

SE(3) Equivalent Graph Attention Network as an Energy-Based Model for Protein Side Chain Conformation

Deqin Liu^{1,†}, Sheng Chen^{1,†}, Shuangjia Zheng², Sen Zhang¹, Yuedong Yang^{1,*}

School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
Global Institute of Future Technology (GIFT), Shanghai Jiao Tong University, Shanghai, China

Abstract—Protein design energy functions have been developed over decades by leveraging physical forces approximation and knowledge-derived features. However, manual feature engineering and parameter tuning might suffer from knowledge bias. Learning potential energy functions fully from crystal structure data is promising to automatically discover unknown or high-order features contributing to the protein’s energy. Here we propose a novel data-driven energy-based model based on SE(3)-equivariant model for protein conformation, namely GraphEBM. By combining with the graph attention network, GraphEBM improve the message passing on the chemical bond and capture the interatomic interaction and overlap. GraphEBM was benchmarked on the local rotamer recovery task and found to outperform both Rosetta and the state-of-the-art deep learning based methods. Furthermore, GraphEBM also yielded promising results on combinatorial side chain optimization, improving 13.8% χ_1 rotamer recovery to the Atom Transformer method on average.

Index Terms—protein conformation, protein side chain, energy-based model, graph attention network, deep learning

I. INTRODUCTION

Proteins, composed of amino acids, fold into stable conformations with the lowest free energy as per Anfinsen’s thermodynamic hypothesis [1], [2]. This led to the use of potential energy functions in protein structure prediction [3] and design [4]. However, optimizing these functions can be challenging due to the rough energy landscape [5]. To address this, researchers developed statistical potential methods [6]–[8] that combine data-driven energetic terms with physical force fields. Over time, these energy functions have incorporated extensive feature engineering [9], [10]. Deep learning has been shown to have the ability to capture the hidden high-order dependencies between source and target [11]. A number of deep-learning based methods including our previous works have successfully leveraged the deep learning methods in the field of protein design [12], [13] and protein structure prediction [14]. Therefore, it is promising to learn protein energy function fully from crystal structure data by deep learning methods.

Du et.al. used Transformer as an energy-based model [15] for protein side chain conformation. However, their architecture is not equivariant to the SE(3) group transformation,

meaning it may not output the same energy value after a protein conformation’s rigid rotation or translation. SE(3) group equivariance has been a principle of modern machine learning on molecule-related tasks [16], [17]. Several SE(3) equivariant architectures have been developed for protein design [18] but they focused on residue-wise backbone structures instead of atomic conformation. The directional message passing neural network (DimeNet) [19], [20] is an atomic resolution SE(3) equivariant architecture for small molecular graphs. However, DimeNet has not been refined for protein-related tasks and it suffered from training gradient exploding for the sampled conformation without physical constraints.

Here we propose GraphEBM, to our best knowledge, the first SE(3) equivariant energy-based model for protein side chain conformation. We tested GraphEBM on the side chain rotamer recovery task through two different sampling strategies. On average, for both sampling strategies, GraphEBM outperformed two well-known energy functions of Rosetta [10] and the state-of-the-art deep learning based method [15]. As a further study, we then applied GraphEBM on combinatorial side chain optimization for a fixed backbone [21]. Starting from the protein conformation yielded by PULCHRA [22], we simply adopted a naive strategy to solve the combinatorial optimization problem. To this end, 22.2% optimized side chain conformations were obtained by GraphEBM on average, showing the potential application of GraphEBM on the general problem settings of protein rational design.

II. METHODS

A. Model architecture

Our model GraphEBM described by Fig 1, improves the DimeNet++ and extends its capabilities to the energy prediction of proteins. We update the embedding layer for residue types which can’t be embedded before and introduce GAT and MLP to learn the atom bonds ignored in DimeNet++. The RBF and SBF focus on the distance and angles between atoms. The RBF is defined by

$$\tilde{e}_{\text{RBF},n}^{(ji)}(d) = \sqrt{\frac{2}{c}} \frac{\sin \frac{n\pi}{c} d}{d} \quad (1)$$

[†] Co-first authors

^{*} Correspondence author: Yuedong Yang(yangyd25@mail.sysu.edu.cn)

