

AlphaFold2 and Deep Learning for Elucidating Enzyme Conformational Flexibility and Its Application for Design

Guillem Casadevall, Cristina Duran, and Sílvia Osuna*



Cite This: *JACS Au* 2023, 3, 1554–1562



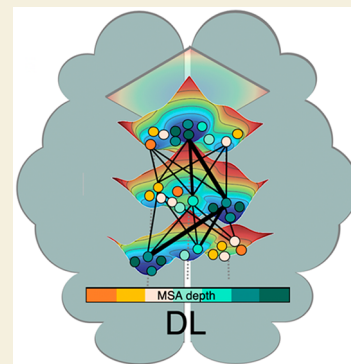
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: The recent success of AlphaFold2 (AF2) and other deep learning (DL) tools in accurately predicting the folded three-dimensional (3D) structure of proteins and enzymes has revolutionized the structural biology and protein design fields. The 3D structure indeed reveals key information on the arrangement of the catalytic machinery of enzymes and which structural elements gate the active site pocket. However, comprehending enzymatic activity requires a detailed knowledge of the chemical steps involved along the catalytic cycle and the exploration of the multiple thermally accessible conformations that enzymes adopt when in solution. In this Perspective, some of the recent studies showing the potential of AF2 in elucidating the conformational landscape of enzymes are provided. Selected examples of the key developments of AF2-based and DL methods for protein design are discussed, as well as a few enzyme design cases. These studies show the potential of AF2 and DL for allowing the routine computational design of efficient enzymes.



KEYWORDS: *AlphaFold2, conformational heterogeneity, free energy landscape, enzyme design, deep learning*

1. INTRODUCTION

The 60-year problem of knowing the folded structure from the primary sequence of proteins (and enzymes) was thought to be solved by the recent success of AlphaFold2 (AF2).^{1–3} AF2 is a deep-learning (DL) algorithm that incorporates novel neural network architectures based on the evolutionary, physical, and geometric constraints of protein structures and is able to predict with high levels of accuracy the three-dimensional structure of proteins. AF2 is recognized as one of the milestones in protein structure prediction and has boosted the application of DL methods for many other applications.⁴ Despite the impressive performance of AF2 algorithms in predicting the native lowest in energy structure of enzymes, knowing the single static folded structure is not sufficient for understanding and engineering function, as recently highlighted.^{5,6} As discussed below, another limitation of these methods is that nonprotein parts (i.e., cofactors, substrates, metal ions) are not predicted.

The three-dimensional structure of the enzymes indeed provides very relevant information on the arrangement of the catalytic machinery and structural elements gating the active-site pocket, but understanding enzymatic function requires the exploration of the ensemble of thermally accessible conformations that enzymes adopt in solution. This ensemble of conformations can be represented in the so-called Free Energy Landscape (FEL, see Figure 1 for FELs at different reaction stages),⁷ which displays the relative stabilities of the thermally accessible conformations, as well as the kinetic barriers separating them. Conformational changes that can directly

impact catalytic function include side-chain conformational changes in the fast time scale, loop motions often playing a key role in substrate binding/product release in slower time scales, and in some cases allosteric transitions that usually correspond to the slowest processes. The evaluation of the conformational landscapes of natural and evolved enzymes has provided relevant new insights. Experimental X-ray structures and associated B-factors,⁸ room-temperature X-ray experiments,^{9,10} and NMR experiments¹¹ have been used to explore the changes in the conformational landscape induced by mutations along several enzyme variants generated with the experimental Directed Evolution technique. From a computational perspective, the reconstruction of the FEL and how this is shifted after mutation provides crucial information for understanding enzyme function (and also for design).⁷

It has been recently shown by different groups that AF2 can be actually tuned to provide multiple conformations of the same protein, which suggests the potential of AF2 for elucidating the conformational landscape of enzymes and proteins.^{12,13} Given the rather low computational cost of AF2, especially if compared to the computationally demanding

Received: April 14, 2023

Revised: May 22, 2023

Accepted: May 22, 2023

Published: June 6, 2023



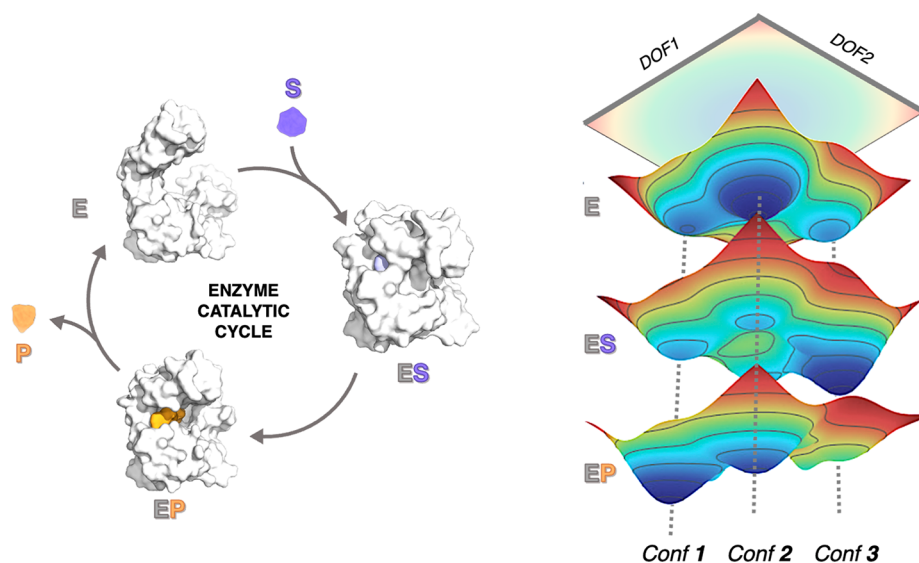


Figure 1. Schematic representation of a catalytic cycle of a model enzyme and associated conformational changes represented in the Free Energy Landscape (FEL) at the different steps: free enzyme (E), enzyme–substrate (ES), and enzyme–product (EP). For FEL reconstruction some key degrees of freedom (DOF) need to be defined, as explained in section 3.

Molecular Dynamics (MD) simulations, its application for assessing the effect of mutations on the conformational landscape is highly appealing. This could impact the development of AF2-based conformationally focused enzyme designs protocols.^{7,14}

Multiple reviews are available in the literature covering the available tools for rationalizing the changes in activity induced by mutations in several enzymes^{15–17} and for rationally designing novel enzymes by means of computational protocols based on Quantum Mechanics (QM), hybrid QM and Molecular Mechanics (QM/MM), Empirical Valence Bond (EVB), MD, and Monte Carlo simulations, or combinations of them.¹⁴ Instead, in this Perspective we will cover a few of the most recent applications of DL strategies for elucidating enzyme conformational flexibility and for its application for enzyme design.

