Efficient Generative Modelling of Protein Structure Fragments using a Deep Markov Model

Christian B. Thygesen [1 2] **Ahmad Salim Al-Sibahi** [1] **Christian S. Steenmanns** [2] **Lys S. Moreta** [1]
Anders B. Sørensen [* 2] **Thomas Hamelryck** [* 1 3]

Abstract

Fragment libraries are often used in protein structure prediction, simulation and design as a means to significantly reduce the vast conformational search space. Current state-of-the-art methods for fragment library generation do not properly account for aleatory and epistemic uncertainty, respectively due to the dynamic nature of proteins and experimental errors in protein structures. Additionally, they typically rely on information that is not generally or readily available, such as homologous sequences, related protein structures and other complementary information. To address these issues, we developed BIFROST, a novel take on the fragment library problem based on a Deep Markov Model architecture combined with directional statistics for angular degrees of freedom, implemented in the deep probabilistic programming language Pyro. BIFROST is a probabilistic, generative model of the protein backbone dihedral angles conditioned solely on the amino acid sequence. BIFROST generates fragment libraries with a quality on par with current state-of-the-art methods at a fraction of the run-time, while requiring considerably less information and allowing efficient evaluation of probabilities.

1. Introduction

Fragment libraries (Jones & Thirup, 1986) find wide application in protein structure prediction, simulation, design and experimental determination (Trevizani et al., 2017; Chikenji et al., 2006; Boomsma et al., 2012). Predicting the fold of

---

[*]Equal contribution [1]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark [2]Evaxion Biotech, Copenhagen, Denmark [3]Department of Biology, University of Copenhagen, Copenhagen, Denmark. Correspondence to: Christian B. Thygesen <christiank.thygesen@di.ku.dk>, Thomas Hamelryck <thamelry@bio.ku.dk>.

a protein requires evaluating a conformational space that is too vast for brute-force sampling to be feasible (Levinthal, 1969). Fragment libraries are used in a divide-and-conquer approach, whereby a full length protein is divided into a manageable sub-set of shorter stretches of amino acids for which backbone conformations are sampled. Typically, sampling is done using a finite set of fragments derived from experimentally determined protein structures. Fragment libraries are used in state-of-the-art protein structure prediction frameworks such as Rosetta (Rohl et al., 2004), I-TASSER (Roy et al., 2010), and AlphaFold (Senior et al., 2019).

Generally, knowledge-based methods for protein structure prediction follow two main strategies: homology (or template-based) modelling (Eswar et al., 2006; Šali & Blundell, 1993; Song et al., 2013) and *de novo* modelling (Rohl et al., 2004). Both approaches assume that the native fold of a protein corresponds to the minimum of a physical energy function and make use of statistics derived from a database of known proteins structures (Alford et al., 2017; Leaver-Fay et al., 2013). Whereas homology modelling relies on the availability of similar structures to limit the search space, knowledge-based *de novo* protocols require extensive sampling of the conformational space of backbone angles (figure 1).
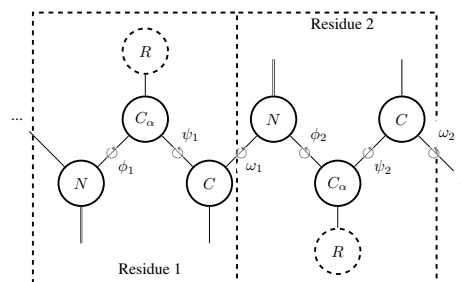

Figure 1. Schematic of the three dihedral angles (ϕ, ψ, and ω) that parameterise the protein backbone. R represents the side chain.

To overcome the shortcomings of either strategy, modelling tools like Rosetta (Rohl et al., 2004) use a combined approach of extensive sampling and prior information. Rosetta employs simulated annealing of backbone conformations

according to an energy function (Alford et al., 2017), while reducing the conformational space by sampling fragments of typically 3 or 9 amino acids at a time (Simons et al., 1997).

Fragments are typically extracted from experimentally determined protein structures in the Protein Data Bank (Berman et al., 2000) and used in prediction based on similarities in sequence and sequence-derived features (Gront et al., 2011; Kalev & Habeck, 2011; Santos et al., 2015; De Oliveira et al., 2015; Trevizani et al., 2017; Wang et al., 2019). Generative probabilistic models of protein backbone angles (Hamelryck et al., 2006; Boomsma et al., 2008; Bhattacharya et al., 2016; Edgoose et al., 1998; Li et al., 2008; Lennox et al., 2010) offer an alternative way to construct fragment libraries and aim to represent the associated epistemic and aleatory uncertainty. In this case, epistemic uncertainty is due to experimental errors from the determination of protein structures, while aleatory or inherent uncertainty is due to the dynamic nature, or flexibility, of proteins (Best, 2017).

Here, we present BIFROST - Bayesian Inference for FRagments Of protein STructures - a deep, generative, probabilistic model of protein backbone angles that solely uses the amino acid sequence as input. BIFROST is based on an adaptation of the Deep Markov Model (DMM) architecture (Krishnan et al., 2017) and represents the angular variables (ϕ and ψ) in a principled way using directional statistics (Mardia & Jupp, 2008). Finally, BIFROST makes it possible to evaluate the probability of a backbone conformation given an amino acid sequence, which is important for applications such as sampling the conformational space of proteins with correct statistical weights in equilibrium simulations (Boomsma et al., 2014).

