Protein structure generation via folding diffusion

Kevin E. Wu[*]
Stanford University
wukevin@stanford.edu

Kevin K. Yang
Microsoft Research
yang.kevin@microsoft.com

Rianne van den Berg
Microsoft Research
rvandenberg@microsoft.com

James Y. Zou
Stanford University
jamesz@stanford.edu

Alex X. Lu
Microsoft Research
lualex@microsoft.com

Ava P. Amini
Microsoft Research
avasoleimany@microsoft.com

Abstract

The ability to computationally generate novel yet physically foldable protein structures could lead to new biological discoveries and new treatments targeting yet incurable diseases. Despite recent advances in protein structure prediction, directly generating diverse, novel protein structures from neural networks remains difficult. In this work, we present a new diffusion-based generative model that designs protein backbone structures via a procedure that mirrors the native folding process. We describe protein backbone structure as a series of consecutive angles capturing the relative orientation of the constituent amino acid residues, and generate new structures by denoising from a random, unfolded state towards a stable folded structure. Not only does this mirror how proteins biologically twist into energetically favorable conformations, the inherent shift and rotational invariance of this representation crucially alleviates the need for complex equivariant networks. We train a denoising diffusion probabilistic model with a simple transformer backbone and demonstrate that our resulting model unconditionally generates highly realistic protein structures with complexity and structural patterns akin to those of naturally-occurring proteins. As a useful resource, we release the first open-source codebase and trained models for protein structure diffusion.

1 Introduction

Proteins are critical for life, playing a role in almost every biological process, from relaying signals across neurons (Zhou et al., 2017) to recognizing microscopic invaders and subsequently activating the immune response (Mariuzza et al., 1987), from producing energy for cells (Bonora et al., 2012) to transporting molecules along cellular highways (Dominguez & Holmes, 2011). Misbehaving proteins, on the other hand, cause some of the most challenging ailments in human healthcare, including Alzheimer's disease, Parkinson's disease, Huntington's disease, and cystic fibrosis (Chaudhuri & Paul, 2006). As a consequence of their importance, proteins have been extensively studied as a therapeutic medium (Kamionka, 2011; Dimitrov, 2012) and constitute a rapidly growing segment of approved therapies (H Tobin et al., 2014). Thus, the ability to computationally generate novel yet physically foldable protein structures could open the door to discovering novel ways to harness cellular pathways and eventually lead to new treatments targeting yet incurable diseases.

[*] Work done during an internship at Microsoft Research

Many works have tackled the problem of generating new protein structures, but have generally run into challenges with creating diverse yet realistic folds. Traditional approaches typically apply heuristics to assemble fragments of experimentally profiled proteins into structures (Schenkelberg & Bystroff, 2016; Holm & Sander, 1991). This approach is limited by the boundaries of expert knowledge and available data. More recently, deep generative models have been proposed. However, due to the incredibly complex structure of proteins, these commonly do not directly generate protein structures, but rather constraints (such as pairwise distance between residues) that are heavily post-processed to obtain structures (Anand et al., 2019; Lee & Kim, 2022). Not only does this add complexity to the design pipeline, but noise in these predicted constraints can also be compounded during post-processing, resulting in unrealistic structures – that is, if the constraints are at all satisfiable to begin with. Other generative models rely on complex equivariant network architectures or loss functions to learn to generate a 3D point cloud that describes a protein structure (Anand & Achim, 2022; Trippe et al., 2022; Luo et al., 2022; Eguchi et al., 2022). Such equivariant architectures can ensure that the probability density from which the protein structures are sampled is invariant under translation and rotation. However, translation- and rotation-equivariant architectures are often also symmetric under reflection, leading to violations of fundamental structural properties of proteins like chirality (Trippe et al., 2022). Intuitively, this point cloud formulation is also quite detached from how proteins biologically fold – by twisting to adopt energetically favorable configurations (Šali et al., 1994; Englander et al., 2007).

Inspired by the *in vivo* protein folding process, we introduce a generative model that acts on the *inter-residue angles* in protein backbones instead of on Cartesian atom coordinates (Figure 1). This treats each residue as an independent reference frame, thus shifting the equivariance requirements from the neural network to the coordinate system itself. For generation, we use a denoising diffusion probabilistic model (diffusion model, for brevity) (Ho et al., 2020; Sohl-Dickstein et al., 2015) with a vanilla transformer parameterization without any equivariance constraints. Diffusion models train a neural network to start from noise and iteratively "denoise" it to generate data samples. Such models have been highly successful in a wide range of data modalities from images (Saharia et al., 2022; Rombach et al., 2022) to audio (Rouard & Hadjeres, 2021; Kong et al., 2021), and are easier to train with better modal coverage than methods like generative adversarial networks (GANs) (Dhariwal & Nichol, 2021; Nichol & Dhariwal, 2021). We present a suite of validations to quantitatively demonstrate that unconditional sampling from our model directly generates realistic protein backbones – from recapitulating the natural distribution of protein inter-residue angles, to producing overall structures with appropriate arrangements of multiple structural building block motifs. We show that our generated backbones are diverse and designable, and are thus biologically plausible protein structures. Our work demonstrates the power of biologically-inspired problem formulations and represents an important step towards accelerating the development of new proteins and protein-based therapies.

2 Related work

2.1 Generating new protein structures

Many generative deep learning architectures have been applied to the task of generating novel protein structures. Anand et al. (2019) train a GAN to sample pairwise distance matrices that describe protein backbone arrangements. However, these pairwise distance matrices must be corrected, refined, and converted into realizable backbones via two independent post-processing steps, the Alternating Direction Method of Multipliers (ADMM) and Rosetta. Crucially, inconsistencies in these predicted constraints can render them unsatisfiable or lead to significant errors when reconstructing the final protein structure. Sabban & Markovsky (2020) use a long short-term memory (LSTM) GAN to generate the (ϕ, ψ) dihedral angles. However, their network relies on downstream post-processing to filter, refine, and fold predicted 3D structures, partly due to the fact that these two dihedrals do not sufficiently specify backbone structure. Eguchi et al. (2022) propose a variational auto-encoder with equivariant losses to generate 3D coordinates for protein backbones. However, their work only targets immunoglobulin proteins and also requires refinement through Rosetta. Non-deep learning methods have also been explored. Schenkelberg & Bystroff (2016) apply heuristics to ensembles of similar sequences to make relatively small perturbations to known protein structures, while Holm & Sander (1991) use a database search to find and assemble existing protein fragments that might fit a new scaffold structure. These approaches' reliance on known proteins and hand-engineered heuristics limit them to relatively small deviations from naturally-occurring proteins.

2.1.1 Diffusion models for protein structure generation

Several recent works have proposed extending diffusion models towards generating protein structures. These predominantly perform diffusion on the 3D Cartesian coordinates of the residues themselves. For example, Trippe et al. (2022) use an E(3)-equivariant graph neural network to model the coordinates of protein residues. Anand & Achim (2022) adopt a hybrid approach where they train an equivariant transformer with invariant point attention (Jumper et al., 2021); this model generates the 3D coordinates of C_α atoms, the amino acid sequence, and the angles defining the orientation of side chains. Another recent work by Luo et al. (2022) performs diffusion for generating antibody fragments' structure and sequence by modeling 3D coordinates using an equivariant neural network. Note that these prior works all use some form of equivariance to translation, rotation, and/or reflection due to their formulation of diffusion on Cartesian coordinates. Another method, ProteinSGM, (Lee & Kim, 2022) implements a score-based diffusion model that generates image-like square matrices describing pairwise angles and distances between all residues in an amino acid chain. However, this set of values is highly over-constrained, and must be used as a set of *input constraints* for Rosetta's folding algorithm (Yang et al., 2020), which in turn produces the final folded output. This is a similar approach to Anand et al. (2019), and is likewise subject to the aforementioned concerns regarding complexity, satisfiability, and cleanliness of predicted constraints. Our work instead uses a minimal set of angles required to specify a protein backbone, and thus directly generates structures without relying on additional methods for refinement. Unfortunately, none of these prior works have publicly-available code, model weights, or generated examples at the time of this writing. Thus, our ability to perform direct qualitative and quantitative comparisons is limited.

2.2 Diffusion models for small molecules

A related line of work focuses on creating and modeling small molecules, typically in the context of drug design, using similar generative approaches. These small molecules average 44 atoms in size (Jing et al., 2022). Compared to proteins, which average several hundred residues and thousands of atoms (Tiessen et al., 2012), the relatively small size of small molecules makes them easier to model. The E(3) Equivariant Diffusion Model (Hoogeboom et al., 2022) uses an equivariant transformer to design small molecules by diffusing on their coordinates in Euclidean space. Other works have explored torsional diffusion, i.e., modelling the angles that specify a small molecule, to sample from the space of energetically favorable molecular conformations (Jing et al., 2022). This work still requires an $SE(3)$-equivariant model as the input to their model is still a 3D point cloud. In contrast, our problem formulation allows us to work entirely in terms of relative angles.

