Using artificial sequence coevolution to predict disulfide-rich peptide structures with experimental connectivity in AlphaFold

Gabriella J. Gerlach
Computational and Systems Biology Department
University of Pittsburgh
Pittsburgh, PA
gjg21@pitt.edu

John M. Nicoludis
Department of Structural Biology
Genentech, Inc.
South San Francisco, CA 94080
nicoludj@gene.com

Abstract

We present a novel approach for embedding contact information in Alphafold to predict structures of disulfide-rich peptides (DRPs) with experimental disulfide connectivity. While AlphaFold generates accurate DRP structure prediction in most cases, it sometimes fails at predicting the specific connectivity pattern of the multiple disulfide bonds. Here, we take advantage of the principles of sequence coevolution to directly embed specific connectivity patterns within the MSA by mutating highly conserved cysteines in subsets of the MSA. This approach can be used to incorporate experimental disulfide connectivity patterns from mass spectrometry into DRP structure prediction. Lastly, after minimization of predicted structures by molecular dynamics, we find that predicted DRP structures with native connectivity display favorable peptide properties compared to non-native connectivities, suggesting our approach may be useful for determining the native connectivity of DRPs from sequence alone.

1 Introduction

Disulfide-rich peptides (DRPs) are a class of short peptides characterized by the presence of multiple disulfide bonds that provide stability and protection from proteolysis (1–3). DRPs are used by various organisms as toxins, host defense peptides, and peptide hormones and commonly target ion channels and receptors with high potency and often exquisite subtype selectivity. They are an appealing therapeutic modality because of the intrinsic druglike properties including stability, favorable pharmacodynamics and established bioactivity.

Structural characterization of these peptides can be challenging as they are often recalcitrant to crystallization, too small for cryo-electron microscopy, and disulfide connectivity is difficult to determine by NMR. Mass spectrometry-based methods can be used to determine native disulfide connectivity (2, 4, 5), but not a three dimensional structure. Molecular dynamics simulations can be used for ab initio folding of DRPs (6, 7), but even small peptides require computationally intensive timescales of simulations (micro- to milli-second) with current computing strategies. Developing methods for DRP structure prediction from sequence and MS-derived connectivity data would enable high-throughput models for biotechnology applications such as peptide-protein docking or protein engineering to improve properties such as stability or target affinity.

AlphaFold can also be used to predict cyclic peptide and DRP structure as was demonstrated recently (8–11). AlphaFold performs best on shorter peptides with high secondary structure, but also works well on DRPs. AlphaFold, however, can sometimes incorrectly predict the native disulfide connectivity of DRPs and thus provide incorrect structural models for the applications above.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

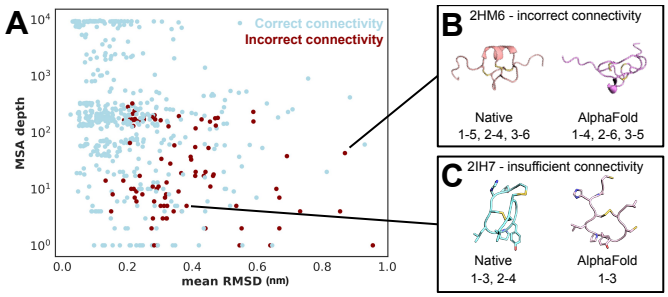


Figure 1: AlphaFold predicts non-native connectivities in DRPs. (A) In roughly 15% of our DRP dataset, AlphaFold predicts the incorrect connectivity. These incorrect predictions are either due to (**B**) different connectivity patterns or (**C**) insufficient connectivity. Connectivity patterns indicate which cysteines in the sequence form disulfides such that "1-3, 2-4" indicates that the first cysteine in the sequence forms a disulfide with the third, and the second with the fourth.

Modifying AlphaFold has been a successful strategy for numerous applications, such as predicting alternative conformations (12, 13), predicting structures of head-to-tail cyclized peptides (9), and using experimental constraints in protein structure prediction (14). Thus, we developed a method to incorporate disulfide connectivity data into AlphaFold structure predictions.

A key component of this work is the use of MSAs as input to AlphaFold, which provide data for learning epistasis caused by structural proximity, also known as protein sequence coevolution (15–17). Structure prediction methods that relied on statistical models of sequence coevolution to generate pairwise distance constraints, like EVFold (18) and GREMLIN (19), were precursors to the use of sequence information in deep learning protein structure prediction models (20–25), such as in the EvoFormer block of AlphaFold (11). In this work, we artificially generated sequence coevolution signals in the input MSA of AlphaFold to give the impression of coevolution between distinct pairs of cysteines in DRPs, guiding AlphaFold to predict the specified disulfide connectivity.