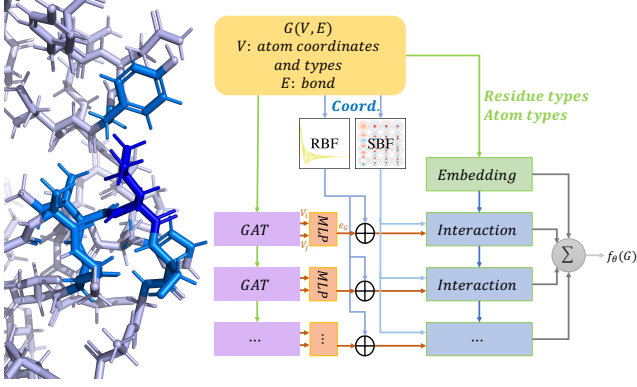


Fig. 1. Details on GraphEBM's architecture: The structural context ($\leq 5\text{\AA}$) of a given residue is represented by a graph G which contains nodes V represented by atom types and coordinates and edges E with atom bonds; e_{RBF} , e_{SBF} and e_G is edges where the distance $\leq 10\text{\AA}$ between two atoms. Different modules receive the information labelled by color; The output of GraphEBM is the summary of the Embedding and Interaction module.

where $n \in [1 \dots N_{RBF}]$ denotes the order of RBF, c denotes cutoff distance to consider their interactions and d denotes the distance between atom i and j . And the SBF defined by

$$\tilde{a}_{\text{SBF},ln}^{(kj,ji)}(d, \alpha) = \sqrt{\frac{2}{c^3 j_{l+1}^2(z_{ln})}} j_l\left(\frac{z_{ln}}{c}d\right) Y_l^0(\alpha) \quad (2)$$

where $l \in [1 \dots N_{SBF}]$ denotes the order of Bessel functions, j_l denotes the l -order Bessel function, z_{ln} denotes the n -th root of the l -order Bessel function and Y_l^0 denotes the Spherical harmonics. Considering the bonds have an important influence on the physical properties of atoms, we should use a Graph Neural Network to catch the topological structure and residue type information. The attention mechanism has proven to be very effective so far, so we introduce GAT to aggregate the residue types and atom types information by message passing on atom bonds. The GAT can be described by

$$\begin{aligned} e_{ij} &= a(\mathbf{W}h_i, \mathbf{W}h_j) \\ \alpha_{ij} &= \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})} \\ h'_i &= \parallel \sigma\left(\sum_{k=1}^K \alpha_{ij}^k \mathbf{W}^k h_j\right) \end{aligned} \quad (3)$$

where h_i is the input feature of node i , a is a shared attentional mechanism, $\mathbf{W}$ is a weight matrix, K is the number of attention heads, $\mathcal{N}_i$ denotes the set of neighbors of node i , σ denotes the activate function and $\parallel$ denotes the concatenation operation.

B. Smooth factor

In Fourier-based calculations, the multiplicative inverse of the polynomial is crucial and indispensable for precision. In this work, sampling a side-chain conformation is random, and is not constrained by physical laws. This sampling strategy will generate some atoms which are so close that the Bessel function overflows or causes exploding gradient. Inspired by

Laplace smooth, the distance in RBF and SBF can add a λ factor for smoothness and stability in training procedures.

$$\tilde{e}_{\text{RBF},n}^{(ji)}(d) = \tilde{e}_{\text{RBF},n}^{(ji)}(d + \lambda) \quad (4)$$

$$\tilde{a}_{\text{SBF},ln}^{(kj,ji)}(d, \alpha) = \tilde{a}_{\text{SBF},ln}^{(kj,ji)}(d + \lambda, \alpha) \quad (5)$$

III. EXPERIMENTS

A. Dataset

The dataset is the same with [15], which contains high-resolution PDB structures and removes similar proteins in the test dataset. The training dataset has 12473 structures, and the test dataset has 121 structures.

B. Evaluation

The energy function should distinguish the conformation closest to the native conformation from samples. The comparison methods consist of Deep Learning methods and the Rosetta energy functions. We rerun the Rosetta to predict the side-chain conformations of the test dataset using Rosetta score12 and Rosetta ref2015 [10] energy functions. And we compare it with the Atom Transformer [15] which is the state-of-the-art model in this task. We use the same test sampling strategy as Atom Transformer to reimplement the sampling strategy in Rosetta. The energy function scores every conformation sampled and selects a conformation with the lowest energy. When all χ angles of the selected conformation are within 20° of the ground truth, the rotamer is recovered correctly. For a more detailed analysis, we used the classification from [15] to define buried residues (≥ 24), medium residues (*others*) and surface residues (< 16) by the number of neighbors within $10\text{\AA}$ of the residue's C_β .

