30] Controllable Protein Design by Prefix-Tuning Protein Language Models (Jiawei Luo, 2023) View paper
- Computational Methods for Structure Discretization (4 papers)
  - [10] The open-source Masala software suite: Facilitating rapid methods development for synthetic heteropolymer design (T. Zaborniak, 2025) View paper
  - [34] Tuning interval Branch-and-Prune for protein structure determination (Bradley Worley, 2018) View paper
  - [36] Discretization orders for protein side chains (V. Costa, 2014) View paper
  - [39] 4.13: Predicting Structure and Function of Biomolecules Through Natural Language Processing Tools (L Hallee, n.d.) View paper

Narrative

Core task: protein structure tokenization and generation. The field has organized itself around several complementary strategies for representing three-dimensional protein structures in forms amenable to machine learning. Discrete Structure Tokenization Methods focus on converting continuous coordinates into symbolic vocabularies—ranging from geometric clustering approaches like Geometric Byte Pair[2] to learned codebooks in works such as FoldToken2[9] and Bio2Token[7]. Continuous Structure Representation and Embedding takes an alternative path, learning smooth latent spaces through autoencoders and flow-based models that preserve geometric information without hard discretization. Multimodal Protein Language Models bridge sequence and structure by integrating both modalities, as seen in Language Protein Structure[1] and ProtTeX[5], while Protein Language Model Architectures and Training explores the underlying neural frameworks—including transformer variants and state-space models like Long-context Mamba[31]. Evaluation and Benchmarking of Tokenization, exemplified by Tokenization Benchmarking[27], provides systematic comparisons of these diverse encoding schemes, and Structure Prediction and Reconstruction addresses the inverse problem of generating plausible three-dimensional conformations from learned representations.

A central tension runs through the field between preserving fine-grained geometric detail and achieving compact, generalizable representations. Discrete tokenization methods often face trade-offs between vocabulary size and reconstruction fidelity, with works like Balancing Locality Reconstruction[3] explicitly addressing this challenge. In contrast, continuous embedding approaches—including ProteinAE[22] and the original Flow Autoencoders[0]—sidestep hard quantization by learning smooth latent manifolds, typically using variational or flow-based objectives to maintain structural coherence. Flow Autoencoders[0] sits squarely within this continuous paradigm, employing normalizing flows to map protein backbones into tractable distributions, closely aligned with ProteinAE[22] in its emphasis on differentiable, geometry-preserving encodings. Compared to hybrid approaches like Balancing Locality Reconstruction[3], which negotiate between discrete tokens and local geometric constraints, Flow Autoencoders[0] prioritizes end-to-end continuity, offering a complementary perspective on how to compress and generate structural diversity without categorical boundaries.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. ProteinAE: Protein Diffusion Autoencoders for Structure Encoding

Authors: Li Shaoning, Zhuo Le, Shaoning Li, Wang Yu-song, Le Zhuo, et al. (17 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Developing effective representations of protein structures is essential for advancing protein science, particularly for protein generative modeling. Current approaches often grapple with the complexities of the SE(3) manifold, rely on discrete tokenization, or the need for multiple training objectives, all of which can hinder the model optimization and generalization. We introduce ProteinAE, a novel and streamlined protein diffusion autoencoder designed to overcome these challenges by directly m...

Relationship Analysis

Both papers belong to the Flow-Based and Diffusion Autoencoders category, using flow matching or diffusion processes to learn continuous latent representations of protein structures. They share the core approach of employing flow-based autoencoders with diffusion losses for protein structure tokenization, operating on 3D coordinates rather than SE(3)-invariant representations, and using Diffusion Transformers (DiT) architectures. The key differences are that Kanzi focuses on discrete tokenization via vector quantization (FSQ) for downstream autoregressive generation, while ProteinAE emphasizes continuous latent representations with a bottleneck design for latent diffusion modeling, and ProteinAE explicitly incorporates length/dimension downsampling for compression efficiency.

Contributions Analysis

Overall novelty summary. The paper introduces Kanzi, a flow-based diffusion autoencoder for protein structure tokenization that replaces vector-quantized codebooks with continuous latent representations. Within the taxonomy, it resides in the 'Flow-Based and Diffusion Autoencoders' leaf under 'Continuous Structure Representation and Embedding', sharing this leaf with only one sibling paper (Flow Autoencoders). This places Kanzi in a relatively sparse research direction—only two papers occupy this specific methodological niche—suggesting the flow-matching approach to structure tokenization remains underexplored compared to the more crowded discrete tokenization branches.

