GDCh Highlight Angewandte International Edition Chemie www.angewandte.org

How to cite: *Angew. Chem. Int. Ed.* **2025**, *64*, e202421686
doi.org/10.1002/anie.202421686

Proteins

Structure Prediction and Computational Protein Design for Efficient Biocatalysts and Bioactive Proteins

*Rebecca Buller, Jiri Damborsky, Donald Hilvert, and Uwe T. Bornscheuer**

Dedicated to the 2024 Nobel laureates in Chemistry David Baker, Demis Hassabis, and John Jumper for their pioneering work in computational protein design and protein structure prediction.

Abstract: The ability to predict and design protein structures has led to numerous applications in medicine, diagnostics and sustainable chemical manufacture. In addition, the wealth of predicted protein structures has advanced our understanding of how life's molecules function and interact. Honouring the work that has fundamentally changed the way scientists research and engineer proteins, the Nobel Prize in Chemistry in 2024 was awarded to David Baker for computational protein design and jointly to Demis Hassabis and John Jumper, who developed AlphaFold for machine-learning-based protein structure prediction. Here, we highlight notable contributions to the development of these computational tools and their importance for the design of functional proteins that are applied in organic synthesis. Notably, both technologies have the potential to impact drug discovery as any therapeutic protein target can now be modelled, allowing the de novo design of peptide binders and the identification of small molecule ligands through in silico docking of large compound libraries. Looking ahead, we highlight future research directions in protein engineering, medicinal chemistry and material design that are enabled by this transformative shift in protein science.

Protein structure prediction and protein design, two sides of the same coin, are rapidly transforming how we develop drugs, detect chemicals and synthesize small molecules.[1] In 1957, the first protein structure was elucidated by X-ray crystallography,[2] establishing a foundation for understanding protein function and enabling their rational optimization through engineering. Building on the 200,000 protein structures (https://www.rcsb.org) solved by experimental technologies over the last 65 years, AlphaFold2 has revealed the tertiary structures of millions of additional proteins since its release in 2020. Today, open-access databases, such as the AlphaFold Protein Structure Database created in partnership with EMBL's European Bioinformatics Institute, contain more than 200 million protein structures, including nearly all catalogued proteins known to science (https://alphafold.ebi.ac.uk).[3]

The availability of experimental and computationally predicted data for protein structures has likewise shaped scientists' ability to create folded proteins from scratch. Fascinated by how polypeptide chains fold into well-defined and complex 3D shapes, William DeGrado, Stephen Mayo and David Baker pioneered the in silico design of new-to-nature protein structures. Following the construction of simple systems such as four-helix bundles[4] and zinc-finger proteins[5] – designed based on rules learned from experimentally determined protein structures – David Baker and his team further enabled the development of physics and artificial intelligence-driven computational tools that are now facilitating protein design efforts in laboratories all over the world.

[*] Prof. Dr. R. Buller
Competence Center for Biocatalysis, Institute of Chemistry and Biotechnology,
Zurich University of Applied Sciences,
Einsiedlerstrasse 31, 8820 Wädenswil (Switzerland)

Prof. Dr. J. Damborsky
Loschmidt Laboratories, Dept. of Experimental Biology and RECETOX, Faculty of Science,
Masaryk University
Kamenice 5, 625 00 Brno, (Czech Republic)
and
International Clinical Research Centre, St. Anne's University Hospital,
Pekarska 53, Brno, (Czech Republic)

Prof. Dr. D. Hilvert
Laboratory of Organic Chemistry,
ETH Zürich
8093 Zürich (Switzerland)

Prof. Dr. U. T. Bornscheuer
Biotechnology & Enzyme Catalysis, Institute of Biochemistry,
University of Greifswald,
Felix-Hausdorff-Str. 4, 17489 Greifswald (Germany)
E-mail: uwe.bornscheuer@uni-greifswald.de
bornsche@uni-greifswald.de

Physics-based de novo protein design

The design of Top7, the first computationally created protein with a completely new fold, was a landmark achievement in protein science that has expanded the scope of protein engineering from simply modifying natural proteins to creating entirely new ones. Using their computational software suite Rosetta, David Baker's team designed the 93 residue-long α/β-protein with near-atomic accuracy both in terms of backbone structure and placement of the amino acid side chains.[6] Twenty years later, Rosetta-based de novo proteins include protein-based binders, receptors, virus mimics, nanostructures for diverse materials, vaccines and delivery vehicles.[7]

Rosetta[8] models the three-dimensional structures of proteins and other biological macromolecules[9], utilising physics-based energy functions and advanced conformational sampling algorithms to predict a protein's most energetically favourable configuration given only its primary amino acid sequence. Rosetta's energy functions aim to accurately capture the various physical interactions that stabilise protein structures, including van der Waals forces, hydrogen bonding, electrostatic interactions, and solvation effects.[10] Of note, Rosetta can also incorporate experimental restraints from techniques such as X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy to further guide and refine its structure predictions.[11] Its modular architecture and the extensive developer community "Rosetta Commons" have enabled the continuous expansion of its functionality through the incorporation of new algorithms and application-specific protocols. Thanks to these efforts, the first drug developed through computational protein design, the COVID-19 vaccine SKYCovione[12], was recently approved in South Korea and Great Britain and has been granted emergency use listing by the WHO[13], highlighting the importance of protein design for pandemic preparedness.

Beyond its protein structure prediction capabilities, Rosetta has been extended to tackle a wide range of applications in structural biology and protein engineering, including its use as a computational engine in software tools for protein stabilization, such as FRESCO[14], PROSS[15] and FireProt,[16] in antibody and biomolecular materials engineering, protein-protein and protein-ligand molecular docking, and, excitingly, enzyme design.[17]

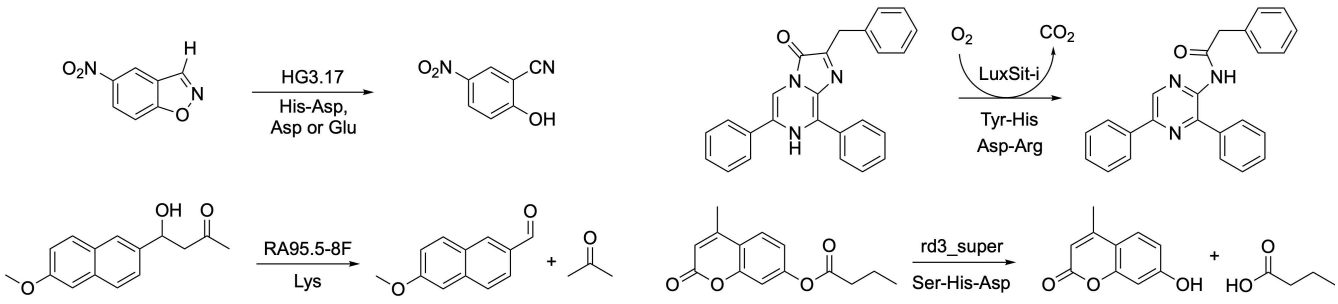
Enzyme design using natural protein scaffolds