C. Local rotamer recovery

TABLE I
ROTAMER RECOVERY ACCURACY OVER THE TEST DATASET

Discrete sampling strategy				
Model	Avg	Buried	Medium	Surface
Rosetta score12 (rotamer-trials)	73.1	90.7	78.4	59.7
Rosetta ref2015 (rotamer-trials)	75.1	91.5	80.4	62.5
Atom Transformer	70.2	91.3	73.7	58.2
Atom Transformer (ensemble)	71.5	91.2	75.3	59.5
GraphEBM	76.0	87.0	78.3	69.2
Continuous sampling strategy				
Model	Avg	Buried	Medium	Surface
Rosetta score12 (rt-min)	73.2	91.0	78.2	60.2
Rosetta ref2015 (rt-min)	75.8	91.9	80.8	62.5
Atom Transformer	73.1	91.1	79.3	58.3
Atom Transformer (ensemble)	74.1	91.1	80.3	59.5
GraphEBM	78.4	90.6	81.4	70.2

Table I compares our model with two versions of Rosetta energy functions and the Atom Transformer, as reported by [15]. The comparison was conducted on the same test dataset of 121 proteins using the rotamer-trials mover and rt-min mover in Rosetta. Our model outperforms both the Rosetta energy functions and the Atom Transformer. It shows a significant

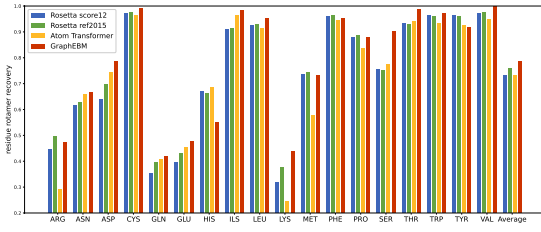


Fig. 2. Residues rotamer recovery

improvement in the recovery of surface residues, a challenging task due to the loss of physical constraints compared to buried residues. This improvement is likely due to GraphEBM's additional incorporation of geometric information compared to other methods. The performance of our model on buried residues is not as strong as other models, which we attribute to the granularity of the sampling strategy. However, in the medium category, our model surpasses others. Despite the challenges in tasks like surface residue recovery, our model consistently outperforms other methods by a significant margin.

Fig 2 shows our model's recovery rate for each residue type, excluding ALA and GLY. Our model generally matches or surpasses other models, including the Rosetta ref2015 energy function. However, it performs less effectively on ARG, GLN, GLU, and LYS rotamer recovery, similar to other methods. These residues, likely to be on the surface, have high hydropathy indices [23].

D. The χ_1 and χ_2 angles distribution

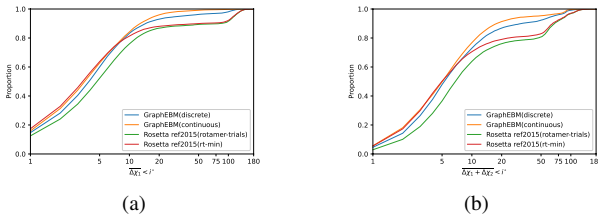
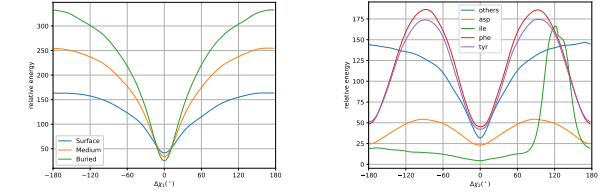


Fig. 3. The distribution of $\Delta\chi_1$ and $\Delta\chi_2$.

Fig 3.a shows the distribution of the $\Delta\chi_1$. The χ_1 angle is the most precise angle of the native conformation, so we visualize the angle proportion. The performance of GraphEBM is almost close to Rosetta when $\Delta\chi_1$ is small. But the Rosetta has a more extreme distribution. It is worth noting that the proportion of GraphEBM approaches 1 faster than Rosetta when $\Delta\chi_1 > 10^\circ$. Fig 3.b has the same trend of $\Delta\chi_1 + \Delta\chi_2$. And those figures show the different sampling strategies have a greater impact on the model accuracy because the same strategy's models have the same performance while $\Delta\chi$ is small.