3 Method

3.1 Simplified framing of protein backbones using internal angles

Proteins are variable-length chains of amino acid residues. There are a total of 20 canonical amino acids, all of which share the same three-atom $N - C_\alpha - C$ backbone, but have varying side chains attached to the C_α atom (typically denoted R, see illustration in Figure 1). These residues assemble to form polymer chains typically hundreds of residues long (Tiessen et al., 2012). These chains of amino acids fold into 3D structures, taking on a shape that largely determines the protein's functions. These folded structures can be described on four levels: primary structure, which simply captures the linear sequence of amino acids; secondary structure, which describes the *local* arrangement of amino acids and includes structural motifs like α-helices and β-sheets; tertiary structure, which describes the full spatial arrangement of all residues; and quaternary structure, which describes how multiple different amino acid chains come together to form larger complexes (Sun et al., 2004).

We propose a simplified framing of protein backbones that follows the biological intuition of protein folding while removing the need for complex equivariant networks. Rather than viewing a protein backbone of length N amino acids as a cloud of 3D coordinates (i.e., $x \in \mathbb{R}^{N\times3}$ if modeling only C_α atoms, or $x \in \mathbb{R}^{3N\times3}$ for a full backbone) as prior works have done, we view it as a sequence of six internal, consecutive angles $x \in [-\pi, \pi)^{(N-1)\times6}$. That is, each vector of six angles describes

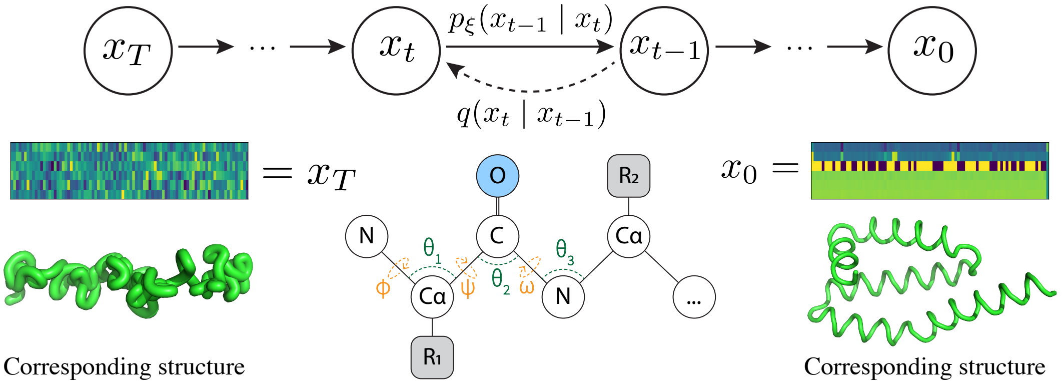


Figure 1: We perform diffusion on six angles as illustrated in the schematic in the bottom center (also defined in Table 1). Three of these are dihedral torsion angles (orange), and three are bond angles (green). We start with an experimentally observed backbone described by angles x_0 and iteratively add Gaussian noise via the forward noising process q until the angles are indistinguishable from a wrapped Gaussian at x_T. We use these examples to learn the "reverse" denoising process p_ξ.

Table 1: Internal angles used to specify protein backbone structure. Some of these involve multiple residues, indicated via i index subscripts. These are illustrated in Figure 1.

Angle	**Description**
ψ	Dihedral torsion about $N_i - C\alpha_i - C_i - N_{i+1}$
ω	Dihedral torsion about $C\alpha_i - C_i - N_{i+1} - C\alpha_{i+1}$
ϕ	Dihedral torsion about $C_i - N_{i+1} - C\alpha_{i+1} - C_{i+1}$
θ_1	Bond angle about $N_i - C\alpha_i - C_i$
θ_2	Bond angle about $C\alpha_i - C_i - N_{i+1}$
θ_3	Bond angle about $C_i - N_{i+1} - C\alpha_{i+1}$

the relative position of all backbone atoms in the *next* residue given the position of the *current* residue. These six angles are defined precisely in Table 1 and illustrated in Figure 1. These internal angles can be easily computed using trigonometry, and subsequently converted back to 3D Cartesian coordinates by iteratively adding atoms to the protein backbone as described in Parsons et al. (2005), keeping bond distances fixed to average lengths (see Appendix A.1, Figure S1).

This internal angle formulation has several key advantages. Most importantly, since each residue forms its own independent reference frame, there is no need to use an equivariant neural network. No matter how the protein is rotated or shifted, the angle of the *next* residue given the *current* residue never changes. This allows us to use a simple transformer as the backbone architecture; in fact, we demonstrate that our model fails when substituting our shift- and rotation-invariant internal angle representation with Cartesian coordinates, keeping all other design choices identical (see Appendix A.2, Figure S2). This internal angle formulation also closely mimics how proteins actually fold by twisting into more energetically stable conformations.

3.2 Denoising diffusion probabilistic models

Denoising diffusion probabilistic models (or diffusion models, for short) leverage a Markov process $q(x_t \mid x_{t-1})$ to corrupt a data sample x_0 over T discrete timesteps until it is indistinguishable from noise at x_T. A diffusion model $p_\xi(x_{t-1} \mid x_t)$ parameterized by ξ is trained to reverse this forward noising process, "denoising" pure noise towards samples that appear drawn from the native data distribution (Sohl-Dickstein et al., 2015). Diffusion models were first shown to achieve good generative performance by Ho et al. (2020); we adapt this framework for generating protein backbones, introducing necessary modifications to work with periodic angular values.

We modify the standard Markov forward noising process that adds noise at each discrete timestep t to sample from a wrapped normal instead of a standard normal (Jing et al., 2022):

$$q(x_t \mid x_{t-1}) = \mathcal{N}_{\text{wrapped}}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \propto \sum_{k=-\infty}^{\infty} \exp\left(\frac{-\|x_t - \sqrt{1-\beta_t}x_{t-1} + 2\pi k\|^2}{2\beta_t^2}\right)$$

where $\beta_t \in (0,1)_{t=1}^{T}$ are set by a variance schedule. We use the cosine variance schedule (Nichol & Dhariwal, 2021) with $T = 1000$ timesteps:

$$\beta_t = \text{clip}\left(1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}, 0.999\right) \quad \bar{\alpha}_t = \frac{f(t)}{f(0)} \quad f(t) = \cos\left(\frac{t/T + s}{1+s} \cdot \frac{\pi}{2}\right)$$

where $s = 8 \times 10^{-3}$ is a small constant for numerical stability. We train our model for $p_\xi(x_{t-1}|x_t)$ with the simplified loss proposed by Ho et al. (2020), using a neural network $\text{nn}_\xi(x_t, t)$ that predicts the noise $\epsilon \sim \mathcal{N}(0, I)$ present at a given timestep (rather than the denoised mean values themselves). To handle the periodic nature of angular values, we introduce a function to "wrap" values within the range $[-\pi, \pi)$: $w(x) = ((x + \pi) \bmod 2\pi) - \pi$. We use w to wrap a smooth L1 loss (Girshick, 2015) L_w, which behaves like L1 loss when error is high, and like an L2 loss when error is low; we set the transition between these two regimes at $\beta_L = 0.1\pi$. While this loss is not as well-motivated as torsional losses used by Jing et al. (2022), we find that it achieves strong empirical results.

$$d_w = w\left(\epsilon - \text{nn}_\xi\left(w\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon\right), t\right)\right)$$
$$L_w = \begin{cases} 0.5\frac{d_w^2}{\beta_L} & \text{if } |d_w| < \beta_L \\ |d_w| - 0.5\beta_L & \text{otherwise} \end{cases}$$

During training, timesteps are sampled uniformly $t \sim U(0, T)$. We normalize all angles in the training set to be zero mean by subtracting their element-wise angular mean μ; validation and test sets are shifted by this same offset.

Figure 1 illustrates this overall training process, including our previously described internal angle framing. The internal angles describing the folded chain x_0 are corrupted until they become indistinguishable from random angles, which results in a disordered mass of residues at x_T; we sample points along this diffusion process to train our model nn_ξ. Once trained, the reverse process of sampling from p_ξ also requires modifications to account for the periodic nature of angles, as described in Algorithm 1. The variance of this reverse process is given by $\sigma_t = \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t}$.

Algorithm 1 Sampling from p_ξ with FoldingDiff

```
$x_T \sim w(\mathcal{N}(0, I))$    ▷ Sample from a wrapped Gaussian
for $t = T, \ldots, 1$ do
    $z = \mathcal{N}(0, I)$ if $t > 1$ else $z = 0$
    $x_{t-1} = w\left(\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\text{nn}_\xi(x_t, t)\right) + \sigma_t z\right)$    ▷ Wrap sampled values about $[-\pi, \pi)$
end for
return $w(x_0 + \mu)$    ▷ Un-shift generated values by original mean shift
```

This sampling process can be intuitively described as refining internal angles from an unfolded state towards a folded state. As this is akin to how proteins fold *in vivo*, we name our method FoldingDiff.