Creating proteins with customized catalytic properties, including activities that do not occur in nature or are rare, offers scientists unique tools for the sustainable synthesis of chemicals. With this goal in mind, Baker's team developed in silico workflows to equip protein scaffolds with catalytic function (Figure 1a). Commonly, the design process starts from an idealized template of a minimal active site, consisting of a quantum mechanically calculated structure of the target reaction's rate-limiting transition state and amino acid side chains spatially arranged to stabilize it, resulting in a so-called "theozyme".[18] In initial examples, this structure was then docked into structurally characterized protein scaffolds, and the resulting active sites were repacked to maximize stability. The designs were ranked, and the most promising candidates were tested experimentally.

First successes in creating catalytically active proteins included a Kemp eliminase, capable of carrying out a proton-abstraction from the substrate 5-nitrobenzisoxazole,[19] and a retroaldolase[20] (Scheme 1). The first Kemp eliminase[19] had only modest activity, but this was improved starting from a different design[21] and by using directed evolution.[22] In the best cases, the resulting catalysts rivaled natural enzymes in terms of their activity and (enantio)-specificity. In all cases, optimization of the early in silico-designed enzymes required extensive protein engineering because active sites were too simplistic and active residue placement was insufficiently accurate.

Rosetta-based methods later enabled the creation of new-to-nature enzymes for Diels–Alder cycloadditions (Figure 1a)[23] and the Morita–Baylis–Hillman reaction.[24] More recently, a computationally designed formolase was created to catalyse a carboligation reaction, directly fixing the one-carbon unit formaldehyde into the three-carbon unit dihydroxyacetone.[25] This formolase was employed in a chemoenzymatic cascade to make starch from CO_2 via methanol and formaldehyde.[26]

Complementing Rosetta's ability to install completely new chemistry into protein scaffolds, another use of the



Scheme 1. Examples of designed enzymes: Kemp eliminase[19,22b,48] (top left) and a retro aldolase (bottom left)[20] were designed using natural protein scaffolds while a luciferase[46] (top right) and serine hydrolase[49] (bottom right) were completely designed de novo; key amino acids are indicated together with the name of a well-performing design.

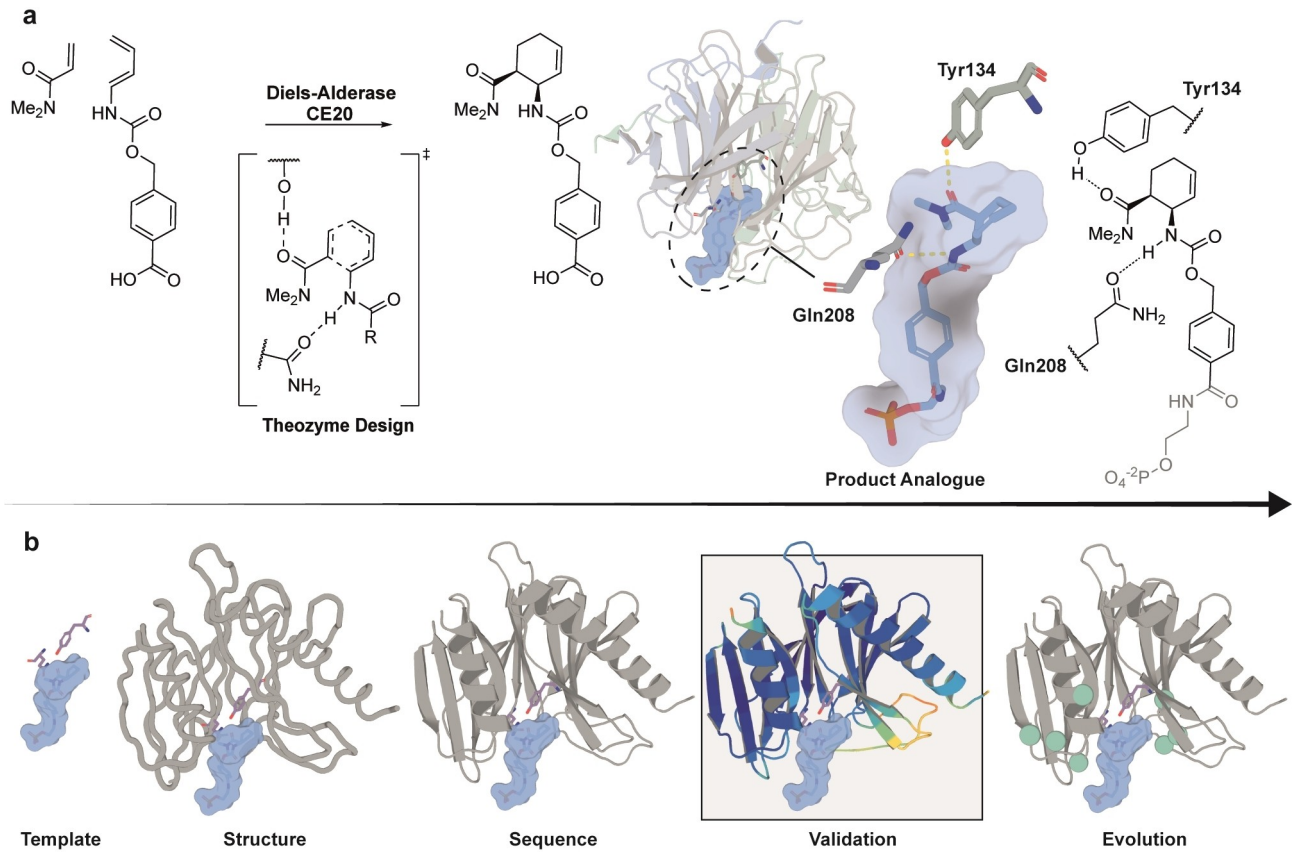


Figure 1. Computational protein design for new-to-nature function. (a) The design process for the Diels–Alderase CE20 using an existing protein scaffold.[23] (b) For the purpose of illustration, the CE20 active site was used as a template for de novo protein scaffold generation with RFDiff.

bioinformatic tools is to help enhance the promiscuous activity of wildtype enzymes. For example, this approach was profitably employed to optimize a pantothenate kinase used in an in vitro biocatalytic cascade for the manufacture of the HIV drug islatravir.[27]

Machine learning-based de novo protein design

Only a few years after these first computationally designed enzymes were described, the power and accuracy of computational protein design were further increased by the incorporation of artificial intelligence (AI) approaches, including diffusion models and neural networks.