(a) Different position energy curve (b) Energy curve for the amino acids Asp, Phe, Tyr, Ile and others about χ_2

Fig. 4. Relative energy curve

E. Energy visualization

The buried side chains are more tightly packed than others due to their fewer degrees of freedom, as noted by Richardson [24]. This characteristic is reflected in the Buried/Medium/Surface energy curve in Fig 4.a, which shows a steeper response to variations away from the native conformation. Interestingly, some residues like Tyr, Asp, and Phe exhibit symmetry about χ_2 . This symmetry is manifested as a 180° periodicity in Fig 4.b. However, not all amino acids display this kind of symmetry. For instance, the ILE amino acid presents a significant energy barrier at 120 degrees. This is because atomic overlap or conflict occurs at this angle, violating physical laws. This observation demonstrates the sensitivity of our model to conformations that defy physical principles. Such a finding is not reported in the Atom Transformer, further validating the superior ability of our model to capture atomic reactions.

F. Combinational side chain optimization

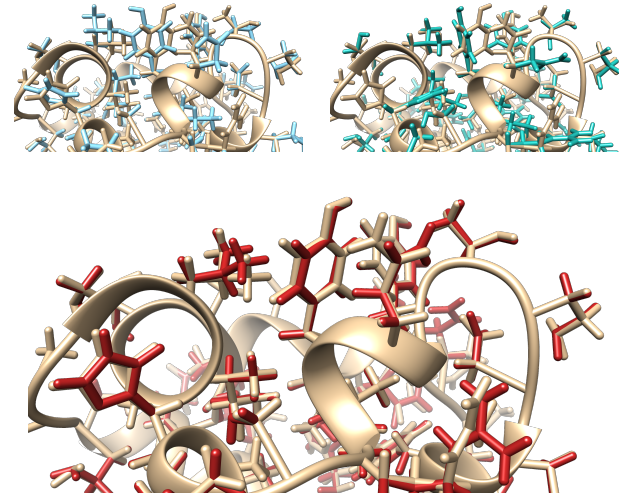


Fig. 5. The PDB deposited structure, the PULCHRA predicted all-atom model, Atom Transformer and our side chain conformation refined model are colored tan, sky-blue, sea-green and brick-red respectively. .

For testing our model on Combinational side chain optimization, we run the PULCHRA [22] to generate an initial

side-chain conformation with only backbone. PULCHRA is a geometry method and very fast for side chain generation. But, its recovery rate is bad even only considering χ_1 . The recovery strategy is to iterate residue by residue and select the optimal until every residue is stable. GraphEBM tries to recover the side chain from the initial conformation by PULCHRA and GraphEBM improves the recovery rate from 53.5% to 75.7% in χ_1 and for 38.3% to 62.2% in $\chi_1 + \chi_2$ the same, while the recovery rate of Atom Transformer is only 59.5% in χ_1 and 38.8% in $\chi_1 + \chi_2$. In conclusion, our model's ability to capture physical laws and apply them in a simple iteration strategy allows it to outperform other methods such as PULCHRA and Atom Transformer in side chain conformation optimization.

For a more intuitive display, we visualize the surface of a protein(PDBID:1TUKA). In Fig 5, the side-chain from PULCHRA and Atom Transformer are disorganized and unaligned to the ground truth, but the refined side-chain is more regular and close to the truth.

IV. CONCLUSION

We proposed the first SE(3) equivariant energy-based model for protein side chain conformation. This model refines the message aggregating architecture of DimeNet by combining it with the Graph Attention Network (GAT). To address the training gradient exploding problem of DimeNet, we introduced a smooth factor in the Bessel function. Both theoretical and experimental analyses were conducted to understand its influence on the performance of gradient descent optimization. But this is still limited by the sampling strategy because the whole protein recovery task is a combinatorial optimization problem. Based on simple sampling strategy cannot solve that problem well. We think the future work is to generate the side chain conformation.

V. CODE AVAILABILITY

All code is available at github.com/biomed-AI/GraphEBM.

VI. ACKNOWLEDGMENT

This study has been supported by the National Key R&D Program of China (Grant No. 2022YFF1203100), the NSFC under Grant 12126610, R&D project of Pazhou Lab (Huangpu) under Grant 2023K0606.

