PLOS ONE

RESEARCH ARTICLE

PolyFold: An interactive visual simulator for distance-based protein folding

Andrew J. McGehee[1], Sutanu Bhattacharya[1☯], Rahmatullah Roche[1☯], Debswapna Bhattacharya[1,2]*

1 Department of Computer Science and Software Engineering, Auburn University, Auburn, AL, United States of America, **2** Department of Biological Sciences, Auburn University, Auburn, AL, United States of America

☯ These authors contributed equally to this work.
* bhattacharyad@auburn.edu

Citation: McGehee AJ, Bhattacharya S, Roche R, Bhattacharya D (2020) PolyFold: An interactive visual simulator for distance-based protein folding. PLoS ONE 15(12): e0243331. https://doi.org/10.1371/journal.pone.0243331

Editor: Yang Zhang, University of Michigan, UNITED STATES

Received: August 14, 2020

Accepted: November 18, 2020

Published: December 3, 2020

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was partially supported by the National Science Foundation CAREER Award DBI-1942692 to DB, the National Science Foundation grant IIS-2030722 to DB, and the National Institute of General Medical Sciences Maximizing Investigators' Research Award (MIRA) R35GM138146 to DB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Recent advances in distance-based protein folding have led to a paradigm shift in protein structure prediction. Through sufficiently precise estimation of the inter-residue distance matrix for a protein sequence, it is now feasible to predict the correct folds for new proteins much more accurately than ever before. Despite the exciting progress, a dedicated visualization system that can dynamically capture the distance-based folding process is still lacking. Most molecular visualizers typically provide only a static view of a folded protein conformation, but do not capture the folding process. Even among the selected few graphical interfaces that do adopt a dynamic perspective, none of them are distance-based. Here we present PolyFold, an interactive visual simulator for dynamically capturing the distance-based protein folding process through real-time rendering of a distance matrix and its compatible spatial conformation as it folds in an intuitive and easy-to-use interface. PolyFold integrates highly convergent stochastic optimization algorithms with on-demand customizations and interactive manipulations to maximally satisfy the geometric constraints imposed by a distance matrix. PolyFold is capable of simulating the complex process of protein folding even on modest personal computers, thus making it accessible to the general public for fostering citizen science. Open source code of PolyFold is freely available for download at https://github.com/Bhattacharya-Lab/PolyFold. It is implemented in cross-platform Java and binary executables are available for macOS, Linux, and Windows.

Introduction

Computational protein structure prediction has witnessed remarkable progress in the recent past due to advances in folding new proteins from scratch using sufficiently accurate estimation of the inter-residue distance matrix [1–4]. A distance matrix encodes a protein's three-dimensional (3D) structure through inter-residue spatial proximity information that can be converted to physical constraints in order to drive the ab initio folding process with minimal conformational search [5,6]. Consequently, distance-based protein folding has gained a lot of attention, fueling considerable research efforts [7–11]. However, the lack of a dedicated visualization system that can dynamically capture the distance-based folding process precludes the

Competing interests: The authors have declared that no competing interests exist.

possibility of obtaining a visual understanding of its nature. Currently popular molecular visualization tools like PyMol and UCSF Chimera [12,13] typically provide only a static view of a folded protein conformation, but do not capture the folding process. Recent graphical interfaces such as the PyRosetta Toolkit [14] and InteractiveROSETTA [15] adopt dynamic perspectives, but they are built exclusively for the ROSETTA molecular modeling suite [16], which primarily relies on a fragment-based approach for protein folding. InteractiveROSETTA has many sophisticated ROSETTA-based features, including APIs to incorporate various distance restraints. However, a standalone visualizer that provides insights to distance-based folding for researchers is still lacking. Beyond the realm of expert-oriented visualization tools, the interactive graphical interface Foldit Standalone [17] makes it possible for non-experts to manipulate protein structures in the context of the popular scientific discovery game Foldit [18], which itself is based on ROSETTA and thus not distance-based. A dedicated distance-based visual folding simulator with a simple to use interface will not only provide a central platform for researchers to delve deeper into the folding process and gain critical insights, but will also make the latest technological advances in protein folding and molecular modeling easily accessible to non-experts, while still being scientifically accurate.

We have developed a brand-new standalone GUI called PolyFold for visually simulating the distance-based protein folding process. PolyFold provides several user-friendly controls for running powerful distance matrix optimization algorithms, including gradient descent [19] and simulated annealing [20,21], with on-demand customization and interactive manipulations. Through real-time rendering of a live interaction map with smooth color ramping to capture the distance matrix alongside its compatible 3D conformation color-coded to highlight the secondary structural geometry, PolyFold makes it possible to dynamically view the folding process. Additionally, an interactive movement panel provides the ability to structurally manipulate the molecule. PolyFold does not require familiarity with protein biochemistry and provides an easily accessible platform for elucidating the distance-based protein folding process.

