# 1141ML-Week 11 Final Project Toy Model Construction

Reconstructing Ancient Life through Molecular Diffusion Models

分子擴散模型重建古生命

313652018 王宣瑋

September 2025

## 1 Core Concepts from Reference Works

### (1) EDM: Equivariant Diffusion for Molecules

EDM[1] employs an Equivariant Graph Neural Network (EGNN) as the denoising model, enforcing the property:

$$s_\theta(Rx, t) = R\, s_\theta(x, t),$$

which guarantees SE(3)-equivariance.

The model is trained using the Denoising Score Matching (DSM) loss:

$$\mathcal{L}_{\text{DSM}} = \mathbb{E}_{x,\,\epsilon,\,t} \left\| s_\theta(x_t, t) + \frac{\epsilon}{\sigma_t} \right\|^2,$$

which is both simple to implement and mathematically stable for score-based diffusion.

### (2) GEODIFF (Xu et al., ICML 2022)

GEODIFF[3] proposes a complete diffusion framework operating directly in the 3D molecular conformation space. Its key ideas include:

- Forward diffusion is performed by adding Gaussian noise.
- Reverse diffusion is parameterized by an **SE(3)-equivariant denoiser**.
- The entire Markov chain is constructed to be equivariant (a geometric Markov kernel).

The forward noising process is defined as:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(\sqrt{\alpha_t}\, x_{t-1},\ \beta_t I)\,.$$

### (3) RFdiffusion (Nature 2023)

RFdiffusion[2] converts the RoseTTAFold protein structure prediction network into a 3D diffusion denoiser. Although the full architecture is considerably large, it conveys several essential insights:

- **3D diffusion = denoising network + iterative refinement.**
- **Self-conditioning** significantly stabilizes the generation process.
- The diffusion formulation is entirely consistent with GEODIFF and EDM; only the denoiser architecture is replaced by RoseTTAFold.

## 2 Objective of the Toy Model: 3D Point Cloud Generation

In this toy model, the goal is not to generate real molecular or protein structures, but rather to generate a *simplified three-dimensional point cloud*. By working with this simplified setting, the model can learn three essential capabilities that are required for more advanced 3D generative models in the future:

1. **Handling three-dimensional data.**
2. **Performing denoising within an SE(3)-equivariant architecture.**
3. **Executing iterative reverse diffusion sampling.**

This abstraction allows the toy model to focus on the geometric and mathematical foundations shared by GEODIFF, EDM, and RFdiffusion, while avoiding the complexities of real molecular constraints.

## 3 Toy Model Design

### (1) Data: Simplified 3D Point Clouds

We generate two clusters of 3D Gaussian blobs to mimic simple molecular-like shapes:

$$x \sim 0.5\mathcal{N}(\mu_1, I) \; + \; 0.5\mathcal{N}(\mu_2, I),$$

where $\mu_1, \mu_2 \in \mathbb{R}^3$ denote the centers of the two clusters. This abstraction allows us to focus on geometric diffusion without any chemical constraints.

### (2) Forward Diffusion Process

Following the formulation used in GEODIFF, the forward noising process is defined as:

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sigma_t \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I),$$

where $\bar{\alpha}_t$ represents the cumulative product of noise decay coefficients and $\sigma_t$ controls the noise scale.

### (3) Denoiser (Score Network)

We adopt a simplified Equivariant Graph Neural Network (EGNN) to serve as the denoising model. The message passing depends only on radial distances:

$$r_{ij} = \|x_i - x_j\|,$$

and the SE(3)-equivariant coordinate update rule is:

$$x_i' = x_i + \sum_j \phi(r_{ij}, h_i, h_j)\,(x_i - x_j),$$

where $\phi$ is a learned message function and $h_i$ represents node features.

### (4) DSM Training Loss

Training follows the Denoising Score Matching (DSM) objective, shared by EDM and GEODIFF:

$$\mathcal{L}_{\text{DSM}} = \mathbb{E}\left[\left\| s_\theta(x_t, t) + \frac{\epsilon}{\sigma_t} \right\|^2\right].$$

### (5) Reverse Sampling Procedure

Generation begins from pure Gaussian noise:

$$x_T \sim \mathcal{N}(0, I),$$

and proceeds iteratively with the score network estimating the denoising direction:

$$x_{t-1} = x_t + \beta_t \, s_\theta(x_t, t) + \sqrt{\beta_t}\, z, \qquad z \sim \mathcal{N}(0, I).$$

This iterative process gradually refines the noisy sample into a structured 3D point cloud.

## 4 Possible Extensions After Completing the Toy Model

Successfully implementing the toy model establishes the core mathematical and computational foundations required for future 3D generative modeling. Once the SE(3)-equivariant diffusion framework on simplified point clouds is functional, several meaningful and technically feasible extensions naturally follow:

### (1) Extending to Molecular Geometry (QM9 Dataset)

A direct next step is to move from abstract point clouds to small organic molecules. This involves incorporating atom types (C, N, O, H) and simple geometric constraints such as typical bond lengths and angles. Applying the diffusion pipeline to the QM9 dataset enables the model to learn realistic molecular conformations, forming a bridge between purely geometric generation and chemically meaningful structures.

### (2) Introducing Basic Chemical Priors

After mastering QM9-scale molecules, one can incorporate additional physics-based constraints:

- preferred bond lengths and bond angle distributions,
- valence-based restrictions,
- simple energy terms or geometry regularizers.

These priors reduce unrealistic samples and move the model closer to physically valid molecular design.

### (3) Generating Simplified Protein Backbones

A more ambitious yet achievable extension is to generate coarse protein backbone geometries. Instead of producing full atomic structures, the model can focus on generating the $(\mathrm{N}, \mathrm{C}_\alpha, \mathrm{C})$ backbone coordinates for short fragments (20–50 residues). This step introduces protein-like geometry while preserving computational tractability, functioning as a conceptual bridge toward RFdiffusion-style protein modeling.

# References


[1] Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pages 8867–8887. PMLR, 2022.

[2] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

[3] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.