The taxonomy reveals that most structure tokenization work clusters in 'Discrete Structure Tokenization Methods', particularly 'Vector-Quantized Autoencoder Approaches' (five papers) and 'Geometry-Constrained Tokenization' (two papers). Kanzi diverges from these by avoiding explicit codebooks and geometric invariance constraints, instead learning smooth latent spaces through flow matching. Its closest conceptual neighbors are continuous embedding methods like ProteinAE and the original Flow Autoencoders, yet it differs by framing tokenization as a diffusion process rather than pure variational or normalizing-flow objectives. This positions Kanzi at the boundary between continuous representation learning and the broader tokenization ecosystem.

Among eleven candidates examined, one paper was identified as potentially refuting the core contribution of a flow-based tokenizer, while nine others were non-refutable or unclear. The simplification contribution (replacing frame-based representations with global coordinates and standard attention) was examined against one candidate with no refutation found. The reconstruction metric contribution was not examined against any candidates. These statistics reflect a limited search scope—top-K semantic matches plus citation expansion—rather than exhaustive coverage. The flow-based tokenization approach appears less contested in the examined literature, though the small candidate pool (eleven total) limits confidence in this assessment.

Given the sparse occupancy of the flow-based autoencoder leaf and the limited overlap found among eleven examined candidates, Kanzi appears to occupy a relatively novel methodological position within the surveyed literature. However, the analysis is constrained by the search scope: only top-K semantic neighbors were examined, and the taxonomy itself captures thirty-nine papers across the broader field. A more exhaustive search—particularly within diffusion-based generative modeling and continuous embedding methods—might reveal additional overlapping work not surfaced by semantic similarity alone.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Kanzi: a flow-based protein structure tokenizer

Description: The authors introduce Kanzi, a novel protein structure tokenizer that uses a flow matching autoencoder architecture. Unlike existing tokenizers that rely on SE(3)-invariant components and complex losses, Kanzi operates on global coordinates with standard attention and uses a single flow matching loss for training.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Flexibility-Conditioned Protein Structure Design with Flow Matching

URL: View paper

Brief Assessment

Flexibility-Conditioned Design[47] focuses on conditioning protein structure generation on flexibility profiles using flow matching, not on developing flow-based tokenizers for protein structures. The candidate addresses a different problem domain (flexibility-conditioned design vs. structure tokenization).

---

2. Co-design protein sequence and structure in discrete space via generative flow

URL: View paper

Brief Assessment

Co-design Discrete Space[44] focuses on joint sequence-structure co-design using discrete flow in a multi-modal framework, not on developing a flow-based structure tokenizer architecture. The candidate addresses a different problem space (co-design) rather than tokenization methodology.

---

3. Design of ligand-binding proteins with atomic flow matching

URL: View paper

Brief Assessment

Ligand-binding Design[46] focuses on designing ligand-binding proteins using flow matching for protein-ligand complexes, not on protein structure tokenization. The candidate addresses a different problem (protein-ligand binding design) rather than creating a tokenizer for protein structures.

---

4. Proteina: Scaling Flow-based Protein Structure Generative Models

URL: View paper

Brief Assessment

Proteina Scaling[43] focuses on flow-based protein structure generation, not tokenization. While both use flow matching, Proteina generates protein backbones directly from noise, whereas Kanzi tokenizes structures into discrete representations for downstream tasks.

---

5. ProteinAE: Protein Diffusion Autoencoders for Structure Encoding

URL: View paper

Prior Art Analysis

ProteinAE[22] demonstrates that a flow-based autoencoder approach for protein structure encoding was developed independently and contemporaneously. Both papers present nearly identical core concepts: using flow matching/diffusion autoencoders to encode protein structures into latent representations, operating on global coordinates rather than SE(3)-invariant components, and training with a single flow matching loss instead of complex multi-objective losses. The architectural approaches and motivations are remarkably similar, suggesting that the novelty claim of being the first to propose this specific combination cannot be sustained.

Evidence

Evidence 1 - **Rationale**: Both papers describe the same core innovation: a flow/diffusion-based autoencoder that operates on global coordinates (E(3)) rather than SE(3)-invariant representations, trained with a single diffusion/flow loss. This demonstrates prior work exists with the same fundamental approach. - **Original**: we present kanzi, a flow-based tokenizer for tokenization and generation of protein structures. kanzi consists of an autoencoder trained with a flow matching loss. we show that this approach simplifies several aspects of protein structure tokenizers: frame-based representations can be replaced with ... - **Candidate**: we propose proteinae, a protein diffusion autoencoder designed for effective and efficient structure encoding and generation. specifically, proteinae operates in a non-equivariant manner, and conducts autoencoding protein backbone atoms (c α, n, c, o) directly on e(3), avoiding the discretization. i...