2. THE ROLE OF CONFORMATIONAL DYNAMICS IN ENZYME FUNCTION AND EVOLUTION

Enzymes present highly preorganized active-site pockets with the catalytic residues perfectly arranged for efficiently stabilizing the transition state(s) of a specific reaction.^{18,19} This preorganization is complemented by the enzyme ability to adopt multiple conformations of importance for substrate binding and/or product release. The importance of conformational flexibility was clearly shown with the design of catalytic antibodies presenting an ideal complementary structure to the transition state, which showed a clearly inferior catalytic activity with respect to enzymes.²⁰ This shows that efficient catalysis requires not only transition-state stabilization but also the optimization of the conformational ensemble.^{14,21} In fact, the ability of enzymes to adapt and evolve toward novel functions either by natural or laboratory evolution has been connected to their inherent conformationally rich dynamic nature.^{7,22–25} Enzymes display a high degree of flexibility and versatility as shown by their promiscuous side activities²⁶ and also their tolerance to evolve toward novel functions.^{7,22–25}

Along the enzymatic cycle the following steps take place: (1) first, the substrate(s) bind to the catalytic pocket, which often

require and/or induce a change in the conformation of loops and flexible domains regulating the access to the active site;^{27,28} (2) the substrate(s) are activated to facilitate productive formation of the Enzyme–Substrate (ES) complex; (3) this is followed by the stabilization of the transition state(s) for the formation of multiple reaction intermediates and product(s); (4) finally, once the Enzyme–Product (EP) complex is formed the product(s) are released from the pocket, which is often accompanied by conformational changes that initiate the next round of the catalytic cycle. All of these steps are essential for maximizing catalytic activity by optimized throughput of the overall pathway. The binding of the substrate for ES formation can also modulate the conformational landscape as shown for the multienzyme complex pyruvate dehydrogenase complex.²⁹ In Figure 1, the conformational changes that take place along the catalytic itinerary of the enzyme adenylate kinase (AdK) is shown as a model. The catalytic cycle involves the conformational change from open to closed structures of a lid that covers the active site. The computational evaluation of the chemical steps along the catalytic itinerary (steps 2 and 3) require the use of QM, hybrid QM/MM, and EVB, which are too expensive to be applied for analyzing the conformational changes taking place through the cycle and the processes of substrate binding and product release (steps 1 and 4).^{7,14,15} This explains the large available number of computational approaches developed along the years. Current computational strategies put mostly the focus to only some of the above-mentioned features, in part explaining the often low success in achieving high levels of enzymatic activity.¹⁴

3. COMPUTATIONAL RECONSTRUCTION OF THE FREE ENERGY LANDSCAPE

The ensemble of conformations that enzymes adopt in solution can be represented in the free energy landscape (FEL). The free energy (G) is proportional to the negative logarithm of the population distribution in $k_B T$ units; thus, a maximum in this distribution is a minimum in the FEL. The FEL therefore provides crucial information on the thermodynamics (i.e.,

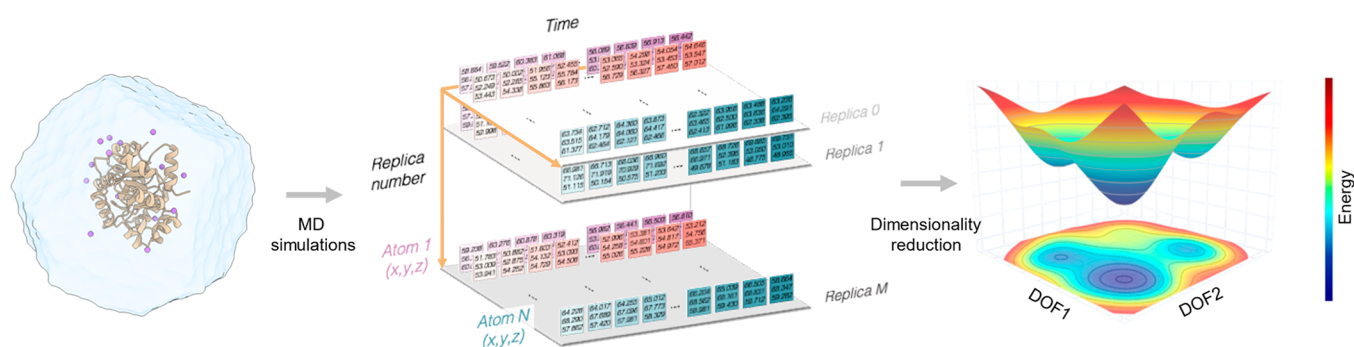


Figure 2. Schematic representation of the Free Energy Landscape (FEL) reconstruction process. The high dimensional data of the MD simulations needs to be reduced and projected into a set of key collective degrees of freedom (DOF) for probability distribution calculation to reconstruct the FFEL.

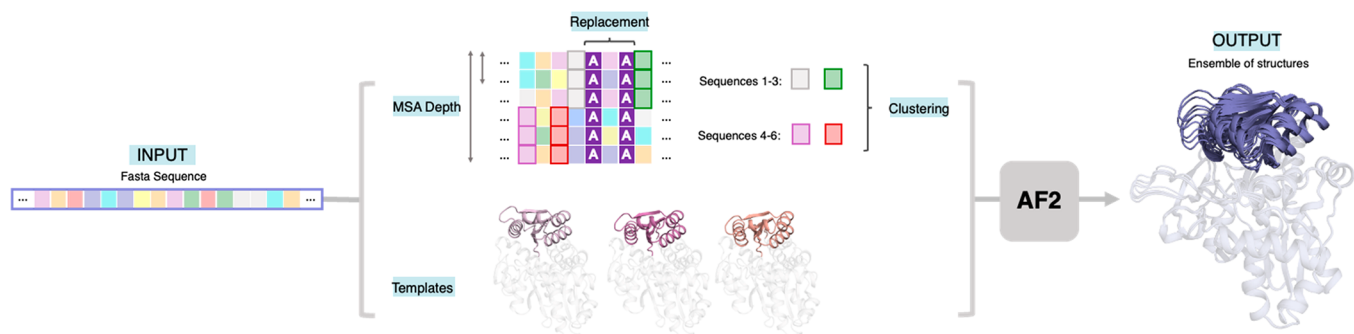


Figure 3. Overview of strategies developed for predicting alternate states with AlphaFold2 (AF2). As done in del Alamo et al.¹² the Multiple Sequence Alignment (MSA) depth can be altered, some of the MSA positions can be masked as shown by Stein and McHaourab,¹³ and the MSA can be clustered as in the Kern and Ovchinnikov preprint paper.³⁷ The provided set of templates can also be changed as done in some cases in del Alamo et al.¹² and also in our recent publication.³⁴

which are the lowest in energy conformations at a given set of conditions) and the kinetics for the conformational transitions. These energy barriers separating the different minima will determine the time scale of the conformational exchange: fast conformational changes occur in the picosecond to microsecond time scales (this is the case of loop motions crucial for enzyme catalysis), whereas slow motions will take place in millisecond to seconds.