PolyFold features

As shown in **Fig 1**, the PolyFold GUI consists of three main panels: a live interaction map panel for visualizing the target distance matrix (upper triangle) and the distance matrix currently realized (lower triangle) with real-time updates, a dynamic structural display panel rendering the 3D conformation of the protein molecule compatible with the current distance matrix, and a movement panel that permits users to interactively manipulate the molecule. The core of PolyFold is implemented in Java, and the GUI controls make extensive use of the JavaFX application library. The code is cross-platform and builds and runs on macOS, Linux, and Windows. Pre-packaged binaries are also available for plug-and-play execution (see section 2 in **S1 Text**).

PolyFold includes two optimizers for distance-based folding: gradient descent and simulated annealing, the former operating in Cartesian space and the latter in angular space. Users can launch interactive versions of both optimizers, which dynamically update the display as they run and can be cancelled prior to completion. Cascaded runs which either alternate or repeat optimizers, such as the repeated gradient descent used in AlphaFold [7], are also possible. The parameters of both optimizers are fully configurable (see section 3.3 in **S1 Text**). It is worth mentioning that PolyFold's optimization engine is designed for real-time and interactive visualization of the distance-based optimization as opposed to physics-based Molecular Dynamics (MD) simulations, which are often used for intermediate state or pathway analysis.

PolyFold's custom interactive manipulations have been specifically implemented for real-time manipulation of a molecule. Users are able to manipulate the molecular geometry in real-

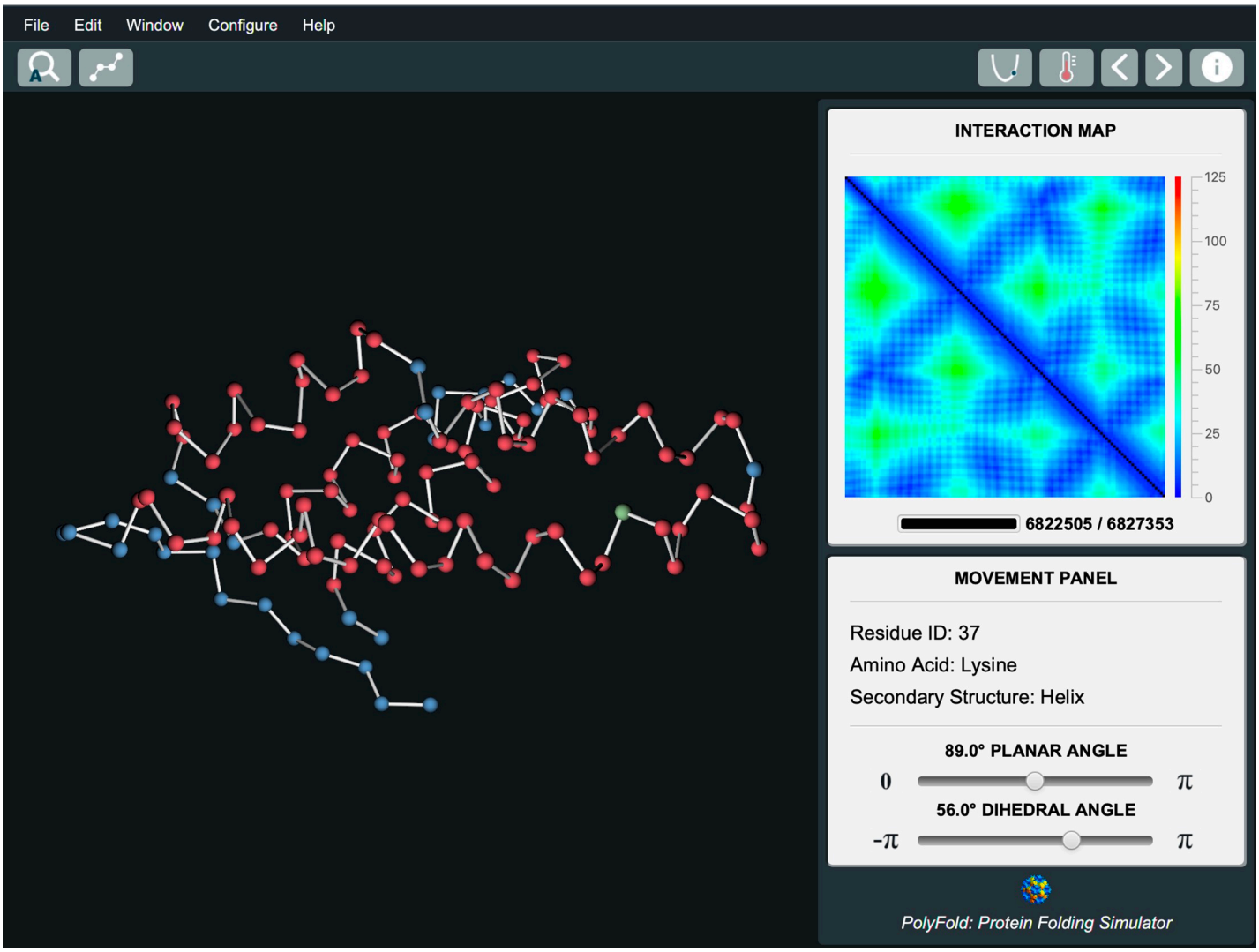

Fig 1. A representative PolyFold distance-based folding session for the amino terminal domain of enzyme I from escherichia coli (PDB ID: 1zym), with real-time display of the interaction map and its compatible 3D structure. Residue 37 is selected for manipulation.