3.3 Modeling and dataset

For our reverse (denoising) model $p_\xi(x_t, t)$, we adopt a vanilla bidirectional transformer architecture (Vaswani et al., 2017) with relative positional embeddings as described in Shaw et al. (2018). Our six-dimensional input is linearly upscaled to the model's embedding dimension ($d = 384$). To incorporate the timestep t into this model, we generate random Fourier feature embeddings (Tancik et al., 2020) as done in Song et al. (2020) and add these embeddings to each upscaled input. To convert the transformer's final per-position representations to our six outputs, we apply a regression head consisting of a densely connected layer, followed by GELU activation (Hendrycks & Gimpel, 2016), layer normalization, and finally a fully connected layer outputting our six values. We train this network with the AdamW optimizer (Loshchilov & Hutter, 2019) over 10,000 epochs, with a learning rate that linearly scales from 0 to 5×10^{-5} over 1,000 epochs, and back to 0 over the final 9,000 epochs. Validation loss appears to plateau after $\approx$ 1,400 epochs; additional training does not improve validation loss, but appears to lead to a poorer diversity of generated structures. We thus take a model checkpoint at 1,484 epochs for all subsequent analyses.

We train our model on the CATH dataset, which provides a "de-duplicated" set of proteins spanning a wide range of functions where no two chains share more than 40% sequence identity over 60% overlap (Sillitoe et al., 2015). We exclude any chains with fewer than 40 residues. Chains longer than 128 residues are randomly cropped to a 128-residue window at each epoch. Using a random 80/10/10 training/validation/test split, we have 24,316 training backbones, 3,039 validation backbones, and 3,040 test backbones. Expanding this training set is a target for future work.

4 Experiments

4.1 Generating protein internal angles

After training our model, we check that it is able to recapitulate the correct marginal distributions of dihedral and bond angles in proteins. We unconditionally generate 10 backbone chains each for every length from 50 to 128, which results in a total of 780 generated backbone chains. We plot the distributions of all six angles, aggregated across these 780 structures, and compare each distribution to that of test set structures less than 128 residues in length (Figures 2, S4). We observe that, across all angles, the generated distribution almost exactly recapitulates the test distribution. This is true both for angles that are nearly Gaussian with low variance $(\omega, \theta_1, \theta_2, \theta_3)$ as well as for angles with highly complex, high-variance distributions (ϕ, ψ). Compared to similar plots generated from other protein diffusion methods (e.g., Figure 1 in Anand & Achim (2022), reproduced with permission in Figure S5), we qualitatively observe that our method produces a much tighter distribution that more closely matches the natural distribution of bond angles.

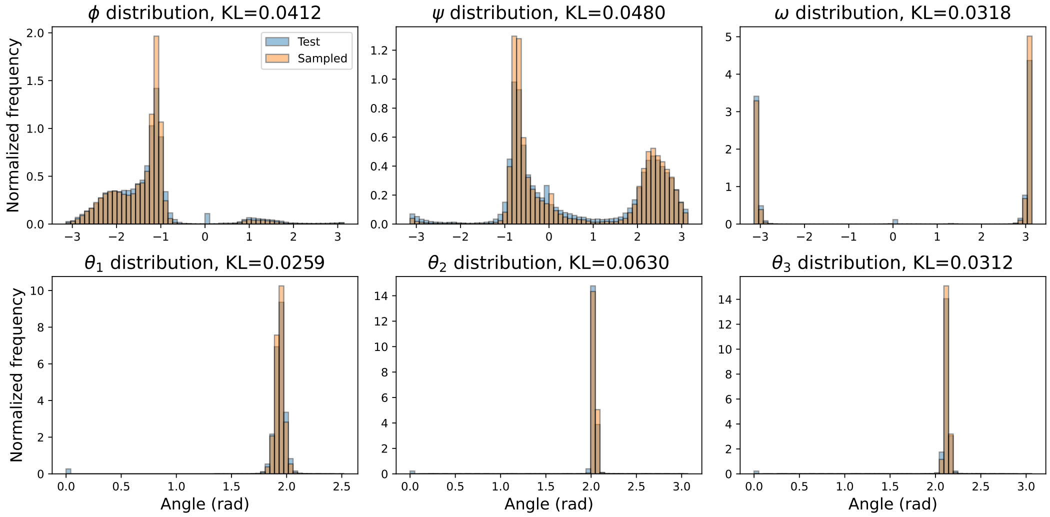
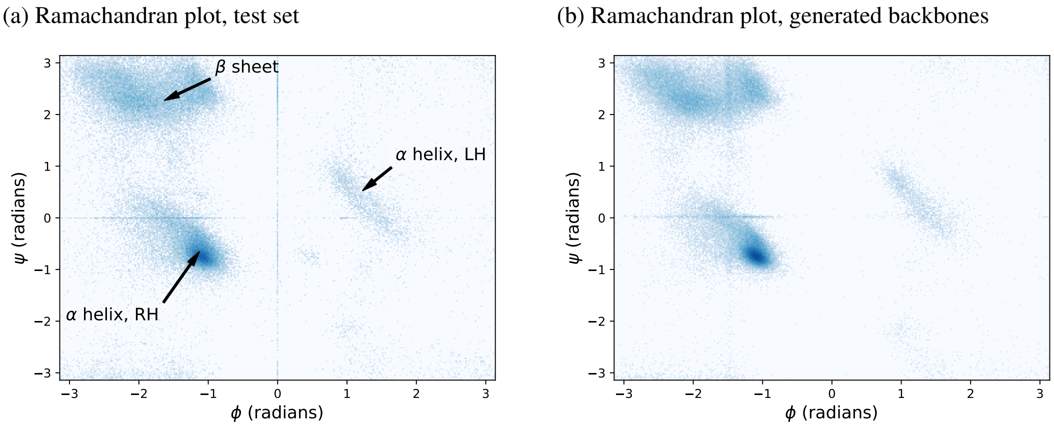


Figure 2: Comparison of the distributions of angular values in held-out test set and in generated samples. Top row shows dihedral angles (torsional angles involving 4 atoms), and bottom row shows bond angles (involving 3 atoms). KL divergence is calculated between $D_{KL}(\text{sampled}||\text{test})$. Figure S4 shows the cumulative distribution function (CDF) corresponding to these histograms.

However, looking at individual distributions of angles alone does not capture the fact that these angles are not independently distributed, but rather exhibit significant correlations. A Ramachandran plot, which plots the frequency of co-occurrence between the dihedrals (ϕ, ψ), is commonly used to illustrate these correlations between angles (Ramachandran & Sasisekharan, 1968). Figure 3 shows the Ramachandran plot for chains with fewer than 128 residues in the test set, as well as that for our 780 generated structures. The Ramachandran plot for natural structures (Figure 3a) contains three major concentrated regions corresponding to right-handed α helices, left-handed α helices, and β sheets. All three of these regions are recapitulated in our generated structures (Figure 3b). In other words, FoldingDiff is able to generate all three major secondary structure elements in protein backbones. Furthermore, we see that our model correctly learns that right-handed α helices are much more common than left-handed α helices (Cintas, 2002). Prior works that use equivariant networks, such as Trippe et al. (2022), cannot differentiate between these two types of helices due to network equivariance to reflection. This concretely demonstrates that our internal angle formulation leads to improved handling of chirality (i.e., the asymmetric nature of proteins) in generated backbones.


(a) Ramachandran plot, test set

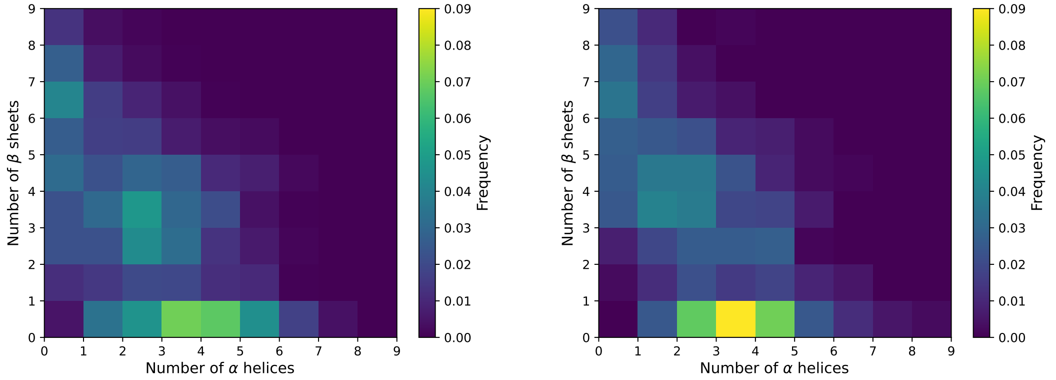
(b) Ramachandran plot, generated backbones

Figure 3: Ramachandran plots comparing the (ϕ, ψ) dihedral angles for test set (3a) and generated protein backbones (3b). Each major region of this plot indicates a different secondary structure element, as indicated in panel 3a. All three main structural elements are recapitulated in our generated backbones, along with some less common angle combinations. Lines are artifacts of null values.

