

Perspective

Bridging Algorithms and Biocatalysis: Perspectives on AI-Supported Enzyme Engineering

Rosa Teijeiro-Juiz ¹, Thomas Brück ² and Bernhard Loll ^{1,*}

¹ Laboratory of Structural Biochemistry, Institute of Chemistry and Biochemistry, Freie Universität Berlin, 14195 Berlin, Germany; rosa.teijeiro.juiz@fu-berlin.de

² Werner Siemens-Chair of Synthetic Biotechnology, TUM School of Natural Sciences, Technical University of Munich (TUM), 85748 Garching, Germany; brueck@tum.de

* Correspondence: loll@chemie.fu-berlin.de

Abstract

The combination of computational and experimental methods has become indispensable for optimization and rational enzyme design. Recently, the development of artificial intelligence (AI)-based tools has further streamlined enzyme engineering pipelines, enabling more accurate designs, while reducing the number of variants required for experimental validation. However, due to the intricate complexity of enzymatic systems, significant challenges must be addressed before we take the next step to fully optimize the use of these AI-guided enzyme design methodologies. These challenges include un-curated datasets, the need to consider both the static and dynamic structure of enzymes, and the requirement for effective interdisciplinary collaborations to ensure the integration of computational and experimental approaches. Here, we present recent advances in AI-based computational enzyme design, discussing the main challenges in the field and how a combination with classical physics-based methods could help overcome them. We further explore novel trends that could completely modulate the future of protein design and provide our outlook on the key concepts and future opportunities that will shape the next steps of enzyme design.

Keywords: enzyme design; protein engineering; artificial intelligence; protein stability; de novo design; non-canonical amino acids

1. Introduction

Enzymes serve as highly efficient catalysts mediating the acceleration of chemical reactions happening in all living organisms [1–4]. The precision and increased catalytic rate enzymes grant over the reaction outcome has motivated researchers to employ these proteins with the goal of overcoming some of the challenges emerging from traditional chemical processes [5]. In the last two decades, biocatalysis has become crucial for green and sustainable industrial biotechnological processes within biopharmaceutical, food or biofuel sectors, among others [6–9]. During these industrial processes, enzymes must resist extreme environments, ranging from high temperatures to non-natural pH conditions, or withstand organic solvents. These unnatural conditions could compromise the stability of the enzymes and hinder their ability to optimally catalyze and perform the reaction of interest [10]. Thereby, tailoring enzymatic systems for the catalysis of specific chemical reactions, sometimes under rather harsh industrial conditions, has become a main task for biochemists/biotechnologists.



Academic Editor: Alessio Peracchi

Received: 10 June 2026

Revised: 1 July 2026

Accepted: 2 July 2026

Published: 4 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Traditionally, researchers based novel enzyme design on available structural information of the respective protein. Enzyme function could rationally be altered by targeting specific amino acid residues, unlocking a desired change in the properties of the enzyme. Historically, this was performed through a directed evolution approach, which involves random mutagenesis on the gene of interest, to then screen and select enzyme variants that show improvement or the desired properties [11,12]. While this methodology has demonstrated positive results, directed evolution generates thousands of variants that then need to be experimentally validated, turning this process into a cost and time expensive approach for enzyme design [13,14]. More recently, the rise of computational approaches, which study dynamics and mechanics of the system, in combination with this traditional evolutionary method, has produced outstanding results, managing to design enzymes which, in some cases, could even rival the efficiency of natural catalytic proteins [15].

Current advances in AI, specifically in deep learning approaches for prediction of protein structure and sequence design, have transformed and accelerated our ability to accurately engineer biological systems [16,17]. By learning from curated databases, AI methods can predict which alterations in the sequence of the protein would allow for the improvement of different properties like stability or activity. However, despite the rapid progress in the protein design field, enzyme engineering remains challenging due to the complexity of enzymatic systems [18–20].

In this perspective article, we discuss the current advantages and limitations AI-guided enzyme design offers, highlight main challenges that remain unsolved, and outline emerging trends in the field that may shape the future of enzyme engineering.

2. Challenges in Predicting and Designing Enzymes

Enzymatic systems are extensively complex. Holding a deep understanding of the structure and function of the biocatalyst is required, but not enough, for a complete rational and optimized enzyme design approach [14]. Earlier protein design approaches, as applied for the Kemp elimination, were based on computed transition state geometries [21]. The protein variants were generated by error-prone PCR. Recently, a fully computational workflow improved the catalytic activity of the Kemp eliminase [22]. The synergy of computational and experimental methods is imperative, and newly developed AI-driven tools are gaining traction due to their ability to drastically improve the accuracy when engineering new enzyme variants [16]. Despite all the advances and improvements in the design of biocatalysts, scientists are still facing a wide array of challenges [23–25]. In this section we want to focus on some of the major bottlenecks in the computational–experimental pipelines for enzyme design strategies (Figure 1).

One of the biggest challenges AI tools face is the lack of quality structural and functional data [26,27]. An elevated number of enzyme systems lack high-resolution crystal structures, and, while in recent years advances in structure prediction have provided scientists with the ability to predict enzyme models that can then be employed for enzyme design analysis, it is crucial to remember that in the end, these are just models, not real, verified, experimental data [27]. In the absence of proper comprehensive functional and structural datasets, how reliable are the trained machine learning (ML) models for enzyme design? Moreover, not only structural datasets are missing, but there is also a lack of standardized kinetic constants (k_{cat} , K_M) in the published datasets [28]. Accessibility to this information and the possibility of feeding these data into algorithms would potentially improve their predictive power towards enzymatic activity. In addition, there is a dataset bias towards successful variants, most frequently completely ignoring negative results and resulting in a data imbalance that can affect the algorithm's predictions [29,30]. Feeding negative results into AI-driven tools would allow them to learn which changes in the

enzyme do not bring the desired properties to the system. Without the incorporation of negative results, a model might have the tendency to become overly optimistic and may predict many proteins as active even though they are not. In general, inclusion of negative results could teach the AI decision making between success and failures. The reliability of AI predictions is ultimately constrained by the quality and diversity of available training data, which is why putting focus into obtaining more structural and functional data, as well as documenting negative results, should be a major focus in the future [31].