https://doi.org/10.1371/journal.pone.0243331.g001

time with a simultaneous update to the live interaction map by selecting a residue and updating its pseudo planar and dihedral angles by dragging sliders. This feature can be particularly useful for multi-domain proteins by folding the full-length structure via distance-based optimization and subsequently adjusting the relative domain orientations by manipulating the domain linkers. PolyFold also keeps track of a history of modified states for undoing, redoing, saving, loading, and restoring the molecule to an unfolded extended state during various stages of interactive manipulations or in-built optimizations. Structures can be translated, rotated, scaled, and auto-zoomed as needed (see section 3.1 and 3.2 in **S1 Text**).

A PolyFold session can be started by supplying a distance matrix similar to the biannual Critical Assessment of protein Structure Prediction (CASP) [22–25] experiments' residue-residue (RR) format along with secondary structures (see section 4.1 in **S1 Text**). Intermediate session states can be saved to anonymous save states for quick recall or saved to named save states for lengthier sessions. Further, structures can be saved in Protein Data Bank (PDB) [26] format and restored in a later session. Prior to saving a structure, PolyFold performs secondary

structure-assisted geometric chirality checking using a heuristic cost function for identifying the correct chirality as the sum over tetrapeptides in α-helices and β-sheets [27]. While PolyFold can work for large proteins, reasonably sized structures (length < 500 residues) are currently supported for seamless rendering.

Case study

To examine the accuracy and robustness of PolyFold for distance-based folding, we study the folding of a Ribosomal protein 1ctfA of length 68 residues [28] from near-native distance matrices as well as noisy distance matrices. In all cases, we run PolyFold by employing a single run of simulated annealing with a random seed of 0 followed by three repeated runs of gradient descent with PolyFold's default parameters.

We first perform distance-based folding by feeding a near-native distance matrix into PolyFold in CASP RR format after computing the floors and ceilings of the true inter-residue distances of the target protein 1ctfA. That is, the near-native distance matrix supplied to PolyFold specifies the distances to be within 1Å of the true real-valued distances, thus simulating a reconstruction scenario. As shown in **Fig 2**, PolyFold successfully reconstructs the structure of the target protein with a very high TM-score [29] of 0.92, demonstrating the effectiveness of PolyFold's in-built optimizers.

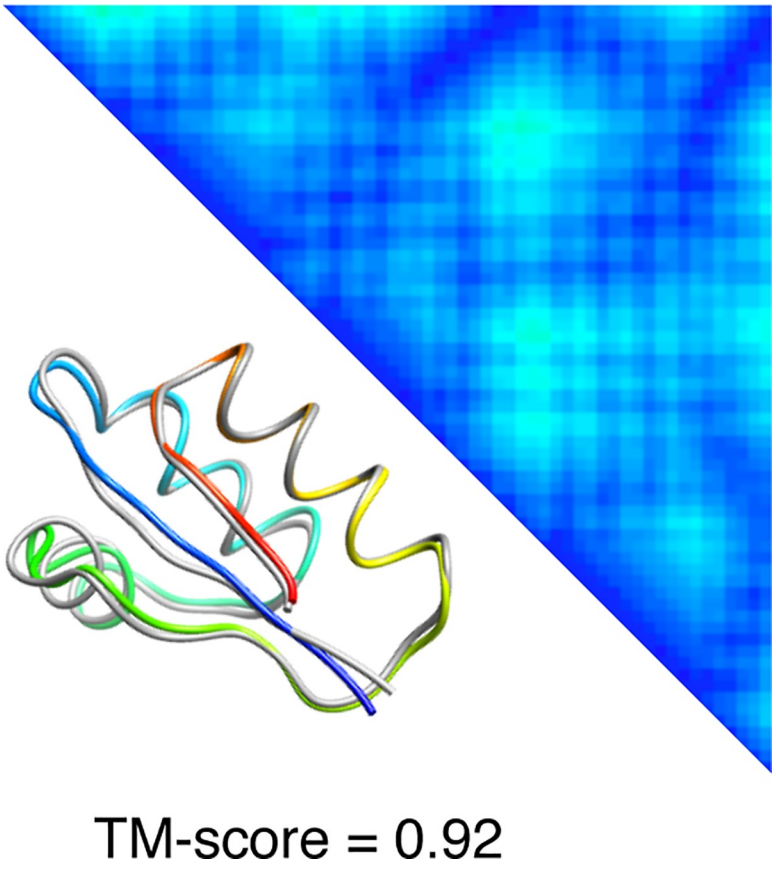

Fig 2. PolyFold distance-based reconstruction for the target 1ctfA with a near-native distance matrix. The upper diagonal shows the inter-residue distance matrix, and the lower diagonal shows the structural superimposition between the PolyFold predicted model (in rainbow) and the experimental structure of the target (in gray).