The goal of diffusion models is to learn a diffusion process for a given dataset such that the process can generate new elements that are distributed similarly to the original dataset. RoseTTAFold Diffusion (RFdiffusion) leverages these models to generate novel protein structures.[28] RFdiffusion starts with a random or partially guided protein backbone. Noise is applied, and then, in an iterative process, the model removes the noise, allowing the protein structure to emerge, while respecting structural constraints and desired features.[28] The model is trained on large datasets of known protein structures to understand and generate realistic protein folds and side-chain orientations.[29] The output is a new protein structure that can be tailored for specific binding sites, stability, co-factors, structural features and functions (Figure 1b).

RFdiffusion was validated through computational and experimental methods, demonstrating its ability to generate soluble, stable and functional protein structures,[28] including novel binding proteins, enzyme mimetics, self-assembling nanostructures and virus-like particles. Selected designs were synthesized, and X-ray crystallography and cryo-electron microscopy were used to verify that the experimentally generated proteins closely matched the design models, while functional assays confirmed that the designed proteins had the intended functions.[28] Another recently introduced all-atom diffusion model for protein design, named Protpardelle, codesigns the backbone, sequence, and sidechains of a protein together.[30]

Inverse folding

Message-Passing Neural Network (MPNN) is a deep learning architecture for protein sequence design developed by the team of Tommi Jaakkola from MIT[31] and later optimized and experimentally validated by David Baker and co-workers.[32] The model operates by taking a protein backbone structure as input and predicting optimal amino

acid sequences that would fold into that structure through a graph neural network. The architecture uses both forward and backward passes to capture bidirectional dependencies between amino acids, incorporating geometric and chemical constraints inherent in protein structures.[32] ProteinMPNN demonstrates high reliability in designing stable proteins, achieving sequence recovery rates >50 % for many protein families. However, like all computational design tools, its reliability depends on the quality of input structural information and performs best when designing proteins like those in its training set. ProteinMPNN-ddG, which exhibits improved accuracy, was recently developed by the company Peptone[33] and can be used for protein engineering applications, including stability optimization, interface design, and novel scaffold creation.

Validating ProteinMPNN's generality, hundreds of designed proteins covering diverse topologies and sizes have been synthesized and experimentally tested.[32] Many of these proteins are stable and correctly folded, as confirmed by biophysical assays and, in some cases, by solving their structures through crystallography or cryo-EM. Moreover, ProteinMPNN has been used to optimize sequences for known structures such as TEV protease and myoglobin.[34]

Tools such as ProteinMPNN[32], RFdiffusion[35] and AlphaFold[36], as well as related software including AlphaProteo[37], PPIformer[38], PINDER[39], and BindCraft[40] now routinely allow the computational design of novel proteins that bind to predetermined targets. Harnessing these improved technologies, designer proteins have been created that enable targeted protein degradation[41], bind peptide hormones with subnanomolar affinity[42], or may even serve as synthetic sensors resembling those in the nose[43], opening up intriguing possibilities in diagnostics. In addition, the design of soluble analogues of integral membrane proteins has been recently demonstrated by Bruno Correia and co-workers from EPFL and the Swiss Institute of Bioinformatics by combining ProteinMPNN with a re-trained $MPNN_{sol}$ and AlphaFold2.[44] Finally, LigandMPNN, a neural network that explicitly models all non-protein components of biomolecular systems[45], was found to significantly outperform Rosetta and ProteinMPNN on native backbone sequence recovery for residues interacting with small molecules, nucleotides, and metals.

Enzyme design using de novo protein scaffolds

While Nature's protein scaffolds can be reprogrammed to catalyse non-native chemistry (see above), designed protein structures offer greater control of the protein backbone, active site and overall structure to fit desired reaction mechanisms. Avoiding evolutionary constraints, de novo protein scaffolds can be designed to contain minimal structural complexity while exhibiting improved stability and solubility, facilitating the implementation of complex, multi-step reactions.

Illustrating the power of de novo scaffolds for enzyme design, luciferases[46] were created that selectively oxidize diphenylterazine, which possesses a high quantum yield, red-shifted emission and favourable in vivo pharmacokinetics compared to natural substrates.[47] The protocol combined RosettaDesign with hallucinations, ProteinMPNN, gene synthesis and site-saturation mutagenesis, providing an active site containing an arginine residue to stabilize the anionic state of diphenylterazine required for the chemiluminescent reaction and optimized pocket polarity for the single-electron transfer process with triplet molecular oxygen. The best design achieved a catalytic efficiency (k_{cat}/K_M) of 10^6 M^{-1} s^{-1}, which is in the range of natural luciferases.[46]

Expanding Rosetta's capabilities further, Baker and co-workers recently combined RFdiffusion with a newly developed ensemble generation method (ChemNet) to design serine[49] and zinc-dependent[49–50] hydrolases. Assessing active site geometry and preorganization at each step of the ester hydrolysis, de novo designed hydrolases were created that catalyse ester hydrolysis with catalytic efficiencies (k_{cat}/K_M) up to 2.3×10^4 M^{-1} s^{-1}. These studies represent a significant step towards the de novo design of enzymes for applications in the chemical and pharmaceutical industries.[7]

Protein structure prediction from sequence

A longstanding 'holy grail' in biochemistry and structural biology has been predicting a protein's tertiary structure from its amino acid sequence.[51] While gradual advances have been made, the methods developed over the past few decades struggled to achieve atomic accuracy, especially when no similar protein structure was known. In 2020, Demis Hassabis, John Jumper and the team from DeepMind introduced AlphaFold – a novel neural network-based model that can consistently predict protein structures with atomic accuracy, even in the absence of homologous structures.[36] In the CASP14 competition[52], AlphaFold demonstrated exceptional performance (90 % accuracy), rivalling experimental methods and outperforming other computational approaches including Rosetta-based methods.[53]

AlphaFold2

AlphaFold's success is attributed to its innovative architecture (Figure 2), which integrates physical and biological principles related to protein structure. By incorporating multiple sequence alignments (MSA), AlphaFold leverages evolutionary information to refine its predictions.[36] Together with related tools such as RoseTTAFold[54], ESMFold[55] and ColabFold,[56] AlphaFold2's ability to predict structures with near-experimental accuracy has opened up new avenues of research and applications in structural biology[57], protein[58] and vaccine design[59], as well as for the prediction of protein function[60] and protein-protein interaction.[61]