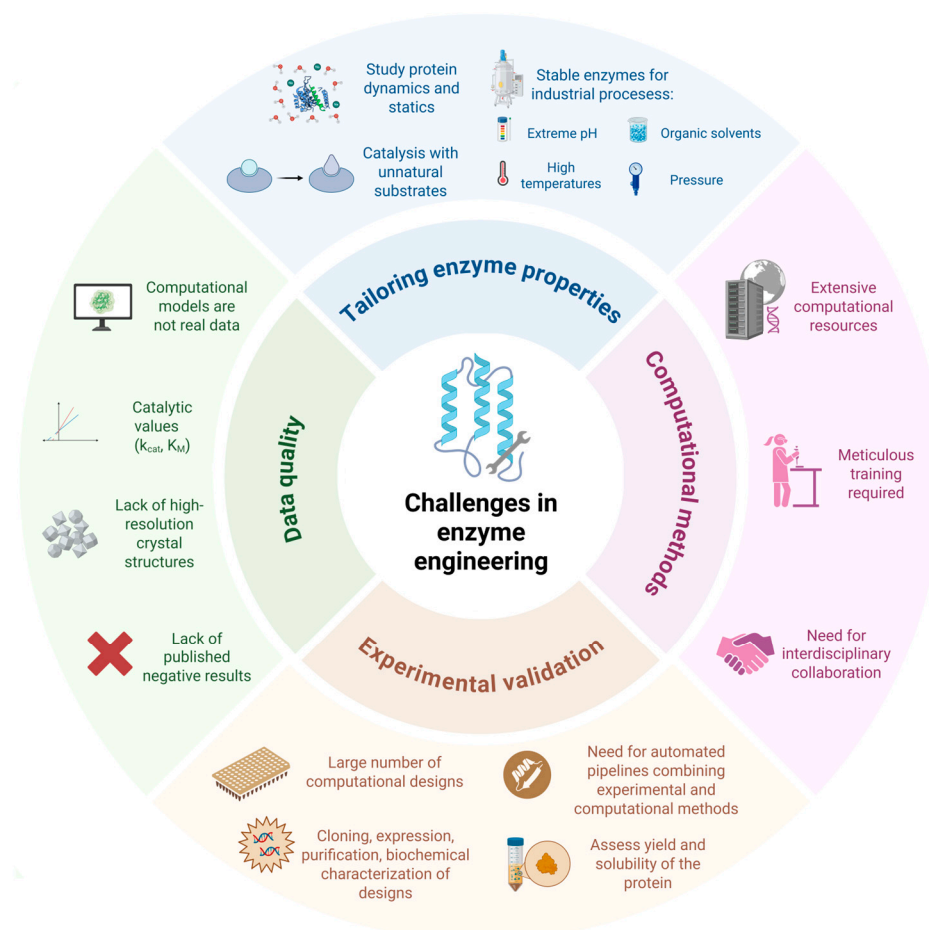


Figure 1. Current challenges in enzyme design for experiments, which can be sub-divided in for major categories: tailoring enzyme properties, computational methods, experimental validation, and data quality. The figure highlights that enzyme engineering is a multidisciplinary field where progress depends on balancing accurate data, advanced computational modeling, effective experimental validation, and the ability to tailor enzyme properties for specific applications. Overcoming these challenges is essential for developing robust enzymes for industrial, environmental, and biomedical applications. The data quality is important for training data sets for AI tools. Ideally, these data should include negative results. Based on the experimental data, transition state barriers could be calculated by MD or QM/MM methods. Protein or fold designs with increased stability and activity could be generated with AI-based tools such as AF3, ESMFold, or RoseTTAFold. The interaction with experimental biochemistry is important to validated to computer-based novel designs. This might include determination of the protein melting temperature to confirm an increased protein stability or for instance an enzymatic characterization of the kinetic parameters. Ideally, such experimental data flow back into the training sets.

Since computational approaches have become a main component of enzyme design pipelines, multiple tools and algorithms have been developed with the goal of improving the rational tailoring of biocatalysts [16,32]. Some of these methods, apart from being time-consuming, also require extensive computational resources. Moreover, to use some of

the more complex tools, meticulous training is required. For trained experimentalists who lack experience in computational software, it can be very tedious to properly learn how to use these tools. The increasing methodological complexity of computational tools creates barriers for experimental researchers, highlighting the need for public accessible platforms and interdisciplinary collaboration.

As previously stated, tailoring robust enzymes for biotechnological industrial processes is currently a point of interest within the field of protein design [33]. In order to withstand industrial conditions, such as extreme pH, high temperatures, organic solvents, or high pressure, enzymes must be stable enough, so that they do not lose activity or even completely denature [34,35]. However, one of the main challenges we currently face when designing industrial enzymes is the ability to, while increasing the stability of the system, maintain sufficient flexibility in the scaffold, for the reaction to commence. Thereby, another key focus point for the future of enzyme design is to combine engineering strategies that analyze both dynamics and statics of the entire protein, establishing relationships between stability and catalytic efficiency [36].

Following up the growing interest in enzyme design towards industrial biotechnology, another major bottleneck lies in the ability to control the substrate selectivity and specificity [37]. Due to economic reasons, or with the objective of obtaining novel products of the reaction, industrial biocatalytic processes occasionally rely on the usage of unnatural substrates [38]. However, predicting how enzymes will accommodate and transform different substrates remains challenging because of the complex relation between active-site architecture, protein dynamics, and reaction mechanisms [39]. Even minor structural differences in substrates can lead to changes in binding modes or catalytic efficiency [40]. Deep structural and chemical knowledge of the enzyme–substrate complex could allow for the prediction of the reaction outcome, yet current computational approaches still struggle to reliably capture these effects [41]. Addressing this limitation will be critical for expanding the applicability of enzyme design towards industrially relevant transformation involving unnatural substrates.

Finally, despite the rapid advances in computational prediction, experimental validation remains a challenge in enzyme design workflows. Computational prediction tools tend to generate large number of candidate-designs and structural models. Depending on the enzyme properties which are target for the optimization, translating this into experimental results might require intensive cloning, expression, purification and biochemical characterization [42–44]. Moreover, success rates are often limited by factors that are difficult to predict computationally, such as expression yield or solubility. Consequently, the development of high-throughput screening platforms and automated workflows that further combine computational design with experimental data, which ideally will be then fed back into the algorithms guiding the computational design, will be essential to fully exploit the potential of AI-driven enzyme engineering.