2 Results

2.1 AlphaFold predicts non-native connectivity patterns for DRPs

We first tested a set of 624 DRPs (**Supplemental Information**) with two or three disulfides to determine the accuracy of AlphaFold (**Fig. 1A**). AlphaFold performed well at predicting the structures based on RMSD in the vast majority of examples but in many cases disulfide connectivity patterns were predicted incorrectly (**Fig. 1B**) or insufficiently (meaning not all cysteines formed any disulfides, **Fig. 1C**). Of the 624 DRPs, 93 (15%) had incorrect connectivity predicted in at least 3 out of 5 predictions from AlphaFold. In many cases, the RMSD of non-native connectivity models is lower than examples with native connectivity, suggesting RMSD is an insufficient metric to evaluate folding accuracy of DRPs. In DRPs with two disulfides, 13 out of 47 incorrectly predicted structures had insufficient disulfide formation, while the remaining 35 had incorrect pairing. In DRPs with three disulfides, 12 out of 46 had insufficient disulfide formation, while 34 had incorrect pairing. These results comport with another study benchmarking AlphaFold on peptide structure prediction (8). MSA depth was a factor in the accuracy of the prediction (connectivity for all peptides with MSA depths larger than 283 was predicted accurately), suggesting novel peptides would present more challenges for accurate structure prediction with AlphaFold.

2.2 Artificial coevolution in MSAs can enforce disulfide connectivity

Sequence coevolution can be detected from MSAs and used to provide proximity information in AlphaFold. In the case of DRPs the cysteine residues are highly conserved in MSAs, which means they do not provide information on the pairwise connectivity of cysteines as all cysteines show equal coevolution with each other. To provide connectivity restraints in the form of sequence coevolution, the MSA can be altered such that pairs of cysteines in subsets of sequences were conserved while others were randomly mutated (**Fig. 2A-B**). Thus, one can provide experimental connectivity data from MS to constrain the AlphaFold output.

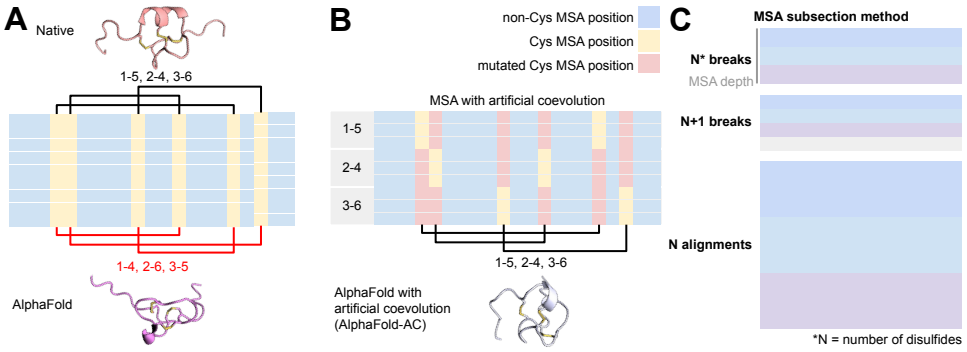


Figure 2: Schematic of artificial coevolution to direct disulfide connectivity in DRPs. (**A**) The conservation of cysteine residues (yellow) in MSAs of DRPs makes it challenging for AlphaFold to predict the native connectivity of DRPs as all pairs of cysteines show equal coevolution. (**B**) In contrast, by mutating residue positions in subsets of the MSA, artificial coevolution can be embedding in the MSA, providing data for AlphaFold to predict the native connectivity. (**C**) Schematic showing the different MSA subsection methods used. The MSA was either broken into N or N+1 subsections (where N is the number of disulfides), or the MSA was repeated N times (N alignments).

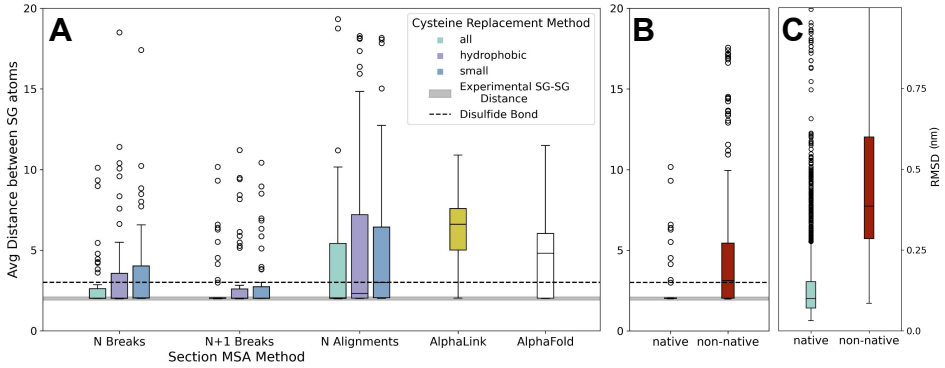


Figure 3: Artificial coevolution approaches improve RMSD and native connectivity of DRPs. (**A**) Different cysteine replacement methods and MSA sectioning methods show different extents of improvement in the distance between SG atoms of native cysteine pairs. This improves DRP connectivity prediction overall compared to AlphaFold or using the native cysteine pairs as constraints in AlphaLink (14). AlphaFold predictions with artificial coevolution of native disulfide connectivity show lower SG atom distance (**B**) and overall lower RMSD (**C**) compared to non-native connectivities.