While AlphaFold2 and related tools have had a tremendous impact on structural biology and protein design, fewer successful applications in drug discovery have been described.[62] In retrospective docking studies, AlphaFold2

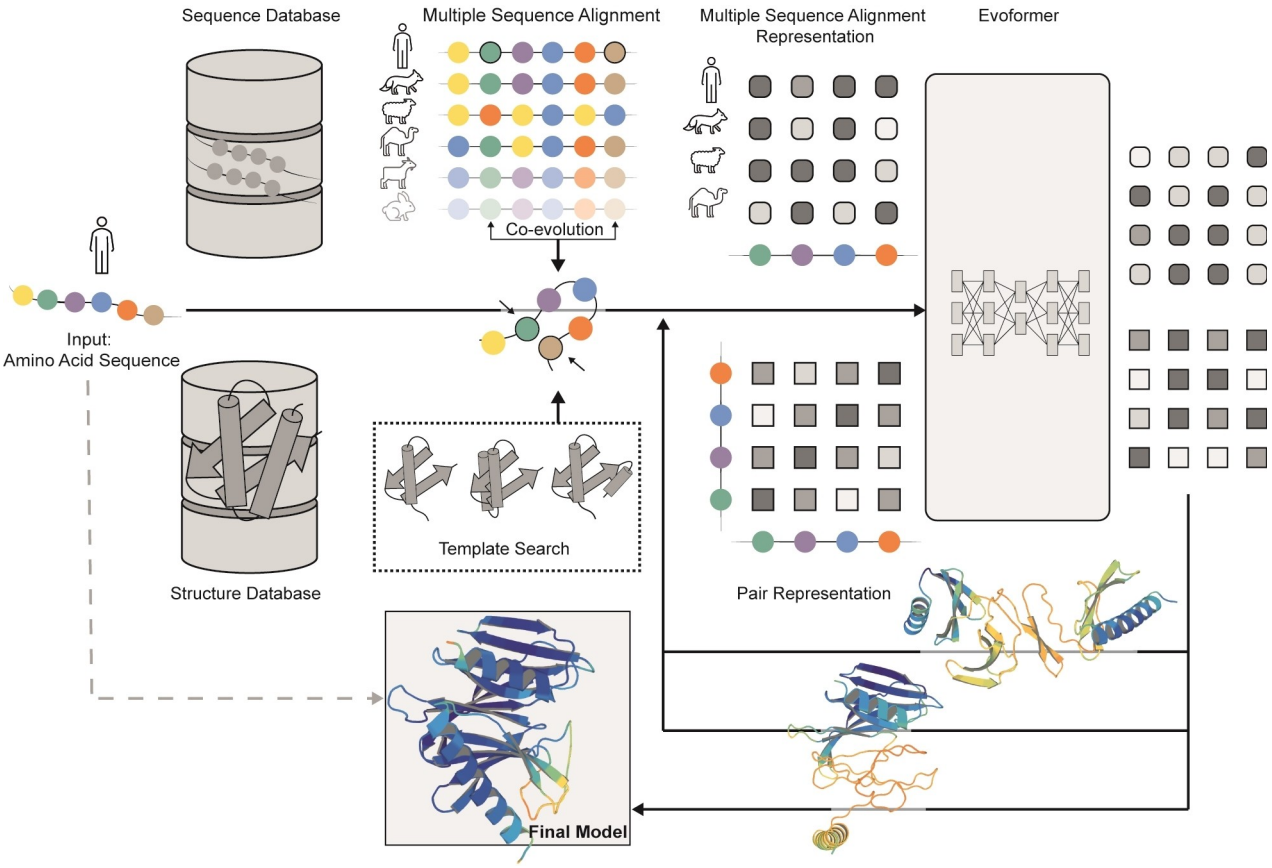


Figure 2. Model architecture of AlphaFold2. AlphaFold2 uses the input amino acid sequence to query several databases of protein sequences and constructs a multiple sequence alignment (MSA). In parallel, AlphaFold2 attempts to identify proteins that may have a similar structure to the input, so-called templates and constructs an initial pair representation of the structure, modelling which amino acids are likely to be in contact with each other. Next, AlphaFold2 passes the MSA and the templates through a transformer, named Evoformer, that quickly identifies which pieces of information are most informative. Finally, the obtained information is passed to a structure module, which takes the refined MSA and pair representations and leverages them to construct a three-dimensional model of the protein. The network generates a final structure in a single step without any optimization algorithm, which makes AlphaFold2 distinct from previous models. After generating a final structure, all the information – MSA and pair representations, and the predicted structure – are passed back to the Evoformer blocks, allowing iterative refinement of the prediction.

models can fall short in reliably reproducing binding modes or differentiating active compounds from decoys when compared to the same calculations done with experimental structures. This limitation can reduce the effectiveness of AlphaFold2 models in simulations intended for identifying potential drug candidates, as binding accuracy is crucial in distinguishing functional compounds from inactive ones. However, in prospective studies, the models have shown notable utility. For instance, therapeutic targets such as the σ2 receptor and the serotonin 2 A receptor were selected for ligand discovery using AlphaFold2 models before the proteins' experimental structures were available. Large libraries of small molecules were docked against both AlphaFold2-generated and later-released experimental structures for these receptors. Remarkably, prospective docking against AlphaFold2 models was just as effective in identifying potential ligands as docking against the actual experimental structures. The fact that the use of modelled versus experimental protein structure also led to the identification of different hit molecules indicates that a subset of the AlphaFold2 models may represent low-energy alternative receptor conformations that can equally guide the discovery of new ligands, enlarging molecular space.[63] In a similar drug discovery approach, a small molecule ligand was designed for cyclin-dependent kinase 20, a potential therapeutic target to treat hepatocellular carcinoma, for which no experimental structure was available. Using a generative chemistry platform (Chemistry42), ligands could be designed and synthesised within 30 days from target selection, exhibiting appreciable binding constants in the micromolar range.[64] Beyond guiding the design of ligands for specific structurally uncharacterised target proteins, AlphaFold structures of the human proteome have also been utilised to identify and describe druggable pockets. Looking forward, this type of overview could help prioritise chemical probe development.[65]

AlphaFold3

Further optimization of the AlphaFold2 design architecture has yielded AlphaFold3,[66] a more versatile and efficient tool for biomolecular modelling, extending its capabilities beyond those of its predecessors. The most crucial innovation in AlphaFold3 is its ability to model a wide range of biomolecular complexes. Unlike AlphaFold2, specifically tailored for modelling proteins and protein-protein complexes, AlphaFold3 allows for the inclusion of modified residues, metals, ligands and nucleic acids.