3. AI Tools Available for Enzyme Design

The revolution AI-guided tools have brought into the enzyme design field originates from their ability to optimize multiple properties simultaneously, including structure, stability and catalytic function. In contrast to traditional approaches, which often rely on experimental screening, AI methods enable the a priori exploration of sequence–structure–function relationships [45]. Rather than replacing established methodologies, such as directed evolution, these approaches expand the methodology for enzyme design and enable new opportunities for rational enzyme engineering. In this section we present a collection of AI tools which can help experimentalists guide their enzyme design strategies. For a more detailed comparison between these tools and when to use them, readers can

refer to the cited reviews. More general reviews introducing the methodology of protein folding and design have been recently published [22,46–51].

3.1. Structure Prediction

As early as 1973, Anfinsen postulated that the amino acid sequence of a protein encodes for its three-dimensional structure [45]. The recognition of AI-based structure prediction by the 2024 Nobel Prize in Chemistry further highlights the impact of these approaches for protein design. The possibility of predicting the structure of proteins, and therefore, of enzymes, has completely changed the status quo of the field. In the past, researchers were limited by the vast number of designs which needed to be tested before finding one that would fold in the correct way [52]. Now, this can be reduced by utilizing AlphaFold3 (AF3) [53] predictions to estimate which designs have a higher probability of folding in the correct way. This model enables high-accuracy prediction of biomolecular complexes starting from polymer sequences, residue modifications and ligand SMILES. While AF3 presents some limitations, such as the output model not respecting chirality or the low accuracy of the prediction for certain targets, the results are still outstanding, managing to provide high-quality prediction that matches the later observed experimental structural data [54].

AF3 set an unprecedented revolution for protein structure prediction, but it is no longer the only available tool to study the folding of biomacromolecules. Lately, methods such as Chai-1 [55] or Boltz-2 [56] have emerged; they allow for the prediction of full biomolecular complexes through an open-source code. Both of these novel methods report computationally fast accurate results for structure prediction, and while they also present limitations, they must be considered as challengers for AF3. Another example of this newly developed structure prediction methods is RoseTTaFold All-Atom, created by David Baker's group [57]. This AI-guided tool allows for the prediction of full assemblies and is also available open-source.

Overall, as previously mentioned, a major bottleneck in protein and therefore enzyme science is the lack of good, validated datasets [58]. Despite their role in advancing enzyme design, it is critical to remember that many of these functional prediction tools remain insufficiently validated experimentally, and their predictive power across diverse enzyme classes is still uncertain.

3.2. Function Prediction and Structure-to-Function Design

While structure prediction does not equal function prediction in protein systems, it remains true that both properties are heavily linked [59,60]. Therefore, another major interest in the enzyme design field is development of AI tools which can assist in the design of sequences that will enable the catalysis of a specific chemical reaction, allowing for the prediction of the enzyme function. Currently, a different array of AI tools is able to assist with the structure-to-function design. For the investigation of the substrate affinity of enzymes, researchers can make use of tools, such as FusionESP [61], EZSpecificity [62] or Catnip [63], the latest specifically being used for the α -ketoglutarate/Fe(II) non-heme iron enzymatic family. These computational methods do not present experimental validation, and therefore for the moment, the accuracy of the prediction is disputable.

Regarding the direct prediction of catalysis, tools such as SelenzymeRF [64], ALDELE [65] and CLEAN [66] can assist. While the latter focuses on the prediction of function for uncharacterized enzymes, ALDELE focuses on the prediction of the potential catalysis facilitated by an enzyme. SelenzymeRF, while also focusing on the prediction of catalysis, ranks candidate enzymes in order to provide information about the potential of each system to catalyze the

reaction. Both SelenzymeRF and CLEAN have been experimentally validated, while ALDELE does not provide experimental data for the moment.

Finally, to accurately design a protein sequence which will eventually fold into an enzyme capable of catalyzing a specific reaction, efforts to develop AI tools to assist with this matter have also come up with different methods. EnzyACT [67] can potentially assist with the prediction of the impact single or multiple mutations happening in the active site may have on the enzyme. On a bigger level, Progen [68] is able to completely generate artificial protein sequences with a predictable function. Other tools like PLACER [69] can enable researchers to assess the accuracy and pre-organization of the active site. These last two AI-guided methods have been experimentally validated and are currently assisting in the accurate design of novel enzymes.

3.3. Optimization of Stability

Protein stability represents a critical parameter for enzyme performance, specifically for industrial applications that require robust enzymes which can catalyze reactions under non-physiological conditions [70]. Enzymes must balance two competing requirements: stability versus activity. The enzyme must maintain a correct fold and adopt a rigid structure under physiological or industrial conditions. On the other hand, activity includes the freedom to adopt different conformations and interactions that facilitate catalysis. In consequence, features that promote activity can destabilize the protein, while features that enhance stability may restrict the motions needed for catalysis. Notably, amino acid residues involved in catalysis are often placed in energetically unfavorable positions to stabilize the transition state. Mutations introduced to improve catalytic efficiency often can end up reducing the structural stability of the system [71,72]. Different AI-based stability predictors have arisen as valuable tools for the assessment of mutational effects on the stability of enzymes.

ProStab [73], ProstaNet [74] ThermoMPNN [75] and Pythia [76] are AI models that allow researchers to study the stabilizing and de-stabilizing effects of point-mutations in the target enzyme. The latter acts as a self-supervised learning model while the other two carry out supervised learning. While these tools have shown consistent experimental data related to their predictions, there also exists an array of other computational methods that, while not being experimentally validated yet, also show potential as predictors to study the thermostability of protein systems. Some of these tools include SPURS [77], ESMtherm [78], ProSTAGE [79], SPIRED-Stab [80] or Stability Oracle [81].

4. Integration of AI Tools and Traditional Physics-Based Methods

While AI methods are completely transforming enzyme engineering, bringing more accurate results, classical physics-based computational methods remain essential for understanding and predicting enzyme behavior at an atomic level. Techniques such as molecular docking, molecular dynamics (MD) simulations, and quantum mechanics/molecular mechanics (QM/MM) provide complementary insights to AI-driven models [82–84]. Rather than presenting these two approaches as competing methodologies, AI and physics-based computational design tools should be viewed as combinatorial components of modern enzyme design workflows.