Using a test dataset of 48 DRPs with two or three disulfides (**Supplemental Information**) enriched for examples where AlphaFold failed to predict the experimental connectivity (33 incorrect, 15 correct), we determined whether embedding connectivity using artificial coevolution (AlphaFold-AC) improves predictions. Several methods to embed the pairwise coevolutionary signals by subsetting (or multiplying) the MSA in different ways (**Fig. 2C**) and mutating the non-paired cysteines to different subsets of amino acids were tested. Breaking up the MSA into N+1 equally-sized subsets (where N is the number of disulfides) and mutating the non-paired cysteines to any amino acid as opposed to just hydrophobic or just small amino acids was the best method for embedding the desired connectivity into DRP structure predictions in our dataset (**Fig. 3**). This strategy greatly decreased the sulfur-sulfur distance of the native connectivity (SG-SG distance) to nearly the known disulfide bond distance of 2.05 Åand less than the disulfide bond cutoff in the PyMOL (3 Å) in the majority of cases. There were only 8 DRPs for which the mean SG-SG distance was over 4 Å, which all had MSAs with depths under 10 sequences, suggesting embedding artificial coevolution requires enough sequence diversity. In summary, we rectified the connectivity in 25 out of 33 AlphaFold predictions using our approach.

For a three-disulfide peptide, there are 15 potential connectivities, only one of which is native. To test whether this method could distinguish native vs. non-native connectivity, we predicted structures using the non-native disulfide connectivity as well (**Fig. 3B**). Even in non-native connectivities, these constraints resulted in decreased SG-SG average distances, but were still significantly larger than native connectivities, suggesting this approach may be helpful in determining the native connectivity of a DRP.

2.3 Comparison to other methodology

We also explored other approaches to providing connectivity constraints to AlphaFold, namely AlphaLink (14) and relative positional encoding in ColabDesign (9, 26, 27). AlphaLink requires restraints provided as mean±standard deviation for the residue-residue C-α distance. Using a distance of 5.6±0.7 Å for the native connectivity had minimal, if not negative, impact of the desired disulfide bond distance for our test set (**Fig. 3A**). This model was trained on synthetic crosslinking data meant for larger distances than disulfides which may explain why this approach failed, though could be optimized using a similar training scheme on disulfide bonds.

The relative positional encoding of DRPs followed the framework of cyclic peptides, where a relative position of 1 was used between cysteines involved in a particular disulfide pair (**Fig. S1A**). Then, the Floyd algorithm (28, 27) was used to fill in the remaining pairwise distances and then sign corrected for directionality. We also included a disulfide loss function described by (29). While this method positioned the correct cysteines in close proximity, the predicted structures contained severe clashes (**Fig. S1B**). Further optimization of the loss functions for DRP structure prediction in ColabDesign would hopefully resolve these issues, as recently suggested (27). It is important to note that these methods are based on cyclic peptides, regardless of the presence of additional disulfides, suggesting that relative position encoding is especially useful for these cases but is less so for non-cyclic DRPs.

2.4 Peptide properties may help distinguish native vs. non-native connectivities

Our method succeeds at positioning desired cysteines in near disulfide-bonding range, but sometimes fell outside the scope of real disulfide bond lengths. Thus, Rosetta's Disulfidize mover and FastRelax were used to form the desired disulfides and these models were subsequently used in Gaussian-accelerated Molecular Dynamics (GaMD) simulations to further minimize the predicted structures for 25 ns. We then compared peptide properties of predicted DRP structures at each stage of this process: modified AlphaFold, after disulfidization, after MD equilibration, and after 25 ns of GaMD simulation (**Fig. 4**). The radius of gyration did not differentiate native vs. non-native connectivities from AlphaFold, but interestingly, peptides with native connectivies had a lower radius of gyration on average after disulfidization, equilibration and simulation. Rosetta energy was lowest after disulfidization, most likely because the Rosetta energy score functions were used to do a FastRelax of the disulfidized peptide. In the other steps, Rosetta energy was lower in peptides with native connectivities. The mean surface-accessible surface area was noticeably decreased in the AlphaFold prediction, yet did not differentiate native and non-native connectivities in other steps. Lastly, the helical content seemed to be greater, while coil content and sheet content were lower, in the native connectivity. While overall these trends are modest, they suggest that peptide properties may be useful in determining the native connectivity of DRPs.

3 Discussion

In this study we used the principles of protein sequence coevolution to artificially alter MSAs at cysteine positions to constrain AlphaFold predictions of DRPs with defined disulfide connectivity. This approach, though simple, is quite effective at improving structure prediction and also illustrates that the MSA input to AlphaFold's EvoFormer blocks provide data to learn residue proximity information in a similar fashion to statistical models of sequence coevolution. In addition, providing the native disulfide connectivity pattern (as opposed to non-native connectivity) in our approach is more likely to result in a predicted structure that has the desired connectivity and in favorable peptide properties (Rosetta energy, radius of gyration, coil content) compared to non-native connectivity predictions. These results suggest that native connectivity might be effectively predicted by comparing predicted structures with all connectivities in a deep learning framework, though further work is required to demonstrate this.