2. Background and related work

Probabilistic, generative models of local protein structure Most generative, probabilistic models of local protein structure are Hidden Markov Models (HMMs) that represent structure and sequence based on the assumption of a Markovian structure (Hamelryck et al., 2012). The first such models did not include the amino acid sequence (Edgoose et al., 1998), discretised the angular variables (Bystroff et al., 2000), or used continuous, but lossy representations (Camproux et al., 1999; Hamelryck et al., 2006), making sampling of conformations with atomic detail problematic. These early models are thus probabilistic but only approximately "generative" at best. TorusDBN (Boomsma et al., 2008) was the first joint model of backbone angles and sequence that properly accounted for the continuous and angular nature of the data. Others introduced richer probabilistic models of local protein structure including Dirichlet Process mixtures of HMMs (DPM-HMMs) Lennox et al. (2010) and Conditional Random Fields (CRFs) (Zhao et al.,

2010; 2008). As far as we know, BIFROST is the first deep generative model of local protein structure that aims to quantify the associated aleatory and epistemic uncertainty using an (approximate) Bayesian posterior.

Deep Markov Models The DMM, introduced in (Krishnan et al., 2017), is a generalisation of the variational autoencoder (VAE) (Kingma & Welling, 2014) for sequence or time series data. Related stochastic sequential neural models were reported by Fraccaro et al. (2016) and Chung et al. (2015). Published applications of DMMs include natural language processing tasks (Khurana et al., 2020), inference of time series data (Zhi-Xuan et al., 2020), and human pose forecasting (Toyer et al., 2017). Our application of the DMM and the modifications made to the standard model will be described in section 3.3.

3. Methods

3.1. Data set

BIFROST was trained on a data set of fragments derived from a set of 3733 proteins from the *cullpdb* data set (Wang & Dunbrack, 2005). Quality thresholds were (i) resolution < 1.6Å, (ii) R-factor < 0.25, and (iii) a sequence identity cutoff of 20%. For the purpose of reliable evaluation, sequences with $> 20\%$ identity to CASP13 targets were removed from the dataset.

Fragments containing angle-pairs in disallowed regions of the Ramachandran plot (Ramachandran et al., 1963) were removed using the Ramalyze function of the crystallography software PHENIX (Liebschner et al., 2019). The resulting data set consisted of ~ 186000 9-mer fragments. Prior to training, the data was randomly split into train, test, and validation sets with a $60/20/20\%$ ratio.

3.2. Framework

The presented model was implemented in the deep probabilistic programming language Pyro, version 1.3.0 (Bingham et al., 2019) and Pytorch version 1.4.0 (Paszke et al., 2019). Training and testing were carried out on a machine equipped with an Intel Xeon CPU E5-2630 and Tesla M10 GPU. The model trains on a single GPU and converges after 150 epochs for a total training time of approximately 34 hours.

3.3. Model

BIFROST consists of a DMM (Krishnan et al., 2017) with an architecture similar to an Input-Output HMM (IO-HMM) (Bengio & Frasconi, 1995). The model employs the Markovian structure of an HMM, but with continuous, as opposed to discrete, latent states (**z**) and with *transition and emission neural networks* instead of transition and emission matrices.

Consequently, the latent states are iteratively transformed using the transition neural network, such that the value of the current latent state depends on the previous state and the (processed) amino acid information at that position (figure 2).

Observed angles (ϕ and ψ) are generated from the latent state sequence by applying an *emitter neural network* at each position (figure 2). Since the backbone angle ω is most often narrowly distributed around $180°$, this degree of freedom is not included in the current version of BIFROST.
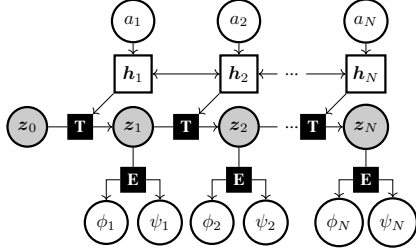

Figure 2. The BIFROST model. Grey nodes are latent random variables, white circular nodes are observed variables, white rectangular nodes represent hidden states from a bidirectional Recurrent Neural Network (RNN) H, and black squares represent neural networks. E and T denote the emitter and the transition network, respectively.

The structure of the model is shown in figure 2. For notational simplicity, the sequence of ϕ and ψ pairs will be denoted by $\mathbf{x}$. The joint distribution of the latent variable $\mathbf{z}$ and the angles $\mathbf{x}$ conditioned on the amino acid sequence $\mathbf{a}$ with length N of the graphical model in figure 2 factorises as

$$p(\mathbf{z}, \mathbf{x}|\mathbf{a}) = \prod_{n=1}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{h}_n(\mathbf{a}))p(\mathbf{x}_n|\mathbf{z}_n) \quad (1)$$

where $\mathbf{h}_n(\mathbf{a})$ is the deterministic hidden state generated at position n by a bidirectional RNN H with parameters $\boldsymbol{\theta}_H$ running across the amino acid sequence. The bidirectional RNN incorporates information from amino acids upstream and downstream of position n. The initial latent state $\mathbf{z}_0$ is treated as a trainable parameter and is thus shared for all sequences.