Physics-based simulations offer the advantage of modeling molecular and atomic interactions using established physical principles, allowing the exploration of the system's dynamics and energy [85,86]. Dynamics and energy data are critical for the assessment of enzyme catalysis as small energetic differences and transition states can determine catalytic efficiency and activity [87,88]. In the end, just predicting enzyme structure with AI-driven tools will only provide structural data, but knowing how this structure will behave during a

reaction in a specific environment is equally important. Therefore, integrating mechanistic simulations with AI predictions can improve the accuracy of enzyme design pipelines.

4.1. Molecular Dynamics Simulations

MD simulations are widely used to analyze protein flexibility, substrate binding modes, stability and behavior under different conditions. Unlike static structural prediction, MD provides with time/conformation data, allowing researchers to capture motion that may be related with the catalytic activity of the system [89–91]. This is particularly interesting for enzymatic activity studies that depend on rare conformational states that are not evident from predicted structural models.

In enzyme engineering, MD simulations can assist for the evaluation of structural stability of designed variants, identifying flexible regions suitable for mutagenesis, and refining active site geometries predicted by AI-based tools [92,93]. However, MD simulations are computationally demanding and sensitive to force-field accuracy, which may restrict the application in large-scale screening workflows [94].

4.2. Molecular Docking

Molecular docking remains a commonly employed method for ligand binding mode prediction and for estimating substrate specificity in enzyme systems [95–97]. Docking approaches are computationally efficient and can rapidly scan large libraries of new designs, appearing as an interesting approach in early-stage design workflow. Nevertheless, traditional docking methods rely on simplified scoring functions and often treat proteins as rigid or semi-flexible entities, which can limit their accuracy when modeling enzymatic reactions that involve conformational rearrangements. On top of this, docking scores do not directly correspond to catalytic activity, as binding affinity alone does not capture transition state stabilization or reaction kinetics. These limitations highlight the need to interpret docking results cautiously and to complement them with higher-resolution simulations and, if possible, experimental validation [98–101]. Presenting the wide array of available docking methods is out the scope of this article, but readers can refer to the cited reviews for further information [95,97,102].

4.3. Quantum Mechanics/Molecular Mechanics Methods

QM/MM methods represent the most rigorous computational framework for studying enzymatic reaction mechanisms. By treating the reactive region using QM while modeling the surrounding environment with MM, researchers are able to investigate bond formation and cleavage, charge redistribution, and transition state stabilization within enzymatic systems [103,104]. QM/MM calculations have played a key role in investigating catalytic mechanisms and identifying key catalytic residues responsible for activity. In the context of enzyme design, these methods can guide the rational mutagenesis of catalytic residues, evaluate the effects of these alterations on the reaction, and validate AI-based predictions [105,106]. However, their high computational cost and methodological complexity limit their application to relatively small systems or focused studies rather than large-scale screening of enzymes.

5. Emerging Opportunities in Enzyme Engineering

So far, we have focused on some of the more traditional enzyme properties and processes and how AI tools can help tailor them, but of course as science advances, new opportunities come up in the field of enzyme engineering, opening doors for novel catalysis and design methods.

5.1. Unnatural Amino Acids

One interesting new addition to the field, is the possibility of utilizing non-canonical amino acids (ncAAs) for rational tailoring of enzymes. Traditionally, catalytic system design was limited to the use of 20 natural amino acids, but since introduction of synthetic biology methods that allow for the incorporation of unnatural amino acids into proteins, the possibilities are unlimited [107–109]. AI-guided design may be particularly powerful when combined with expanded genetic codes, enabling catalytic functionalities beyond those accessible to natural evolution. The wide library of available ncAAs offers the possibility to tailor enzymes very specifically. If, rationally, we believe the introduction of certain functional groups in the active site of the enzyme will allow for interesting changes in the reaction, but none of the 20 natural amino acids would be able to provide this functional group in that specific position, it is possible that one of the newly developed ncAAs can do this instead.

5.2. De Novo Design

De novo design of enzymes represents another rapid growing trend in the field of enzyme engineering. In contrast to traditional methodologies, de novo design creates new enzymes with pre-defined catalytic functions employing computational tools [19,110–114]. Generative AI approaches, including protein language models, hallucination methods and diffusion models, are reshaping the field. These methods allow us to generate backbone conformations optimized around functional motifs, catalytic residues, or transition-state geometries, facilitating the creation of enzymes to perform specific reactions [115–117]. Conditional design strategies, in which constraints or active-site geometries are predefined, coexist with unconditional approaches, which focus on generating entirely new folds without a starting scaffold.

Despite these advances, de novo design also faces challenges. While there have been cases of de novo designed enzymes that afterwards showed impressive experimental results [118,119], the reality is that generally the catalytic efficiency of these newly designed proteins has a tendency to be lower than in natural enzymes. Currently protein design works best for enzymes that have been very well characterized in terms of active site architectures, kinetics, transition state barriers. Often those enzymes like the Kemp eliminase have been already extensively studied by directed evolution or error-prone PCR [21,22]. This reflects the limitations of accurately modeling reactions from scratch. Moreover, generative models are trained on databases of naturally occurring enzymes, generating a biased exploration of sequence space towards naturally evolved proteins. In the future, it will be crucial to combine this approach with previously mentioned physics-based methods to bridge the gap between structural design and catalytic performance.

6. Conclusions and Future Perspectives

AI is rapidly transforming the field of enzyme engineering, providing unprecedented capabilities for protein structure prediction, sequence design, and functional optimization. However, despite these advances, achieving truly predictive enzyme design remains a significant scientific challenge. Enzymatic catalysis is determined by complex energetic functions, conformational dynamics, and environmental interactions that are not yet fully captured by current computational models. As a result, experimental validation and mechanistic understanding remains essential components of enzyme engineering workflows.

One of the key conclusions of the recent progress is that structure prediction alone is insufficient to enable reliable enzyme design. Future advances will likely depend on deeper integration of AI-driven tools with physics-based modeling approaches capable of describing catalytic mechanisms and protein dynamics. The combination of these computa-

tional frameworks with improved sampling strategies may significantly enhance predictive power of AI methods while preserving interpretability regarding the dynamics and activity of the system.

Another critical factor that will shape the future of AI-guided enzyme design is data availability. The development of standardized, high-quality datasets containing experimental structural data, kinetic parameters, stability measurements, and negative experimental results will be essential for improving model generalization and reliability. In parallel, advances in high-throughput screening are expected to enable full computational-experimental pipelines, accelerating the discovery of functional biocatalysts.