In AlphaFold2, protein structures were represented by associating a rigid body frame with each amino acid, focusing on the Cα-atoms, with side chains parameterised by χ-angles. This protein-specific approach does not generalise to non-proteins. AlphaFold3 models systems as collections of individual atoms, each with independent global coordinates and free from rigid constraints.[66] This allows the model to learn the chemical structure of molecules through training rather than parameterised output constraints improving its ability to capture protein-ligand interactions.

Future Directions

We are heading toward a future where designing custom proteins will be as accessible as ordering from a vending machine, enabling scientists to specify proteins by function, size, and binding characteristics.[67] However, to access this reality, unresolved challenges include reliably creating new catalytic functions, understanding proteins' conformational changes, and designing multi-functional molecular machines.[1a,68] The complexity is due, in part, to the intricate dynamics of proteins, which often require precise folding and interaction patterns that are difficult to model fully. Despite these hurdles, AI-driven tools like RFdiffusion and AlphaFold continue to make strides, showing promise in designing enzymes capable of catalyzing complex reaction trajectories.[49–50]

The fully programmable design of dynamic proteins will require careful engineering of energy landscapes to enable controlled transitions between multiple states, considering factors like protein-ligand interactions, allosteric interactions, solvent effects, and entropic contributions. To model the intricate interplay between local flexibility and global conformational changes, first models such as AF-Cluster[69] or Flexpert-Design were developed. Understanding and creating dynamic proteins will be further supported by the application of data-driven techniques to analyze molecular dynamic simulations.[70] Well-curated databases of trajectories are appearing and will play an essential role in training and validating these models (see: https://bioexcel.eu/molecular-dynamics-databases-at-the-cusp-of-an-upward-trajectory/). Looking ahead, these developments will be further supported by improvements in quantum computing technologies which will allow us to solve certain problems faster than classical computers.[71] This long-awaited event will showcase the 'quantum advantage' in biocatalysis, demonstrating the use of a quantum computer to solve a practical problem that is impossible to address with classical computers.

While in silico protein design plays an increasingly important role in many applications, directed evolution (recognized with a Nobel prize in 2018 to Frances H. Arnold[72]) will remain a key technology to improve the properties of designed and natural enzymes, including activity, stereoselectivity, expression yield, thermal stability, tolerance to organic solvents, and substrate specificity (see Highlight from 2018[73]). Looking forward, enzyme engineering will become even more powerful by integrating protein structure prediction and design tools as well as through the application of machine learning.[74] The latter has progressed from its initial use of multivariate statistical analyses of a small number of enzyme variants[75] to the application of artificial neural networks.[76] However, development limitations through lack of data are becoming more obvious[77] and further improvements in the methods will depend on our capacity to generate, curate and store high-quality biological data in the same way as the community of structural biologists has achieved for protein structures in the Protein Data Bank.

We are on the cusp of the "century of the protein": Integrated approaches that combine computational power with empirical testing will drive the field of protein science, and particularly biocatalysis, forward such that the goal of routine, custom-designed protein generation is not just a possibility but an impending reality.

Acknowledgements

We thank D. Schaub for help with Figure preparation. Funding: R.B. thanks NCCR Catalysis, a National Centre of Competence in Research funded by the Swiss National Science Foundation (Grant no. 180544); DH thanks the ETH Zurich for support. UTB thanks the German Research Foundation, DFG (NFDI4CAT). JD is supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreements No. 857560 and 101136607. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Keywords: AlphaFold · Computational protein design · Nobel prize · Protein engineering · Protein structure prediction

[1] a) R. Buller, S. Lutz, R. J. Kazlauskas, R. Snajdrova, J. C. Moore, U. T. Bornscheuer, *Science* **2023**, *382*, eadh8615; b) U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.

[2] a) J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, D. C. Phillips, *Nature* **1958**, *181*, 662–666; b) M. F. Perutz, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will, A. C. T. North, *Nature* **1960**, *185*, 416–422.

[3] Z. Yang, X. Zeng, Y. Zhao, R. Chen, *Signal Transd. Targ. Ther.* **2023**, *8*, 115.

[4] L. Regan, W. F. DeGrado, *Science* **1988**, *241*, 976–978.

[5] B. I. Dahiyat, S. L. Mayo, *Science* **1997**, *278*, 82–87.

[6] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, *Science* **2003**, *302*, 1364–1368.

[7] A. E. Chu, T. Lu, P.-S. Huang, *Nat. Biotechnol.* **2024**, *42*, 203–215.

[8] a) G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker, K. N. Houk, *Angew. Chem. Int. Ed.* **2013**, *52*, 5700–5025; b) H. Kries, R. Blomberg, D. Hilvert, *Curr. Opin. Chem. Biol.* **2013**, *17*, 221–228.

[9] C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, D. Baker, in *Methods Enzymol.*, Vol. 383, Academic Press, **2004**, pp. 66–93.

[10] R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, Jr., R. Das, D. Baker, B. Kuhlman, T. Kortemme, J. J. Gray, *J. Chem. Theory Comp.* **2017**, *13*, 3031–3048.

[11] a) F. DiMaio, N. Echols, J. J. Headd, T. C. Terwilliger, P. D. Adams, D. Baker, *Nat. Meth.* **2013**, *10*, 1102–1104; b) J. A.-O. Koehler Leman, G. Künze, *Int. J. Mol. Sci.* **2024**, *24*, 7835; c) B. Frenz, A. C. Walls, E. H. Egelman, D. Veesler, F. DiMaio, *Nat. Meth.* **2017**, *14*, 797–800.

[12] D. Baker, G. Church, *Science* **2024**, *383*, 349–349.

[13] https://www.skbioscience.com/en/news/news 01 01?mode=view&id=214&, **2024**.

[14] H. J. Wijma, M. Fürst, D. B. Janssen, *Methods Mol. Biol.* **2018**, *1685*, 69–85.

[15] J. J. Weinstein, A. Goldenzweig, S. Hoch, S. J. Fleishman, *Bioinformatics* **2021**, *37*, 123–125.

[16] M. Musil, A. Jezik, J. Horackova, S. Borko, P. Kabourek, J. Damborsky, D. Bednar, *Brief. Bioinform.* **2024**, *25*, bbad425.