Enzymes can be captured in different conformational states by means of X-ray, room-temperature, and time-resolved X-ray, cryo-EM, NMR, and biophysical techniques can be applied for providing complementary kinetic information.³⁰ These multiple conformations of the same enzyme deposited in the protein data bank (PDB) played an important role in AF2 training but also for the AF2 application for assessing the conformational heterogeneity of biological systems (as discussed below). Computational methods are particularly appropriate for reconstructing the FEL: MD simulations sample the population distribution by integrating Newton's laws of motion. By defining a reduced set of collective degrees of freedom (DOFs) the high dimensional data obtained in the MD runs can be projected for probability distribution calculation and thus FEL reconstruction (see Figure 2). The selection of the reduced set of DOFs can be made manually or automatically by means of different dimensionality reduction schemes.^{7,31} The accurate exploration of the conformational changes for FEL reconstruction requires extensive MD simulations, and depending on the time scale of the conformational transitions enhanced sampling techniques need to be applied.^{7,14,15} These techniques have a high

computational cost associated with them (from weeks to many months of simulations), which limits the applicability of these strategies for computationally designing and ranking enzyme designs.

4. APPLICATION OF AF2 FOR CAPTURING CONFORMATIONAL HETEROGENEITY

The standard AF2 protocol requires the primary sequence of the enzyme, a multiple sequence alignment (MSA) generated with information on evolutionary related proteins, and the 3D coordinates of a small number of homologous structures named templates (see Figure 3). Although AF2 was designed to predict single *static* structures, some recent papers have shown that by reducing the depth of the input MSAs used in the AF2 algorithm (in addition to decrease the number of recycles) accurate models in multiple conformations can be generated.^{12,13,32} In particular, del Alamo and coworkers showed that multiple conformations of transporters and G-protein-coupled receptors can be obtained by altering the AF2 pipeline and providing a reduced number of MSA sequences (as low as 16 sequences only).¹² They generated up to 50 different models of each protein receptor for each MSA size, as opposed to the standard AF2 protocol that provides conformationally homogeneous and nearly identical models. Interestingly, they observed limited conformational sampling for proteins that were contained in the AF2 training set. In another study, Stein and McHaourab reported a universal method for biasing the models generated by AF2 based on the replacement of specific residues within the MSA to alanine or another residue.¹³ AF2 was used to generate initial models, and

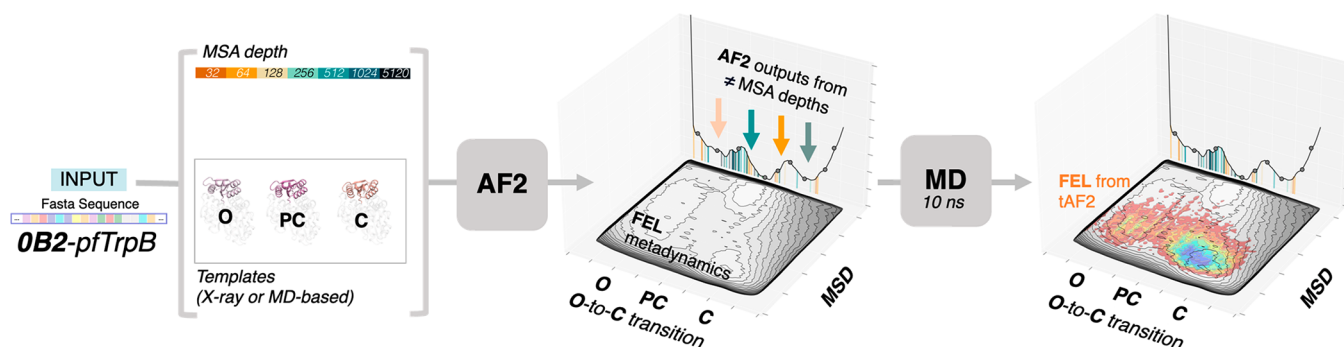


Figure 4. Our template-based AF2 (tAF2) approach for estimating the conformational heterogeneity. Different Multiple Sequence Alignment (MSA) depths and set of templates taken from either selected X-ray structures or Molecular Dynamics (MD) snapshots are provided to AF2.³⁴ The multiple output models generated by AF2 at the different MSA depths shown as vertical lines in the central plot are then subjected to short MD simulations for FEL reconstruction. The new FEL generated from the 10 ns MD simulations starting from the ca. 1000 AF2 outputs at different MSA is shown in a blue to red colormap on top of the computationally reconstructed FEL obtained via well-tempered multiple-walker metadynamics simulations (in gray).^{33,34} The x and y axis of the reconstructed FELs indicate the Open-to-Closed (O-to-C) transition of the COMM domain of TrpB that covers the active site, and the Mean Square Deviation (MSD) from the path of the generated O-to-C structures, respectively.^{33–35} The input sequence is the OB2-*pf*TrpB variant.³⁸ Reproduced with permission from ref 34. Copyright 2022 John Wiley & Sons, Inc.

the MSA was modified based on possible contact points in the initial structures, prior structural information, or regions of uncertainty within the main structure. They found that the replacement of certain amino acid columns to alanine or other residues turns the attention of the network to other parts of the MSA allowing for AF2 to find alternative conformations based on other coevolved residues. One of the provided examples is AdK that undergoes a large-scale conformational change of a lid and a flap that gate the active site, as revealed by the unbound and inhibitor-bound crystal structures (see Figure 1). By masking some residues and replacing them to alanine, closed and open structures of AdK were obtained. Although none of the AF2 open structures reached the level of opening of the crystal structure (PDB: 4AKE), the set of generated AF2 models displaying a different level of closure showed the potential of the methodology for predicting alternate conformations describing the conformational heterogeneity of the systems.