Emerging technologies are also expanding the conceptual boundaries of enzyme engineering. In the coming years, ncAA-based enzyme design is expected to play an increasingly central role in expanding the functional and structural repertoire of biocatalysts. Since experimental screening of ncAA-containing libraries remains both expensive and time-consuming, AI-guided modeling will become indispensable for prioritizing promising designs and building efficient, iterative engineering workflows.

As generative models and experimental screening technologies further advance, we can anticipate a transition from manual, case-by-case protein design toward more automated enzyme engineering systems.

Looking ahead, we hypothesize the next 3 years will see an exponential growth in the number of designed enzymes entering the market. The development of fully automated enzyme design pipelines will allow experimental researchers to employ complex computational tools and generate tailored de novo enzymes for specific applications. However, we also expect that a significant number of these designs will still not be fully functional, as fully reproducing the complexity of enzymatic catalysis is still beyond the capabilities of current computational methods. Therefore, we believe that, while enabling researchers to more accurately tailor enzyme engineering, AI tools will never be able to replace human expertise. Specialists in molecular dynamics, catalysis and experimental validation will remain essential to interpret the designs, optimize them and assess their functionality.

Ultimately, the future of enzyme engineering lies not in AI alone but in the synergistic combination of data-driven algorithms, physical principles, and experimental innovation. As these components continue to evolve together, the prospect of designing enzymes with tailored activities, novel chemistries, and industrial robustness may become an achievable reality.

Author Contributions: R.T.-J., T.B., and B.L. were involved in conceptualization; investigation; visualization; writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: R.T.-J. was funded by the European Union's Horizon Europe research and innovation 765 program under the Marie Skłodowska-Curie grant agreement No 101072686—BioCatCodeExpander.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: T.B. gratefully acknowledges funding by the Werner Siemens foundation for establishing the field of Synthetic Biotechnology at the Technical University of Munich. We acknowledge continuous support by Carl Helmer. The graphical abstract was generated with ChatGPT (version GPT-5-5).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	artificial intelligence
AF	AlphaFold3
MD	molecular dynamics
QM/MM	quantum mechanics/molecular mechanics
ncAAs	non-canonical amino acids

References

- Agarwal, P.K. Enzymes: An integrated view of structure, dynamics and function. *Microb. Cell Fact.* **2006**, *5*, 2. [[CrossRef](#)] [[PubMed](#)]
- Rehm, F.; Chen, S.; Rehm, B. Enzyme Engineering for In Situ Immobilization. *Molecules* **2016**, *21*, 1370. [[CrossRef](#)] [[PubMed](#)]
- Samish, I. *Achievements and Challenges in Computational Protein Design*; Springer: New York, NY, USA, 2017; pp. 21–94.
- Buchner, E. Alkoholische Gärung ohne Hefezellen. *Berichte Dtsch. Chem. Ges.* **1897**, *30*, 117–124. [[CrossRef](#)]
- Buller, R.; Lutz, S.; Kazlauskas, R.J.; Snajdrova, R.; Moore, J.C.; Bornscheuer, U.T. From nature to industry: Harnessing enzymes for biocatalysis. *Science* **2023**, *382*, eadh8615. [[CrossRef](#)] [[PubMed](#)]
- Bell, E.L.; Finnigan, W.; France, S.P.; Green, A.P.; Hayes, M.A.; Hepworth, L.J.; Lovelock, S.L.; Niikura, H.; Osuna, S.; Romero, E.; et al. Biocatalysis. *Nat. Rev. Methods Prim.* **2021**, *1*, 46. [[CrossRef](#)]
- Mesbah, N.M. Industrial Biotechnology Based on Enzymes From Extreme Environments. *Front. Bioeng. Biotechnol.* **2022**, *10*, 870083. [[CrossRef](#)] [[PubMed](#)]
- Karigar, C.S.; Rao, S.S. Role of Microbial Enzymes in the Bioremediation of Pollutants: A Review. *Enzym. Res.* **2011**, *2011*, 1–11. [[CrossRef](#)]
- Weise, N.J.; France, S.P. Advancing Biocatalysis Education: Sustaining the Future of Industrial Biotechnology. *JACS Au* **2025**, *5*, 2932–2938. [[CrossRef](#)] [[PubMed](#)]
- Ming, Y.F.; Wang, W.K.; Yin, R.; Zeng, M.; Tang, L.; Tang, S.Z.; Li, M. A review of enzyme design in catalytic stability by artificial intelligence. *Brief. Bioinform.* **2023**, *24*, bbad065. [[CrossRef](#)]
- Dalby, P.A. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* **2011**, *21*, 473–480. [[CrossRef](#)] [[PubMed](#)]
- Song, Z.D.; Zhang, Q.F.; Wu, W.H.; Pu, Z.J.; Yu, H.R. Rational design of enzyme activity and enantioselectivity. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1129149. [[CrossRef](#)] [[PubMed](#)]
- Bustos-Jaimes, I.; Montiel, C. Trends and Challenges in Directed Evolution. *Curr. Chem. Biol.* **2008**, *2*, 50–59. [[CrossRef](#)]
- Ding, Y.Y.; Perez-Ortiz, G.; Peate, J.; Barry, S.M. Redesigning Enzymes for Biocatalysis: Exploiting Structural Understanding for Improved Selectivity. *Front. Mol. Biosci.* **2022**, *9*, 908285. [[CrossRef](#)] [[PubMed](#)]
- Hossack, E.J.; Hardy, F.J.; Green, A.P. Building Enzymes through Design and Evolution. *Acs Catal.* **2023**, *13*, 12436–12444. [[CrossRef](#)]
- Khan, M.F.; Khan, M.T. AI-Driven Enzyme Engineering: Emerging Models and Next-Generation Biotechnological Applications. *Molecules* **2025**, *31*, 45. [[CrossRef](#)] [[PubMed](#)]
- Wen, S.; Zheng, W.; Bornscheuer, U.T.; Wu, S. Generative artificial intelligence for enzyme design: Recent advances in models and applications. *Curr. Opin. Green Sustain. Chem.* **2025**, *52*, 101010. [[CrossRef](#)]
- Xie, W.J.; Warshel, A. Harnessing generative AI to decode enzyme catalysis and evolution for enhanced engineering. *Natl. Sci. Rev.* **2023**, *10*, nwad331. [[CrossRef](#)] [[PubMed](#)]
- Listov, D.; Fleishman, S.J. De novo enzyme design: Controlling structure to design function. *Curr. Opin. Struct. Biol.* **2026**, *98*, 103252. [[CrossRef](#)] [[PubMed](#)]
- Listov, D.; Goverde, C.A.; Correia, B.E.; Fleishman, S.J. Opportunities and challenges in design and optimization of protein function. *Nat. Rev. Mol. Cell Biol.* **2024**, *25*, 639–653. [[CrossRef](#)] [[PubMed](#)]
- Röthlisberger, D.; Khersonsky, O.; Wollacott, A.M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J.L.; Althoff, E.A.; Zanghellini, A.; Dym, O.; et al. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453*, 190–195. [[CrossRef](#)] [[PubMed](#)]
- Listov, D.; Vos, E.; Hoffka, G.; Hoch, S.Y.; Berg, A.; Hamer-Rogotner, S.; Dym, O.; Kamerlin, S.C.L.; Fleishman, S.J. Complete computational design of high-efficiency Kemp elimination enzymes. *Nature* **2025**, *643*, 1421–1427. [[CrossRef](#)] [[PubMed](#)]
- Landwehr, G.M.; Bogart, J.W.; Magalhaes, C.; Hammarlund, E.G.; Karim, A.S.; Jewett, M.C. Accelerated enzyme engineering by machine-learning guided cell-free expression. *Nat. Commun.* **2025**, *16*, 865. [[CrossRef](#)] [[PubMed](#)]
- Frushicheva, M.P.; Cao, J.; Chu, Z.T.; Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16869–16874. [[CrossRef](#)] [[PubMed](#)]