Evidence 2 - **Rationale**: Both papers explicitly state the same key design choices: operating directly on 3D backbone coordinates, using a single flow/diffusion loss, and avoiding SE(3)-invariant components. The approaches are functionally identical. - **Original**: kanzi uses a flow matching loss to train an autoencoder that tokenizes protein structures. this flow loss simplifies model training by replacing the collection of symmetry-invariant reconstruction losses that are commonly used to train protein structure tokenizers. kanzi operates directly on the 3d ... - **Candidate**: proteinae: a simple and effective protein diffusion autoencoder.we introduce a non-equivariant autoencoder based on diffusion transformers that operates directly on backbone atom coordinates ine(3). it learns a continuous, compact latent representation using a single flow-matching loss, avoiding the...

Evidence 3 - **Rationale**: Both papers identify the same limitations of prior work (SE(3) complexity, discrete tokenization, complex losses) and propose the same solution direction, indicating parallel development of the same core idea. - **Original**: current approaches to protein structure tokenization rely on bespoke components that are invariant to spatial symmetries, but that are challenging to optimize and scale. we present kanzi, a flow-based tokenizer for tokenization and generation of protein structures. - **Candidate**: researchers have made several attempts to encode protein structures with autoencoders. pioneering efforts in this area include the esm3 vq-v ae tokenizer (hayes et al., 2025) and the dplm-2 lookup-free quantization (lfq) tokenizer (wang et al., 2024; 2025), which were among the first to convert cont...

---

6. La-proteina: Atomistic protein generation via partially latent flow matching

URL: View paper

Brief Assessment

La-proteina[42] focuses on fully atomistic protein generation via partially latent flow matching, not on protein structure tokenization. The candidate uses flow matching for joint generation of sequences and full-atom structures, while the original contribution addresses tokenization of protein structures using a flow matching autoencoder.

---

7. Sequence-Augmented SE(3)-Flow Matching For Conditional Protein Backbone Generation

URL: View paper

Brief Assessment

Sequence-Augmented Flow[49] focuses on conditional protein backbone generation using SE(3)-flow matching with sequence conditioning, not on structure tokenization. The candidate addresses a different problem (generation conditioned on sequences) rather than tokenization with flow-based autoencoders.

---

8. All-atom inverse protein folding through discrete flow matching

URL: View paper

Brief Assessment

Discrete Flow Matching[41] focuses on inverse protein folding (sequence design from structure) using discrete flow matching, while Kanzi addresses protein structure tokenization (encoding structures into discrete tokens). These are fundamentally different tasks with different objectives and architectures.

---

9. Outsourced diffusion sampling: Efficient posterior inference in latent spaces of generative models

URL: View paper

Brief Assessment

Outsourced Diffusion Sampling[48] focuses on posterior inference in latent spaces of generative models across multiple domains (images, proteins), not on protein structure tokenization. The candidate uses flow-based models for conditional sampling tasks, while the original introduces a specific tokenizer architecture for protein structures.

---

10. Design of peptides with non-canonical amino acids using flow matching

URL: View paper

Brief Assessment

Non-canonical Peptides[45] focuses on designing peptides with non-canonical amino acids using flow matching for single-residue structure prediction, not on protein structure tokenization or autoencoder architectures for discrete token generation.

Contribution 2: Simplification of protein structure tokenization

Description: The authors demonstrate that their flow-based approach eliminates the need for SE(3)-invariant architectural components, frame-based representations, and collections of complex reconstruction losses that are standard in existing protein tokenizers, replacing them with simpler alternatives while maintaining or improving performance.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Deep generative modeling of atomistic systems

URL: View paper

Brief Assessment

Deep Generative Modeling[40] is a thesis focused on generative modeling of atomistic systems including molecular conformations and protein design, but does not address protein structure tokenization, frame-based representations, or the specific architectural simplifications claimed by the original paper.

Contribution 3: Reconstruction Fr´echet Protein Structure Distance (rFPSD) metric

Description: The authors propose rFPSD, a new distribution-level metric for evaluating protein structure tokenizers. This metric extends prior work on generative evaluation to the reconstruction task, providing broader information about tokenization performance beyond point-wise metrics like RMSD.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Appendix: Text Similarity Detection

Textual similarity detection checked 21 papers and found 2 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Sequence-Augmented SE(3)-Flow Matching For Conditional Protein Backbone Generation