https://doi.org/10.1371/journal.pone.0243331.g002

Next, we investigate the effect of feeding noisy distance matrices [30,31] into PolyFold's structural optimization engine by systematically introducing zero-mean Gaussian noise having standard deviations of σ = 1, 2, and 4Å into the true distance matrix of the target protein 1ctfA. This is accomplished by first calculating the inter-residue pair-wise distances from the near-native PDB file. We then create a pool of residue pairs (i, j) where $|i-j| > 6$. Next, we either uniform randomly select 50% of pairs or select 100% of pairs from the pool to be modified. We refer to this selection percentage as the "noise level." The calculated near-native distances are then modified with zero-mean Gaussian noise with the specified standard deviation. As shown in **Fig 3**, we observe that PolyFold's optimizations are fairly noise-tolerant, predicting the correct fold with a TM score > 0.5 [32] in all cases except in the most extreme case with noise level 100% and a standard deviation of 4Å (**S1**–**S6 Movies**). When noisy distance matrices with σ = 1Å are fed into PolyFold, it achieves TM-scores of 0.89 and 0.86 for 50% and 100% noise levels, respectively. By doubling the noise to σ = 2Å, PolyFold still predicts correct folds with TM-scores ≥ 0.6 for both 50% and 100% noise levels. Finally, PolyFold predicts the correct fold with a TM-score of 0.5 using a quite noisy distance matrix with σ = 4Å and a noise level of 50%, demonstrating its robustness in distance-based folding when using noisy distance matrices.

Benchmarking

While PolyFold is primarily an interactive visual simulator for distance-based protein folding as opposed to a structure prediction method, we assess PolyFold's predictive modeling performance using a benchmark set of six small proteins ranging in length from 43 to 76 residues that have been the subject of previous studies [33,34]. For predictive modeling using PolyFold, we predict secondary structures by running SPIDER3 [35] locally. We then feed the predicted secondary structures together with distance matrices to PolyFold at varied resolutions ranging from near-native to noisy and predicted maps. In all cases, we employ two cascaded runs of PolyFold's gradient descent optimization, both for 65,000 iterations, with the first run using a step size of 0.005 and the second run using a step size of 0.0001. The optimized structural models are subsequently saved in PDB format for assessment. First, we feed near-native distance matrices within 1Å of the true real-valued distances along with predicted secondary structures into PolyFold and evaluate the accuracy of the predicted models. Next, we assess the predictive performance of PolyFold using noisy distance matrices. We follow the same strategy of introducing zero-mean Gaussian noise as discussed before to generate noisy distance matrices for the benchmark set. We feed the noisy distances matrices having standard deviations of σ = 1, 2, and 4Å at 50% and 100% noise levels together with the predicted secondary structures into PolyFold and evaluate the folding performance. Finally, we investigate the predictive ability of PolyFold when predicted distance matrices and predicted secondary structures are supplied as input. For each protein target in the benchmark set, we predict inter-residue distance maps by feeding the multiple sequence alignments (MSA) [36] of the target proteins into trRosetta [10] and then supply the predicted distances maps together with the predicted secondary structures into PolyFold to evaluate the accuracies of the predicted models. trRosetta [10] is a state-of-the-art deep learning-based protein structure prediction method that predicts inter-residue distances and orientation (dihedral and planer angles), which are subsequently transformed into restraints to generate 3D structures using energy minimization. From the standpoint of folding, both trRosetta and PolyFold rely on gradient-based optimization. However, trRosetta-based folding utilizes both distance and orientation information, whereas PolyFold uses only distance information. For a fair performance comparison with PolyFold, we, therefore, employ trRosetta-based folding using only predicted distance maps but no orientation information.