Inspired by these previous publications showing AF2's ability to sample additional conformational states, we developed a template-based AF2 approach to assess the conformational heterogeneity and how this is altered by mutations on the β subunit of several tryptophan synthase enzymes (TrpB, see Figure 4).^{33–35} As done in the work of del Alamo et al.,¹² we tested the effect of reducing the provided number of sequences in the MSA, but we additionally assessed how AF2 predictions are altered when different templates displaying multiple conformational states are provided.³⁴ We tested the template-based AF2 pipeline by providing either X-ray based or conformations extracted from MD simulations as templates. With these settings AF2 revealed major differences in the conformational landscapes among the analyzed systems. Interestingly, this was further demonstrated by running multiple short MD simulations from the set of AF2 structures and reconstructing the associated FELs (Figure 4). The comparison of the generated FEL from the template-based AF2 predictions were in line with the computationally expensive FELs generated with well-tempered multiple-walker metadynamics simulations. This is exciting as it shows the potential of AF2 for rapidly and accurately assessing the FELs of different systems, which could be applied for conforma-

tionally driven enzyme design approaches.^{7,14} The multiple outputs obtained via AF2 at different MSA depths were also recently combined with Reweighted Autoencoded Variational Bayes for Enhanced (RAVE) sampling.³⁶

In a recent preprint paper, Kern and Ovchinnikov showed that clustering the input MSA by sequence similarity allows AF2 to visit multiple conformational states of some metamorphic proteins known to display large conformational changes.³⁷ They also identified two mutations that according to AF2 predictions could switch the circadian rhythm protein KaiB between the two major conformational states. Their developed methodology was also applied for searching for alternative conformational states in other protein families and found a putative alternate state for the oxidoreductase DsbE. These computational predictions were, however, not tested experimentally in the published preprint paper.

5. APPLICATION OF AF2 AND OTHER DEEP LEARNING TECHNIQUES FOR PROTEIN AND ENZYME DESIGN

Inspired by the AF2 approach, other DL techniques have been recently developed for elucidating the folded structure of enzymes and providing some metrics to be potentially used for protein design. The field is advancing fast, and the number of DL strategies developed especially for protein design is constantly increasing. In this section, we aim to provide a brief overview of the most representative techniques developed, and we put special emphasis to those strategies that are particularly relevant for enzyme design. Some recent reviews focused on structure-based protein design with DL strategies are available,^{39,40} as well as a review related to the design of more stable enzymes.⁴¹

These available strategies for structure prediction can be classified depending on the number of input parameters used: those that require the input query sequence, MSA, and set of templates for accurate predictions and those that predict the folded structure based on the input sequence only. Similarly to AF2, the RoseTTAFold (RF) algorithm developed almost at the same time as AF2 requires an MSA and a set of initial templates to make accurate predictions of the folded structure. RF showed improved accuracy toward protein–protein

complex prediction as compared to AF2 and AF2 multi-mer.^{42,43} OpenFold2 was also developed to replicate the AF2 algorithm and make it accessible to the structural biology community.⁴⁴ AlphaLink was also introduced to incorporate experimental distance restraint information, thus generating a modified version of the AF2 network architecture.⁴⁵

Sequences contain implicit information about the enzyme structure and function, as the position of each amino acid in the sequence is determined by the spatial arrangement and the possible interactions established between them. The main advantage is the comparison of sequences is computationally cheap (at least as compared to physics-based approaches) and provides crucial information about the most frequent residues at each position, conservation score, and correlated mutation pairs that have emerged during evolution. Covarying mutations have been associated with function, tertiary contacts, and binding. It has also been shown that the use of language models previously used for Natural Language Processing (NLP) could be applied in the context of the biology language to generate “content-aware” data representations from large-scale sequence data sets. This is the case of ESM-2, which corresponds to the largest language model of protein sequences developed to date.⁴⁶ ESMFold was then developed, which was found to perform end-to-end folded structure predictions with similar accuracies to AF2, *albeit* at an order of magnitude faster.⁴⁶ OmegaFold (OF) is another end-to-end structure prediction algorithm developed that combines a pretrained language model and a geometrical transformer model for reconstructing the structure.⁴⁷ Similarly to ESMfold, OF only needs the input sequence and is 10-fold faster than AF2 and RF. More importantly, OF was found to do a better job in predicting the folded structure of orphan proteins, i.e., those proteins that do not have any assigned functional family.

Apart from the different methodologies developed to predict the folded structure of proteins, different NLP and deep-learning architectures have been developed to generate new non-natural sequences. These different strategies have targeted different objectives that range from generating new sequences for maintaining some natural activities^{48,49} to imagining new folds and sophisticated symmetric assemblies,⁵⁰ among others.

The generative language models ProGen and ProGen2 trained on millions of raw protein sequences were developed to generate de novo artificial proteins that express well and maintain enzymatic function.^{51,52} ProtGPT2 is an unsupervised language model that can generate new sequences based on the principles of natural ones.⁵³ Similarly, variational autoencoders trained on a data set of luciferase-like oxidoreductases were also used to generate new sequences maintaining the luciferase activity.⁴⁹ ProteinGAN, which is based on a self-attention-based variant of the generative adversarial network, learns natural protein sequences for generating new functional variants.⁵⁴ The conditional language model ZymCTRL trained on the BRENDA database of enzymes has also been recently developed, which is able to provide new artificial enzymes within a user-defined Enzyme Classification (EC)-based enzymatic class.⁵⁵ Language models have also been used to obtain a set of sequences that are likely to fold into a given desired structure. This is, for instance, the case for recently developed LM-Design⁵⁶ and ProteinDT⁵⁷. Yu and co-workers have recently developed CLEAN based on contrastive learning that is able to assign EC numbers to a given sequence.⁵⁸

The transform-restrained Rosetta (trRosetta) was developed by the Baker lab in 2020 to design a variety of proteins by randomly modifying the starting sequences to find sharply predicted residue–residue interdistance maps.⁵⁹ The combination of trRosetta and the physics-based Rosetta was shown to provide more funneled energy landscapes: trRosetta was used to disfavor alternative states, and high-resolution Rosetta was used for creating a deep energy minimum at the designed target structure.⁶⁰ Small β barrel proteins and proteins with discontinuous functional sites were developed with trRosetta.^{61,62} Recently, Dauparas and co-workers developed a method called ProteinMPNN, which is a graph neural network that was found to rescue previously failed designs targeted with Rosetta or AF2.⁶³ ProteinMPNN was recently applied to generate de novo luciferases.⁶⁴ MutComput is a convolutional neural network (CNN) that was successfully applied for designing new hydrolases for poly(ethylene terephthalate) depolymerization.⁶⁵ Another more recent CNN for protein sequence design was provided by Anand et al. to generate a de novo TIM-barrel protein backbone.⁶⁶ Holographic CNNs have also been developed to learn the shape of protein micro-environments to predict the impact of mutations on stability and binding of protein complexes.⁶⁷