[17] J. K. Leman, B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth, B. Basanta, B. J. Bender, K. Blacklock, J. Bonet, S. E. Boyken, P. Bradley, C. Bystroff, P. Conway, S. Cooper, B. E. Correia, B. Coventry, R. Das, R. M. De Jong, F. DiMaio, L. Dsilva, R. Dunbrack, A. S. Ford, B. Frenz, D. Y. Fu, C. Geniesse, L. Goldschmidt, R. Gowthaman, J. J. Gray, D. Gront, S. Guffy, S. Horowitz, P.-S. Huang, T. Huber, T. M. Jacobs, J. R. Jeliazkov, D. K. Johnson, K. Kappel, J. Karanicolas, H. Khakzad, K. R. Khar, S. D. Khare, F. Khatib, A. Khramushin, I. C. King, R. Kleffner, B. Koepnick, T. Kortemme, G. Kuenze, B. Kuhlman, D. Kuroda, J. W. Labonte, J. K. Lai, G. Lapidoth, A. Leaver-Fay, S. Lindert, T. Linsky, N. London, J. H. Lubin, S. Lyskov, J. Maguire, L. Malmström, E. Marcos, O. Marcu, N. A. Marze, J. Meiler, R. Moretti, V. K. Mulligan, S. Nerli, C. Norn, S. Ó'Conchúir, N. Ollikainen, S. Ovchinnikov, M. S. Pacella, X. Pan, H. Park, R. E. Pavlovicz, M. Pethe, B. G. Pierce, K. B. Pilla, B. Raveh, P. D. Renfrew, S. S. R. Burman, A. Rubenstein, M. F. Sauer, A. Scheck, W. Schief, O. Schueler-Furman, Y. Sedan, A. M. Sevy, N. G. Sgourakis, L. Shi, J. B. Siegel, D.-A. Silva, S. Smith, Y. Song, *Nat. Meth.* **2020**, *17*, 665–680.

[18] D. J. Tantillo, C. Jiangang, K. N. Houk, *Curr. Opin. Chem. Biol.* **1998**, *2*, 743–750.

[19] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, *Nature* **2008**, *453*, 190–195.

[20] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas III, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, *Science* **2008**, *319*, 1387–1391.

[21] H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, S. L. Mayo, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3790–3795.

[22] a) O. Khersonsky, G. Kiss, D. Röthlisberger, O. Dym, S. Albeck, K. N. Houk, D. Baker, D. S. Tawfik, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10358–10363; b) R. Blomberg, H. Kries, D. M. Pinkas, P. R. E. Mittl, M. G. Grütter, H. K. Privett, S. L. Mayo, D. Hilvert, *Nature* **2013**, *503*, 418–421.

[23] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker, *Science* **2010**, *329*, 309–313.

[24] S. Bjelic, L. G. Nivon, N. Celebi-Olcum, G. Kiss, C. F. Rosewall, H. M. Lovick, E. L. Ingalls, J. L. Gallaher, J. Seetharaman, S. Lew, G. T. Montelione, J. F. Hunt, F. E. Michael, K. N. Houk, D. Baker, *ACS Chem. Biol.* **2013**, *8*, 749–757.

[25] J. B. Siegel, A. L. Smith, S. Poust, A. J. Wargacki, A. Bar-Even, C. Louw, B. W. Shen, C. B. Eiben, H. M. Tran, E. Noor, J. L. Gallaher, J. Bale, Y. Yoshikuni, M. H. Gelb, J. D. Keasling, B. L. Stoddard, M. E. Lidstrom, D. Baker, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3704–3709.

[26] T. Cai, H. Sun, J. Qiao, L. Zhu, F. Zhang, J. Zhang, Z. Tang, X. Wei, J. Yang, Q. Yuan, W. Wang, X. Yang, H. Chu, Q. Wang, C. You, H. Ma, Y. Sun, Y. Li, C. Li, H. Jiang, Q. Wang, Y. Ma, *Science* **2021**, *373*, 1523–1527.

[27] M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra-Garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. McLaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel, A. Rodriguez-Granillo, S. A. Robaire, E. C. Sherer, M. D. Truppo, A. M. Whittaker, D. Verma, L. Xiao, Y. Xu, H. Yang, *Science* **2019**, *366*, 1255–1259.

[28] J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, *Nature* **2023**, *620*, 1089–1100.

[29] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucl. Acids Res.* **2000**, *28*, 235–242.

[30] A. E. Chu, J. Kim, L. Cheng, G. El Nesr, M. Xu, R. W. Shuai, P.-S. Huang, *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2311500121.

[31] J. Ingraham, V. K. Garg, R. Barzilay, T. Jaakkola, in *Neural Information Processing Systems Foundation*, Vol. 32 (Eds.: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett), **2019**, pp. 15741–15752.

[32] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, D. Baker, *Science* **2022**, *378*, 49–56.

[33] O. Dutton, S. Bottaro, M. Invernizzi, I. Redl, A. Chung, F. Hoffmann, L. Henderson, S. Ruschetta, F. Airoldi, B. M. J.

Owens, P. Foerch, C. Fisicaro, K. Tamiola, *bioRxiv* **2024**, 10.1101/2024.1106.1115.599145.

[34] K. H. Sumida, R. Núñez-Franco, I. Kalvet, S. J. Pellock, B. I. M. Wicky, L. F. Milles, J. Dauparas, J. Wang, Y. Kipnis, N. Jameson, A. Kang, J. De La Cruz, B. Sankaran, A. K. Bera, G. Jiménez-Osés, D. Baker, *J. Am. Chem. Soc.* **2024**, *146*, 2054–2061.

[35] J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, *bioRxiv* **2022**, 2022.2012.2009.519842.

[36] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.

[37] V. Zambaldi, D. La, A. E. Chu, H. H. Patani, A. E. Danson, T. O. C. Kwan, T. Frerix, R. G. Schneider, D. Saxton, A. Thillaisundaram, Z. Wu, I. Moraes, O. Lange, E. Papa, G. Stanton, V. Martin, S. Singh, L. H. Wong, R. Bates, S. A. A. Kohl, J. Abramson, A. W. Senior, Y. Alguel, M. Y. Wu, I. M. Aspalter, K. Bentley, D. L. V. Bauer, P. Cherepanov, D. Hassabis, P. Kohli, R. Fergus, J. Wang, *arxiv.org* **2024**, 10.48550/arXiv.42409.08022.

[38] A. Bushuiev, R. Bushuiev, P. Kouba, A. Filkin, M. Gabrielova, M. Gabriel, J. Sedlar, T. Pluskal, J. Damborsky, S. Mazurenko, J. Sivic, *arxiv.org* **2024**, 10.48550/arXiv.42310.18515.

[39] D. Kovtun, M. Akdel, A. Goncearenco, G. Zhou, G. Holt, D. Baugher, D. Lin, Y. Adeshina, T. Castiglione, X. Wang, C. Marquet, M. McPartlon, T. Geffner, E. Rossi, G. Corso, H. Stärk, Z. Carpenter, E. Kucukbenli, M. Bronstein, L. Naef, *bioRxiv* **2024**, 10.1101/2024.1107.1117.603980.