REFERENCES

- [1] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973.
- [2] C. Anfinsen and H. Scheraga, "Experimental and theoretical aspects of protein folding," *Advances in protein chemistry*, vol. 29, pp. 205–300, 1975.
- [3] C. A. Rohl, C. E. Strauss, K. M. Misura, and D. Baker, "Protein structure prediction using rosetta," in *Methods in enzymology*. Elsevier, 2004, vol. 383, pp. 66–93.
- [4] B. Kuhlman and P. Bradley, "Advances in protein structure prediction and design," *Nature Reviews Molecular Cell Biology*, vol. 20, no. 11, pp. 681–697, 2019.
- [5] T. Lazaridis and M. Karplus, "'new view' of protein folding reconciled with the old through multiple unfolding simulations," *Science*, vol. 278, no. 5345, pp. 1928–1931, 1997.
- [6] S. Tanaka and H. A. Scheraga, "Medium-and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins," *Macromolecules*, vol. 9, no. 6, pp. 945–950, 1976.
- [7] M. J. Sippl, "Calculation of conformational ensembles from potentials of mean force: an approach to the knowledge-based prediction of local structures in globular proteins," *Journal of molecular biology*, vol. 213, no. 4, pp. 859–883, 1990.
- [8] T. Lazaridis and M. Karplus, "Effective energy functions for protein structure prediction," *Current opinion in structural biology*, vol. 10, no. 2, pp. 139–145, 2000.
- [9] F. E. Boas and P. B. Harbury, "Potential energy functions for protein design," *Current opinion in structural biology*, vol. 17, no. 2, pp. 199–204, 2007.
- [10] R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, R. Das, D. Baker, B. Kuhlman, T. Kortemme, and J. J. Gray, "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design," *Journal of Chemical Theory and Computation*, vol. 13, no. 6, pp. 3031–3048, 2017.
- [11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [12] S. Chen, Z. Sun, L. Lin, Z. Liu, X. Liu, Y. Chong, Y. Lu, H. Zhao, and Y. Yang, "To improve protein sequence profile prediction through image captioning on pairwise residue distance map," *Journal of chemical information and modeling*, vol. 60, no. 1, pp. 391–399, 2019.
- [13] J. Ingraham, V. Garg, R. Barzilay, and T. Jaakkola, "Generative models for graph-based protein design," *Advances in neural information processing systems*, vol. 32, 2019.
- [14] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.
- [15] Y. Du, J. Meier, J. Ma, R. Fergus, and A. Rives, "Energy-based models for atomic-resolution protein conformations," *arXiv:2004.13167 [cs, q-bio, stat]*, 2020.
- [16] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, "Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds," *arXiv preprint arXiv:1802.08219*, 2018.
- [17] O.-E. Ganea, X. Huang, C. Bunne, Y. Bian, R. Barzilay, T. S. Jaakkola, and A. Krause, "Independent SE(3)-Equivariant Models for End-to-End Rigid Protein Docking," in *International Conference on Learning Representations*, 2021.
- [18] M. McPartlon, B. Lai, and J. Xu, "A deep se (3)-equivariant model for learning inverse protein folding," *bioRxiv*, 2022.
- [19] J. Klicpera, J. Groß, and S. Günnemann, "Directional Message Passing for Molecular Graphs," *arXiv:2003.03123 [physics, stat]*, 2020.
- [20] J. Klicpera, S. Giri, J. T. Margraf, and S. Günnemann, "Fast and uncertainty-aware directional message passing for non-equilibrium molecules," *arXiv preprint arXiv:2011.14115*, 2020.
- [21] L. Holm and C. Sander, "Fast and simple monte carlo algorithm for side chain optimization in proteins: application to model building by homology," *Proteins: Structure, Function, and Bioinformatics*, vol. 14, no. 2, pp. 213–223, 1992.
- [22] P. Rotkiewicz and J. Skolnick, "Fast procedure for reconstruction of full-atom protein models from reduced representations," *Journal of computational chemistry*, vol. 29, no. 9, pp. 1460–1465, 2008.
- [23] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.
- [24] J. S. Richardson and D. C. Richardson, "Principles and patterns of protein conformation," in *Prediction of protein structure and the principles of protein conformation*. Springer, 1989, pp. 1–98.