4.2 Analyzing generated structures

We have shown that our model generates realistic distributions of angles and that our generated joint distributions capture secondary structure elements. We now demonstrate that the overall structures specified by these angles are biologically reasonable. Recall that naturally occurring protein structures contain multiple secondary structure elements. We use P-SEA (Labesse et al., 1997) to count the number of secondary structure elements in each test-set backbone of fewer than 128 residues, and plot the frequency of α/β co-occurrence counts in Figure 4a. Figure 4b repeats this analysis for our generated structures, which frequently contain multiple secondary structure elements just as naturally-occurring proteins do. FoldingDiff thus appears to generate rich structural information.

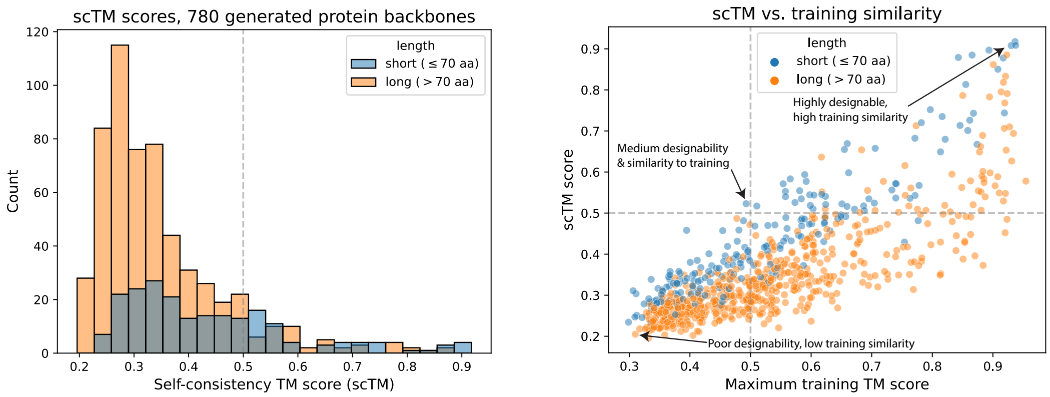
Beyond demonstrating that FoldingDiff's generated backbones contain reasonable structural motifs, it is also important to show that they are designable – meaning that we can find a sequence of amino acids that can fold into the designed backbone structure. After all, a novel protein structure is not useful if we cannot physically realize it. Previous works evaluate this *in silico* by predicting possible amino acids that fold into a generated backbone and checking whether the predicted structure for these sequences matches the original backbone. Following this general procedure, for a generated structure s, we use the ESM-IF1 inverse folding model (Hsu et al., 2022) to generate 8 different amino acid sequences. We then use OmegaFold (Wu et al., 2022) to predict the 3D structures $\hat{s}_1, \ldots, \hat{s}_8$ corresponding to each of these sequences. We use TMalign (Zhang & Skolnick, 2005), which evaluates structural similarity between backbones, to score each of these 8 structures against the original structure s. The maximum score $\max_{i \in [1,8]} \mathrm{TMalign}(s, \hat{s}_i)$ is the self-consistency TM (scTM) score. A scTM score of $\geq$ 0.5 is considered to be in the same fold, and thus is "designable." We repeat this process for each of our 780 generated backbones. This evaluation is similar to previous evaluations done by Trippe et al. (2022) and Lee & Kim (2022), except that we use OmegaFold instead of AlphaFold (Jumper et al., 2021). OmegaFold is designed to


(a) Secondary structure co-occurrence, test

(b) Secondary structure co-occurrence, generated

Figure 4: 2D histograms describing co-occurrence of secondary structures in test backbones (4a) and generated backbones (4b). Axes indicate the number of secondary structure present in a chain; color indicates the frequency of a specific combination of secondary structure elements. Our generated structures mirror real structures with multiple α helices, multiple β sheets, and a mixture of both.

be used without multiple sequence alignments (MSAs), and performs similarly to AlphaFold while generalizing better to orphan proteins that may not have such evoluationary annotations (Wu et al., 2022). Furthermore, given that prior works use AlphaFold without MSA information in their evaluation pipelines (Trippe et al., 2022; Lee & Kim, 2022), OmegaFold appears to be a more appropriate method for scTM evaluation.



(a) Backbone designability by length

(b) Designability compared to training set similarity

Figure 5: Of our 780 generated backbones, ranging in length from 50-128 residues, we observe that 111 are designable $(\text{scTM} \geq 0.5)$. Shorter structures of less than 70 amino acids tend to have higher scTM scores than longer structures (5a). We also see that generated backbones that are more similar to training examples (greater maximum training TM score) tend to have better designability (5b). The three structures indicated by arrows are illustrated in Figure S6.

Overall, we find that 111 of our 780 structures (14.2%) are designable with an scTM score ≥ 0.5 (Figure 5a), which is higher than the value of 11.8% reported by Trippe et al. (2022). Our improved designability comes from a significantly higher proportion of short structures (≤ 70 residues) being designable (57/210 for ours compared to 36/210 in Trippe et al. (2022), $p = 0.014$, Chi-square test). There is no significant difference in designability for longer structures of 71-128 residues (54/570 for ours compared to 51/570 in Trippe et al. (2022), $p = 0.76$, Chi-square test). While ProteinSGM (Lee & Kim, 2022) reports an even higher designability proportion of 50.5%, this metric is not directly comparable, as ProteinSGM generates sets of *constraints* that are then used to fold the final structure with Rosetta, rather than generating backbone structures directly as with our approach or that of Trippe et al. (2022). It is unsurprising that a structure produced by a dedicated protein folding tool produces greater designability. The authors themselves note that this Rosetta “post-processing” significantly improves the viability of their structures. To additionally contextualize our scTM scores, we benchmark against a naive baseline structure sampling method that preserves the overall distribution of protein bond angles but destroys their positional relationships; FoldingDiff significantly outperforms this baseline (see Appendix A.3, Figure S3).

We additionally evaluate the similarity of each generated backbone to any training backbone by taking the maximum TM score across the entire training set. The maximum training TM-score is significantly correlated with scTM score (Spearman’s $r = 0.79$, $p = 1.2 \times 10^{-165}$, Figure 5b), indicating that structures more similar to the training set tend to be more designable. However, this does not suggest that we are merely memorizing the training set; doing so would result in a distribution of training TM scores near 1.0, which is not what we observe. We note that ProteinSGM (Lee & Kim, 2022) reports a distribution of training set TMscores much closer to 1.0; this suggests a greater degree of memorization and may contribute to their high reported scTM designability ratio.

Selected examples of our generated backbones and corresponding OmegaFold predictions of various lengths are visualized using PyMOL (Schrödinger, LLC, 2015) in Figure 6. Interestingly, we find that of our 111 designable backbones, only 4 contain β sheets as annotated by P-SEA. Conversely, of our 669 backbones with scTM < 0.5, 545 contain β sheets. This suggests that generated structures with β sheets may be less designable ($p < 1.0 \times 10^{-5}$, Chi-square test). It is unclear whether this is due to our model generating poor backbones in this condition, or because ESM-IF1 and OmegaFold struggle with β sheets. We additionally cluster our designable backbones and observe a large diversity of structures (Figure S8). This suggests that our model is not simply generating small variants of a handful of core structures, which prior works appear to do (Figure S9).

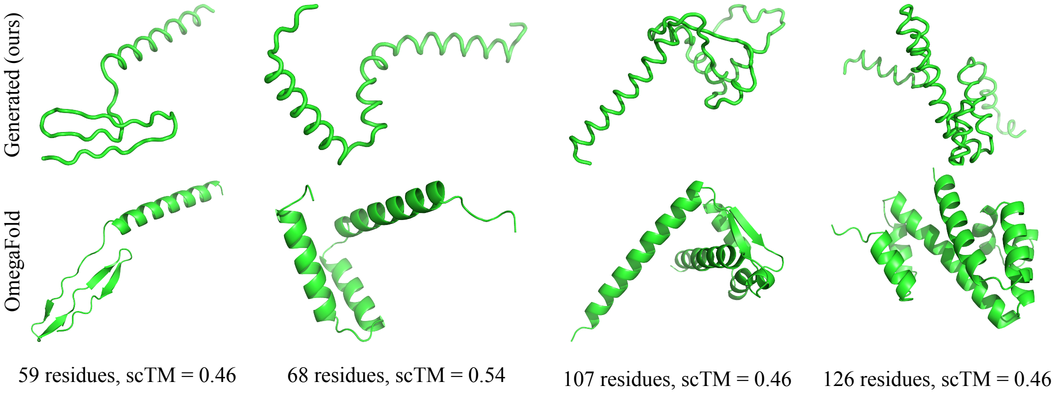


Figure 6: Selected generated protein backbones of varying length that are approximately designable ($\mathrm{scTM} \approx 0.5$). Top row shows our directly generated backbones; bottom row shows OmegaFold predicted structure for residues inferred by ESM-IF1 to produce our generated backbone. Structures contain both α helices (coils, columns 1-4) and β sheets (ribbons, columns 1 and 3), and each appears meaningfully different from its most similar training example (Figure S7).