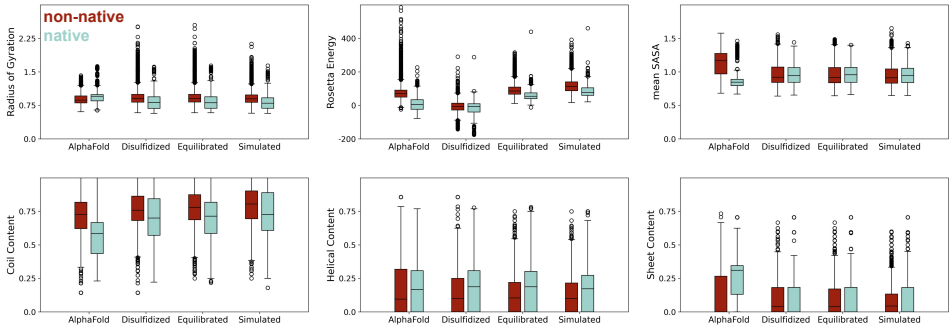


Figure 4: DRP structure predictions with native connectivity have improved peptide properties. Comparison of the DRP structure prediction properties (radius of gyration, Rosetta energy, mean SASA, and coil, helical and sheet content) at different steps of the pipeline: after AlphaFold-AC prediction **AlphaFold-AC**, after disulfidization with PyRosetta **Disulfidized**, after equilibration in OpenMM **Equilibrated**, and after a 25 ns GaMD production run **Simulated**.

We compared other methods for directing interactions in AlphaFold (9, 14, 27) to our approach. Artificial coevolution was straightforward to implement and successful at providing connectivity constraints for DRPs to AlphaFold. The other approaches may prove to be equally or more successful at this task if specifically oriented to this problem (27), though was not as successful in our hands with admittedly limited effort.

Others have successfully manipulated MSAs to modify AlphaFold predictions (12, 13). These efforts have focused on 'row-wise' manipulations of MSAs, while this study focuses on 'column-wise' manipulations that embed artificial coevolution directly. Our results demonstrate these manipulations can also successfully modify AlphaFold predictions in intended ways. The extent to which artificial coevolution can modify AlphaFold predictions should be explored further, though we believe more discretion should be placed on this approach than row-wise approaches, as it involves mutating real protein sequences, which could move MSA representations outside of the distributions of natural protein sequences that were used to train AlphaFold.

4 Methods

Dataset DRP structures were collected from RCSB based on having two or three disulfides in a peptide of less than 60 residues in length, resulting in a dataset of 624 DRPs. After running AlphaFold on this set of peptides, we created a smaller dataset of 48 DRPs enriched in peptides whose native connectivity was not predicted by AlphaFold for subsequent work. See **5.1** for list of PDB IDs.

Artificial coevolution in MSAs After running AlphaFold without any connectivity restraints, MSAs were randomly subset into either N, N+1 or multiplied N times (where N is the number of disulfides). In the case of N subsets, cysteines were maintained in each subset for one of the disulfides in the peptide, while the other positions were mutated. In the case of N+1 subsets, one subset was left as is without manipulation of the alignment. In the case of multiplying the alignment by N, each multiplication was treated as its own subset as in the N subsets case. Positions were randomly mutated to either any amino acid, only hydrophobic amino acids (A, C, F, I, L, M, V, W, Y) or small amino acids (A, C, L, S, T, V). Then, this MSA was used to predict structures in AlphaFold without templates following previous examples (12).

Post processing and simulations of predicted DRP structure After AlphaFold-AC prediction, we used the Disulfidize Mover in PyRosetta (30) to enforce the disulfide bonds, followed by a standard FastRelax protocol to improve peptide geometry. We then used OpenMM (31) to run 25 ns of Gaussian-accelerated Molecular Dynamics (GaMD) of each peptide according to established protocols (32, 33).

References

[1] Glenn F King. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opinion on Biological Therapy*, 11(11):1469–1484, 2011. ISSN 1471-2598. doi: 10.1517/14712598.2011.621940.

[2] Miriam Gongora-Benitez, Judit Tulla-Puche, and Fernando Albericio. Multifaceted Roles of Disulfide Bonds. Peptides as Therapeutics. *Chemical Reviews*, 114(2):901–926, 2014. ISSN 0009-2665. doi: 10.1021/cr400031z.

[3] Conan K. Wang and David J. Craik. Designing macrocyclic disulfide-rich peptides for biotechnological applications. *Nature Chemical Biology*, 14(5):417–427, 2018. ISSN 1552-4450. doi: 10.1038/s41589-018-0039-y.

[4] Jeffrey J. Gorman, Tristan P. Wallis, and James J. Pitt. Protein disulfide bond determination by mass spectrometry. *Mass Spectrometry Reviews*, 21(3):183–216, 2002. ISSN 0277-7037. doi: 10.1002/mas.10025.