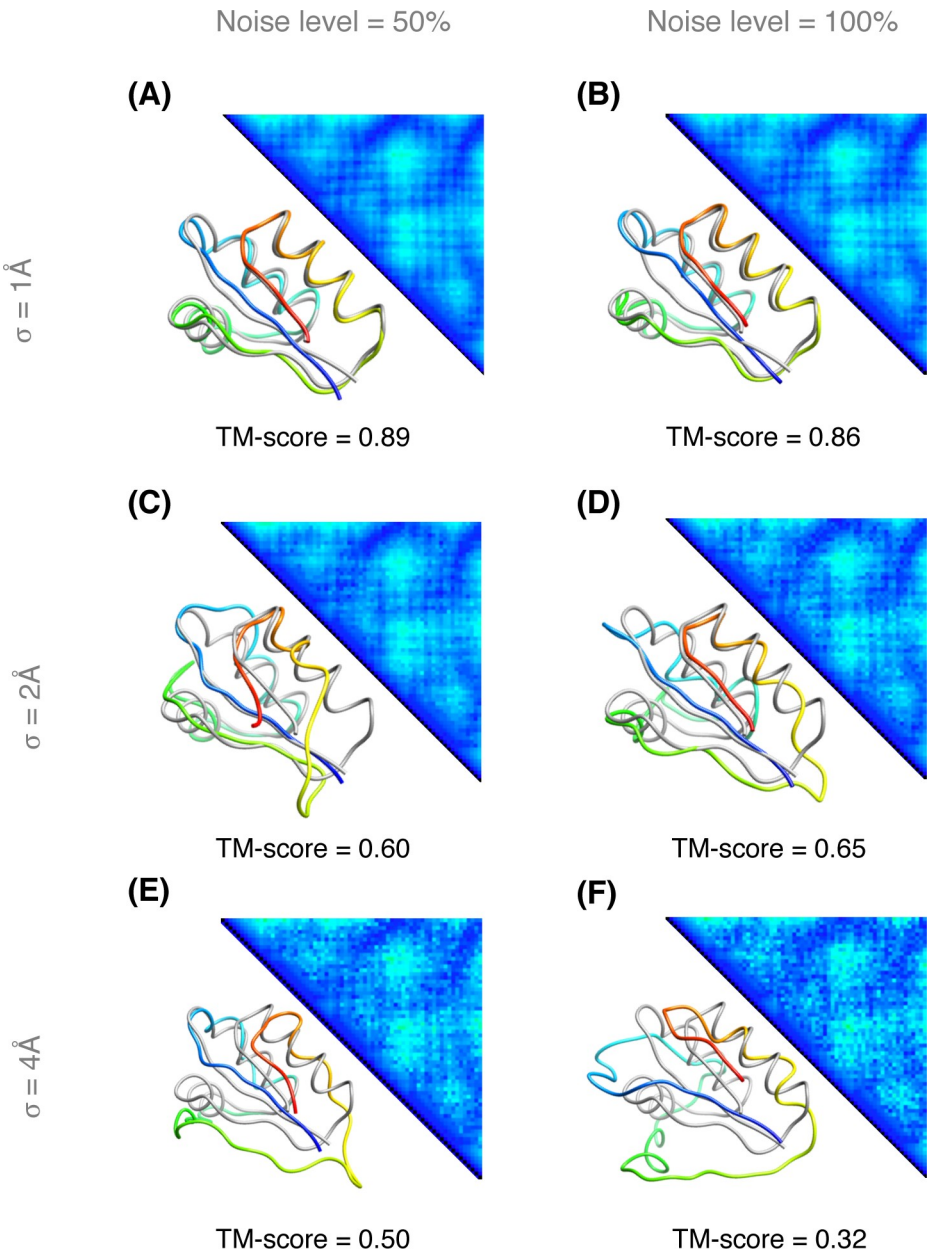

Fig 3. PolyFold distance-based folding for the target 1ctfA with noisy distance matrices. The upper diagonal shows the noisy inter-residue distance matrix by introducing zero-mean Gaussian noise into the true distance matrix with various standard deviations (σ) and noise levels. The lower diagonal shows the structural superimposition between the PolyFold predicted model (in rainbow) and the native structure of the target (in gray). (A) Noise level of 50% and σ of 1Å, (B) Noise level of 100% and σ of 1Å. (C) Noise level of 50% and σ of 2Å, (D) Noise level of 100% and σ of 2Å. (E) Noise level of 50% and σ of 4Å, (F) Noise level of 100% and σ of 4Å.

https://doi.org/10.1371/journal.pone.0243331.g003

We run trRosetta-based distance-only folding locally by setting the parameter ('—no-orient') that uses the same trRosetta-predicted distance maps supplied to PolyFold, albeit without orientation. Additionally, we compare the predictive modeling performance of PolyFold with two state-of-the-art protein structure prediction pipelines: I-TASSER [37,38] and Robetta [39]. We submit jobs to the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) after excluding homologous templates with 30% sequence identity cutoff with the target

protein and collect the top predicted model for each target protein. We submit jobs to the Robetta structure prediction server (https://robetta.bakerlab.org/) by selecting the 'AB only' option to use the Rosetta fragment assembly method for *ab initio* folding [16] and collect the top predicted model for each target protein. Of note, unlike the head-to-head comparison between the distance-based folding using PolyFold and trRosetta, a direct comparison between PolyFold and I-TASSER or Robetta is not fair because I-TASSER and Robetta have clear advantages in their use of template and/or fragment information as well as other structural features such as solvent accessibility. Furthermore, both I-TASSER and Robetta servers employ a full-fledged structure prediction pipeline by performing time-consuming conformational sampling to generate a large pool of structural decoys followed by optimal decoy selection and all-atom refinement. By contrast, PolyFold does not have such advantages since it employs computationally inexpensive distance matrix optimization over a single session while operating on a singular structure without having access to other structural features such as templates or fragments and does not perform all-atom refinement. Nonetheless, the comparison between PolyFold and I-TASSER or Robetta offers some interesting insights.