5 Conclusion

In this work, we present a novel parameterization of protein backbone structures that allows for simplified generative modeling. By considering each residue to be its own reference frame, we describe a protein using the resulting relative internal angle representation. We show that a vanilla transformer can then be used to build a diffusion model that generates high-quality, biologically plausible, diverse protein structures. These generated backbones better respect protein chirality and exhibit greater designability compared to prior works that use equivariance assumptions.

While we demonstrate promising results with our model, there are several limitations to our work. Although formulating a protein as a series of angles enables us to use simpler models without equivariance mechanisms, this framing allows accumulated errors to significantly alter the overall generated structure. This takes two forms: small errors iteratively compounding into larger errors over many residues, and single large errors drastically modifying the structure. Additionally, some generated structures exhibit collisions where the generated structure crosses through itself. Future work could explore methods to avoid these pitfalls using geometrically-informed architectures such as those used in Wu et al. (2022). Our generated structures are still of relatively short lengths (up to 128 residues) compared to natural proteins which typically have several hundred residues. We also do not handle multi-chain complexes or ligand interactions, and are only able to generate static structures that do not capture the dynamic nature of proteins. Additional future work could incorporate amino acid sequence generation in parallel with structure generation, along with guided generation using protein function or domain annotations. In summary, our work provides an important step in using biologically-inspired problem formulations for generative protein design.

Code availability and reproducibility

All code for training our model and performing downstream analyses is available at `https://github.com/microsoft/foldingdiff`. Trained model weights used for generating all results in this manuscript are available there as well.

Author Contributions

K.E.W., K.K.Y., A.X.L., and A.P.A. initiated, conceived, and designed the work. K.E.W. performed modeling and analyses, with input from all authors. R.vdB., K.K.Y., and A.P.A. provided guidance on diffusion models and their implementation. A.P.A., K.K.Y., J.Z., and A.X.L. provided guidance on evaluation methods. A.P.A., K.K.Y., and A.X.L. supervised the research. All authors wrote the manuscript.

Acknowledgments

We thank Brian Trippe, Jason Yim, Namrata Anand, and Tudor Achim for permission to reuse their figures.

References

Namrata Anand and Tudor Achim. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. *arXiv preprint arXiv:2205.15019*, 2022.

Namrata Anand, Raphael Eguchi, and Po-Ssu Huang. Fully differentiable full-atom protein backbone generation. In *DGS@ICLR*, 2019.

Massimo Bonora, Simone Patergnani, Alessandro Rimessi, Elena De Marchi, Jan M Suski, Angela Bononi, Carlotta Giorgi, Saverio Marchi, Sonia Missiroli, Federica Poletti, et al. ATP synthesis and storage. *Purinergic Signalling*, 8(3):343–357, 2012.

Tapan K Chaudhuri and Subhankar Paul. Protein-misfolding diseases and chaperone-based therapeutic approaches. *The FEBS Journal*, 273(7):1331–1349, 2006.

Pedro Cintas. Chirality of living systems: a helping hand from crystals and oligopeptides. *Angewandte Chemie International Edition*, 41(7):1139–1145, 2002.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

Dimiter S Dimitrov. Therapeutic proteins. *Therapeutic Proteins*, pp. 1–26, 2012.

Roberto Dominguez and Kenneth C Holmes. Actin structure and function. *Annual Review of Biophysics*, 40:169, 2011.

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.

S Walter Englander, Leland Mayne, and Mallela MG Krishna. Protein folding and misfolding: mechanism and principles. *Quarterly Reviews of Biophysics*, 40(4):1–41, 2007.

Ross Girshick. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, 2015.

Peter H Tobin, David H Richards, Randolph A Callender, and Corey J Wilson. Protein engineering: a new frontier for biological therapeutics. *Current Drug Metabolism*, 15(7):743–756, 2014.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (GELUs). *arXiv preprint arXiv:1606.08415*, 2016.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Liisa Holm and Chris Sander. Database algorithm for generating protein backbone and side-chain co-ordinates from a Cα trace: application to model building and detection of co-ordinate errors. *Journal of Molecular Biology*, 218(1):183–194, 1991.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In *International Conference on Machine Learning*, pp. 8867–8887. PMLR, 2022.

Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In *International Conference on Machine Learning*, pp. 8946–8970. PMLR, 2022.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *arXiv preprint arXiv:2206.01729*, 2022.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021.

Mariusz Kamionka. Engineering of therapeutic proteins production in Escherichia coli. *Current Pharmaceutical Biotechnology*, 12(2):268–274, 2011.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*, 2021.

Gilles Labesse, N Colloc'h, Joël Pothier, and J-P Mornon. P-SEA: a new efficient assignment of secondary structure from Cα trace of proteins. *Bioinformatics*, 13(3):291–295, 1997.

Jin Sub Lee and Philip M. Kim. ProteinSGM: Score-based generative modeling for de novo protein design. *bioRxiv*, 2022. doi: 10.1101/2022.07.13.499967.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *bioRxiv*, 2022. doi: 10.1101/2022.07.10.499510.

RA Mariuzza, SEV Phillips, and RJ Poljak. The structural basis of antigen-antibody recognition. *Annual Review of Biophysics and Biophysical Chemistry*, 16(1):139–159, 1987.

Alexander Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pp. 8162–8171. PMLR, 2021.

Jerod Parsons, J Bradley Holmes, J Maurice Rojas, Jerry Tsai, and Charlie EM Strauss. Practical conversion from torsion space to cartesian space for in silico protein synthesis. *Journal of Computational Chemistry*, 26(10):1063–1068, 2005.

GN Ramachandran and V Sasisekharan. Conformation of polypeptides and proteins. *Advances in Protein Chemistry*, 23:283–437, 1968.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

Simon Rouard and Gaëtan Hadjeres. CRASH: Raw audio score-based generative modeling for controllable high-resolution drum sound synthesis. *arXiv preprint arXiv:2106.07431*, 2021.

Sari Sabban and Mikhail Markovsky. RamaNet: Computational de novo helical protein backbone design using a long short-term memory generative neural network. *bioRxiv*, pp. 671552, 2020.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, et al. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*, 2022.

Andrej Šali, Eugene Shakhnovich, and Martin Karplus. How does a protein fold. *Nature*, 369(6477): 248–251, 1994.

Christian D Schenkelberg and Christopher Bystroff. Protein backbone ensemble generation explores the local structural space of unseen natural homologs. *Bioinformatics*, 32(10):1454–1461, 2016.

Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*, 2018.

Ian Sillitoe, Tony E Lewis, Alison Cuff, Sayoni Das, Paul Ashford, Natalie L Dawson, Nicholas Furnham, Roman A Laskowski, David Lee, Jonathan G Lees, et al. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Research*, 43(D1): D376–D381, 2015.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Peter D Sun, Christine E Foster, and Jeffrey C Boyington. Overview of protein structural and functional folds. *Current Protocols in Protein Science*, 35(1):17–1, 2004.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020.

Martha M Teeter. Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. *Proceedings of the National Academy of Sciences*, 81 (19):6014–6018, 1984.

Axel Tiessen, Paulino Pérez-Rodríguez, and Luis José Delaye-Arredondo. Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. *BMC Research Notes*, 5(1):1–23, 2012.

Brian L Trippe, Jason Yim, Doug Tischer, Tamara Broderick, David Baker, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Ruidong Wu, Fan Ding, Rui Wang, Rui Shen, Xiwen Zhang, Shitong Luo, Chenpeng Su, Zuofan Wu, Qi Xie, Bonnie Berger, Jianzhu Ma, and Jian Peng. High-resolution de novo structure prediction from primary sequence. *bioRxiv*, 2022. doi: 10.1101/2022.07.21.500999.

Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3):1496–1503, 2020.

Yang Zhang and Jeffrey Skolnick. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33(7):2302–2309, 2005.

Qiangjun Zhou, Peng Zhou, Austin L Wang, Dick Wu, Minglei Zhao, Thomas C Südhof, and Axel T Brunger. The primed snare–complexin–synaptotagmin complex for neuronal exocytosis. *Nature*, 548(7668):420–425, 2017.