[5] Jude C. Lakbub, Joshua T. Shipman, and Heather Desaire. Recent mass spectrometry-based techniques and considerations for disulfide bond characterization in proteins. *Analytical and Bioanalytical Chemistry*, 410(10):2467–2484, 2018. ISSN 1618-2642. doi: 10.1007/s00216-017-0772-1.

[6] Hao Geng, Fangfang Chen, Jing Ye, and Fan Jiang. Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. *Computational and Structural Biotechnology Journal*, 17:1162–1170, 2019. ISSN 2001-0370. doi: 10.1016/j.csbj.2019.07.010.

[7] Sergei A. Izmailov, Ivan S. Podkorytov, and Nikolai R. Skrynnikov. Simple MD-based model for oxidative folding of peptides and proteins. *Scientific Reports*, 7(1):9293, 2017. doi: 10.1038/s41598-017-09229-7.

[8] Eli Fritz McDonald, Taylor Jones, Lars Plate, Jens Meiler, and Alican Gulsevin. Benchmarking AlphaFold2 on peptide structure prediction. *Structure*, 31(1):111–119.e2, 2023. ISSN 0969-2126. doi: 10.1016/j.str.2022.11.012.

[9] Stephen A. Rettie, Katelyn V. Campbell, Asim K. Bera, Alex Kang, Simon Kozlov, Joshmyn De La Cruz, Victor Adebomi, Guangfeng Zhou, Frank DiMaio, Sergey Ovchinnikov, and Gaurav Bhardwaj. Cyclic peptide structure prediction and design using AlphaFold. *bioRxiv*, page 2023.02.25.529956, 2023. doi: 10.1101/2023.02.25.529956.

[10] Jun Zha, Jinjing Li, Shihui Fan, Zengping Duan, Yibing Zhao, and Chuanliu Wu. An evolution-inspired strategy to design disulfide-rich peptides tolerant to extensive sequence manipulation. *Chemical Science*, 12(34):11464–11472, 2021. ISSN 2041-6520. doi: 10.1039/d1sc02952e.

[11] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Andrew J Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. ISSN 0028-0836. doi: 10.1038/s41586-021-03819-2.

[12] Diego del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, 11:e75751, 2022. doi: 10.7554/elife.75751.

[13] Hannah K. Wayment-Steele, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Prediction of multiple conformational states by combining sequence clustering with AlphaFold2. *bioRxiv*, page 2022.10.17.512570, 2022. doi: 10.1101/2022.10.17.512570.

[14] Kolja Stahl, Andrea Graziadei, Therese Dau, Oliver Brock, and Juri Rappsilber. Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nature Biotechnology*, pages 1–10, 2023. ISSN 1087-0156. doi: 10.1038/s41587-023-01704-z.

[15] Tyler N. Starr and Joseph W. Thornton. Epistasis in protein evolution. *Protein Science*, 25(7): 1204–1218, 2016. ISSN 0961-8368. doi: 10.1002/pro.2897.

[16] Sergey Ovchinnikov, Hetunandan Kamisetty, and David Baker. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife*, 3:e02030, 2014. doi: 10.7554/elife.02030.

[17] Debora S. Marks, Lucy J. Colwell, Robert Sheridan, Thomas A. Hopf, Andrea Pagnani, Riccardo Zecchina, and Chris Sander. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS ONE*, 6(12):e28766, 2011. doi: 10.1371/journal.pone.0028766.

[18] Robert Sheridan, Robert J. Fieldhouse, Sikander Hayat, Yichao Sun, Yevgeniy Antipin, Li Yang, Thomas Hopf, Debora S. Marks, and Chris Sander. EVfold.org: Evolutionary Couplings and Protein 3D Structure Prediction. *bioRxiv*, page 021022, 2015. doi: 10.1101/021022.

[19] Hetunandan Kamisetty, Sergey Ovchinnikov, and David Baker. Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *Proceedings of the National Academy of Sciences*, 110(39):15674–15679, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1314045110.

[20] Sheng Wang, Siqi Sun, Zhen Li, Renyu Zhang, and Jinbo Xu. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLoS Computational Biology*, 13(1): e1005324, 2017. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1005324.

[21] Yang Liu, Perry Palmedo, Qing Ye, Bonnie Berger, and Jian Peng. Enhancing Evolutionary Couplings with Deep Convolutional Neural Networks. *Cell Systems*, 6(1):65–74.e3, 2018. ISSN 2405-4712. doi: 10.1016/j.cels.2017.11.014.

[22] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020. ISSN 0028-0836. doi: 10.1038/s41586-019-1923-7.

[23] Nicholas Bhattacharya, Neil Thomas, Roshan Rao, Justas Dauparas, Peter K Koo, David Baker, Yun S Song, and Sergey Ovchinnikov. Interpreting Potts and Transformer Protein Models Through the Lens of Simplified Attention. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 27:34–45, 2021.

[24] Roshan Rao, Jason Liu, Robert Verkuil, Joshua Meier, John F Canny, Pieter Abbeel, Tom Sercu, and Alexander Rives. MSA Transformer. *bioRxiv*, page 2021.02.12.430858, 2021. doi: 10.1101/2021.02.12.430858.