Table 1 reports the predictive modeling performance of PolyFold using distance matrices at varied resolutions compared to I-TASSER and Robetta, as well as a head-to-head comparison between PolyFold and distance-only trRosetta, both using the same predicted distance matrices. Using predicted secondary structures and near-native distance matrices, PolyFold attains a mean TM-score of 0.73, which is higher than both I-TASSER and Robetta having mean TM-scores of 0.72 and 0.67, respectively. Moreover, PolyFold's accuracy range (maximum TM-score of 0.96, minimum TM-score of 0. 49) is better than that of I-TASSER (maximum TM-score 0.9, minimum TM-score 0.42), and Robetta (maximum TM-score 0.82, minimum TM-score 0.42. That is, PolyFold-based predictive modeling using near-native distance matrices delivers better performance than I-TASSER and Robetta. When noisy distance matrices (σ = 1Å, noise level = 50%) are fed into PolyFold, the mean TM-score becomes 0.67, the same as the mean TM-score of Robetta. As we increase σ and noise levels of the input distance matrices, the mean TM-scores steadily decrease. This is expected, and it demonstrates the robustness of the PolyFold's optimization engine. When predicted secondary structures and predicted distance matrices are fed into PolyFold, it attains a mean TM-score of 0.39, which is better than the distance-only trRosetta having a mean TM-score of 0.35. The better average performance of PolyFold compared to distance-only trRosetta underscores the

Table 1. Predictive modeling performance on the benchmark dataset using PolyFold with SPIDER3 predicted secondary structures and near-native, noisy, and predicted maps. I-TASSER and Robetta *ab-initio* modeling results, obtained by submitting jobs directly to their web servers, as well as distance-only trRosetta results, obtained by running it locally with parameter settings ('—no-orient'), are also reported. In all cases, the mean, maximum and minimum TM-scores of the top predicted models are reported. Values in bold represents the best performance.

Methods	Mean	Maximum	Minimum
PolyFold w/ near-native maps	**0.73**	**0.96**	**0.49**
I-TASSER	0.72	0.9	0.42
Robetta *ab-initio*	0.67	0.82	0.42
PolyFold w/ noisy maps (σ = 1Å, noise level = 50%)	0.67	0.87	0.49
PolyFold w/ noisy maps (σ = 1Å, noise level = 100%)	0.66	0.87	0.42
PolyFold w/ noisy maps (σ = 2Å, noise level = 50%)	0.56	0.78	0.41
PolyFold w/ noisy maps (σ = 2Å, noise level = 100%)	0.55	0.72	0.32
PolyFold w/ noisy maps (σ = 4Å, noise level = 50%)	0.33	0.44	0.24
PolyFold w/ noisy maps (σ = 4Å, noise level = 100%)	0.26	0.31	0.2
PolyFold w/ predicted maps	0.39	0.49	0.3
trRosetta (distance-only)	0.35	0.63	0.27

https://doi.org/10.1371/journal.pone.0243331.t001

effectiveness of PolyFold's gradient-based optimization. Interestingly, predicted distance matrices lead to better average accuracy of the resulting structural models in PolyFold than the noisy distance matrices with σ = 4Å with noise level 50% and 100%; whereas the use of noisy distance matrices with σ = 2Å results in an average TM score > 0.5 for both 50% and 100% noise levels, thus outperforming modeling with predicted distance matrices. That is, the quality of the predicted distance matrices possibly lies somewhere in between the qualities of the noisy distance matrices at σ = 2Å and σ = 4Å. The results indicate that PolyFold's optimization engine is sensitive to subtle changes in the quality of the input distance matrix and therefore may be suitable for studying the impact of noisy and predicted distance matrices in protein modeling to investigate which parts of the protein are over or under-restrained. These regions of the distance matrix could then be modified as appropriate in order to improve predictive modeling performance. Furthermore, PolyFold's fully customizable optimization engine enables users to experiment how various optimization parameters such as the step size of gradient descent might affect the resulting structural models in real time. This may help with modeling flexible regions such as loops that may be under-restrained in a predicted distance matrix. In summary, PolyFold is robust, versatile, and practically useful for predictive protein modeling.

Conclusions

PolyFold offers a real-time visual simulator for capturing the optimization processes of distance-based protein folding in a dynamic and interactive interface. Being robust and resilient to noise in distance matrices, PolyFold provides a versatile platform for visualizing distance-based protein folding. In the future, PolyFold may be extended to incorporate more features into the GUI for improved user experience such as multi-directional rotations of the structure or to interactively manipulate and possibly de-noise predicted distance matrices. In conclusion, PolyFold's fully configurable, robust structural optimization and manipulation engine coupled with its easy-to-use intuitive graphical interface make it accessible to both researchers and non-experts, enabling scientists to gain new insights into protein folding and facilitating broader participation.