The transition densities are given by a multivariate Gaussian distribution,

$$p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{h}_n(\mathbf{a})) = \\ \mathcal{N}(\boldsymbol{\mu}_T(\mathbf{z}_{n-1}, \mathbf{h}_n(\mathbf{a})), \boldsymbol{\Sigma}_T(\mathbf{z}_{n-1}, \mathbf{h}_n(\mathbf{a}))) \quad (2)$$

where the mean vector ($\boldsymbol{\mu}_T$) and the (diagonal) covariance matrix ($\boldsymbol{\Sigma}_T$) are given by a neural network T parameterised by $\boldsymbol{\theta}_T$.

The emission densities are given by a *bivariate periodic student-T distribution* (Pewsey et al., 2007) (section 3.5) such that

$$p(\mathbf{x}_n|\mathbf{z}_n) = \\ \mathcal{T}(\mathbf{x}_n|\nu_E(\mathbf{z}_n), \boldsymbol{\mu}_E(\mathbf{z}_n), \boldsymbol{\Sigma}_E(\mathbf{z}_n)) \quad (3)$$

where the single, shared degree of freedom (ν_E), the vector of two means ($\boldsymbol{\mu}_E$), and the 2×2 diagonal covariance matrix ($\boldsymbol{\Sigma}_E$) of the distribution are given by a neural network E parameterised by $\boldsymbol{\theta}_E$.
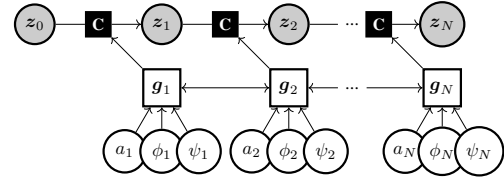

Figure 3. Variational distribution for approximating the posterior. Grey nodes are latent random variables, white circular nodes are observed variables, white rectangular nodes represent hidden states from a bidirectional RNN G, while black squares represent neural networks. C denotes the combiner network.

3.4. Estimation

In order to perform inference of the intractable posterior, we introduce a variational distribution or *guide q* (Kingma & Welling, 2019) (figure 3), which makes use of a *combiner neural network C* parameterised by $\boldsymbol{\zeta}_C$,

$$q(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{a}, \mathbf{x}) = \\ \mathcal{N}(\boldsymbol{\mu}_C(\mathbf{z}_{n-1}, \mathbf{g}_n(\mathbf{a}, \mathbf{x})), \boldsymbol{\Sigma}_C(\mathbf{z}_{n-1}, \mathbf{g}_n(\mathbf{a}, \mathbf{x}))) \quad (4)$$

where $\boldsymbol{g}_n(\boldsymbol{a}, \boldsymbol{x})$ is the deterministic hidden state generated at position n by a bidirectional RNN G with parameters $\boldsymbol{\zeta}_G$ running across the amino acid sequence $\boldsymbol{a}$ and the angles $\boldsymbol{x}$.

For the parameters of the neural networks ($\boldsymbol{\zeta}_C, \boldsymbol{\zeta}_G, \boldsymbol{\theta}_T, \boldsymbol{\theta}_E, \boldsymbol{\theta}_H$), point estimates are obtained using Stochastic Variational Inference (SVI), which optimises the Evidence Lower Bound (ELBO) using stochastic gradient descent (SGD) (Kingma & Welling, 2014; 2019). The ELBO variational objective is given by

$$\mathcal{L}_{\boldsymbol{\theta}, \boldsymbol{\zeta}}(\boldsymbol{x}) = \\ \mathbb{E}_{q_{\boldsymbol{\zeta}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{a})}\left[\log p_{\boldsymbol{\theta}}(\boldsymbol{z}, \boldsymbol{x}|\boldsymbol{a}) - \log q_{\boldsymbol{\zeta}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{a})\right] \quad (5)$$

where $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_C, \boldsymbol{\zeta}_G)$ and $\boldsymbol{\theta} = (\boldsymbol{z}_0, \boldsymbol{\theta}_T, \boldsymbol{\theta}_E, \boldsymbol{\theta}_H)$ are the parameters of the guide and the model, respectively.

3.5. Periodic student T distribution

As angle-pairs are periodic values, i.e. distributed on a torus (Boomsma et al., 2008), they need to be modelled by an appropriate periodic distribution. Traditionally, angles are assumed distributed according to the von Mises distribution, which is defined by a mean that can be any real number and a concentration parameter, which can be any positive number. SVI showed poor performance when the von Mises distribution was used. Here, we circumvent this by representing the likelihood of the angles by a student T distribution that is wrapped around a circle (Pewsey et al., 2007). This allows for appropriate modelling of the periodicity of the angles, while being more robust with regards to outlier issues than the von Mises distribution due to the wider tails of the T distribution. It should be noted as well that Pewsey et al. (2007) showed that the wrapped student T distribution can approximate the von Mises distribution closely.

3.6. Neural network architecture overview

The overall architecture is based on the originally proposed DMM (Krishnan et al., 2017) with modifications. The main difference is the addition of an RNN H in the model that processes the amino acid sequence $\mathbf{a}$, thus providing explicit conditioning on the amino acid sequence. A similar architecture was used by Fraccaro et al. (2016) for time series. In the guide, a second RNN G is used that processes the angles and the amino acid sequence during training. The initial values for both RNNs are treated as trainable parameters. In addition to the RNNs, the model contains an emitter network E and a transition network T, while the guide relies on a combiner network C.

