

Protein structure prediction in the era of AI: challenges and limitations when applying to *in-silico* force spectroscopy

Priscila S. F. C. Gomes¹, Diego E. B. Gomes¹, Rafael C. Bernardi^{1*}

¹Department of Physics, College of Sciences and Mathematics, Auburn University, Auburn, AL, 36849

* Correspondence:

Corresponding Author
rcbernardi@auburn.edu

Keywords: artificial intelligence, protein folding, adhesins, staph infection, molecular dynamics, force spectroscopy. (Min.5-Max. 8)

Abstract

Mechanoactive proteins are essential for a myriad of physiological and pathological processes. Guided by the advances in single-molecule force spectroscopy (SMFS), we have reached a molecular-level understanding of how several mechanoactive proteins respond to mechanical forces. However, even SMFS has its limitations, including the lack of detailed structural information during force-loading experiments. That is where molecular dynamics (MD) methods shine, bringing atomistic details with femtosecond time-resolution. However, MD heavily relies on the availability of high-resolution structures, which is not available for most proteins. For instance, the Protein Data Bank currently has 192K structures deposited, against 231M protein sequences available on Uniprot. But many are betting that this gap might become much smaller soon. Over the past year, the AI-based AlphaFold created a buzz on the structural biology field by being able to, for the first time, predict near-native protein folds from their sequences. For some, AlphaFold is causing the merge of structural biology with bioinformatics. In this perspective, using an *in silico* SMFS approach, we investigate how reliable AlphaFold structure predictions are to investigate mechanical properties of staph bacteria adhesins proteins. Our results show that AlphaFold produce extremally reliable protein folds, but in many cases is unable to predict high-resolution protein complexes accurately. Nonetheless, the results show that AlphaFold can revolutionize the investigation of these proteins, particularly by allowing high-throughput scanning of protein structures. Meanwhile, we show that the AlphaFold results need to be validated and should not be employed blindly, with the risk of obtaining an erroneous protein mechanism.

1 Introduction

Over the past year, the artificial intelligence (AI)-based software AlphaFold created a buzz on the structural biology field. For the first time, a software was able to predict near-native protein folds from their genetic sequence (Jumper et al., 2021b). DeepMind's AlphaFold transformed, in principle, the protein structure solving problem that has been around for the past 50 years into a trivial task. The number of research papers and preprints citing the method soared since its code was released in July 2021 (Callaway, 2022), with the accompanying article achieving about 1,000 citations (according to Google Scholar) in its first year.

The success of AlphaFold, and the analog RoseTTAFold approach (Baek et al., 2021) that appeared a few months later, is partially due to their open-source nature, which makes them readily and freely available to anyone who is interested in trying these software. Furthermore, by pairing it with the European Bioinformatics Institute (EBI), AlphaFold has taken structural biology to the next level, allowing big consortiums to perform protein structure prediction to entire genomes, including human, mouse, *Saccharomyces* and *E. coli* (Tunyasuvunakool et al., 2021). The resulting structures were made available on a database maintained by the EBI, containing almost a million structures: <https://alphafold.ebi.ac.uk>.

The broad spread use of AI-based structure prediction leads us to ask the question: How reliable are the structures predicted by such models? Despite the growing number of success stories (Jumper et al., 2021a, 2021b; Mosalaganti et al., 2021; Skolnick et al., 2021; Hartmann et al., 2022; Varadi et al., 2022), researchers are accumulating evidence showing that AI-based structure prediction methods are still not perfect (Perrakis and Sixma, 2021; Outeiral et al., 2022), and that there is ample room for improvement. In other words, some results suggest that both AlphaFold and RoseTTAFold are qualitatively great, but in many cases, they lack the level of details that is important to understand a protein function (Akdel et al., 2021; Eisenstein, 2021; Callaway, 2022).

High-resolution protein structures are also crucial for drug-discovery. The ability to readily access the structure of any protein of the human genome is very attractive to those developing new drug compounds. Using an AI-based tool to predict how drugs bind to these proteins is an even larger challenge that will probably not be overcome soon due to the limited publicly available data for small molecule binding (Mullard, 2021). In addition to that, AlphaFold lacks the precision to predict structural changes in consequence of mutations (Buel and Walters, 2022).

Working as a “computational microscope” molecular dynamics (MD) simulations are a unique tool to investigate biomolecules’ behavior with atomic resolution (Lee et al., 2009; Dror et al., 2012; Perilla et al., 2015). However, as most computational chemistry methods, the quality of MD results relies heavily on the quality of the initial biomolecule structure (Bernardi and Pascutti, 2012; Heo and Feig, 2018; Melo et al., 2018). If AI-based structure prediction software are able to predict protein folds to the atomic level, MD simulations should be able to profit from these structures and give similar results to those obtained with experimentally determined structures.