Different protocols based on the use of AF2 for predicting the structure of the generated sequences and use the output AF2 metrics for the design of new proteins have also been developed. The AlphaDesign computational framework was constructed to enable the rapid prediction of completely novel protein monomers starting from random sequences.⁶⁸ The potential application of AlphaDesign for designing proteins that bind to prespecific target proteins was also shown. AF2 was also used for the rapid and accurate fixed backbone design of sequences that are strongly predicted to fold to a specific backbone.⁶⁹ The Baker lab combined ProteinMPNN with AF2 to design closed repeat proteins with central pockets⁵⁰ and generate symmetric protein assemblies.⁵⁰ Similarly, RF instead of AF2 was used for designing high-affinity protein binders⁷⁰ or proteins with prespecified functional motifs.⁷¹ RF has also the potential to predict the effect of mutations on protein function.⁷²

The RF-based diffusion model (named RFdiffusion) has been recently developed by the Baker lab.⁷³ RFdiffusion can very rapidly and accurately design topology-constrained protein monomers, protein binders, symmetric oligomers, metal-binding proteins, and even enzyme scaffolds containing specific active-site residues.⁷³ The performance of RFdiffusion outperforms hallucination in terms of success rate, accuracy, and speed. Even though RFdiffusion does not explicitly consider the substrate molecule, it can be implicitly modeled using an external potential to guide the generation of the active-site pocket.

As mentioned in the **Introduction**, catalytic function requires substrate binding and product release, and in many cases enzymatic activity is dependent on cofactor and metal ion binding. In this direction, different strategies based on DL have also been generated to dock ligands, substrates, and missing cofactors into potential pockets. AlphaFill uses sequence and structural similarity to include the missing organic molecules and metal ions into the AF2 models.⁷⁴ The diffusion generative model DiffDock was designed to dock small molecules into potential protein pockets. This strategy was shown to outperform previous traditional and DL docking protocols.⁷⁵ Meller and co-workers also developed an AF2-based strategy to

find cryptic pockets.⁷⁶ DL has also been applied for finding potential location sites of transition metals in proteins (Metal1D and Metal3D).⁷⁷ The coevolution based MetalNet pipeline has also been recently created to predict potential metal-binding sites.⁷⁸

6. OUTLOOK AND FUTURE PERSPECTIVES

Enzyme catalysis is a complex multidimensional process that requires the optimal sequence and structure for allowing substrate(s) binding, catalyzing the chemical steps and product(s) release, and optimizing the multiple conformations needed for developing its function. This high complexity makes the task of enzyme design, especially toward non-natural reactions or substrates in high efficiencies very challenging. The selected examples highlighted in this review show the potential of DL techniques to generate new functional variants mostly within the allowed biological constraints of the sequence space. The application of DL strategies for computational enzyme design for any target reaction and non-natural substrate is only at its beginning. For many years, the lack of precision in incorporating the desired active-site residues into protein scaffolds in computational enzyme design has been considered one of the many limitations of the overall process. This point, however, seems to be solved with the recent RosettaFold-based diffusion model developed by the Baker lab. The incorporation of the QM-based models of the enzyme active site into new non-natural scaffolds specifically designed to hold the functional motifs in place might no longer be the limiting factor, but instead predicting which scaffolds might be more appropriate for the optimization of the conformational ensemble for efficient catalysis will most likely be essential. Considering the huge advances especially in the field of structure prediction and protein design seen these recent years, the combination of DL methods with physics-based approaches will play a key role the coming years for finding optimal solutions for the rational and routine design of highly efficient and stable enzymes for non-natural reactions and substrates.

AUTHOR INFORMATION

Corresponding Author

Silvia Osuna – Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, 17003 Girona, Spain; ICREA, 08010 Barcelona, Spain; orcid.org/0000-0003-3657-6469; Email: silvia.osuna@udg.edu

Authors

Guillem Casadevall – Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, 17003 Girona, Spain

Cristina Duran – Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, 17003 Girona, Spain

Complete contact information is available at: <https://pubs.acs.org/10.1021/jacsau.3c00188>

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. CRediT: **Guillem Casadevall** writing-review &

editing; **Cristina Duran** writing-review & editing; **Silvia Osuna** writing-review & editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the Generalitat de Catalunya for the consolidated group TCBioSys (SGR 2021 00487) and grant projects PID2021-129034NB-I00 and PDC2022-133950-I00 funded by Spanish MICIN. S.O. is grateful for the funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-2015-StG-679001, ERC-2022-POC-101112805, and ERC-2022-CoG-101088032) and the Human Frontier Science Program (HFSP) for project grant RGP0054/2020. G.C. was supported by a research grant from ERC-StG (ERC-2015-StG-679001) and HFSP RGP0054/2020. C.D. was supported by the Spanish MINECO for a PhD fellowship (PRE2019-089147).

ABBREVIATIONS

AF2, AlphaFold2; DL, Deep Learning; FEL, Free Energy Landscape; MD, Molecular Dynamics

REFERENCES

- Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577* (7792), 706–710.
- Ourmazd, A.; Moffat, K.; Lattman, E. E. Structural biology is solved — now what? *Nat. Methods* **2022**, *19* (1), 24–26.
- Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.
- Callaway, E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* **2020**, *588* (7837), 203–204.
- Clementi, C. Fast track to structural biology. *Nat. Chem.* **2021**, *13* (11), 1032–1034.
- Jones, D. T.; Thornton, J. M. The impact of AlphaFold2 one year on. *Nat. Methods* **2022**, *19* (1), 15–20.
- Maria-Solano, M. A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S. Role of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun.* **2018**, *54* (50), 6622–6634.
- Campbell, E.; Kaltenbach, M.; Correy, G. J.; Carr, P. D.; Porebski, B. T.; Livingstone, E. K.; Afriat-Jurnou, L.; Buckle, A. M.; Weik, M.; Hollfelder, F.; Tokuriki, N.; Jackson, C. J. The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **2016**, *12* (11), 944–950.
- Fraser, J. S.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (39), 16247–16252.
- Broom, A.; Rakotoharisoa, R. V.; Thompson, M. C.; Zarifi, N.; Nguyen, E.; Mukhametzanov, N.; Liu, L.; Fraser, J. S.; Chica, R. A. Ensemble-based enzyme design can recapitulate the effects of