Detected in: Contribution: contribution_1

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Flow Autoencoders are Effective Protein Tokenizers View paper
- [1] Learning the language of protein structure View paper
- [2] Protein Structure Tokenization via Geometric Byte Pair Encoding View paper
- [3] Balancing locality and reconstruction in protein structure tokenizer View paper
- [4] Bilingual language model for protein sequence and structure View paper
- [5] ProtTeX: Structure-In-Context Reasoning and Editing of Proteins with Large Language Models View paper
- [6] Unified Cross-Scale 3D Generation and Understanding via Autoregressive Modeling View paper
- [7] Bio2Token: All-atom tokenization of any biomolecular structure with Mamba View paper
- [8] Single-sequence protein structure prediction using supervised transformer protein language models View paper
- [9] FoldToken2: Learning compact, invariant and generative protein structure language View paper
- [10] The open-source Masala software suite: Facilitating rapid methods development for synthetic heteropolymer design View paper
- [11] Tokenizing foldable protein structures with machine-learned artificial amino-acid vocabulary View paper
- [12] Elucidating the Design Space of Multimodal Protein Language Models View paper
- [13] DPLM-2: A Multimodal Diffusion Protein Language Model View paper
- [14] LM2Protein: A Structure-to-Token Protein Large Language Model View paper
- [15] From Static Structures to Ensembles: Studying and Harnessing Protein Structure Tokenization View paper
- [16] The Continuous Language of Protein Structure View paper
- [17] Tokenizing Loops of Antibodies View paper
- [18] GhostFold: Accurate protein structure prediction using structure-constrained synthetic coevolutionary signals View paper
- [19] GCP-VQVAE: A Geometry-Complete Language for Protein 3D Structure View paper
- [20] Integrating Functional Knowledge into Protein Design: A Novel Approach to Tokenization and Noise Injection for Function-Aware Protein Language Models View paper
- [21] HD-Prot: A Protein Language Model for Joint Sequence-Structure Modeling with Continuous Structure Tokens View paper
- [22] ProteinAE: Protein Diffusion Autoencoders for Structure Encoding View paper
- [23] Transformers trained on proteins can learn to attend to Euclidean distance View paper
- [24] Distilling Structural Representations into Protein Sequence Models View paper
- [25] Life-Code: Central Dogma Modeling with Multi-Omics Sequence Unification View paper
- [26] Conformational ensembles for protein structure prediction View paper
- [27] Protein Structure Tokenization: Benchmarking and New Recipe View paper
- [28] Pairing interacting protein sequences using masked language modeling View paper
- [29] SaProt: Protein Language Modeling with Structure-aware Vocabulary View paper
- [30] Controllable Protein Design by Prefix-Tuning Protein Language Models View paper
- [31] Long-context Protein Language Modeling Using Bidirectional Mamba with Shared Projection Layers View paper
- [32] FoldToken3: Fold Structures Worth 256 Words or Less View paper
- [33] P(all-atom) Is Unlocking New Path For Protein Design View paper
- [34] Tuning interval Branch-and-Prune for protein structure determination View paper
- [35] FoldToken4: Consistent & Hierarchical Fold Language View paper
- [36] Discretization orders for protein side chains View paper
- [37] Tokenized and continuous embedding compressions of protein sequence and structure. View paper
- [38] ProteinConformers: Benchmark Dataset for Simulating Protein Conformational Landscape Diversity and Plausibility View paper
- [39] 4.13: Predicting Structure and Function of Biomolecules Through Natural Language Processing Tools View paper
- [40] Deep generative modeling of atomistic systems View paper
- [41] All-atom inverse protein folding through discrete flow matching View paper
- [42] La-proteina: Atomistic protein generation via partially latent flow matching View paper
- [43] Proteina: Scaling Flow-based Protein Structure Generative Models View paper
- [44] Co-design protein sequence and structure in discrete space via generative flow View paper
- [45] Design of peptides with non-canonical amino acids using flow matching View paper
- [46] Design of ligand-binding proteins with atomic flow matching View paper
- [47] Flexibility-Conditioned Protein Structure Design with Flow Matching View paper
- [48] Outsourced diffusion sampling: Efficient posterior inference in latent spaces of generative models View paper
- [49] Sequence-Augmented SE(3)-Flow Matching For Conditional Protein Backbone Generation View paper
- [50] Assessing generative model coverage of protein structures with SHAPES View paper
- [51] Annealed fractional LÃ©vyâ□□Itâ□ diffusion models for protein generation View paper
- [52] FrÃ©chet distance for curves, revisited View paper
- [53] Across atoms to crossing continents: Application of similarity measures to biological location data View paper
- [54] Protein structureâ□□structure alignment with discrete FrÃ©chet distance View paper
- [55] The discrete FrÃ©chet distance with applications View paper
- [56] Protein chain pair simplification under the discrete FrÃ©chet distance View paper
- [57] FrÃ©chet distance with speed limits View paper
- [58] FrÃ©chet distance of surfaces: Some simple hard cases View paper

- [59] Surface plasmon resonance unveils diffusion fingerprints of biomolecular mixtures in ocular fluid models. View paper