A particularly powerful way of using MD simulations is by using it hand-in-hand with experimental methods. In the past few years, taking advantage of steered MD protocols, our group has pioneered what we call *in silico* single-molecule force spectroscopy (*in silico* SMFS) (Bernardi et al., 2019; Sedlak et al., 2019, 2020). In this technique, steered MD (SMD) simulations are used in a wide-sampling approach to perform dozens to thousands of “pulling experiments”, in an analogy to what is done experimentally using atomic force microscopes (AFM). Allied to AFM-based SMFS, SMD has been successfully used to investigate a myriad of mechanically relevant biomolecular systems, including avidin:biotin (Grubmüller et al., 1996; Izrailev et al., 1997; Merkel et al., 1999), titin (Gao et al., 2002), human fibronectin (Gao et al., 2002), aquaporins (de Groot et al., 2009), among others.

The development of an *in silico* SMFS methodology, allowed us to go even further and to fine-tune mechanical properties of protein folds (Verdorfer et al., 2017). Besides protein design, our methodology allowed us to discover ultrastable protein complexes, and to decipher their intricate mechanostability mechanisms (Schoeler et al., 2014; Bernardi et al., 2019; Liu et al., 2020). Among these ultrastable protein complexes, the ones formed by *Staphylococci* bacteria when adhering to humans are particularly interesting (Herman-Bausier and Dufrêne, 2018). These bacteria adhere to their hosts through proteins called adhesins (Dufrêne and Viljoen 2020). Adhesins play critical roles

during infection, especially during the early step of adhesion when cells are exposed to mechanical stress. During the first steps of staph infection, the interactions between adhesins and proteins of the human extracellular matrix are a key virulence factor for these bacteria (Otto, 2008), and a crucial first step of biofilm formation (Latasa et al., 2006). These staph biofilms are associated with more than half of all nosocomial infections (Jamal et al., 2018), with *Staphylococcus epidermidis* and *S. aureus* listed as the most common pathogens (Otto, 2008; Schilcher and Horswill, 2020).

To demonstrate the advantages and limitations of AI-based protein structure prediction methods, in this perspective article we used AlphaFold to predict the structures of several *S. aureus* adhesins from the adhesion domain superfamily. First, a bioinformatics analysis was performed to select a diverse set of adhesin sequences of different *S. aureus* strains that were then used as input for AlphaFold, when structural models were generated. Then, we employed our *in silico* SMFS methodology to characterize the mechanical properties of these proteins, comparing the results to those obtained with traditional structure biology methods.

2 Application: Adhesin folding domains

2.1 How good is AlphaFold to model full length adhesins?

After selecting 48 *S. aureus* adhesins from the adhesion superfamily and we used AlphaFold 2 through the VMD's (Humphrey et al., 1996) QwikFold plugin (Gomes et al., 2022) batch mode to construct the models for 48 full length apo adhesin protein models. Overall, AlphaFold 2 consistently predicted the canonical folds for N2 and N3 domains for all proteins and the homologous B repeats according to each protein domain organization (Foster and Höök, 1998; Ganesh et al., 2011; Foster et al., 2013) (Fig. S1 and Table S1). Interestingly, for the collagen binding adhesin (Can), AlphaFold 2 predicted 4 additional B domains instead of the 3 described on the protein fold organization. As expected, domains such N1, the serine aspartate or fibronectin binding repeats, as well as signal sequences, were predicted as disordered.

An example of an AlphaFold prediction for the serine-aspartate repeat-containing protein E (SdrE) is shown at Figure 1. The software predicted the Ig-like N2 and N3 domains in addition to B1, B2 and B3 repeats (Fig. 1A). The N and C-ter regions normally comprise disordered regions, such as peptide signals and the SD repeats, in the case of the serine aspartate repeat proteins (Fig. 1A). A comparison between the crystal structure for SdrE (PDB ID:5WTA) containing the N2 and N3 domains and the model revealed a root mean square deviation (RMSD) of 1.31 Å for the same region (Fig. 1B), indicating that the model is a good approximation for the crystallographic structure of the Ig-like domains. This was expected since the SdrE crystal structure, among other crystals for adhesins and similar folds, were present among the structures present on AlphaFold's training set.

The per-residue model quality can be evaluated by pLDDT scores. In our studies, the pLDDT scores varied from ~20 to 90 (Fig 1C) ranging from the disordered to folded regions of the proteins, which were predicted with high-quality. The confidence of the prediction can be accessed through the prediction alignment error (PAE) plots, which indicates the expected distance error in Angstrom (Fig. 1D). PAE shows low error values for the N2, N3 (big blue square) and the B domains (three small squares), corroborating the pLDDT scores for the same region and indicating high-confidence for the prediction of the mentioned domains.