[25] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. ISSN 0036-8075. doi: 10.1126/science.ade2574.

[26] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6): 679–682, 2022. ISSN 1548-7091. doi: 10.1038/s41592-022-01488-1.

[27] Chenhao Zhang, Chengyun Zhang, Tianfeng Shang, Xinyi Wu, and Hongliang Duan. Highfold: accurately predicting cyclic peptide monomers and complexes with AlphaFold. *bioRxiv*, page 2023.08.27.554979, 2023. doi: 10.1101/2023.08.27.554979.

[28] Robert W. Floyd. Algorithm 97: Shortest path. *Communications of the ACM*, 5(6):345, 1962. ISSN 0001-0782. doi: 10.1145/367766.368168.

[29] Sergey Ovchinnikov @sokrypton. disulfide-hallucination in ColabDesign. *GitHub*, commit: 31a60b72688bcaeda756a0a0600f62970101dd0a, 2023. URL `https://github.com/sokrypton/ColabDesign/commits/main/af/examples/disulfide_design.ipynb`.

[30] Sidhartha Chaudhury, Sergey Lyskov, and Jeffrey J. Gray. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics*, 26(5):689–691, 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq007.

[31] Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, 13(7): e1005659, 2017. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1005659.

[32] Yinglong Miao, Victoria A. Feher, and J. Andrew McCammon. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation*, 11(8):3584–3595, 2015. ISSN 1549-9618. doi: 10.1021/acs.jctc.5b00436.

[33] Matthew M. Copeland, Hung N. Do, Lane Votapka, Keya Joshi, Jinan Wang, Rommie E. Amaro, and Yinglong Miao. Gaussian Accelerated Molecular Dynamics in OpenMM. *The Journal of Physical Chemistry B*, 126(31):5810–5820, 2022. ISSN 1520-6106. doi: 10.1021/acs.jpcb.2c03765.

5 Supplemental Information

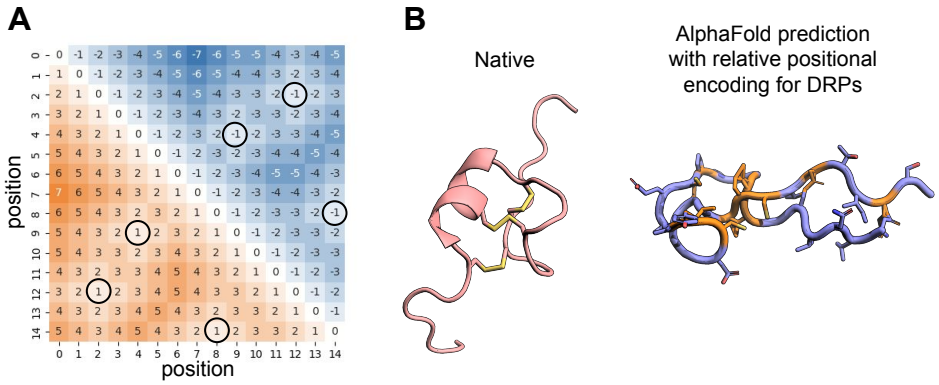
5.1 Dataset

PDB IDs and chain ID of 624 DRPs with 4 or 6 cysteines. DRPs that were incorrectly predicted by AlphaFold are highlighted in red. DRPs in the smaller dataset of 48 DRPs are underlined.