25. Yang, J.; Li, F.Z.; Arnold, F.H. Opportunities and Challenges for Machine Learning-Assisted Enzyme Engineering. *Acs Cent. Sci.* **2024**, *10*, 226–241. [[CrossRef](#)] [[PubMed](#)]
26. Fan, F.J.; Shi, Y. Effects of data quality and quantity on deep learning for protein-ligand binding affinity prediction. *Bioorg. Med. Chem.* **2022**, *72*, 117003. [[CrossRef](#)]
27. Derry, A.; Carpenter, K.A.; Altman, R.B. Training data composition affects performance of protein structure analysis algorithms. *Pac. Symp. Biocomput.* **2022**, *27*, 10–21. [[PubMed](#)]
28. Zhai, J.C.; Qi, X.G.; Cai, L.J.; Liu, Y.; Tang, H.C.; Xie, L.; Wang, J.M. NNKcat: Deep neural network to predict catalytic constants (K) by integrating protein sequence and substrate structure with enhanced data imbalance handling. *Brief. Bioinform.* **2025**, *26*, bbaf212. [[CrossRef](#)] [[PubMed](#)]
29. Fang, J. The role of data imbalance bias in the prediction of protein stability change upon mutation. *PLoS ONE* **2023**, *18*, e0283727. [[CrossRef](#)] [[PubMed](#)]
30. Moorhoff, F.; Zhang, Y.Z.; Qiu, S.Z.; Dong, W.J.; Medina-Ortiz, D.; Zhao, J.; Davari, M.D. Machine Learning-Driven Enzyme Mining: Opportunities, Challenges, and Future Perspectives. *Acs Catal.* **2026**, *16*, 12–30.
31. Maloney, M.P.; Coley, C.W.; Genheden, S.; Carson, N.; Helquist, P.; Norrby, P.-O.; Wiest, O. Negative Data in Data Sets for Machine Learning Training. *J. Org. Chem.* **2023**, *88*, 5239–5241. [[CrossRef](#)] [[PubMed](#)]
32. Tripp, A.; Braun, M.; Wieser, F.; Oberdorfer, G.; Lechner, H. Click, Compute, Create: A Review of Web-based Tools for Enzyme Engineering. *Chembiochem* **2024**, *25*, e202400092. [[CrossRef](#)] [[PubMed](#)]
33. Dou, Z.X.; Sun, Y.Q.; Jiang, X.K.; Wu, X.Y.; Li, Y.J.; Gong, B.; Wang, L.S. Data-driven strategies for the computational design of enzyme thermal stability: Trends, perspectives, and prospects. *Acta Biochim. Biophys. Sin.* **2023**, *55*, 343–355. [[CrossRef](#)] [[PubMed](#)]
34. Hu, Y.; Lv, Y.; Tang, H.; Wen, L.; Cheng, Y.H.; Xu, Z.; Zhang, T.; Wu, H. An Overview of Industrial Enzyme Engineering to Achieve a pH Adaptability-Activity Trade-Off: Current Status and Future Perspectives. *J. Agr. Food Chem.* **2025**, *73*, 20547–20562. [[CrossRef](#)]
35. Farhan, M.; Hasani, I.W.; Khafaga, D.S.R.; Ragab, W.M.; Kazi, R.N.A.; Aatif, M.; Muteeb, G.; Fahim, Y.A. Enzymes as Catalysts in Industrial Biocatalysis: Advances in Engineering, Applications, and Sustainable Integration. *Catalysts* **2025**, *15*, 891. [[CrossRef](#)]
36. Welborn, V.V.; Head-Gordon, T. Computational Design of Synthetic Enzymes. *Chem. Rev.* **2019**, *119*, 6613–6630.
37. Kasperkiewicz, P.; Poreba, M.; Groborz, K.; Drag, M. Emerging challenges in the design of selective substrates, inhibitors and activity-based probes for indistinguishable proteases. *FEBS J.* **2017**, *284*, 1518–1539. [[CrossRef](#)] [[PubMed](#)]
38. Kuo, Y.-M.; Henry, R.A.; Andrews, A.J. Measuring specificity in multi-substrate/product systems as a tool to investigate selectivity in vivo. *Biochim. Biophys. Acta (BBA)—Proteins Proteom.* **2016**, *1864*, 70–76. [[CrossRef](#)]
39. Ribeiro, A.J.M.; Riziotis, I.G.; Borkakoti, N.; Thornton, J.M. Enzyme function and evolution through the lens of bioinformatics. *Biochem. J.* **2023**, *480*, 1845–1863. [[CrossRef](#)] [[PubMed](#)]
40. Amin, S.R.; Erdin, S.; Ward, R.M.; Lua, R.C.; Lichtarge, O. Prediction and experimental validation of enzyme substrate specificity in protein structures. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E4195–E4202. [[CrossRef](#)] [[PubMed](#)]
41. Mondal, D.; Snodgrass, H.M.; Gomez, C.A.; Lewis, J.C. Non-Native Site-Selective Enzyme Catalysis. *Chem. Rev.* **2023**, *123*, 10381–10431. [[CrossRef](#)] [[PubMed](#)]
42. Jiang, L.; Althoff, E.A.; Clemente, F.R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J.L.; Betker, J.L.; Tanaka, F.; Barbas, C.F.; et al. De novo computational design of retro-aldol enzymes. *Science* **2008**, *319*, 1387–1391. [[CrossRef](#)] [[PubMed](#)]
43. St-Jacques, A.D.; Rodriguez, J.M.; Eason, M.G.; Foster, S.M.; Khan, S.T.; Damry, A.M.; Goto, N.K.; Thompson, M.C.; Chica, R.A. Computational remodeling of an enzyme conformational landscape for altered substrate selectivity. *Nat. Commun.* **2023**, *14*, 6058. [[CrossRef](#)] [[PubMed](#)]
44. Zheng, N.; Cai, Y.C.; Zhang, Z.H.; Zhou, H.M.; Deng, Y.; Du, S.; Tu, M.; Fang, W.; Xia, X.L. Tailoring industrial enzymes for thermostability and activity evolution by the machine learning-based iCASE strategy. *Nat. Commun.* **2025**, *16*, 604. [[CrossRef](#)] [[PubMed](#)]
45. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230. [[CrossRef](#)] [[PubMed](#)]
46. Albanese, K.I.; Barbe, S.; Tagami, S.; Woolfson, D.N.; Schiex, T. Computational protein design. *Nat. Rev. Methods Prim.* **2025**, *5*, 13. [[CrossRef](#)]
47. Koh, H.Y.; Zheng, Y.Z.; Yang, M.; Arora, R.; Webb, G.I.; Pan, S.R.; Li, L.; Church, G.M. AI-driven protein design. *Nat. Rev. Bioeng.* **2025**, *3*, 1034–1056. [[CrossRef](#)]
48. Kosonocky, C.W.; Alamdari, S.; Yang, K.K.; Amini, A.P. Closing the loop: Experimentally validated methods in artificial intelligence-driven protein design. *Curr. Opin. Struct. Biol.* **2026**, *98*, 103272. [[CrossRef](#)] [[PubMed](#)]
49. Stocco, F.; Garibbo, M.; Ferruz, N. Steering generative models for protein design: Aligning and conditioning strategies. *Curr. Opin. Struct. Biol.* **2026**, *98*.
50. Ferruz, N.; Höcker, B. Controllable protein design with language models. *Nat. Mach. Intell.* **2022**, *4*, 521–532. [[CrossRef](#)]
51. Teijeiro-Juiz, R.; Jamróg, G.; Ringel, M.; Brück, T.; Di Geronimo, B.; Loll, B. Applications and limitations of AI tools in enzyme design. In *Protein Science*; Wiley: Hoboken, NJ, USA, 2026.