2.2 Is AlphaFold Multimer reliable for *in-silico* force spectroscopy experiments?

Staphylococcal adhesins use a conserved “dock, lock, and latch” (DLL) mechanism—in which the host target, usually a peptide on the order of 15 residues, is first bound (dock), then buried (lock) between two immunoglobulin-like (Ig) fold domains N2 and N3 (Ponnuraj et al., 2003), and finally a “latch” connects N3 back to N2 holding the complex in place (Fig 2A). Small conformational changes on the Ig-like N2 and N3 domains could potentially impact force resilience when complexed to peptides if the DLL configuration is lost. Similar to the DLL mechanism, multiple biological phenomena rely on specific protein:protein interactions. Leveraging the initial protein structure prediction model, AlphaFold Multimer (Evans et al., 2022) was developed to predict structures of protein complexes for computational studies.

Here, we tested the reliability of *in silico* SMFS experiments performed with protein structures predicted by AlphaFold Multimer. To this end, we selected 29 adhesin sequences to be modelled in complex with extracellular matrix peptides (Table S2). AlphaFold Multimer was used to construct models for the complexes through the QwikFold (Gomes et al., 2022) interface. Models were ranked by the predicted interface TM (ipTM) scores, used by AlphaFold Multimer, and the best ranked model for each complex was selected for SMD simulations, carried out using NAMD 3 (Phillips et al., 2020), where the adhesins were C-terminal anchored while the peptides were pulled at a constant speed for which we measure the forces upon the dissociation of the complex. Details and parameters are described at the Supplementary Information session. As control experiments, we also initiated SMD simulations using *S. aureus* crystallographic structures of three adhesin:peptide complexes: bone sialoprotein binding protein (BBP), clumping factor A (ClfA) and SdrE.

The predicted complexes were evaluated using pLDDT scores (Fig. 2B). Most of the protein display high quality (pLDDT > 80), with exception of a very small portion of the N-terminal (10 to 15 residues) and a significant region of the C-terminal (last 50 residues, figure 2B insert). The locking strand involved on the DLL mechanism is located on the C-terminal region of the protein structure, so this loss in model quality could impact the usability of the predicted structures in high-resolution experiments such as MD or SMD simulations.

By comparing the RMSD calculated on an equilibration MD versus the general AlphaFold Multimer scores for the best ranked structures are shown at Figure 2C. We noticed that there is a correlation (Pearson correlation of 0.82, $p < 0.005$) between the model stability and the AlphaFold Multimer scores. Therefore we can anticipate that high scored structures present less deviation from its initial configuration, suggesting a more stable or resilient fold. AlphaFold Multimer scoring is based on an interface predicted template modelling (ipTM) score that takes into account protein-protein interactions. It was shown to be more advantageous over the pTM and pLDDT scores used in AlphaFold 2 (Gao et al., 2022). The best ranked models on this case are a good indicator of model confidence based on the RMSD values.

After performing *in silico* SFMS experiments on all 48 complexes, we observed that the peak force profiles ranged from ~600 to 4000 pN, a much broader range than previously simulated SdrE, BBP and ClfA complexes initiated from crystal structures (Fig. 2D). *S. aureus* adhesins are extremely mechanoresistant with rupture forces consistently on the 2000 pN regime (Milles et al., 2018), values which we also reproduced on this study maintaining the same *in silico* SFMS protocol used for all complexes. Considering the drastic difference in rupture forces, we found that the very low values (600-1000 pN) seen for some of the complexes might have come from inaccurate initial structures. Visual inspection of the models with low rupture forces revealed that in most cases the locking strand was modelled in an unfavorable conformation to hold the peptide in the DLL configuration, which explains the observed behavior (data not shown).

To test this hypothesis, we re-modelled those complexes using comparative modelling with Modeller (Eswar et al., 2008) (Table S2). The models were inspected for the presence of the locking strand and simulated according to the same protocol described above (peak force profiles are shown in Figure 2D). For all cases we recover the force resilience, with peaks reaching 2000 to 3000 pN range, confirming that a high-resolution initial structure is necessary to be used for MD and SMD simulations.