Supporting information

S1 Text. A detailed PolyFold user manual.
(PDF)

S1 Movie. PolyFold protein folding simulation for the target 1ctfA using a noisy distance matrix by introducing zero-mean Gaussian noise into the true distance matrix, having a standard deviation of 1Å and noise level of 50%.
(MOV)

S2 Movie. PolyFold protein folding simulation for the target 1ctfA using a noisy distance matrix by introducing zero-mean Gaussian noise into the true distance matrix, having a standard deviation of 1Å and noise level of 100%.
(MOV)

S3 Movie. PolyFold protein folding simulation for the target 1ctfA using a noisy distance matrix by introducing zero-mean Gaussian noise into the true distance matrix, having a standard deviation of 2Å and noise level of 50%.
(MOV)

S4 Movie. PolyFold protein folding simulation for the target 1ctfA using a noisy distance matrix by injecting zero-mean Gaussian noise into the true distance matrix, having a standard deviation of 2Å and noise level of 100%.
(MOV)

S5 Movie. PolyFold protein folding simulation for the target 1ctfA using a noisy distance matrix by introducing zero-mean Gaussian noise into the true distance matrix, having a standard deviation of 4Å and noise level of 50%.
(MOV)

S6 Movie. PolyFold protein folding simulation for the target 1ctfA using a noisy distance matrix by introducing zero-mean Gaussian noise into the true distance matrix, having a standard deviation of 4Å and noise level of 100%.
(MOV)

Acknowledgments

The authors thank the middle and high school students as well as their teachers and parents or guardians participating in the Auburn Engineering Day ("E-day") for using PolyFold and providing feedback on its interface and features.

Author Contributions

Conceptualization: Debswapna Bhattacharya.

Data curation: Andrew J. McGehee, Sutanu Bhattacharya, Rahmatullah Roche, Debswapna Bhattacharya.

Formal analysis: Debswapna Bhattacharya.

Funding acquisition: Debswapna Bhattacharya.

Investigation: Andrew J. McGehee, Sutanu Bhattacharya, Rahmatullah Roche, Debswapna Bhattacharya.

Methodology: Debswapna Bhattacharya.

Project administration: Debswapna Bhattacharya.

Resources: Debswapna Bhattacharya.

Software: Andrew J. McGehee, Rahmatullah Roche, Debswapna Bhattacharya.

Supervision: Debswapna Bhattacharya.

Validation: Andrew J. McGehee, Sutanu Bhattacharya, Rahmatullah Roche, Debswapna Bhattacharya.

Visualization: Andrew J. McGehee, Sutanu Bhattacharya, Rahmatullah Roche, Debswapna Bhattacharya.

Writing – original draft: Andrew J. McGehee, Sutanu Bhattacharya, Rahmatullah Roche, Debswapna Bhattacharya.

Writing – review & editing: Andrew J. McGehee, Sutanu Bhattacharya, Rahmatullah Roche, Debswapna Bhattacharya.

References

1. Abriata LA, Tamò GE, Peraro MD. A further leap of improvement in tertiary structure prediction in CASP13 prompts new routes for future assessments. Proteins: Structure, Function, and Bioinformatics. 2019; 87: 1100–1112. https://doi.org/10.1002/prot.25787 PMID: 31344267
2. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). Proteins: Structure, Function, and Bioinformatics. 2019; 87: 1141–1148. https://doi.org/10.1002/prot.25834 PMID: 31602685
3. Xu J, Wang S. Analysis of distance-based protein structure prediction by deep learning in CASP13. Proteins: Structure, Function, and Bioinformatics. 2019; 87: 1069–1081. https://doi.org/10.1002/prot.25810 PMID: 31471916
4. Hou J, Wu T, Cao R, Cheng J. Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13. Proteins: Structure, Function, and Bioinformatics. 2019; 87: 1165–1178. https://doi.org/10.1002/prot.25697 PMID: 30985027
5. Kloczkowski A, Jernigan RL, Wu Z, Song G, Yang L, Kolinski A, et al. Distance matrix-based approach to protein structure prediction. J Struct Funct Genomics. 2009; 10: 67–81. https://doi.org/10.1007/s10969-009-9062-2 PMID: 19224393
6. Aszódi A, Gradwell MJ, Taylor WR. Global Fold Determination from a Small Number of Distance Restraints. Journal of Molecular Biology. 1995; 251: 308–326. https://doi.org/10.1006/jmbi.1995.0436 PMID: 7643405
7. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020; 577: 706–710. https://doi.org/10.1038/s41586-019-1923-7 PMID: 31942072
8. Xu J. Distance-based protein folding powered by deep learning. PNAS. 2019; 116: 16856–16865. https://doi.org/10.1073/pnas.1821309116 PMID: 31399549
9. Greener JG, Kandathil SM, Jones DT. Deep learning extends de novo protein modelling coverage of genomes using iteratively predicted structural constraints. Nat Commun. 2019; 10: 1–13. https://doi.org/10.1038/s41467-018-07882-8 PMID: 30602773
10. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. PNAS. 2020; 117: 1496–1503. https://doi.org/10.1073/pnas.1914677117 PMID: 31896580
11. Pietal MJ, Bujnicki JM, Kozlowski LP. GDFuzz3D: a method for protein 3D structure reconstruction from contact maps, based on a non-Euclidean distance function. Bioinformatics. 2015; 31: 3499–3505. https://doi.org/10.1093/bioinformatics/btv390 PMID: 26130575
12. PyMOL | pymol.org. [cited 7 Aug 2020]. Available: https://pymol.org/2/.
13. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004; 25: 1605–1612. https://doi.org/10.1002/jcc.20084 PMID: 15264254
14. Adolf-Bryfogle J Jr, RLD. The PyRosetta Toolkit: A Graphical User Interface for the Rosetta Software Suite. PLOS ONE. 2013; 8: e66856. https://doi.org/10.1371/journal.pone.0066856 PMID: 23874400
15. Schenkelberg CD, Bystroff C. InteractiveROSETTA: a graphical user interface for the PyRosetta protein modeling suite. Bioinformatics. 2015; 31: 4023–4025. https://doi.org/10.1093/bioinformatics/btv492 PMID: 26315900
16. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. Chapter nineteen—Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In: Johnson ML, Brand L, editors. Methods in Enzymology. Academic Press; 2011. pp. 545–574. https://doi.org/10.1016/B978-0-12-381270-4.00019-6 PMID: 21187238
17. Kleffner R, Flatten J, Leaver-Fay A, Baker D, Siegel JB, Khatib F, et al. Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. Bioinformatics. 2017; 33: 2765–2767. https://doi.org/10.1093/bioinformatics/btx283 PMID: 28481970
18. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. Nature. 2010; 466: 756–760. https://doi.org/10.1038/nature09304 PMID: 20686574
19. Ruder S. An overview of gradient descent optimization algorithms. arXiv:160904747 [cs]. 2017 [cited 8 Aug 2020]. Available: http://arxiv.org/abs/1609.04747.
20. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by Simulated Annealing. Science. 1983; 220: 671–680. https://doi.org/10.1126/science.220.4598.671 PMID: 17813860