[40] M. Pacesa, L. Nickel, J. Schmidt, E. Pyatova, C. Schellhaas, L. Kissling, A. Alcaraz-Serna, Y. Cho, K. H. Ghamary, L. Vinué, B. J. Yachnin, A. M. Wollacott, S. Buckley, S. Georgeon, C. A. Goverde, G. N. Hatzopoulos, P. Gönczy, Y. D. Muller, G. Schwank, S. Ovchinnikov, B. E. Correia, *bioRxiv* **2024**, 2024.2009.2030.615802.

[41] B. Huang, M. Abedi, G. Ahn, B. Coventry, I. Sappington, C. Tang, R. Wang, T. Schlichthaerle, J. Z. Zhang, Y. Wang, I. Goreshnik, C. W. Chiu, A. Chazin-Gray, S. Chan, S. Gerben, A. Murray, S. Wang, J. O'Neill, L. Yi, R. Yeh, A. Misquith, A. Wolf, L. M. Tomasovic, D. I. Piraner, M. J. Duran Gonzalez, N. R. Bennett, P. Venkatesh, M. Ahlrichs, C. Dobbins, W. Yang, X. Wang, D. D. Sahtoe, D. Vafeados, R. Mout, S. Shivaei, L. Cao, L. Carter, L. Stewart, J. B. Spangler, K. T. Roybal, P. Greisen, Jr., X. Li, G. J. L. Bernardes, C. R. Bertozzi, D. Baker, *Nature* **2024**, 10.1038/s41586-41024-07948-41582.

[42] S. Vázquez Torres, P. J. Y. Leung, P. Venkatesh, I. D. Lutz, F. Hink, H.-H. Huynh, J. Becker, A. H.-W. Yeh, D. Juergens, N. R. Bennett, A. N. Hoofnagle, E. Huang, M. J. MacCoss, M. Expòsit, G. R. Lee, A. K. Bera, A. Kang, J. De La Cruz, P. M. Levine, X. Li, M. Lamb, S. R. Gerben, A. Murray, P. Heine, E. N. Korkmaz, J. Nivala, L. Stewart, J. L. Watson, J. M. Rogers, D. Baker, *Nature* **2024**, *626*, 435–442.

[43] a) S. Berhanu, S. Majumder, T. Müntener, J. Whitehouse, C. Berner, A. K. Bera, A. Kang, B. Liang, N. Khan, B. Sankaran, L. K. Tamm, D. J. Brockwell, S. Hiller, S. E. Radford, D. Baker, A. A. Vorobieva, *Science* **2024**, *385*, 282–288; b) L. An, M. Said, L. Tran, S. Majumder, I. Goreshnik, G. R. Lee, D. Juergens, J. Dauparas, I. Anishchenko, B. Coventry, A. K. Bera, A. Kang, P. M. Levine, V. Alvarez, A. Pillai, C. Norn, D. Feldman, D. Zorine, D. R. Hicks, X. Li, M. G. Sanchez, D. K. Vafeados, P. J. Salveson, A. A. Vorobieva, D. Baker, *Science* **2024**, *385*, 276–282.

[44] C. A. Goverde, M. Pacesa, N. Goldbach, L. J. Dornfeld, P. E. M. Balbi, S. Georgeon, S. Rosset, S. Kapoor, J. Choudhury, J. Dauparas, C. Schellhaas, S. Kozlov, D. Baker, S. Ovchinnikov, A. J. Vecchio, B. E. Correia, *Nature* **2024**, *631*, 449–458.

[45] J. Dauparas, G. R. Lee, R. Pecoraro, L. An, I. Anishchenko, C. Glasscock, D. Baker, *bioRxiv* **2023**, 10.1101/2023.1112.1122.573103.

[46] A. H.-W. Yeh, C. Norn, Y. Kipnis, D. Tischer, S. J. Pellock, D. Evans, P. Ma, G. R. Lee, J. Z. Zhang, I. Anishchenko, B. Coventry, L. Cao, J. Dauparas, S. Halabiya, M. DeWitt, L. Carter, K. N. Houk, D. Baker, *Nature* **2023**, *614*, 774–780.

[47] J.-S. Völler, *Nat. Catal.* **2023**, *6*, 216.

[48] O. Khersonsky, D. Rothlisberger, O. Dym, S. Albeck, C. J. Jackson, D. Baker, D. S. Tawfik, *J. Mol. Biol.* **2010**, *396*, 1025–1042.

[49] A. Lauko, S. J. Pellock, I. Anischanka, K. H. Sumida, D. Juergens, W. Ahern, A. Shida, A. Hunt, I. Kalvet, C. Norn, I. R. Humphreys, C. Jamieson, A. Kang, E. Brackenbrough, A. K. Bera, B. Sankaran, K. N. Houk, D. Baker, *bioRxiv* **2024**, 10.1101/2024.1108.1129.610411.

[50] D. Kim, S. M. Woodbury, W. Ahern, I. Kalvet, N. Hanikel, S. Salike, S. J. Pellock, A. Lauko, D. Hilvert, D. Baker, *bioRxiv* **2024**, 10.1101/2024.1111.1113.623507v623501.

[51] C. B. Anfinsen, *Science* **1973**, *181*, 223–230.

[52] J. Moult, R. Pedersen, K. Judson, K. Fidelis, *Proteins* **1995**, *23*, ii–v.

[53] J. A.-O. Pereira, A. J. Simpkin, M. D. Hartmann, D. A.-O. Rigden, R. M. Keegan, A. N. Lupas, *Proteins* **2021**, *89*, 1687–1699.

[54] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, *Science* **2021**, *373*, 871–876.

[55] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, *Science* **2023**, *379*, 1123–1130.

[56] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, *Nat. Meth.* **2022**, *19*, 679–682.