A Appendix

A.1 Internal angle formulation of protein backbones

A protein backbone structure can be fully specified by a total of 9 values per residue: 3 bond distances, 3 bond angles, and 3 dihedral torsional angles. The three bond angles and dihedrals are described in Table 1, and the three bond distances correspond to $N_i \to C\alpha_i$, $C\alpha_i \to C_i$, and $C_i \to N_{i+1}$ where i denotes residue index. These values enable a protein backbone to be losslessly converted from Cartesian to internal angle representation, and vice versa. To determine which subset of values to use to frame proteins in our model, we take a set of experimentally profiled proteins and translate their coordinates from Cartesian to internal angles and distances and back, measuring the TM score between the initial and reconstructed structures. When excluding an angle or distance, we fix all corresponding values to the mean. The reconstruction TM scores of various combinations of values is illustrated in Figure S1. Of these 9 values, the three bond distances are the least important for reliably reconstructing a structure from Cartesian coordinates to the inter-residue representation and back; they can usually be replaced with constant average values without much impact on the recovered structure. In comparison, removing even two bond angles with relatively little variance (θ_2, θ_3) results in a large loss in reconstruction TM score (third bar). Removing all bond angles and retaining only dihedrals (ϕ, ψ, ω) results in only about half of proteins being able to be reconstructed (last bar). Thus, we model the three dihedrals and the three bond angles (second bar in Figure S1); this simplifies our prediction problem to use only periodic angular values (instead of a mixture of angular and real values) without a substantial loss in the accuracy of described structures. Future work might include additional modeling of these real-valued bond distances.

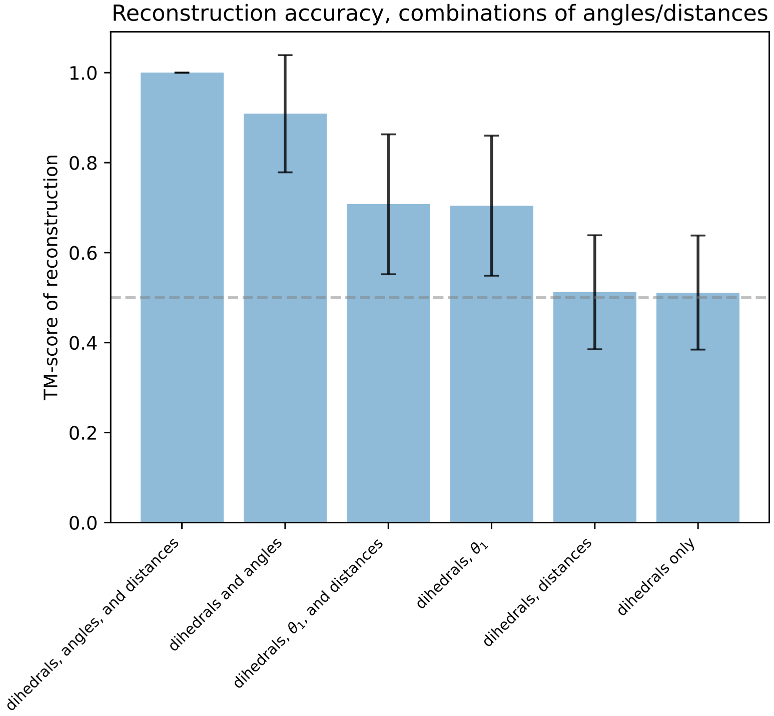


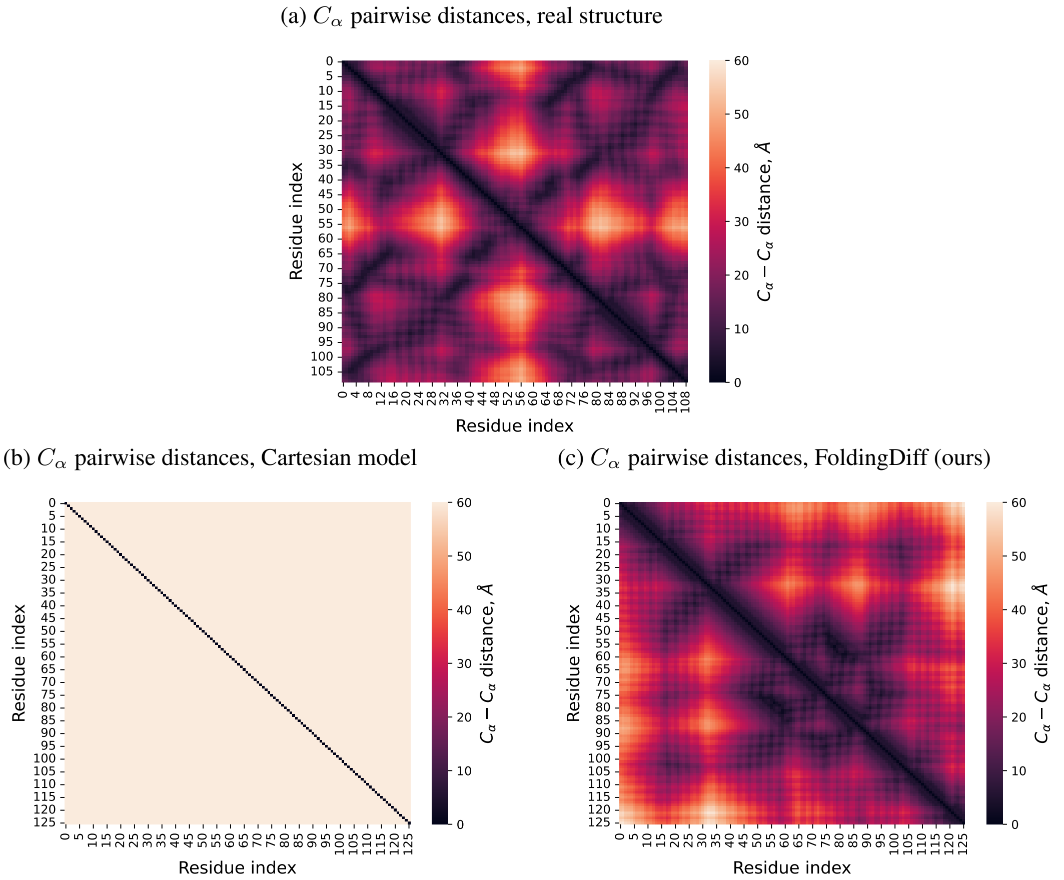
Figure S1: Comparison of various combinations of angles and distances and their ability to reconstruct original protein backbones. Error bars represent standard deviation in TM scores when reconstructing the protein backbone. Dihedrals include ϕ, ψ, ω, and bond angles include $\theta_1, \theta_2, \theta_3$ (Table 1). The first column, with all bond angles, dihedral angles, and bond distances, perfectly reconstructs Cartesian coordinates from internal angles. The second column corresponds to the formulation used in the main text, where we model the 3 dihedrals and 3 bond angles, but keep the 3 bond distances fixed to average values. Other columns replace even more values with their respective means and result in reconstruction TM scores that are too low to be reliably useful.

One challenge with converting between a N-residue set of Cartesian coordinates to a set of $N-1$ angles between consecutive residues is that the latter representation does not capture the first residue's information. To solve this, we use a fixed set of coordinates to "seed" generation of Cartesian coordinates, using the $N-1$ specified angles to build out from this fixed point. For all generations, this fixed point is extracted from the coordinates of the $N - C_\alpha - C$ atoms in first residue on the N-terminus of the PDB structure 1CRN (Teeter, 1984).

A.2 Substituting internal angle formulation for Cartesian coordinates

We perform an "ablation" of our internal angle representation by replacing our framing of proteins as a series of inter-residue internal angles with a simple Cartesian representation of C_α coordinates $x \in \mathbb{R}^{N\times 3}$. Notably, this Cartesian representation is no longer rotation or shift invariant. We train a denoising diffusion model using this Cartesian representation, using the same variance schedule, transformer backbone architecture, and loss function, but sampling from a standard Gaussian and with all usages of our wrapping function w removed. This represents the same modelling approach as our main diffusion model, with only our internal angle formulation removed.

To evaluate the quality of this Cartesian-based diffusion model's generated structures, we calculate the pairwise distances between all C_α atoms in its generated structures and compare these with distance matrices calculated for real proteins and for our internal angle diffusion model's generations. For a real protein, this produces a pattern that reveals close proximity between pairwise residues where the protein is folded inwards to produce a compact, coherent structure (Figure S2a). However, similarly visualizing the C_α pairwise distances in the Cartesian model's generated structures yields no significant proximity or patterns between any residues (Figure S2b). This suggests that the ablated Cartesian model cannot learn to generate meaningful structure. Our internal angle model, on the other hand, produces a visualization that is very similar to that of real proteins (Figure S2c). Simply put, our model's performance drastically degrades when we change only how inputs are represented. This demonstrates the importance and effectiveness of our internal angle formulation.


(a) C_α pairwise distances, real structure

(b) C_α pairwise distances, Cartesian model

(c) C_α pairwise distances, FoldingDiff (ours)

Figure S2: Pairwise distances between all C_α atoms in various protein backbone structures, all of similar length. All panels use the same color scale. S2a illustrates a set of distances for a real protein structure; note the visual patterns that correspond to various secondary structures and potential contacts and interactions between residues. S2b shows these distances for a structure generated by an ablated model that replaces our proposed internal angle representation with Cartesian coordinates, which results in no coherent structural generation. For comparison, our FoldingDiff model produces structures that compactly fold to create many potential contacts, just as real proteins do (S2c). This final distance matrix corresponds to the generated structure in the right-most column in Figure 6.