Emitter architecture The emitter network E parameterises the emission probabilities as stated in equation 3. E is a feedforward neural network containing two initial layers that branch into three. One output branch is a single layer that outputs the degree of freedom of the Student T distribution, which is shared between the two angles. The other two branches output a mean μ and a standard deviation σ for ϕ and ψ, respectively. Each hidden layer of the neural network contained 200 neurons with rectified linear unit (ReLU) activation. Output layers for μ values had no activation, as the periodic distribution automatically transforms values to a range between $-\pi$ and π. Output layers for σ and degrees of freedom ν used softplus activation to ensure positive, real numbered values. The architecture of E is depicted in figure 4.

Transition and combiner architecture The transition network T and the combiner network C specify the transition densities from the previous to the current latent state of the model (equation 2) and the guide (equation 4), respectively. In the original DMM (Krishnan et al., 2017), C was inspired by the Gated Recurrent Unit (GRU) architecture (Cho et al.,
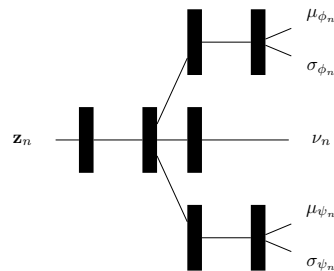

Figure 4. Architecture of the emitter neural network, E. Black rectangles represent ReLU-activated fully connected layers.

2014), while T was a simple feed forward network. Here, both C and T were based on GRU cells to allow for better horizontal information flow (figure 5).
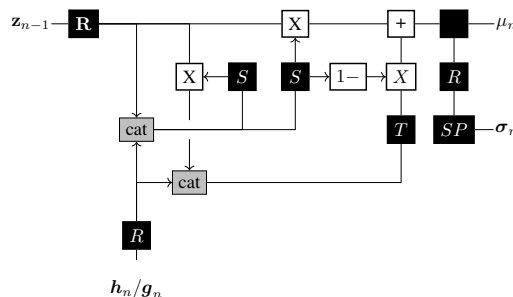

Figure 5. Architecture of the transition T and combiner C neural networks. Black squares represent single neural network layers activated by a ReLU (R), sigmoid (S), tanh (T), softplus (SP) or no activation. White squares represent element-wise mathematical operations. Gray squares represent tensor concatenation. Note that the network takes as input either $\boldsymbol{h}_n$ or $\boldsymbol{g}_n$ obtained from the RNN in the model or the guide, respectively.

The total number of parameters in BIFROST are shown in table 1.

3.7. Hyperparameter optimization

A simple hyperparameter search was performed with the test ELBO as the selection criterion (data not shown). The final model was trained with a learning rate of 0.0003 with a scheduler reducing the learning rate by 90% when no improvement was seen for 10 epochs. Minibatch size was 200. The Adam optimiser was used with a β_1 and β_2 of 0.96 and 0.999 respectively. The latent space dimensionality was 40. All hidden activations (if not specified above) were ReLU activations. We employed norm scaling of the gradient to a norm of 10.0. Finally, early stopping was employed with a patience of 50 epochs.

Neural networks					Z_0	
E	T	C	H	G	p	q
24 805	142 280	142 280	89 200	90 000	40	40
Total parameters:	**488 645**					

Table 1. Number of parameters in BIFROST. E: Emitter, T: Transition, C: Combiner, H: model RNN, G: guide RNN, p: model, q: guide

3.8. Sampling from the model

The BIFROST model (figure 2) is designed with explicit conditioning on amino acid sequences allowing a simple and efficient ancestral sampling approach that eliminates the need for using the guide for predictions. Thus, the guide is used solely for the purpose of model estimation and is discarded upon sampling.

3.9. Fragment library generation and benchmarking

Fragment libraries are a collection of fragments, consisting of typically 3 or 9 amino acids with known backbone angles. Here, we focus on fragments of nine amino acids. For each fragment in a protein, 200 possible backbone conformations are sampled from BIFROST resulting in a set of $(L - 8) \times 200$ fragment candidates, where L is the number of amino acids in the protein. These candidates are compared to the observed fragment by calculating the angular root mean square deviation (RMSD) between the corresponding angles as proposed in Boomsma et al. (2008). The choice of 9-mer fragments and the 200 samples per fragment were made to emulate the default behavior of the Rosetta fragment picker (see below), for fair comparison.

The aggregated quality of fragment libraries are generally represented by two metrics; *precision* and *coverage*. Precision is defined as the fraction of candidates with an RMSD to the observed below a certain threshold, whereas coverage is the fraction of positions covered by at least one candidate with an RMSD below a certain threshold. Evaluating the precision and coverage at increasing thresholds yields two curves, and the quality of the fragment library is quantified by the area under these two curves.

BIFROST was benchmarked against Rosetta's fragment picker (Gront et al., 2011) using the precision and coverage metrics. The fragment picker was run using default parameters, picking 200 fragments per position. Secondary structure predictions were performed using SAM-T08 (Karplus, 2009), PSIPRED (Jones, 1999) and Jufo (Leman et al., 2013). Sequences that were homologous to the targets were excluded (–*nohoms* flag).