[57] a) T. G. Flower, J. H. Hurley, *Protein Sci.* **2021**, *30*, 728–734; b) L. Hu, W. Salmen, B. Sankaran, Y. Lasanajak, D. F. Smith, S. E. Crawford, M. K. Estes, B. V. V. Prasad, *Commun. Biol.* **2022**, *5*, 419; c) I. Skalidis, F. L. Kyrilis, C. Tüting, F. Hamdi, G. Chojnowski, P. L. Kastritis, *Structure* **2022**, *30*, 575–589.e576; d) D. J. Wasilko, B. S. Gerstenberger, K. A. Farley, W. Li, J. Alley, M. E. Schnute, R. J. Unwalla, J. Victorino, K. K. Crouse, R. Ding, P. V. Sahasrabudhe, F. Vincent, R. K. Frisbie, A. Dermenci, A. Flick, C. Choi, G. Chinigo, J. J. Mousseau, J. I. Trujillo, P. Nuhant, P. Mondal, V. Lombardo, D. Lamb, B. J. Hogan, G. S. Minhas, E. Segala, C. Oswald, I. W. Windsor, S. Han, M. Rappas, R. M. Cooke, M. F. Calabrese, G. Berstein, A. Thorarensen, H. Wu, *Nat. Commun.* **2024**, *15*, 7574.

[58] C. A. Goverde, B. Wolf, H. Khakzad, S. Rosset, B. E. Correia, *Protein Sci.* **2023**, *32*, e4653.

[59] D. Zeng, J. Xin, K. Yang, S. Guo, Q. Wang, Y. Gao, H. Chen, J. Ge, Z. Lu, L. Zhang, J. Chen, Y. Chen, N. Xia, in *Viruses*, Vol. 14, **2022**, 10.3390/v14061305.

[60] W. Ma, S. Zhang, Z. Li, M. Jiang, S. Wang, W. Lu, X. Bi, H. Jiang, H. Zhang, Z. Wei, *J. Chem. Inf. Model.* **2022**, *62*, 4008–4017.
[61] P. Trepte, C. Secker, J. Olivet, J. Blavier, S. Kostova, S. B. Maseko, I. Minia, E. Silva Ramos, P. Cassonnet, S. Golusik, M. Zenkner, S. Beetz, M. J. Liebich, N. Scharek, A. Schütz, M. Sperling, M. Lisurek, Y. Wang, K. Spirohn, T. Hao, M. A. Calderwood, D. E. Hill, M. Landthaler, S. G. Choi, J.-C. Twizere, M. Vidal, E. E. Wanker, *Mol. Syst. Biol.* **2024**, *20*, 428–457.
[62] a) J. A. Lees, J. M. Dias, F. Rajamohan, J.-P. Fortin, R. O'Connor, J. X. Kong, E. A. G. Hughes, E. L. Fisher, J. B. Tuttle, G. Lovett, B. L. Kormos, R. J. Unwalla, L. Zhang, A.-M. Dechert Schmitt, D. Zhou, M. Moran, K. A. Stevens, K. F. Fennell, A. E. Varghese, A. Maxwell, E. E. Cote, Y. Zhang, S. Han, *Nat. Commun.* **2023**, *14*, 5938; b) Z. L. Johnson, M. Ammirati, D. J. Wasilko, J. S. Chang, S. Noell, T. L. Foley, H. Yoon, K. Smith, S. Asano, K. Hales, M. Wan, Q. Yang, M. A. Piotrowski, K. A. Farley, T. Gilbert, L. M. Aschenbrenner, K. F. Fennell, J. K. Dutra, M. Xu, C. Guo, A. E. Varghese, J. Bellenger, A. Quinn, C. W. am Ende, G. M. West, M. C. Griffor, D. Bennett, M. Calabrese, C. M. Steppan, S. Han, H. Wu, *Nat. Struct. Mol. Biol.* **2023**, *30*, 22–30.
[63] J. Lyu, N. Kapolka, R. Gumpper, A. Alon, L. Wang, M. K. Jain, X. Barros-Álvarez, K. Sakamoto, Y. Kim, J. DiBerto, K. Kim, I. S. Glenn, T. A. Tummino, S. Huang, J. J. Irwin, O. O. Tarkhanova, Y. Moroz, G. Skiniotis, A. C. Kruse, B. K. Shoichet, B. L. Roth, *Science* **2024**, *384*, eadn6354.
[64] F. Ren, X. Ding, M. Zheng, M. Korzinkin, X. Cai, W. Zhu, A. Mantsyzov, A. Aliper, V. Aladinskiy, Z. Cao, S. Kong, X. Long, B. H. Man Liu, Y. Liu, V. Naumov, A. Shneyderman, I. V. Ozerov, J. Wang, F. W. Pun, D. A. Polykovskiy, C. Sun, M. Levitt, A. Aspuru-Guzik, A. Zhavoronkov, *Chem. Sci.* **2023**, *14*, 1443–1452.
[65] A. Comajuncosa-Creus, G. Jorba, X. Barril, P. Aloy, *Nat. Commun.* **2024**, *15*, 7917.
[66] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, *Nature* **2024**, *630*, 493–500.
[67] S. Reardon, *Nature* **2024**, *635*, 246–248.
[68] S. L. Lovelock, R. Crawshaw, S. Basler, C. Levy, D. Baker, D. Hilvert, A. P. Green, *Nature* **2022**, *606*, 49–58.
[69] H. K. Wayment-Steele, A. Ojoawo, R. Otten, J. M. Apitz, W. Pitsawong, M. Hömberger, S. Ovchinnikov, L. Colwell, D. Kern, *Nature* **2024**, *625*, 832–839.
[70] A. Lappala, *Curr. Opin. Struct. Biol.* **2024**, *89*, 102919.
[71] J. Damborsky, P. Kouba, J. Sivic, D. Bednar, S. Mazurenko, *ChemRxiv* **2024**, 10.26434/chemrxiv-22024-pwkq26435.
[72] F. H. Arnold, *Angew. Chem. Int. Ed.* **2019**, *58*, 14420–14426.
[73] U. T. Bornscheuer, B. Hauer, K. E. Jaeger, U. Schwaneberg, *Angew. Chem. Int. Ed.* **2019**, *58*, 36–40.
[74] D. Patsch, R. Buller, *Chimia* **2023**, *77*, 116–121.
[75] a) J. Damborsky, *Quan. Struct. Act. Rel.* **1997**, *16*, 126–135; b) J. Damborsky, *Protein Eng.* **1998**, *11*, 21–30; c) G. Bucht, P. Wikstrom, K. Hjalmarsson, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1999**, *1431*, 471–482; d) T. Aita, H. Uchiyama, T. Inaoka, M. Nakajima, T. Kokubo, Y. Husimi, *Biopolymers* **2000**, *54*, 64–79.
[76] P. J. Ballester, A. Schreyer, T. L. Blundell, *J. Chem. Inf. Model.* **2014**, *54*, 944–955.
[77] G. Durant, F. Boyles, K. Birchall, C. M. Deane, *Nat. Comp. Sci.* **2024**, *4*, 735–743.

Manuscript received: November 7, 2024
Revised manuscript received: November 22, 2024
Accepted manuscript online: November 25, 2024
Version of record online: December 2, 2024