3 Discussion

Protein structure prediction has been one of the grand challenges in Biology since the 1950's (Dill et al., 2008; Dill and MacCallum, 2012). Several methods have been developed over the past 40 years that span from comparative modeling with the increase of experimentally determined protein structures by X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and cryo-electron microscopy (cryo-EM) (Goh et al., 2016), but little progress was seen on *ab-initio* methodologies that rely solely on the protein sequence. But all of that changed upon the release of AlphaFold in 2021. Although AlphaFold requires only the protein sequence as input, it should not be considered an *ab-initio* method since it is built on 50 years of knowledge of protein structure determination by experimental methods. AlphaFold tremendous success took advantage of both the recent explosion of AI methods, as well as the huge database of protein structure offered by the protein data bank (PDB) (Berman et al., 2000).

However, as nearly any other AI-based tool, AlphaFold is biased towards its training set, meaning that the search for unusual folds is unlikely to provide an accurate result. Despite the software's success on the folded part of most proteins, AlphaFold lacks accuracy for regions where fewer sequences are available for alignment and intrinsically disordered regions, the latter are about one third of the human proteome, present in all proteomes of all kingdoms of life, and of all viral proteomes analyzed so far (Xue et al., 2012; Peng et al., 2015). It also struggles with protein interfaces in homo or hetero-multimers (Evans et al., 2022) and other aspects of protein structures such as co-factors, post-translational modifications and DNA or RNA complexes.

In order to show how revolutionary AlphaFold is for the single-molecule biophysics community, here we put AlphaFold to the test by using it to model full length staph adhesins and estimate how stable are the protein structures. Ignoring the disordered regions, AlphaFold was able to model the Ig-like domains of adhesins as well as other key structural features of these proteins, such as the homologous B domains, for all the tested sequences. With a little refinement from *in-equilibrium* MD simulations, the generated structures could help to investigate the properties of many of the domains that still have an unknown function.

We also tested the newly developed AlphaFold Multimer to model adhesin:peptide complexes from different strains of *S. aureus* involved in biofilm formation. By comparing the force profile obtained from crystallographic structures of the complexes, we showed that AlphaFold Multimer failed to predict important key structural motifs for some of the protein complexes. Particularly, the locking strand of the adhesins, which are essential for interacting and locking the human target peptide in a tight complex with the N2 and N3 domains. It is still unclear why the predicted models worked for some cases and not for others. Limiting the set of templates to the ones where we know that the correct structures are present did not help to improve the results (data not shown). This highlights that its Multimer mode is not yet suitable to be blindly used as a peptide docking approach and the generated models should pass through a manual inspection to be suited for MD simulations.

In summary, AlphaFold is a truly revolutionary tool that is bringing a new level of structural biology to bioinformatics. Although there are many areas where its methodology can be improved, the current

algorithm can be clearly employed to work alongside single-molecule biophysics experiments. It is important to note that, as any other scientific tool, particularly new ones, AlphaFold results cannot be employed blindly. Assessing the quality of the results and the usability of the predicted structures to infer function or mechanism to proteins is still the work of a trained scientist that can bring together data from multiple sources in a careful analysis of protein structure and dynamics.

4 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

5 Author Contributions

PSFG contributed to performing all the simulations, analysing data and writing of the manuscript. DEBG contributed to analysing data and discussion on AI-based methods. RCB coordinated the project, contributed to writing and discussion on *in silico* force spectroscopy, proof-reading, manuscript revision and approval of the submitted version.

6 Funding

This work was supported by the National Science Foundation under Grant MCB-2143787 (CAREER: *In Silico* Single-Molecule Force Spectroscopy).

7 Acknowledgments

We thank Auburn University and the College of Sciences and Mathematics for the computational resources provided by Dr. Bernardi faculty startup funds. We thank Dr. Marcelo Melo for the fruitful discussions.

8 Data Availability Statement

The data that support the findings of this study are available from the corresponding author, Dr. Bernardi, upon reasonable request.