Fragment libraries were generated for all available regular (denoted "T") targets from the latest installment of the

bi-annual protein structure prediction competition Critical Assessment of Techniques for Protein Structure Prediction (CASP13).

3.10. Runtime comparison

In order to compare the runtime of BIFROST to that of the fragment picker, nine proteins of varying lengths were selected. Both tools generated 200 samples per fragment. The experiment was run on the same 32-core machine for both the fragment picker and BIFROST.

4. Results

To show that the model is able to capture general protein backbone behavior, angles were generated conditioned on the sequences of 5000 previously unseen fragments and compared to the observed angles. The model was able to recreate the observed Ramachandran plots with minimal added noise (figure 6).
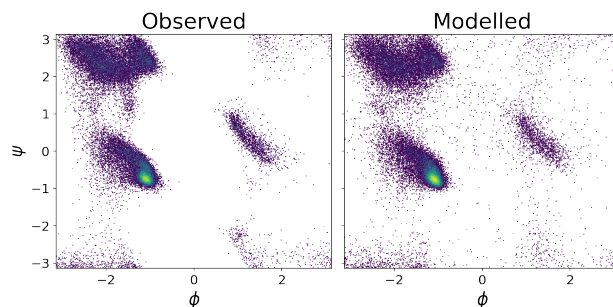

Figure 6. Observed and modelled aggregated Ramachandran plots

While most amino acids show angle distributions similar to the background in figure 6, glycine and proline are exceptions due to the nature of their side chains. The side chain of glycine is a single hydrogen atom, allowing the backbone to be exceptionally flexible, while the side chain of proline is covalently linked to the backbone restraining the conformational space. The modelled distribution of angles for these two unique cases, along with leucine to represent the general case, show that the model is able to capture specific amino acid properties (figure 7).

The left side of figure 8 shows a thin, smoothed coil representation of 100 samples from BIFROST conditioned on example 9-mer fragments that were observed to be either α-helix, β-strand, or coiled. The right side shows distributions of backbone RMSDs of 5000 sampled fragments to the observed structure from BIFROST and as picked by Rosetta's fragment picker.

The RMSDs were generally distributed towards 0Å for the α-helix case, showcasing BIFROSTs ability to predict this
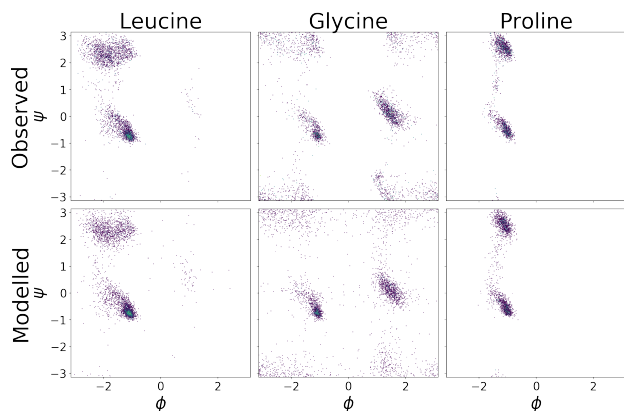
Figure 7. Amino acid specific Ramachandran plots

well defined secondary structure element. The model has more difficulty modelling β-strands and coils. However, the distributions of the RMSDs are nearly identical to those produced by the fragment picker. For coil fragments, the RMSDs were distributed around 3Å reflecting the inherent variability of those fragments.
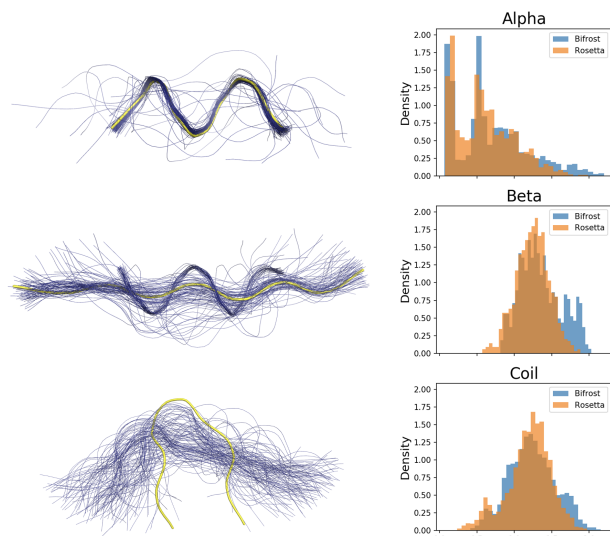

Figure 8. Left: 100 samples of backbone dihedral angles (blue) superimposed on the observed structures (yellow). For clarity, the backbones are represented as thin, smoothed coils instead of traditional cartoon representations. Right: Aggregated RMSDs of BIFROST-sampled conformations and conformations picked by Rosetta's fragment picker for sequences observed as α-helix, β-strand, and coil respectively.