A.3 Baseline method for contextualizing scTM scores

To contextualize the distribution of scTM scores (Figure 5a), we implement a naive angle generation baseline. We take our test dataset, and concatenate all examples into a matrix of $\hat{x} \in [-\pi, \pi)^{\hat{N} \times 6}$, where $\hat{N}$ denotes the total number of angle sets in our test dataset, aggregating across all individual chains. To generate a backbone structure of length l, we simply sample l indices from $U(0, \hat{N})$. This creates a chain that perfectly matches the natural distribution of protein internal angles, while also perfectly reproducing the pairwise correlations, i.e., of dihedrals in a Ramachandran plot, but critically loses the correct *ordering* of these angles. We generate 780 such structures (10 samples for each integer value of $l \in [50, 128)$). This is the same distribution of lengths as the generated set in our main analysis. For each of these, we perform the same scTM evaluation as before. The distribution of scTM scores for these randomly-sampled structures compared to that of FoldingDiff's generated backbones is shown in Figure S3. We observe that this random protein generation method produces significantly poorer scTM scores than our model does ($p = 4.4 \times 10^{-60}$, Mann-Whitney test). This suggests that our model is not simply learning the overall distribution of angles, but is learning the correct *ordering* of angles that comprise a folded protein structure.

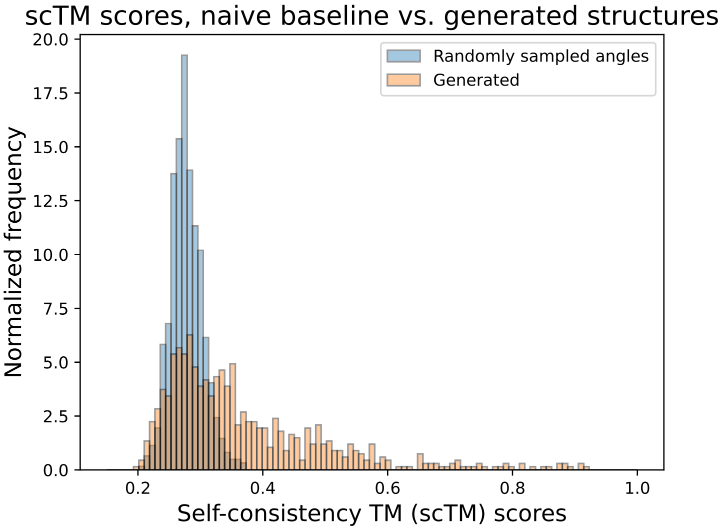


Figure S3: Distribution of scTM scores for our sampled proteins, compared to scTM scores for structures created by randomly shuffling naturally-occurring internal angles. The randomly sampled angles result in no designable structures, despite perfectly capturing the overall distribution and pairwise relations between angles. This suggests our method correctly learns the spatial ordering of angles that folds a valid structure.

A.4 Additional Supplementary Figures

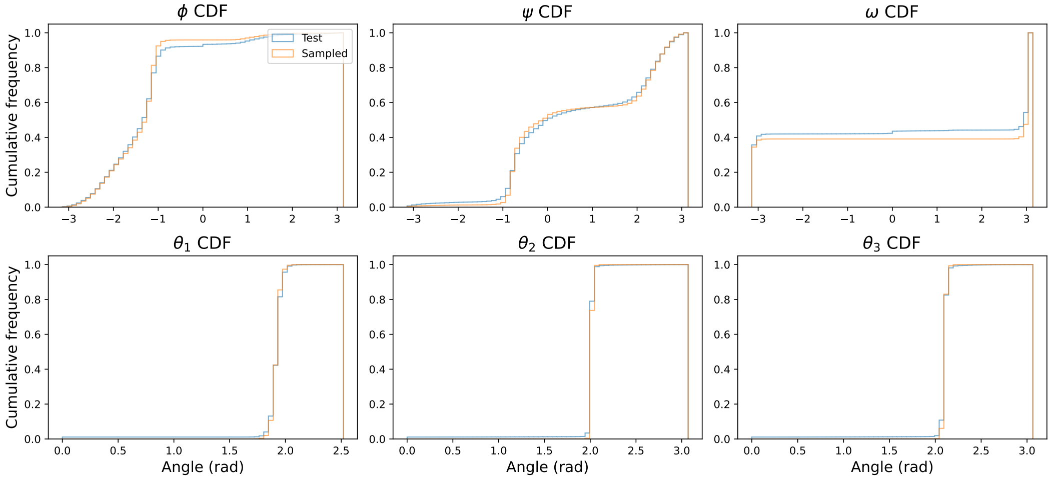


Figure S4: Comparison of the cumulative distribution functions (CDF) of angular values in test set and in generated samples. Top row shows dihedral angles (torsional angles involving 4 atoms), and bottom row shows bond angles (involving 3 atoms). Figure 2 shows the histogram distributions corresponding to these CDFs.

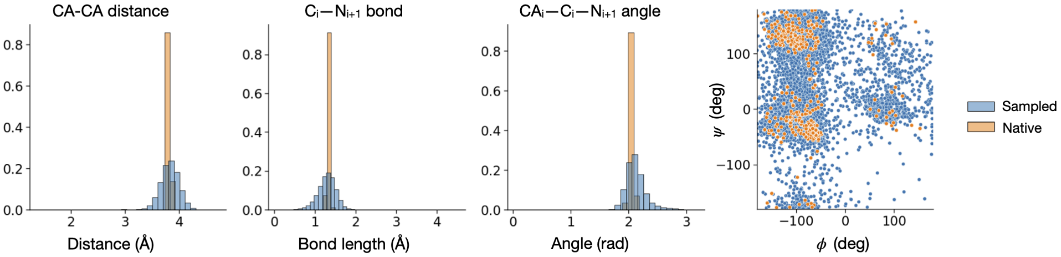
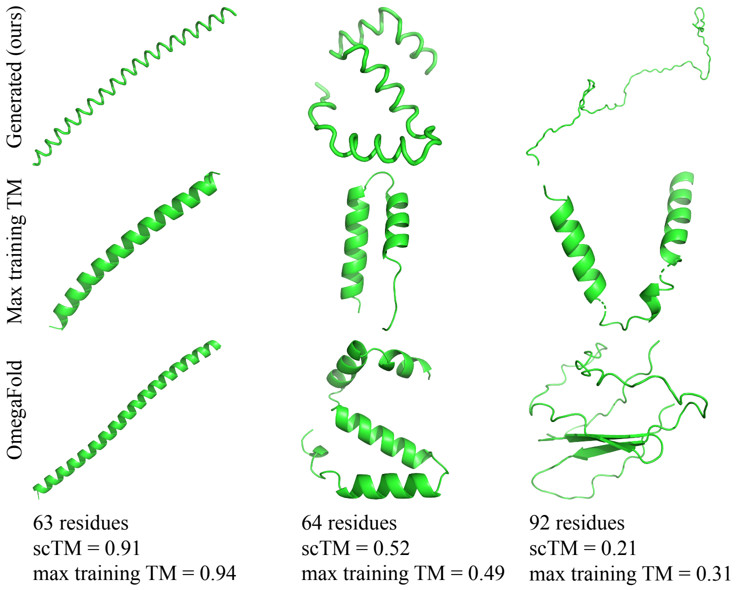


Figure S5: Figure 1B from Anand & Achim (2022), reproduced with permission for ease of reference. Illustrated $C\alpha_i - C_i - N_{i+1}$ bond angle (third plot from the left) corresponds to θ_2 in our formulation. Sampled angles in the work of Anand & Achim (2022) exhibit a much larger spread than the natural distribution of angles, whereas our work matches much more tightly (Figures 2, S4).