- laboratory directed evolution in silico. *Nat. Commun.* **2020**, *11* (1), 4808.
- (11) Otten, R.; Pádua, R. A. P.; Bunzel, H. A.; Nguyen, V.; Pitsawong, W.; Patterson, M.; Sui, S.; Perry, S. L.; Cohen, A. E.; Hilvert, D.; Kern, D. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **2020**, *370* (6523), 1442–1446.
- (12) del Alamo, D.; Sala, D.; McHaourab, H. S.; Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* **2022**, *11*, e75751.
- (13) Stein, R. A.; McHaourab, H. S. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with AlphaFold2. *PLoS Comput. Biol.* **2022**, *18* (8), e1010483.
- (14) Osuna, S. The challenge of predicting distal active site mutations in computational enzyme design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, e1502.
- (15) Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. Computational tools for the evaluation of laboratory-engineered biocatalysts. *Chem. Commun.* **2017**, *53* (2), 284–297.
- (16) Himo, F.; de Visser, S. P. Status report on the quantum chemical cluster approach for modeling enzyme reactions. *Commun. Chem.* **2022**, *5* (1), 29.
- (17) Świderek, K.; Tuñón, I.; Moliner, V. Predicting enzymatic reactivity: from theory to design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (5), 407–421.
- (18) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106* (8), 3210–3235.
- (19) Marti, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.* **2004**, *33* (2), 98–107.
- (20) Winkler, C. K.; Schrittwieser, J. H.; Kroutil, W. Power of Biocatalysis for Organic Synthesis. *ACS Cent. Sci.* **2021**, *7* (1), 55–71.
- (21) Lovelock, S. L.; Crawshaw, R.; Basler, S.; Levy, C.; Baker, D.; Hilvert, D.; Green, A. P. The road to fully programmable protein catalysis. *Nature* **2022**, *606* (7912), 49–58.
- (22) Tokuriki, N.; Tawfik, D. S. Protein Dynamism and Evolvability. *Science* **2009**, *324* (5924), 203–207.
- (23) Petrović, D.; Risso, V. A.; Kamerlin, S. C. L.; Sanchez-Ruiz, J. M. Conformational dynamics and enzyme evolution. *J. R. Soc. Interface* **2018**, *15* (144), 20180330.
- (24) Campbell, E. C.; Correy, G. J.; Mabbitt, P. D.; Buckle, A. M.; Tokuriki, N.; Jackson, C. J. Laboratory evolution of protein conformational dynamics. *Curr. Opin. Struct. Biol.* **2018**, *50*, 49–57.
- (25) Crean, R. M.; Gardner, J. M.; Kamerlin, S. C. L. Harnessing Conformational Plasticity to Generate Designer Enzymes. *J. Am. Chem. Soc.* **2020**, *142* (26), 11324–11342.
- (26) Khersonsky, O.; Tawfik, D. S. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **2010**, *79*, 471–505.
- (27) Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009**, *5* (11), 789–796.
- (28) Hammes, G. G.; Benkovic, S. J.; Hammes-Schiffer, S. Flexibility, Diversity, and Cooperativity: Pillars of Enzyme Catalysis. *Biochem.* **2011**, *50* (48), 10422–10430.
- (29) Prajapati, S.; Haselbach, D.; Wittig, S.; Patel, M. S.; Chari, A.; Schmidt, C.; Stark, H.; Tittmann, K. Structural and Functional Analyses of the Human PDH Complex Suggest a “Division-of-Labor” Mechanism by Local E1 and E3 Clusters. *Structure* **2019**, *27* (7), 1124–1136.
- (30) Baldwin, A. J.; Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **2009**, *5* (11), 808–814.
- (31) Glielmo, A.; Husic, B. E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised Learning Methods for Molecular Simulation Data. *Chem. Rev.* **2021**, *121* (16), 9722–9758.
- (32) Roney, J. P.; Ovchinnikov, S. State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold. *Phys. Rev. Lett.* **2022**, *129* (23), 238101.
- (33) Maria-Solano, M. A.; Iglesias-Fernández, J.; Osuna, S. Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution. *J. Am. Chem. Soc.* **2019**, *141* (33), 13049–13056.
- (34) Casadevall, G.; Duran, C.; Estévez-Gay, M.; Osuna, S. Estimating conformational heterogeneity of tryptophan synthase with a template-based AlphaFold2 approach. *Protein Sci.* **2022**, *31* (10), e4426.
- (35) Maria-Solano, M. A.; Kinateder, T.; Iglesias-Fernández, J.; Sterner, R.; Osuna, S. In Silico Identification and Experimental Validation of Distal Activity-Enhancing Mutations in Tryptophan Synthase. *ACS Catal.* **2021**, *11* (21), 13733–13743.
- (36) Vani, B. P.; Aranganathan, A.; Wang, D.; Tiwary, P. AlphaFold2-RAVE: From Sequence to Boltzmann Ranking. *J. Chem. Theory Comput.* **2023**, DOI: 10.1021/acs.jctc.3c00290.
- (37) Wayment-Steele, H. K.; Ovchinnikov, S.; Colwell, L.; Kern, D. Prediction of multiple conformational states by combining sequence clustering with AlphaFold2. *bioRxiv* **2022**, DOI: 10.1101/2022.10.17.512570.
- (38) Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H. Directed evolution of the tryptophan synthase β -subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (47), 14599–14604.
- (39) Ovchinnikov, S.; Huang, P.-S. Structure-based protein design with deep learning. *Curr. Opin. Chem. Biol.* **2021**, *65*, 136–144.
- (40) Gao, W.; Mahajan, S. P.; Sulam, J.; Gray, J. J. Deep Learning in Protein Structural Modeling and Design. *Patterns* **2020**, *1* (9), 100142.
- (41) Ming, Y.; Wang, W.; Yin, R.; Zeng, M.; Tang, L.; Tang, S.; Li, M. A review of enzyme design in catalytic stability by artificial intelligence. *Brief. Bioinform.* **2023**, 1–19.
- (42) Baek, M.; Baker, D. Deep learning and protein structure modeling. *Nat. Methods* **2022**, *19* (1), 13–14.
- (43) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; Dijk, A. A. v.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373* (6557), 871–876.
- (44) Ahdriz, G.; Bouatta, N.; Kadyan, S.; Xia, Q.; Gerecke, W.; O'Donnell, T. J.; Berenberg, D.; Fisk, I.; Zanichelli, N.; Zhang, B.; Nowaczynski, A.; Wang, B.; Stepniewska-Dziubinska, M. M.; Zhang, S.; Ojewole, A.; Guney, M. E.; Biderman, S.; Watkins, A. M.; Ra, S.; Lorenzo, P. R.; Nivon, L.; Weitzner, B.; Ban, Y.-E. A.; Sorger, P. K.; Mostaque, E.; Zhang, Z.; Bonneau, R.; AlQuraishi, M. OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *bioRxiv* **2022**, DOI: 10.1101/2022.11.20.517210.
- (45) Stahl, K.; Graziadei, A.; Dau, T.; Brock, O.; Rappsilber, J. Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nat. Biotechnol.* **2023**, DOI: 10.1038/s41587-023-01704-z.
- (46) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; Costa, A. d. S.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379* (6637), 1123–1130.
- (47) Wu, R.; Ding, F.; Wang, R.; Shen, R.; Zhang, X.; Luo, S.; Su, C.; Wu, Z.; Xie, Q.; Berger, B.; Ma, J.; Peng, J. High-resolution de novo structure prediction from primary sequence. *bioRxiv* **2022**, DOI: 10.1101/2022.07.21.500999.
- (48) Madani, A.; Krause, B.; Greene, E. R.; Subramanian, S.; Mohr, B. P.; Holton, J. M.; Olmos, J. L.; Xiong, C.; Sun, Z. Z.; Socher, R.; Fraser, J. S.; Naik, N. Deep neural language modeling enables

functional protein generation across families. *bioRxiv* **2021**, DOI: 10.1101/2021.07.18.452833.