BIFROST was benchmarked against Rosetta's fragment picker (Gront et al., 2011) on all publicly available CASP13 regular targets. BIFROST generated fragment libraries with comparable precision and coverage to the fragment picker (figure 9).
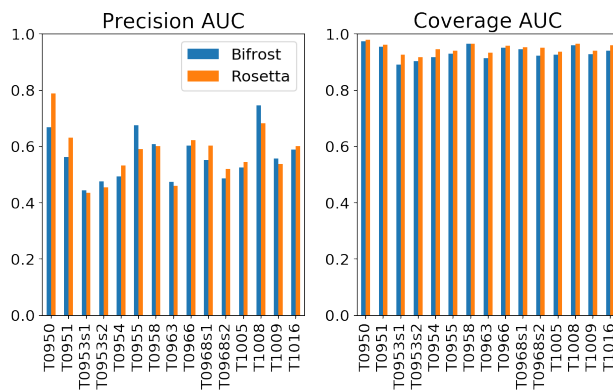

Figure 9. Comparison of fragment libraries generated by BIFROST, relying on just the amino acid sequences, against Rosetta's fragment picker, which uses external information and relies on ensemble predictions of secondary structure.

Finally, BIFROST enables efficient sampling of fragment libraries. The runtime of BIFROST and the fragment picker are compared in figure 10 on a set of nine proteins of varying lengths. Both runtimes roughly scale linearly with protein length, but BIFROST has a smaller constant term than the fragment picker.
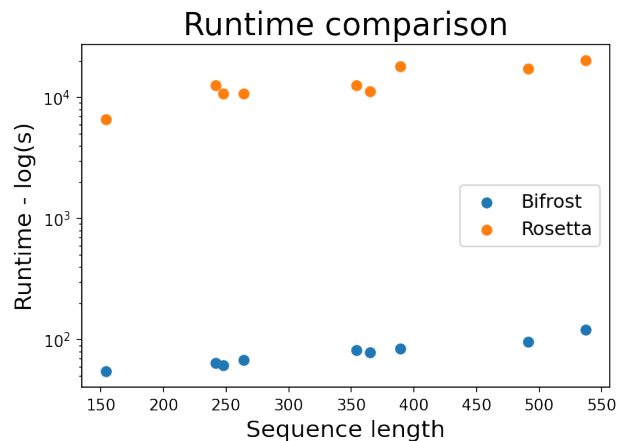

Figure 10. Runtime comparison between Rosetta's fragment picker and BIFROST on a set of nine proteins of varying lengths.

5. Discussion

BIFROST is a deep, generative model of local protein structure conditioned on sequence that provides a probabilistic approach to generating fragment libraries.

The quality of the generated fragment libraries is on par with Rosetta's fragment picker, despite using much less information, such as an ensemble of secondary structure predictors. Due to the probabilistic nature of BIFROST, distributions tend to be slightly wider than those resulting

from picking structural fragments from the PDB based on sequence similarity. This wider distribution plausibly reflects the dynamic nature of protein structure, which is not captured in the experimental data provided by static X-ray structures.

The model was estimated using SVI, relying on the ELBO variational objective. As the ELBO provides a lower bound on the log evidence (Kingma & Welling, 2014), we can evaluate the probability of a specific local structure given the sequence, simply by evaluating the ELBO. Evaluating the probability of fragments is crucial for correct sampling of the conformational space, for example in the case of equilibrium simulations of protein dynamics (Boomsma et al., 2014). The probabilities assigned by BIFROST can be used to decide how often a fragment should be sampled in the folding process. In contrast, existing methods do not provide an explicit measure of fragment confidence.

In this paper the focus was kept on fragments of nine residues for ease of comparison to the fragment picker. However, the DMM architecture of BIFROST allows generation of fragments of arbitrary length but with an observed drop-off in performance as the length of fragments are increased (data not shown).

Existing methods rely heavily on the availability of multiple sequence alignments (MSA) and other information, such as secondary structure predictors. As MSAs are not available for orphan proteins or synthetic proteins, the need for pure sequence based models is evident.

6. Acknowledgements

References

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., and Gray, J. J. The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, jun 2017. ISSN 15499626. doi: 10.1021/acs.jctc.7b00125.

Bengio, Y. and Frasconi, P. An input output HMM architecture. *Neural Information Processing Systems*, pp. 427–434, 1995. ISSN 15322092. doi: 10.1093/europace/euq350.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The Protein Data Bank, jan 2000. ISSN 03051048.

Best, R. B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Current Opinion in Structural Biology*, 42:147–154, feb 2017. ISSN 0959440X. doi: 10.1016/j.sbi.2017.01.006.

Bhattacharya, D., Adhikari, B., Li, J., and Cheng, J. FRAGSION: Ultra-fast protein fragment library generation by IOHMM sampling. *Bioinformatics*, 32(13):2059–2061, 2016. ISSN 14602059. doi: 10.1093/bioinformatics/btw067.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P. A., Horsfall, P., and Goodman, N. D. Pyro: Deep Universal Probabilistic Programming. *J. Mach. Learn. Res.*, 20: 28:1—-28:6, 2019.

Boomsma, W., Mardia, K. V., Taylor, C. C., Ferkinghoff-Borg, J., Krogh, A., and Hamelryck, T. A generative, probabilistic model of local protein structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(26):8932–8937, 2008. ISSN 00278424. doi: 10.1073/pnas.0801715105.