2B5P.A	6PI2.A	5UJG.A	5WCV.A	1EMX.A	4E86.A	2M50.A	2PLZ.A	1BNB.A	1B8W.A	2M86.A	2M99.A
2IH6.A	6PI3.A	2LDF.A	1WM7.A	2KHT.A	1HP3.A	1NIY.A	2NLS.A	1XSW.A	1H5O.A	1SGY.I	5M4V.A
2IH7.A	1WO0.A	6E1L.A	6DHR.A	6CHC.A	1G1P.A	1LMR.A	6BUC.A	5XA6.A	3FP7.J	1SGP.I	1ZJD.B
2LER.A	6QKF.A	1HVW.A	2LAM.A	3LO6.A	1EWS.A	1ROO.A	2I1T.A	2AXK.A	1BDS.A	1R0R.I	3BTD.I
2IFI.A	2M2Y.A	1TT3.A	1CTI.A	2KUX.A	2K1I.A	5OLL.A	1E4S.A	2LR5.A	2LN4.A	1SGD.I	1T8N.B
1G2G.A	2M79.A	2KM9.A	1ACW.A	1I8X.A	1NIX.A	2LZY.A	2NLB.A	1SCO.A	2E3F.A	2NU1.I	1KNT.A
1IM1.A	1PG1.A	1FYG.A	2MW0.A	2MN1.A	1C6W.A	1D1H.A	1PMC.A	1EI0.A	1P0A.A	1CSO.I	1TFX.C
2IFJ.A	1K64.A	1P1P.A	2LUR.A	1PQR.A	6MM4.A	2M4Z.A	1H9I.I	1SXM.A	1I2U.A	1HIC.A	1AAP.A
2M62.A	2M1P.A	2FQC.A	2K7G.A	1ZUV.A	6BTV.A	6AV8.A	2NLC.A	6ATN.A	2MN3.A	1CT2.I	1BTI.A
5T6T.A	1HVZ.A	2LDE.A	2F2I.A	2HM6.A	1R02.A	5T4R.A	2XTT.A	2K2Y.A	1I2V.A	2SGQ.I	3BTQ.I
1E74.A	2LYE.A	5ZNU.A	1LU0.A	1RMK.A	5EPM.C	5TLR.A	1CBH.A	2K2Z.A	1OZZ.A	1HJA.I	1FAN.A
1E75.A	2LZI.A	1MVJ.A	2BTC.I	1W7Z.A	6DMQ.A	2N1N.A	5I1X.A	5WOW.A	2LXO.A	1SGQ.I	1P2J.I
1E76.A	2M2G.A	1FEO.A	2RTZ.A	1PNH.A	2MXM.A	2E2S.A	2NLD.A	2MT8.A	1CIX.A	1SGN.I	2FTM.B
1HY2.H	2MD6.A	1F3K.A	2M9O.A	2LU9.A	2KY3.A	2K9E.A	2MPQ.A	6ATM.C	5JYH.A	2MD0.A	1BPT.A
2IFZ.A	1ZLC.A	1YZ2.A	6Q5Z.A	6D9O.A	2M3J.A	2N6O.A	2N8B.A	2HLG.A	2KMO.A	1SGE.I	1BRB.I
2G6U.A	2ATG.A	1CNN.A	1NB1.A	2GJ0.A	1ZMM.A	2F91.B	6ATL.A	1MTX.A	1P00.A	2NU2.I	1P2Q.B
2JUR.A	1T7H.A	1WQC.A	2F2J.A	1ZA8.A	2M7T.A	1TSK.A	2A9H.E	1HLY.A	1KJ6.A	2GKT.I	2FI3.I
1R8T.A	1KFP.A	1V4Q.A	3E4H.A	6D9P.A	2JTB.A	1RYG.A	1KTX.A	2H1Z.A	1G9P.A	3C9A.D	3P92.E
2JUS.A	6PIN.A	1AV3.A	6CGW.A	6D8Q.A	2MEL.A	1KGM.A	2L2R.A	1H20.A	2E3E.A	3L33.E	1B0C.A
1GNA.A	6PIO.A	1FU3.A	4TTN.A	1G26.A	6BR0.A	2MT7.A	2LZX.A	6AVC.A	1ED0.A	1IY5.A	4TPI.I
1HQQ.H	6PIP.A	5GSF.A	3LVX.A	6D8Y.A	2ME7.A	2N8H.A	2K4U.A	6AY7.A	1AB1.A	3T62.D	1G6X.A
2JUT.A	2NC7.A	1TR6.A	2PM4.A	6D8R.A	2VU8.I	2M9L.A	2MW7.A	2NAW.A	1ORL.A	1CA0.D	7PTI.A
1NOT.A	2M77.A	1IXT.A	4LBB.A	2CK5.A	1WSO.A	1E4R.A	2BMT.A	2NZ3.A	1JMN.A	1KMA.A	1T8O.B
2I28.A	2M78.A	2FQA.A	1NBJ.A	6D8S.A	1QK6.A	1C49.A	1LIR.A	6AY8.A	3C8P.A	2JOT.A	1T8M.B
2EW4.A	2MZ6.A	1OMC.A	6D3T.A	1CMR.A	2N2G.A	1RYV.A	4AOQ.D	1ZFU.A	2V9B.A	1FAK.I	3P95.E
1HXL.C	1MTQ.A	2N5Q.A	2MM5.A	1SCY.A	6GFT.A	1KIO.A	2MLA.A	2LG4.A	1CRN.A	4WXV.C	1P2K.I
2JUQ.A	1IEN.A	2NX6.A	2MM6.A	6D8T.A	2NAJ.A	1WO9.A	2MCR.A	2NY8.X	1JMP.A	2KCN.A	1LD5.A
1HJE.A	5UG3.A	1WQD.A	3LOE.A	2HM3.A	1IE6.A	5T3M.A	1BKT.A	6K50.A	1CCM.A	2NB0.A	1EJM.B
2M61.A	5UG5.A	1KCP.A	3H6C.B	6D93.A	2MH1.A	1C4E.A	1E4Q.A	6DRI.A	3UE7.A	4U32.X	1LD6.A
3ZKT.A	5XIV.A	1D5Q.A	1DF6.A	2ERI.A	1YP8.A	1FWO.A	1BIG.A	6QJB.A	1ATX.A	1SHP.A	1NAG.A
2LU6.1	6IGK.B	6EFE.A	6CEG.A	2HM4.A	6DMM.A	1GL0.I	2MLD.A	6K51.A	1AN1.I	3UOU.B	1T7C.B