(49) Hawkins-Hooker, A.; Depardieu, F.; Baur, S.; Couairon, G.; Chen, A.; Bikard, D. Generating functional protein variants with variational autoencoders. *PLoS Comput. Biol.* **2021**, *17* (2), e1008736.

(50) Wicky, B. I. M.; Milles, L. F.; Courbet, A.; Ragotte, R. J.; Dauparas, J.; Kinfu, E.; Tipps, S.; Kibler, R. D.; Baek, M.; DiMaio, F.; Li, X.; Carter, L.; Kang, A.; Nguyen, H.; Bera, A. K.; Baker, D. Hallucinating symmetric protein assemblies. *Science* **2022**, *378* (6615), 56–61.

(51) Madani, A.; Krause, B.; Greene, E. R.; Subramanian, S.; Mohr, B. P.; Holton, J. M.; Olmos, J. L.; Xiong, C.; Sun, Z. Z.; Socher, R.; Fraser, J. S.; Naik, N. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **2023**, DOI: 10.1038/s41587-022-01618-2.

(52) Nijkamp, E.; Ruffolo, J.; Weinstein, E. N.; Naik, N.; Madani, A. ProGen2: Exploring the Boundaries of Protein Language Models. *arXiv* **2022**, DOI: 10.48550/arXiv.2206.13517.

(53) Ferruz, N.; Schmidt, S.; Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* **2022**, *13* (1), 4348.

(54) Repecka, D.; Jauniskis, V.; Karpus, L.; Rembeza, E.; Rokaitis, I.; Zrimec, J.; Poviloniene, S.; Laurynenas, A.; Viknander, S.; Abuajwa, W.; Savolainen, O.; Meskys, R.; Engqvist, M. K. M.; Zelezniak, A. Expanding functional protein sequence spaces using generative adversarial networks. *Nat. Mach. Intell.* **2021**, *3* (4), 324–333.

(55) Munsamy, G.; Lindner, S.; Lorenz, P.; Ferruz, N. Zymctrl: a conditional language model for the controllable generation of artificial enzymes. In *NeurIPS Machine Learning in Structural Biology Workshop*, 2022.

(56) Zheng, Z.; Deng, Y.; Xue, D.; Zhou, Y.; YE, F.; Gu, Q. Structure-informed Language Models Are Protein Designers. *arXiv* **2023**, DOI: 10.48550/arXiv.2302.01649.

(57) Liu, S.; Zhu, Y.; Lu, J.; Xu, Z.; Nie, W.; Gitter, A.; Xiao, C.; Tang, J.; Guo, H.; Anandkumar, A. A Text-guided Protein Design Framework. *arXiv* **2023**, DOI: 10.48550/arXiv.2302.04611.

(58) Yu, T.; Cui, H.; Li, J. C.; Luo, Y.; Jiang, G.; Zhao, H. Enzyme function prediction using contrastive learning. *Science* **2023**, *379* (6639), 1358–1363.

(59) Anishchenko, I.; Pellock, S. J.; Chidyausiku, T. M.; Ramelot, T. A.; Ovchinnikov, S.; Hao, J.; Bafna, K.; Norn, C.; Kang, A.; Bera, A. K.; DiMaio, F.; Carter, L.; Chow, C. M.; Montelione, G. T.; Baker, D. De novo protein design by deep network hallucination. *Nature* **2021**, *600* (7889), 547–552.

(60) Norn, C.; Wicky, B. I. M.; Juergens, D.; Liu, S.; Kim, D.; Tischer, D.; Koepnick, B.; Anishchenko, I.; Baker, D.; Ovchinnikov, S.; Coral, A.; Bubar, A. J.; Boykov, A.; Pérez, A. U. V.; MacMillan, A.; Lubow, A.; Mussini, A.; Cai, A.; Ardill, A. J.; Seal, A.; Kalantarian, A.; Failer, B.; Lackersteen, B.; Chagot, B.; Haight, B. R.; Taştan, B.; Uitham, B.; Roy, B. G.; Cruz, B. R. D. M.; Echols, B.; Lorenz, B. E.; Blair, B.; Kestemont, B.; Eastlake, C. D.; Bragdon, C. J.; Vardeman, C.; Salerno, C.; Comisky, C.; Hayman, C. L.; Landers, C. R.; Zimov, C.; Coleman, C. D.; Painter, C. R.; Ince, C.; Lynagh, C.; Malaniia, D.; Wheeler, D. C.; Robertson, D.; Simon, V.; Chisari, E.; Kai, E. L. J.; Rezae, F.; Lengyel, F.; Tabotta, F.; Padelletti, F.; Boström, F.; Gross, G. O.; McIlvaine, G.; Beecher, G.; Hansen, G. T.; Jong, G. d.; Feldmann, H.; Borman, J. L.; Quinn, J.; Norrgard, J.; Truong, J.; Diderich, J. A.; Canfield, J. M.; Photakis, J.; Slone, J. D.; Madzio, J.; Mitchell, J.; Stomieroski, J. C.; Mitch, J. H.; Altenbeck, J. R.; Schinkler, J.; Weinberg, J. B.; Burbach, J. D.; Costa, J. C. S. d.; Juarez, J. F. B.; Gunnarsson, J. P.; Harper, K. D.; Joo, K.; Clayton, K. T.; DeFord, K. E.; Scully, K. F.; Gildea, K. M.; Abbey, K. J.; Kohli, K. L.; Stenner, K.; Takács, K.; Poussaint, L. L.; Manalo, L. C.; Withers, L. C.; Carlson, L.; Wei, L.; Fisher, L. R.; Carpenter, L.; Ji-hwan, M.; Ricci, M.; Belcastro, M. A.; Leniec, M.; Hohmann, M.; Thompson, M.; Thayer, M. A.; Gaebel, M.; Cassidy, M. D.; Fagiola, M.; Lewis, M.; Pfützenreuter, M.; Simon, M.; Elmassry, M. M.; Benevides, N.; Kerr, N. K.; Verma, N.; Shannon, O.; Yin, O.; Wolfteich, P.; Gummersall, P.; Thlušcik, P.; Gajar, P.; Triggiani, P. J.; Guha, R.; Innes,

R. B. M.; Buchanan, R.; Gamble, R.; Leduc, R.; Spearing, R.; Gomes, R. L. C. d. S.; Estep, R. D.; DeWitt, R.; Moore, R.; Shneider, S. G.; Zaccanelli, S. J.; Kuznetsov, S.; Burillo-Sanz, S.; Rooney, S.; Vasiliy, S.; Butkovich, S. S.; Hudson, S. B.; Pote, S. L.; Denne, S. P.; Schwegmann, S. A.; Ratna, S.; Kleinfelder, S. C.; Bausewein, T.; George, T. J.; Almeida, T. S. d.; Yeginer, U.; Barmettler, W.; Pulley, W. R.; Wright, W. S.; Willyanto; Lansford, W.; Hochart, X.; Gaiji, Y. A. S.; Lagodich, Y.; Christian, V. Protein sequence design by conformational landscape optimization. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118* (11), e2017228118.