Boomsma, W., Frellsen, J., and Hamelryck, T. *Probabilistic models of local biomolecular structure and their applications*. Springer, 2012. doi: 10.1007/978-3-642-27225-7_10.

Boomsma, W., Tian, P., Frellsen, J., Ferkinghoff-Borg, J., Hamelryck, T., Lindorff-Larsen, K., and Vendruscolo, M. Equilibrium simulations of proteins using molecular fragment replacement and NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America*, 111(38):13852–13857, 2014. ISSN 10916490. doi: 10.1073/pnas.1404948111.

Bystroff, C., Thorsson, V., and Baker, D. HMMSTR: A hidden Markov model for local sequence-structure correlations in proteins. *Journal of Molecular Biology*, 301(1):173–190, 2000. ISSN 00222836. doi: 10.1006/jmbi.2000.3837.

Camproux, A. C., Tuffery, P., Chevrolat, J. P., Boisvieux, J. F., and Hazout, S. Hidden Markov model approach for identifying the modular framework of the protein backbone. *Protein Engineering*, 12(12):1063–1073, 1999. ISSN 02692139. doi: 10.1093/protein/12.12.1063.

Chikenji, G., Fujitsuka, Y., and Takada, S. Shaping up the protein folding funnel by local interaction: Lesson from a structure prediction study. *Proceedings of the National Academy of Sciences*, 103(9):3141–3146, feb 2006. ISSN 0027-8424. doi: 10.1073/pnas.0508195103.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1724–1734. Association for Computational Linguistics (ACL), jun 2014. ISBN 9781937284961. doi: 10.3115/v1/d14-1179.

Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A., and Bengio, Y. A recurrent latent variable model for sequential data. In *Advances in Neural Information Processing Systems*, volume 2015-Janua, pp. 2980–2988, 2015.

De Oliveira, S. H., Shi, J., and Deane, C. M. Building a better fragment library for de novo protein structure prediction. *PLoS ONE*, 10(4), 2015. ISSN 19326203. doi: 10.1371/journal.pone.0123998.

Edgoose, T., Allison, L., and Dowe, D. L. An MML classification of protein structure that knows about angles and sequence. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pp. 585–596, 1998. ISSN 2335-6928.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M., Eramian, D., Shen, M.-y., Pieper, U., and Sali, A. Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, 15(1):5.6.1–5.6.30, sep 2006. ISSN 1934-340X. doi: 10.1002/0471250953. bi0506s15.

Fraccaro, M., Sønderby, S. K., Paquet, U., and Winther, O. Sequential neural models with stochastic layers. In *Advances in Neural Information Processing Systems*, pp. 2207–2215, 2016.

Gront, D., Kulp, D. W., Vernon, R. M., Strauss, C. E., and Baker, D. Generalized fragment picking in rosetta: Design, protocols and applications. *PLoS ONE*, 6(8): 23294, 2011. ISSN 19326203. doi: 10.1371/journal. pone.0023294.

Hamelryck, T., Kent, J. T., and Krogh, A. Sampling realistic protein conformations using local structural bias. *PLoS Computational Biology*, 2(9):e131, sep 2006. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0020131.

Hamelryck, T., Mardia, K., and Ferkinghoff-Borg, J. (eds.). *Bayesian Methods in Structural Bioinformatics*. Statistics for Biology and Health. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-642-27224-0. doi: 10.1007/978-3-642-27225-7.

Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 292(2):195–202, 1999. ISSN 00222836. doi: 10.1006/jmbi.1999.3091.

Jones, T. A. and Thirup, S. Using known substructures in protein model building and crystallography. *The EMBO journal*, 5(4):819–22, apr 1986. ISSN 0261-4189.

Kalev, I. and Habeck, M. HHfrag: HMM-based fragment detection using HHpred. *Bioinformatics*, 27(22): 3110–3116, 2011. ISSN 13674803. doi: 10.1093/ bioinformatics/btr541.

Karplus, K. SAM-T08, HMM-based protein structure prediction. *Nucleic Acids Research*, 37(SUPPL. 2):492–497, 2009. ISSN 03051048. doi: 10.1093/nar/gkp403.

Khurana, S., Laurent, A., Hsu, W.-N., Chorowski, J., Lancucki, A., Marxer, R., and Glass, J. A Convolutional Deep Markov Model for Unsupervised Speech Representation Learning. In *Proceedings of Interspeech*, jun 2020.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014.

Kingma, D. P. and Welling, M. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12(4):307–392, 2019. ISSN 1935-8237. doi: 10.1561/2200000056.

Krishnan, R. G., Shalit, U., and Sontag, D. Structured inference networks for nonlinear state space models. In *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, pp. 2101–2109, sep 2017.

Leaver-Fay, A., O'Meara, M. J., Tyka, M., Jacak, R., Song, Y., Kellogg, E. H., Thompson, J., Davis, I. W., Pache, R. A., Lyskov, S., Gray, J. J., Kortemme, T., Richardson, J. S., Havranek, J. J., Snoeyink, J., Baker, D., and Kuhlman, B. Scientific benchmarks for guiding macromolecular energy function improvement. In *Methods in Enzymology*, volume 523, pp. 109–143. Academic Press Inc., 2013. ISBN 9780123942920. doi: 10.1016/B978-0-12-394292-0.00006-0.