9 References

- Akdel, M., Pires, D. E. v., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., et al. (2021). A structural biology community assessment of AlphaFold 2 applications. *bioRxiv*, 2021.09.26.461876. doi: 10.1101/2021.09.26.461876.
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi: 10.1126/SCIENCE.ABJ8754.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235–242. doi: 10.1093/NAR/28.1.235.
- Bernardi, R. C., Durner, E., Schoeler, C., Malinowska, K. H., Carvalho, B. G., Bayer, E. A., et al. (2019). Mechanisms of Nanonewton Mechanostability in a Protein Complex Revealed by Molecular Dynamics Simulations and Single-Molecule Force Spectroscopy. *J Am Chem Soc* 141, 14752–14763. doi: 10.1021/JACS.9B06776.
- Bernardi, R. C., and Pascutti, P. G. (2012). Hybrid QM/MM Molecular Dynamics Study of Benzocaine in a Membrane Environment: How Does a Quantum Mechanical Treatment of Both Anesthetic and Lipids Affect Their Interaction. *Journal of Chemical Theory and Computation* 8, 2197–2203. doi: 10.1021/ct300213u.
- Buel, G. R., and Walters, K. J. (2022). Can AlphaFold2 predict the impact of missense mutations on structure? *Nat Struct Mol Biol* 29. doi: 10.1038/S41594-021-00714-2.
- Callaway, E. (2022). What’s next for AlphaFold and the AI protein-folding revolution. *Nature* 604, 234–238. doi: 10.1038/D41586-022-00997-5.
- de Groot, B. L., Hub, J. S., and Grubmüller, H. (2009). Dynamics and energetics of permeation through aquaporins. What Do we learn from molecular dynamics simulations? *Handbook of Experimental Pharmacology* 190, 57–76. doi: 10.1007/978-3-540-79885-9_3.
- Dill, K. A., and MacCallum, J. L. (2012). The protein-folding problem, 50 years on. *Science* 338, 1042–1046. doi: 10.1126/SCIENCE.1219021.
- Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008). The Protein Folding Problem. *Annu Rev Biophys* 37, 289. doi: 10.1146/ANNUREV.BIOPHYS.37.092707.153558.
- Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., and Shaw, D. E. (2012). Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys* 41, 429–452. doi: 10.1146/ANNUREV-BIOPHYS-042910-155245.
- Dufrêne, Y. F., and Viljoen, A. (2020). Binding Strength of Gram-Positive Bacterial Adhesins. *Frontiers in Microbiology* 11, 1457. doi: 10.3389/FMICB.2020.01457.
- Eisenstein, M. (2021). Artificial intelligence powers protein-folding predictions. *Nature* 599, 706–708. doi: 10.1038/D41586-021-03499-Y.

- Eswar, N., Eramian, D., Webb, B., Shen, M.-Y., and Sali, A. (2008). Protein structure modeling with MODELLER. *Methods Mol Biol* 426, 145–59. doi: 10.1007/978-1-60327-058-8_8.
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. (2022). Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.10.04.463034. doi: 10.1101/2021.10.04.463034.
- Foster, T. J., Geoghegan, J. A., Ganesh, V. K., and Höök, M. (2013). Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nature Reviews Microbiology* 12:1 12, 49–62. doi: 10.1038/nrmicro3161.
- Foster, T. J., and Höök, M. (1998). Surface protein adhesins of *Staphylococcus aureus*. *Trends in Microbiology* 6, 484–488. doi: 10.1016/S0966-842X(98)01400-0.
- Ganesh, V. K., Barbu, E. M., Deivanayagam, C. C. S., Le, B., Anderson, A. S., Matsuka, Y. v., et al. (2011). Structural and biochemical characterization of *Staphylococcus aureus* clumping factor B/ligand interactions. *Journal of Biological Chemistry* 286, 25963–25972. doi: 10.1074/JBC.M110.217414.
- Gao, M., Craig, D., Vogel, V., and Schulten, K. (2002). Identifying Unfolding Intermediates of FN-III10 by Steered Molecular Dynamics. *Journal of Molecular Biology* 323, 939–950. doi: 10.1016/S0022-2836(02)01001-X.
- Gao, M., Nakajima An, D., Parks, J. M., and Skolnick, J. (2022). AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nature Communications* 2022 13:1 13, 1–13. doi: 10.1038/s41467-022-29394-2.
- Goh, B. C., Hadden, J. A., Bernardi, R. C., Singharoy, A., McGreevy, R., Rudack, T., et al. (2016). Computational Methodologies for Real-Space Structural Refinement of Large Macromolecular Complexes. <http://dx.doi.org/10.1146/annurev-biophys-062215-011113> 45, 253–278. doi: 10.1146/ANNUREV-BIOPHYS-062215-011113.
- Gomes, D. E. B., da Silva Figueiredo Celestino Gomes, P., and C Bernardi, R. (2022). QwikMD 2.0: bridging the gap between sequence, structure, and protein function. *Biophysical Journal* 121, 132a.
- Grubmüller, H., Heymann, B., and Tavan, P. (1996). Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science (1979)* 271, 997–999. doi: 10.1126/SCIENCE.271.5251.997.
- Hartmann, S., Ling, M., Dreyer, L. S. A., Zipori, A., Finster, K., Grawe, S., et al. (2022). Structure and Protein-Protein Interactions of Ice Nucleation Proteins Drive Their Activity. *bioRxiv*, 2022.01.21.477219. doi: 10.1101/2022.01.21.477219.
- Heo, L., and Feig, M. (2018). Experimental accuracy in protein structure refinement via molecular dynamics simulations. *Proc Natl Acad Sci U S A* 115, 13276–13281. doi: 10.1073/PNAS.1811364115.
- Herman-Bausier, P., and Dufrêne, Y. F. (2018). Force matters in hospital-acquired infections. *Science (1979)* 359, 1464–1465. doi: 10.1126/SCIENCE.AAT3764.

- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5.
- Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y., and Schulten, K. (1997). Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical Journal* 72, 1568. doi: 10.1016/S0006-3495(97)78804-0.
- Jamal, M., Ahmad, W., Andleeb, S., Jalil, F., Imran, M., Nawaz, M. A., et al. (2018). Bacterial biofilm and associated infections. *J Chin Med Assoc* 81, 7–11. doi: 10.1016/J.JCMA.2017.07.012.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021a). Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics* 89, 1711–1721. doi: 10.1002/PROT.26257.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021b). Highly accurate protein structure prediction with AlphaFold. *Nature* 2021 596:7873 596, 583–589. doi: 10.1038/s41586-021-03819-2.
- Latasa, C., Solano, C., Penadés, J. R., and Lasa, I. (2006). Biofilm-associated proteins. *Comptes Rendus - Biologies* 329, 849–857. doi: 10.1016/j.crv.2006.07.008.
- Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G., and Schulten, K. (2009). Discovery through the computational microscope. *Structure* 17, 1295–1306. doi: 10.1016/J.STR.2009.09.001.
- Liu, Z., Liu, H., Vera, A. M., Bernardi, R. C., Tinnefeld, P., and Nash, M. A. (2020). High force catch bond mechanism of bacterial adhesion in the human gut. *Nature Communications* 2020 11:1 11, 1–12. doi: 10.1038/s41467-020-18063-x.
- Melo, M. C. R., Bernardi, R. C., Rudack, T., Scheurer, M., Riplinger, C., Phillips, J. C., et al. (2018). NAMD goes quantum: An integrative suite for hybrid simulations. *Nature Methods* 15, 351–354. doi: 10.1038/nmeth.4638.
- Merkel, R., Nassoy, P., Leung, A., Ritchie, K., and Evans, E. (1999). Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* 397, 50–53. doi: 10.1038/16219.
- Milles, L. F., Schulten, K., Gaub, H. E., and Bernardi, R. C. (2018). Molecular mechanism of extreme mechanostability in a pathogen adhesin. *Science (1979)* 359, 1527–1533. doi: 10.1126/SCIENCE.AAR2094.
- Mosalaganti, S., Obarska-Kosinska, A., Siggel, M., Turonova, B., Zimmerli, C. E., Buczak, K., et al. (2021). Artificial intelligence reveals nuclear pore complexity. *bioRxiv*, 2021.10.26.465776. doi: 10.1101/2021.10.26.465776.
- Mullard, A. (2021). What does AlphaFold mean for drug discovery? *Nat Rev Drug Discov* 20, 725–727. doi: 10.1038/D41573-021-00161-0.
- Otto, M. (2008). Staphylococcal biofilms. *Current Topics in Microbiology and Immunology* 322, 207–228. doi: 10.1007/978-3-540-75418-3_10.