52. Zhang, G.H.; Liu, C.Y.; Lu, J.J.; Zhang, S.W.; Zhu, L.Y. The Role of AI-Driven De Novo Protein Design in the Exploration of the Protein Functional Universe. *Biology* **2025**, *14*, 1268. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Addendum: Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *636*, E4. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Kovalevskiy, O.; Mateos-Garcia, J.; Tunyasuvunakool, K. AlphaFold two years on: Validation and impact. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2315002121. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Chai, D.; Boitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhnikov, A.; Wu, K. Chai-1: Decoding the molecular interactions of life. *bioRxiv* **2024**. [\[CrossRef\]](#)
56. Passaro, S.; Corso, G.; Wohlwend, J.; Reveiz, M.; Thaler, S.; Somnath, V.R.; Getz, N.; Portnoi, T.; Roy, J.; Stark, H.; et al. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv* **2025**. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Krishna, R.; Wang, J.; Ahern, W.; Sturmfels, P.; Venkatesh, P.; Kalvet, I.; Lee, G.R.; Morey-Burrows, F.S.; Anishchenko, I.; Humphreys, I.R.; et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **2024**, *384*, ead12528. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Varadi, M.; Tsenkov, M.; Velankar, S. Challenges in bridging the gap between protein structure prediction and functional interpretation. *Proteins-Struct. Funct. Bioinform.* **2025**, *93*, 400–410.
59. Leman, J.K.; Szczerbiak, P.; Renfrew, P.D.; Gligorijevic, V.; Berenberg, D.; Vatanen, T.; Taylor, B.C.; Chandler, C.; Janssen, S.; Pataki, A.; et al. Sequence-structure-function relationships in the microbial protein universe. *Nat. Commun.* **2023**, *14*, 2351.
60. Sadowski, M.I.; Jones, D.T. The sequence-structure relationship and protein function prediction. *Curr. Opin. Struct. Biol.* **2009**, *19*, 357–362. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Du, Z.J.; Fu, W.M.; Guo, X.L.; Caragea, D.; Li, Y.H. FusionESP: Improved Enzyme-Substrate Pair Prediction by Fusing Protein and Chemical Knowledge. *J. Chem. Inf. Model.* **2025**, *65*, 2806–2817. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Zhou, J.H.; Huang, M.L. Decoding enzyme-substrate specificity with EZSpecificity. *Chem. Catal.* **2026**, *6*. [\[CrossRef\]](#)
63. Paton, A.E.; Boiko, D.A.; Perkins, J.C.; Cemalovic, N.I.; Reschützegg, T.; Gomes, G.; Narayan, A.R.H. Connecting chemical and protein sequence space to predict biocatalytic reactions. *Nature* **2025**, *646*, 108–116. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Stoney, R.A.; Hanko, E.K.R.; Carbonell, P.; Breitling, R. SelenzymeRF: Updated enzyme suggestion software for unbalanced biochemical reactions. *Comput. Struct. Biotechnol.* **2023**, *21*, 5868–5876. [\[CrossRef\]](#)
65. Wang, X.W.; Quinn, D.; Moody, T.S.; Huang, M.L. ALDELE: All-Purpose Deep Learning Toolkits for Predicting the Biocatalytic Activities of Enzymes. *J. Chem. Inf. Model.* **2024**, *64*, 3123–3139. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Yu, T.H.; Cui, H.Y.; Li, J.C.; Luo, Y.A.; Jiang, G.D.; Zhao, H.M. Enzyme function prediction using contrastive learning. *Science* **2023**, *379*, 1358–1363. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Li, G.; Zhang, N.; Dai, X.W.; Fan, L. EnzyACT: A Novel Deep Learning Method to Predict the Impacts of Single and Multiple Mutations on Enzyme Activity. *J. Chem. Inf. Model.* **2024**, *64*, 5912–5921. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Madani, A.; Krause, B.; Greene, E.R.; Subramanian, S.; Mohr, B.P.; Holton, J.M.; Olmos, J.J.R.; Xiong, C.M.; Sun, Z.Z.; Socher, R.; et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **2023**, *41*, 1099–1106. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Anishchenko, I.; Kipnis, Y.; Kalvet, I.; Zhou, G.F.; Krishna, R.; Pellock, S.J.; Lauko, A.; Lee, G.R.; An, L.; Dauparas, J.; et al. Modeling protein-small molecule conformational ensembles with PLACER. *Proc. Natl. Acad. Sci. USA* **2025**, *122*, e2427161122. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Phintha, A.; Chaiyen, P. Rational and mechanistic approaches for improving biocatalyst performance. *Chem. Catal.* **2022**, *2*, 2614–2643. [\[CrossRef\]](#)
71. May, M.; Chuah, A.; Lehmann, N.; Goodall, L.; Cho, V.; Andrews, T.D. Functionally constrained human proteins are less prone to mutational instability from single amino acid substitutions. *Nat. Commun.* **2025**, *16*, 2492. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Hernández, I.M.; Dehouck, Y.; Bastolla, U.; López-Blanco, J.R.; Chacón, P. Predicting protein stability changes upon mutation using a simple orientational potential. *Bioinformatics* **2023**, *39*, btad011. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Tan, H.; Wei, X.; Lin, S.; Mao, X.; Chen, J.; Sun, H.; Zhang, Y.; Zhou, Z.; Wei, D.-Q.; Lin, S.; et al. ProStab: Prediction of protein stability change upon mutations by protein language and inverse folding models. *bioRxiv* **2025**. [\[CrossRef\]](#)
74. Liang, T.J.; Sun, Z.Y.; Ishima, R.; Xie, X.Q.; Xue, Y.; Li, W.; Feng, Z.W. ProstaNet: A Novel Geometric Vector Perceptrons-Graph Neural Network Algorithm for Protein Stability Prediction in Single- and Multiple-Point Mutations with Experimental Validation. *Research* **2025**, *8*, 0674. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Dieckhaus, H.; Brocchiacono, M.; Randolph, N.Z.; Kuhlman, B. Transfer learning to leverage larger datasets for improved prediction of protein stability changes. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2314853121. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Sun, J.Y.; Zhu, T.; Cui, Y.L.; Wu, B. Structure-based self-supervised learning enables ultrafast protein stability prediction upon mutation. *Innov.-Amst.* **2025**, *6*, 100750. [\[CrossRef\]](#)