Leman, J. K., Mueller, R., Karakas, M., Woetzel, N., and Meiler, J. Simultaneous prediction of protein secondary structure and transmembrane spans. *Proteins: Structure, Function and Bioinformatics*, 81(7):1127–1140, jul 2013. ISSN 08873585. doi: 10.1002/prot.24258.

Lennox, K. P., Dahl, D. B., Vannucci, M., Day, R., and Tsai, J. W. A Dirichlet process mixture of hidden Markov models for protein structure prediction. *Annals of Applied Statistics*, 4(2):916–942, 2010. ISSN 19326157. doi: 10.1214/09-AOAS296.

Levinthal, C. How to fold graciously. *Mössbauer Spectroscopy in Biological Systems Proceedings*, 24(41):22–24, 1969. ISSN 1041-1135. doi: citeulike-article-id: 380320.

Li, S. C., Bu, D., Xu, J., and Li, M. Fragment-HMM: A new approach to protein structure prediction. *Protein Science*, 17(11):1925–1934, 2008. ISSN 09618368. doi: 10.1110/ps.036442.108.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-w., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D Structural Biology*, 75(10):861–877, oct 2019. ISSN 2059-7983. doi: 10.1107/S2059798319011471.

Mardia, K. V. and Jupp, P. E. *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., Hoboken, NJ, USA, jan 2008. ISBN 9780470316979. doi: 10.1002/9780470316979.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. PyTorch: An imperative style, high-performance deep learning library, 2019. ISSN 23318422.

Pewsey, A., Lewis, T., and Jones, M. C. The wrapped t family of circular distributions. *Australian and New Zealand Journal of Statistics*, 49(1):79–91, 2007. ISSN 1467842X. doi: 10.1111/j.1467-842X.2006.00465.x.

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. Stereochemistry of polypeptide chain configurations, 1963. ISSN 00222836.

Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. Protein structure prediction using rosetta. *Methods in Enzymology*, 383:66–93, 2004. ISSN 00766879. doi: 10.1016/S0076-6879(04)83004-0.

Roy, A., Kucukural, A., and Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4):725–738, 2010. ISSN 17502799. doi: 10.1038/nprot.2010.5.

Šali, A. and Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, dec 1993. ISSN 00222836. doi: 10.1006/jmbi.1993.1626.

Santos, K. B., Trevizani, R., Custodio, F. L., and Dardenne, L. E. Profrager Web Server: Fragment libraries generation for protein structure prediction. In *Proceedings of the International Conference on Bioinformatics and Computational Biology (BIOCOMP)*, pp. 38, 2015.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., and Hassabis, D. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins: Structure, Function and Bioinformatics*, 87(12):1141–1148, 2019. ISSN 10970134. doi: 10.1002/prot.25834.

Simons, K. T., Kooperberg, C., Huang, E., and Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology*, 268(1):209–225, 1997. ISSN 00222836. doi: 10.1006/jmbi.1997.0959.

Song, Y., Dimaio, F., Wang, R. Y. R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. High-resolution comparative modeling with RosettaCM. *Structure*, 21 (10):1735–1742, 2013. ISSN 09692126. doi: 10.1016/j. str.2013.08.005.

Toyer, S., Cherian, A., Han, T., and Gould, S. Human pose forecasting via deep Markov models. In *DICTA 2017 - 2017 International Conference on Digital Image Computing: Techniques and Applications*, volume 2017-Decem, pp. 1–8. Institute of Electrical and Electronics Engineers Inc., jul 2017. ISBN 9781538628393. doi: 10.1109/DICTA.2017.8227441.

Trevizani, R., Dio, F. L. C., Santos, K. B. D., and Dardenne, L. E. Critical features of fragment libraries for protein structure prediction. *PLoS ONE*, 12(1), 2017. ISSN 19326203. doi: 10.1371/journal.pone.0170131.

Wang, G. and Dunbrack, R. L. PISCES: Recent improvements to a PDB sequence culling server. *Nucleic Acids Research*, 33(SUPPL. 2), 2005. ISSN 03051048. doi: 10.1093/nar/gki402.

Wang, T., Qiao, Y., Ding, W., Mao, W., Zhou, Y., and Gong, H. Improved fragment sampling for ab initio protein structure prediction using deep neural networks. *Nature Machine Intelligence*, 1(8):347–355, aug 2019. ISSN 2522-5839. doi: 10.1038/s42256-019-0075-7.

Zhao, F., Li, S., Sterner, B. W., and Xu, J. Discriminative learning for protein conformation sampling. *Proteins: Structure, Function and Genetics*, 73(1):228–240, oct 2008. ISSN 08873585. doi: 10.1002/prot.22057.

Zhao, F., Peng, J., Debartolo, J., Freed, K. F., Sosnick, T. R., and Xu, J. A probabilistic and continuous model of protein conformational space for template-free modeling. *Journal of Computational Biology*, 17(6):783–798, 2010. ISSN 10665277. doi: 10.1089/cmb.2009.0235.

Zhi-Xuan, T., Soh, H., and Ong, D. Factorized inference in deep Markov models for incomplete multimodal time series. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(06):10334–10341, apr 2020. ISSN 2374-3468. doi: 10.1609/aaai.v34i06.6597.