(61) Kim, D. E.; Jensen, D. R.; Feldman, D.; Tischer, D.; Saleem, A.; Chow, C. M.; Li, X.; Carter, L.; Milles, L.; Nguyen, H.; Kang, A.; Bera, A. K.; Peterson, F. C.; Volkman, B. F.; Ovchinnikov, S.; Baker, D. De novo design of small beta barrel proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (11), e2207974120.

(62) Tischer, D.; Lisanza, S.; Wang, J.; Dong, R.; Anishchenko, I.; Milles, L. F.; Ovchinnikov, S.; Baker, D. Design of proteins presenting discontinuous functional sites using deep learning. *bioRxiv* **2020**, DOI: 10.1101/2020.11.29.402743.

(63) Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; Haas, R. J. d.; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **2022**, *378* (6615), 49–56.

(64) Yeh, A. H.-W.; Norn, C.; Kipnis, Y.; Tischer, D.; Pellock, S. J.; Evans, D.; Ma, P.; Lee, G. R.; Zhang, J. Z.; Anishchenko, I.; Coventry, B.; Cao, L.; Dauparas, J.; Halabiya, S.; DeWitt, M.; Carter, L.; Houk, K. N.; Baker, D. De novo design of luciferases using deep learning. *Nature* **2023**, *614* (7949), 774–780.

(65) Lu, H.; Diaz, D. J.; Czarniecki, N. J.; Zhu, C.; Kim, W.; Shroff, R.; Acosta, D. J.; Alexander, B. R.; Cole, H. O.; Zhang, Y.; Lynd, N. A.; Ellington, A. D.; Alper, H. S. Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* **2022**, *604* (7907), 662–667.

(66) Anand, N.; Eguchi, R.; Mathews, I. I.; Perez, C. P.; Derry, A.; Altman, R. B.; Huang, P.-S. Protein sequence design with a learned potential. *Nat. Commun.* **2022**, *13* (1), 746.

(67) Pun, M. N.; Ivanov, A.; Bellamy, Q.; Montague, Z.; LaMont, C.; Bradley, P.; Otwinowski, J.; Nourmohammad, A. Learning the shape of protein micro-environments with a holographic convolutional neural network. *arXiv* **2022**, DOI: 10.1101/2022.10.31.514614.

(68) Jendrusch, M.; Korbel, J. O.; Sadiq, S. K. AlphaDesign: A de novo protein design framework based on AlphaFold. *bioRxiv* **2021**, DOI: 10.1101/2021.10.11.463937.

(69) Moffat, L.; Greener, J. G.; Jones, D. T. Using AlphaFold for Rapid and Accurate Fixed Backbone Protein Design. *bioRxiv* **2021**, DOI: 10.1101/2021.08.24.457549.

(70) Torres, S. V.; Leung, P. J. Y.; Lutz, I. D.; Venkatesh, P.; Watson, J. L.; Hink, F.; Huynh, H.-H.; Yeh, A. H.-W.; Juergens, D.; Bennett, N. R.; Hoofnagle, A. N.; Huang, E.; MacCoss, M. J.; Expósito, M.; Lee, G. R.; Korkmaz, E. N.; Nivala, J.; Stewart, L.; Rogers, J. M.; Baker, D. De novo design of high-affinity protein binders to bioactive helical peptides. *bioRxiv* **2022**, DOI: 10.1101/2022.12.10.519862.

(71) Wang, J.; Lisanza, S.; Juergens, D.; Tischer, D.; Watson, J. L.; Castro, K. M.; Ragotte, R.; Saragovi, A.; Milles, L. F.; Baek, M.; Anishchenko, I.; Yang, W.; Hicks, D. R.; Expósito, M.; Schlichthaerle, T.; Chun, J.-H.; Dauparas, J.; Bennett, N.; Wicky, B. I. M.; Muenks, A.; DiMaio, F.; Correia, B.; Ovchinnikov, S.; Baker, D. Scaffolding protein functional sites using deep learning. *Science* **2022**, *377* (6604), 387–394.

(72) Mansoor, S.; Baek, M.; Juergens, D.; Watson, J. L.; Baker, D. Accurate Mutation Effect Prediction using RoseTTAFold. *bioRxiv* **2022**, DOI: 10.1101/2022.11.04.515218.

(73) Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; De Bortoli, V.; Mathieu, E.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek,

M.; Baker, D. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *bioRxiv* **2022**, DOI: [10.1101/2022.12.09.519842](https://doi.org/10.1101/2022.12.09.519842).

(74) Hekkelman, M. L.; de Vries, L.; Joosten, R. P.; Perrakis, A. AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat. Methods* **2023**, *20* (2), 205–213.

(75) Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. *arXiv* **2023**, DOI: [10.48550/arXiv.2210.01776](https://doi.org/10.48550/arXiv.2210.01776).

(76) Meller, A.; Bhakat, S.; Solieva, S.; Bowman, G. R. Accelerating Cryptic Pocket Discovery Using AlphaFold. *J. Chem. Theory Comput.* **2023**, DOI: [10.1021/acs.jctc.2c01189](https://doi.org/10.1021/acs.jctc.2c01189).

(77) Dürr, S. L.; Levy, A.; Rothlisberger, U. Accurate prediction of transition metal ion location via deep learning. *bioRxiv* **2022**, DOI: [10.1101/2022.08.22.504853](https://doi.org/10.1101/2022.08.22.504853).

(78) Cheng, Y.; Wang, H.; Xu, H.; Liu, Y.; Ma, B.; Chen, X.; Zeng, X.; Wang, X.; Wang, B.; Shiao, C.; Ovchinnikov, S.; Su, X.-D.; Wang, C. Co-evolution-based prediction of metal-binding sites in proteomes by machine learning. *Nat. Chem. Biol.* **2023**, *19*, 548–555.