1B45.A	1EDN.A	1MCT.I	3LO1.A	1Q2K.A	2KTC.A	1WZ5.A	6GGZ.1	2LLD.A	3UE7.B	2F3C.I	1AAL.A
1DG2.A	1SRB.A	5WXE.A	4LB7.D	2B38.A	6MK5.A	1MB6.A	3S3X.D	1LMM.A	2H9X.A	1CGJ.I	4Y0Z.I
2M3I.A	1r9i.A	1DU9.A	4LBF.A	2GWP.A	6KLM.A	2NLE.A	1AXH.A	2M5X.A	5LCS.A	2OVO.A	1JV8.A
6BX9.A	2YEN.A	1WM8.A	3LO2.A	1T0W.A	2N8E.A	1MM0.A	1M2S.A	1L4V.A	3NIR.A	2ERW.A	3BTM.I
1A0M.A	2AJW.A	1F2S.I	4DU0.A	2N3P.A	1TYK.A	1AZJ.A	1BGK.A	1ICA.A	1SH1.A	1OVO.A	1P2I.I
2NAY.A	1TCG.A	2NX7.A	1DFN.A	2N7F.A	5WE3.A	2NLF.A	1R1F.A	2L1Q.A	1Y1B.A	2NC2.A	1T8L.B
1M2C.A	1TCH.A	1MCV.I	2MSO.A	5FZV.A	1LA4.A	2RU0.A	1HP2.A	2E3G.A	1SHI.A	5NX3.D	3BYB.A
1UYA.A	1G9I.I	2NAV.A	1VB8.A	6MJV.A	5WOV.A	1AZK.A	1E4T.A	1HY9.A	1Y1C.A	1CHO.I	1DTX.A
2LXG.A	1SMF.I	2KVX.A	4LB1.D	3QTE.A	1AG7.A	6CKD.A	2K72.A	2MJK.A	1CBN.A	1UUA.A	2M01.A
1UL2.A	4TTL.A	2LET.A	1BH4.A	4RBW.A	2MXO.A	1AZ6.A	1P8B.A	2KNI.A	5IPO.A	1UUB.A	4U30.X
2EFZ.A	1Q2J.A	2ETI.A	2MI9.A	1H9H.I	2PO8.A	2NLG.A	1BAH.A	1FD3.A	4OZK.A	1BRC.I	
6J8E.D	6MJD.A	1I8Y.A	1EYO.A	3I5W.A	1S6X.A	5X39.A	1JU8.A	2NY9.X	1AHL.A	1CGI.I	
2MUH.A	6CGX.A	1PT4.A	3LO4.A	4RBX.A	1ZJQ.A	1J5J.A	2LQA.A	2N9Z.A	1APF.A	1TGS.I	
2H8S.A	2AK0.A	1ZMH.A	4BFH.A	1PJV.A	2A2V.A	5I2P.A	4AOR.D	1WQK.A	2CQ7.A	5NX1.D	
1AKG.A	5UJH.A	4TTO.A	2KUK.A	1TV0.A	2MQF.A	2NLH.A	6AU7.A	2LT8.A	2LWL.A	1M8B.A	
2LZ5.A	1WQE.A	1N1U.A	2RTY.A	5CUI.A	2MXQ.A	2LG5.A	3TVJ.I	2DCV.A	2N71.A	1YKT.B	
2MDQ.A	6NUG.A	2IT8.A	2N1S.A	1ZMP.A	2KIR.A	1KOZ.A	1UT3.A	2DCW.A	1TPM.A	1DTK.A	
1PEN.A	1ORX.A	2GX1.A	5ZV6.A	6D8H.A	1I26.A	2NLP.A	4DJZ.H	2JR3.A	1CT0.I	3D65.I	
1EDP.A	1JLP.A	2KHB.A	2KCG.A	1QK7.A	1HA9.A	2LG6.A	2CK4.A	1WXN.A	2SGF.I	1BUS.A	
2GCZ.A	2LO9.A	1KAL.A	1MMC.A	5CUJ.A	2PTA.A	5WLX.A	1AGT.A	2LMZ.A	2NU0.I	3L3T.E	
1MA2.A	2LOC.A	2JWM.A	2LJS.A	1ZMQ.A	4Z7P.A	2NLQ.A	5JBT.Y	3PIS.D	1CT4.I	1E0F.I	
6AZA.A	1AS5.A	1JJZ.A	4B1Q.P	1CLV.I	1C2U.A	2WH9.A	4BMF.A	1D6B.A	2SGP.I	5OQS.A	

5.2 Relative positional encoding in ColabDesign

For relative positional encoding of DRPs, residue positions pairs linked through a disulfide were annotated as having a distance of 1. The remainder of the relative positional encoding were filled out using the Floyd algorithm.



Supplemental Figure 1: Relative positional encoding of DRPs in ColabDesign. (**A**) Example of a relative positional encoding matrix of a DRP with 1-5, 2-4, 3-6 connectivity. Disulfide linked positions are circled. (**B**) Comparison of a DRP structure and the output from using the relative positional encoding for DRPs in the ColabDesign framework. The predicted structure includes many steric clashes (orange), lacks the disulfide linkages and does not form the known secondary structure elements.