- Outeiral, C., Nissley, D. A., and Deane, C. M. (2022). Current structure predictors are not learning the physics of protein folding. *Bioinformatics* 38, 1881–1887. doi: 10.1093/BIOINFORMATICS/BTAB881.
- Peng, Z., Yan, J., Fan, X., Mizianty, M. J., Xue, B., Wang, K., et al. (2015). Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* 72, 137–151. doi: 10.1007/S00018-014-1661-9.
- Perilla, J. R., Goh, B. C., Cassidy, C. K., Liu, B., Bernardi, R. C., Rudack, T., et al. (2015). Molecular dynamics simulations of large macromolecular complexes. *Curr Opin Struct Biol* 31, 64–74. doi: 10.1016/J.SBI.2015.03.007.
- Perrakis, A., and Sixma, T. K. (2021). AI revolutions in biology. *EMBO Rep* 22, e54046. doi: 10.15252/EMBR.202154046.
- Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. v, Bernardi, R. C., et al. (2020). Scalable molecular dynamics on CPU and GPU architectures with **NAMD**. *The Journal of Chemical Physics* 153, 44130. doi: 10.1063/5.0014475.
- Ponnuraj, K., Bowden, M. G., Davis, S., Gurusiddappa, S., Moore, D., Choe, D., et al. (2003). A “dock, lock, and latch” Structural Model for a Staphylococcal Adhesin Binding to Fibrinogen. *Cell* 115, 217–228. doi: 10.1016/S0092-8674(03)00809-2.
- Schilcher, K., and Horswill, A. R. (2020). Staphylococcal Biofilm Development: Structure, Regulation, and Treatment Strategies. *Microbiology and Molecular Biology Reviews* 84. doi: 10.1128/MMBR.00026-19.
- Schoeler, C., Malinowska, K. H., Bernardi, R. C., Milles, L. F., Jobst, M. A., Durner, E., et al. (2014). Ultrastable cellulosome-adhesion complex tightens under load. *Nature Communications* 2014 5:1 5, 1–8. doi: 10.1038/ncomms6635.
- Sedlak, S. M., Schendel, L. C., Gaub, H. E., and Bernardi, R. C. (2020). Streptavidin/biotin: Tethering geometry defines unbinding mechanics. *Science Advances* 6. doi: 10.1126/SCIADV.AAY5999.
- Sedlak, S. M., Schendel, L. C., Melo, M. C. R., Pippig, D. A., Luthey-Schulten, Z., Gaub, H. E., et al. (2019). Direction Matters: Monovalent Streptavidin/Biotin Complex under Load. *Nano Letters* 19, 3415–3421. doi: 10.1021/acs.nanolett.8b04045.
- Skolnick, J., Gao, M., Zhou, H., and Singh, S. (2021). AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. *Journal of Chemical Information and Modeling* 61, 4827–4831. doi: 10.1021/ACS.JCIM.1C01114.
- TJ, F., and M, H. (1998). Surface protein adhesins of Staphylococcus aureus. *Trends Microbiol* 6, 484–488. doi: 10.1016/S0966-842X(98)01400-0.
- Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* 596, 590–596. doi: 10.1038/S41586-021-03828-1.

- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50, D439–D444. doi: 10.1093/NAR/GKAB1061.
- Verdorfer, T., Bernardi, R. C., Meinhold, A., Ott, W., Luthey-Schulten, Z., Nash, M. A., et al. (2017). Combining in Vitro and in Silico Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. *J Am Chem Soc* 139, 17841–17852. doi: 10.1021/JACS.7B07574.
- Xue, B., Dunker, A. K., and Uversky, V. N. (2012). Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn* 30, 137–149. doi: 10.1080/07391102.2012.675145.