21. Li X. Protein Folding Based on Simulated Annealing Algorithm. Third International Conference on Natural Computation (ICNC 2007). 2007. pp. 256–259. https://doi.org/10.1109/ICNC.2007.583

22. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—round x. Proteins: Structure, Function, and Bioinformatics. 2014; 82: 1–6. https://doi.org/10.1002/prot.24452 PMID: 24344053

23. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. Proteins: Structure, Function, and Bioinformatics. 2016; 84: 4–14. https://doi.org/10.1002/prot.25064 PMID: 27171127

24. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—Round XII. Proteins: Structure, Function, and Bioinformatics. 2018; 86: 7–15. https://doi.org/10.1002/prot.25415 PMID: 29082672

25. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins: Structure, Function, and Bioinformatics. 2019; 87: 1011–1020. https://doi.org/10.1002/prot.25823 PMID: 31589781

26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28: 235–242. https://doi.org/10.1093/nar/28.1.235 PMID: 10592235

27. Lund O, Hansen J, Brunak S, Bohr J. Relationship between protein structure and geometrical constraints. Protein Sci. 1996; 5: 2217–2225. https://doi.org/10.1002/pro.5560051108 PMID: 8931140

28. Jones DT, Buchan DWA, Cozzetto D, Pontil M. PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. Bioinformatics. 2012; 28: 184–190. https://doi.org/10.1093/bioinformatics/btr638 PMID: 22101153

29. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins: Structure, Function, and Bioinformatics. 2004; 57: 702–710. https://doi.org/10.1002/prot.20264 PMID: 15476259

30. Vassura M, Margara L, Di Lena P, Medri F, Fariselli P, Casadio R. FT-COMAR: fault tolerant three-dimensional structure reconstruction from protein contact maps. Bioinformatics. 2008; 24: 1313–1315. https://doi.org/10.1093/bioinformatics/btn115 PMID: 18381401

31. Vassura M, Margara L, Medri F, di Lena P, Fariselli P, Casadio R. Reconstruction of 3D Structures from Protein Contact Maps. In: Măndoiu I, Zelikovsky A, editors. Bioinformatics Research and Applications. Berlin, Heidelberg: Springer; 2007. pp. 578–589. https://doi.org/10.1007/978-3-540-72031-7_53

32. Xu J, Zhang Y. How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics. 2010; 26: 889–895. https://doi.org/10.1093/bioinformatics/btq066 PMID: 20164152

33. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions11Edited by Cohen F. E. Journal of Molecular Biology. 1997; 268: 209–225. https://doi.org/10.1006/jmbi.1997.0959 PMID: 9149153

34. Zhao F, Li S, Sterner BW, Xu J. Discriminative learning for protein conformation sampling. Proteins. 2008; 73: 228–240. https://doi.org/10.1002/prot.22057 PMID: 18412258

35. Heffernan R, Yang Y, Paliwal K, Zhou Y. Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. Bioinformatics. 2017; 33: 2842–2849. https://doi.org/10.1093/bioinformatics/btx218 PMID: 28430949

36. Zhang C, Zheng W, Mortuza SM, Li Y, Zhang Y. DeepMSA: constructing deep multiple sequence alignment to improve contact prediction and fold-recognition for distant-homology proteins. Bioinformatics. 2020; 36: 2105–2112. https://doi.org/10.1093/bioinformatics/btz863 PMID: 31738385

37. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols. 2010; 5: 725–738. https://doi.org/10.1038/nprot.2010.5 PMID: 20360767

38. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nature Methods. 2015; 12: 7–8. https://doi.org/10.1038/nmeth.3213 PMID: 25549265

39. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004; 32: W526–W531. https://doi.org/10.1093/nar/gkh468 PMID: 15215442