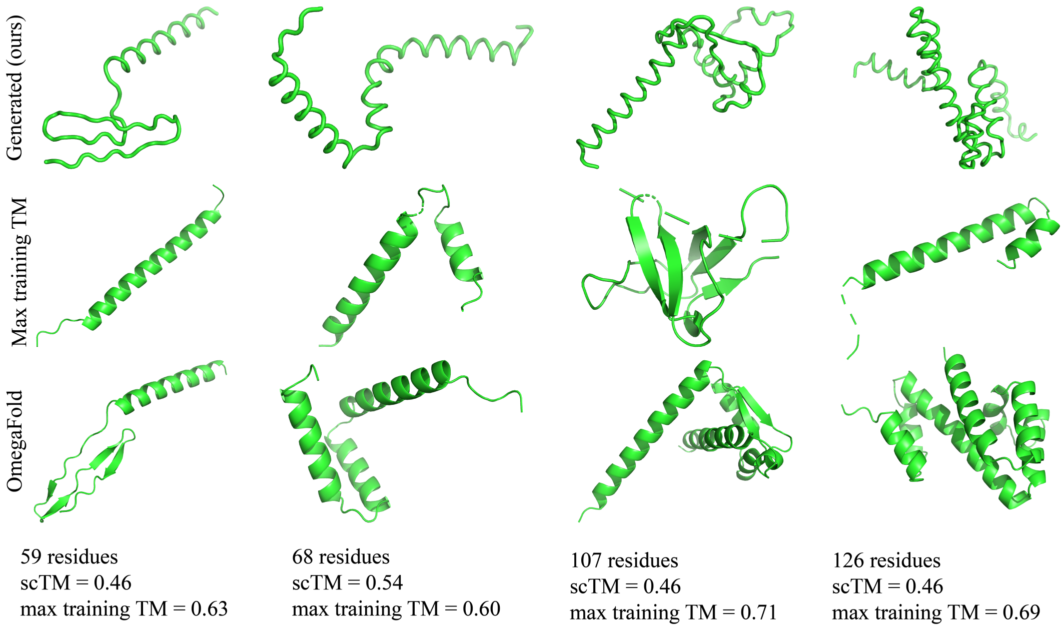


63 residues
scTM = 0.91
max training TM = 0.94

64 residues
scTM = 0.52
max training TM = 0.49

92 residues
scTM = 0.21
max training TM = 0.31

Figure S6: Generated structures representing the full range of designability (scTM) and training similarity scores. The top row indicates the original generated structure, the middle row shows the training structure with the highest TM score, and the bottom row indicates the structure predicted by OmegaFold based on residues predicted to produce our generated structure by ESM-IF1. The first column shows a structure with high designability and high training similarity. The second column shows a structure with designability and training similarity close to 0.5. The third column shows a generated example that is very different from any training chain, but is also not designable.



59 residues
scTM = 0.46
max training TM = 0.63

68 residues
scTM = 0.54
max training TM = 0.60

107 residues
scTM = 0.46
max training TM = 0.71

126 residues
scTM = 0.46
max training TM = 0.69

Figure S7: Structures from Figure 6 illustrated with the training example with the highest TM score (most similar). Figure rows are arranged as in Figure S6. Note that our generated structures are visually quite different compared to the best training set match – in each example, our generated structure contains a completely different arrangement of secondary structure elements than the closest training structure, indicating that they may be more distinct than TM scores alone might suggest.

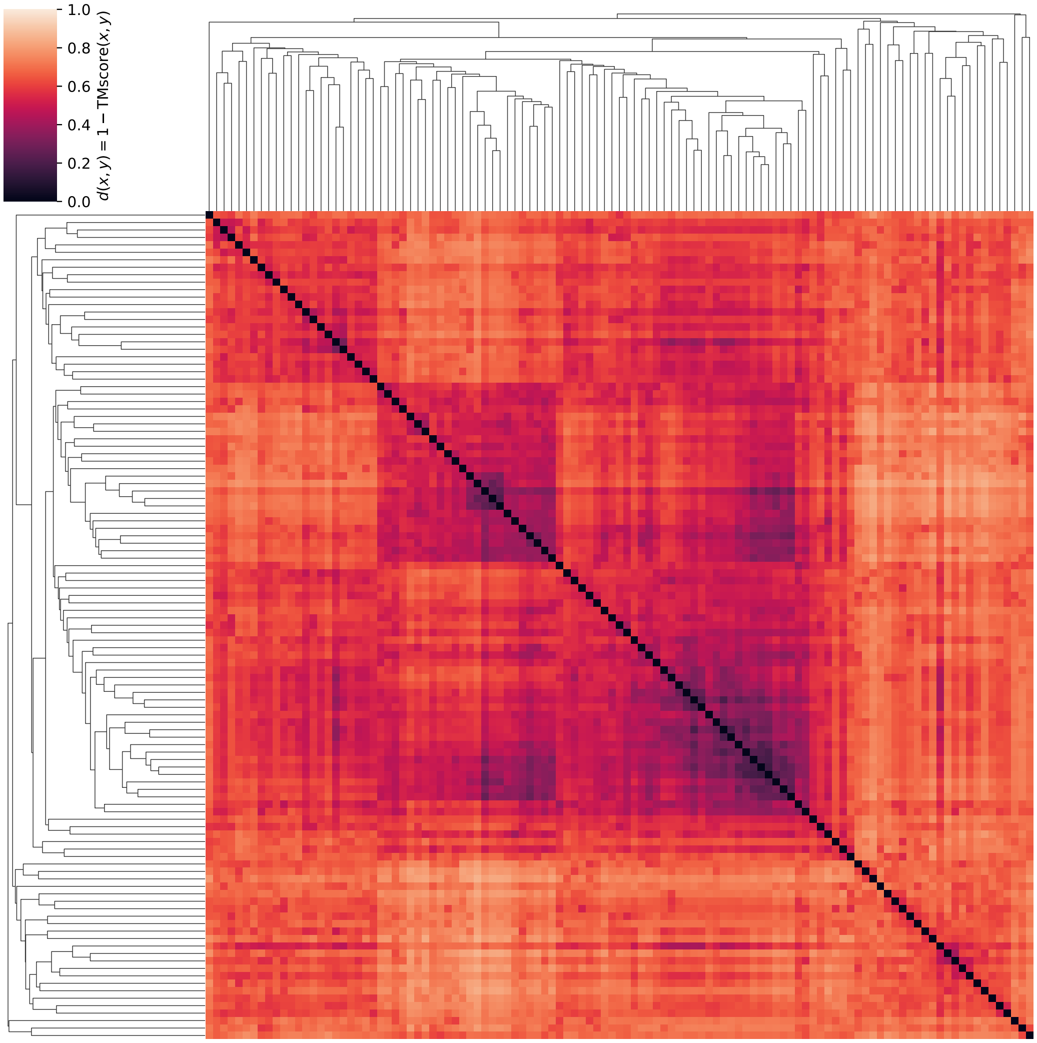


Figure S8: Clustering of our $n = 111$ "designable" generated backbones with scTM scores ≥ 0.5. We use the average distance metric to perform hierarchical clustering on the pairwise distance matrix $d(x, y) = 1 - \text{TMscore}(x, y)$. Dark values corresponding to 0 (or conversely, a TM score of 1) indicate (nearly) identical structures. While there are some loosely related groups of structures, we do not observe clearly delineated groups. This indicates that the designable backbones we generate are diverse and represent a wide range of potential structures. For comparison to prior works, see Figure S9.

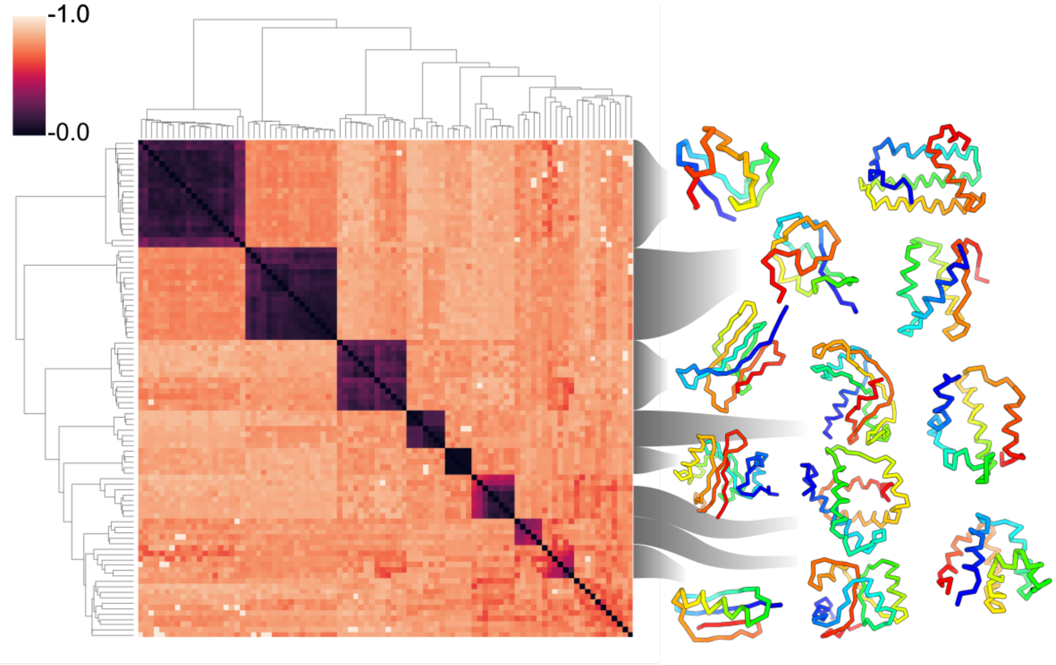


Figure S9: Figure from Trippe et al. (2022), reproduced with permission for ease of reference. These authors similarly cluster unconditionally generated backbones with scTM ≥ 0.5 using $1 -$ TMscore(x, y) as a distance metric. Compared to our identical evaluation, illustrated in Figure S8, we notice that this clustering has a few dark blocks of nearly 0 distance, or a TM score of nearly 1. This suggests that among these designable backbones, many are actually minor variants of a core structure; in actuality, though this work claims to produce 90 designable structures, there seem to be fewer unique structures in this set due to many being near-duplicates.