Figures

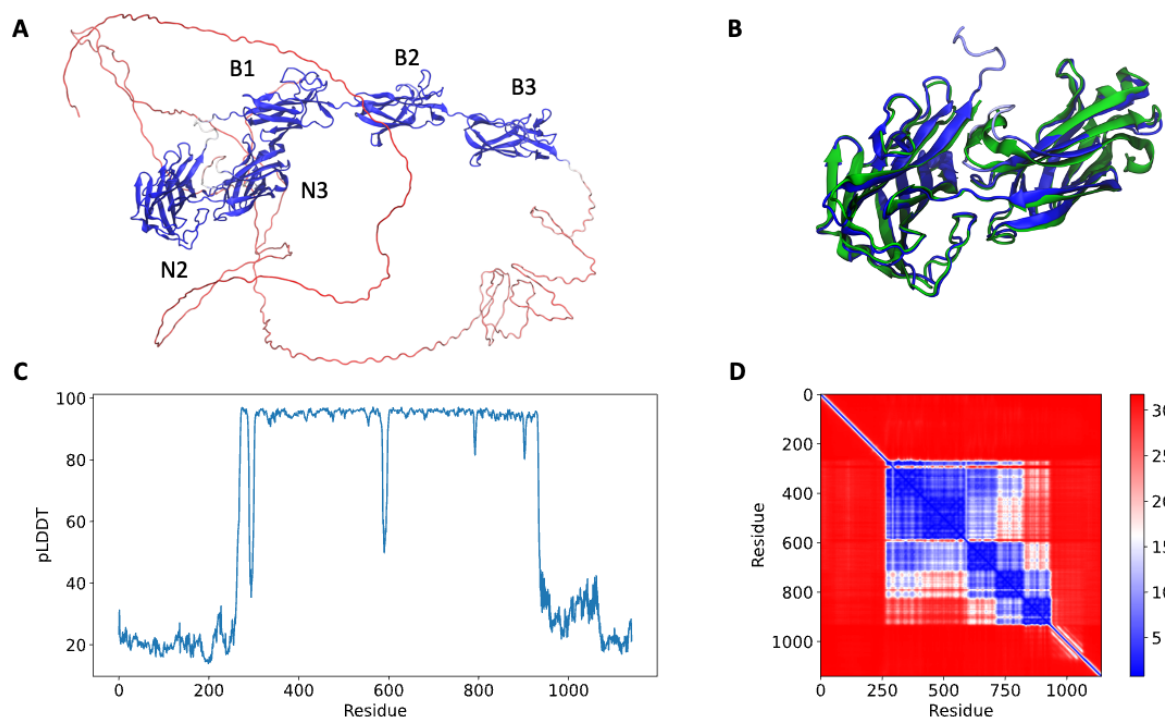


Figure 1. Full-length structure prediction of *S. aureus* serine-aspartate repeat protein (SdrE, Uniprot ID: Q932F7). (A) Top ranked SdrE model is represented in cartoon and its different domains are indicated. The protein is colored by the pLDDT scores generated by AlphaFold 2 where dark blue represents regions with very high quality (pLDDT > 90) and red represents regions with very low quality (pLDDT < 50). (B) Structural alignment between the N2 and N3 regions of the AlphaFold 2 model (dark blue) and SdrE crystallographic structure (cyan, PDB ID: 5WTA). (C) By residue pLDDT scores for the generated SdrE models. (D) Predicted alignment error (PAE) for the best ranked model. The color at (x, y) corresponds to the expected distance error in residue x's position, when the prediction and true structure are aligned on residue y.

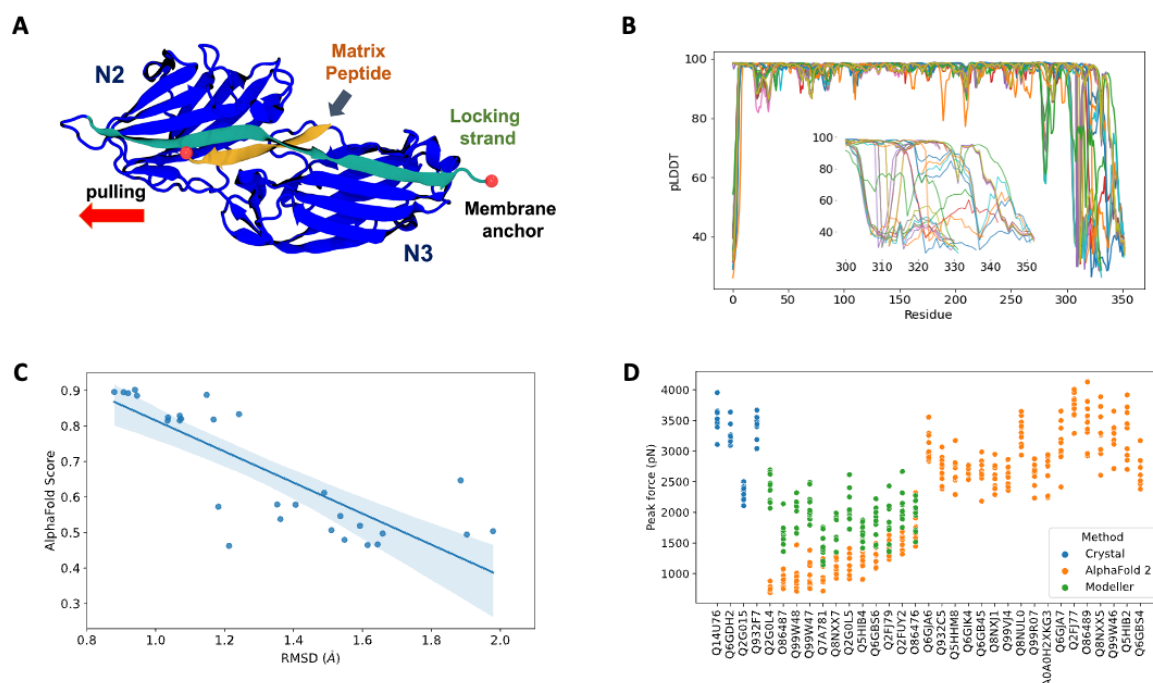


Figure 2. AlphaFold Multimer predictions for *S. aureus* adhesins. (A) Schematic view of adhesin’s Ig-like domain. Peptides from the host extracellular matrix are “locked” on a cleft between the N-ter N2 and N3 domains, snugly accommodated by the “locking strand”, connecting N3 to N2 by β -strand complementation (latch). SMD simulations were performed by keeping the C-ter fixed as it would be anchored to the membrane while the peptide is pulled at the opposite direction by its N-terminal. (B) By residue pLDDT scores for the top ranked model at each complex prediction. The insert shows the variation among the C-ter residues. (C) Comparison between AlphaFold Multimer score (ipTM) and RMSD values for equilibration pre-SMD simulations. (D) Peak Forces registered during SMD simulations for each studied complex. Color code indicates the origin of the departure structure: AlphaFold (orange), Modeller (green) or crystallographic (blue). Description of each accession entry are available at Table S2.