77. Li, Z.; Luo, Y. Rewiring protein sequence and structure generative models to enhance protein stability prediction. In *Research in Computational Molecular Biology*; Springer: Cham, Switzerland, 2025. [\[CrossRef\]](#)
78. Chu, S.K.S.; Narang, K.; Siegel, J.B. Protein stability prediction by fine-tuning a protein language model on a mega-scale dataset. *PLoS Comput. Biol.* **2024**, *20*, e1012248. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Li, G.; Yao, S.J.; Fan, L. ProSTAGE: Predicting Effects of Mutations on Protein Stability by Using Protein Embeddings and Graph Convolutional Networks. *J. Chem. Inf. Model.* **2024**, *64*, 340–347. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Chen, Y.H.; Xu, Y.X.; Liu, D.; Xing, Y.G.; Gong, H.P. An end-to-end framework for the prediction of protein structure and fitness from single sequence. *Nat. Commun.* **2024**, *15*, 7400. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Diaz, D.J.; Gong, C.Y.; Ouyang-Zhang, J.; Loy, J.M.; Wells, J.; Yang, D.V.; Ellington, A.D.; Dimakis, A.G.; Klivans, A.R. Stability Oracle: A structure-based graph-transformer framework for identifying stabilizing mutations. *Nat. Commun.* **2024**, *15*, 6170. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Jurich, C.; Shao, Q.Z.; Ran, X.C.; Yang, Z.J. Physics-based modeling in the new era of enzyme engineering. *Nat. Comput. Sci.* **2025**, *5*, 279–291. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Wang, X.; Wang, J.; Li, W.-L. Machine learning/molecular mechanics enzymology for the next generation of computational enzymatic catalysis. *Chem. Catal.* **2026**, *6*, 101658. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Huang, H.Y.; Guan, X.Y.; Li, W.F.; Zhang, J.; Wang, W. Protein dynamics prediction by integrating biophysics and artificial intelligence. *Curr. Opin. Struct. Biol.* **2026**, *97*, 103223. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Clemente, C.M.; Capece, L.; Mart, M.A. Best Practices on QM/MM Simulations of Biological Systems. *J. Chem. Inf. Model.* **2023**, *63*, 2609–2627. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Duarte, F.; Amrein, B.A.; Blaha-Nelson, D.; Kamerlin, S.C.L. Recent advances in QM/MM free energy calculations using reference potentials. *Biochim. Biophys. Acta (BBA) -Gen. Subj.* **2015**, *1850*, 954–965. [\[CrossRef\]](#)
87. Nam, K.; Wolf-Watz, M. Protein dynamics: The future is bright and complicated! *Struct. Dyn.* **2023**, *10*, 014301.
88. Kamerlin, S.C.; Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **2010**, *78*, 1339–1375. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Hollingsworth, S.A.; Dror, R.O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143. [\[CrossRef\]](#) [\[PubMed\]](#)
90. Karplus, M.; Petsko, G.A. Molecular-Dynamics Simulations in Biology. *Nature* **1990**, *347*, 631–639. [\[PubMed\]](#)
91. Levitt, M. Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* **1983**, *168*, 595–617. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Liao, M.L.; Somero, G.N.; Dong, Y.W. Comparing mutagenesis and simulations as tools for identifying functionally important sequence changes for protein thermal adaptation. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 679–688. [\[PubMed\]](#)
93. Scortica, A.; Capone, M.; Narzi, D.; Frezzini, M.; Scafati, V.; Giovannoni, M.; Angelucci, F.; Guidoni, L.; Mattei, B.; Benedetti, M. A molecular dynamics-guided mutagenesis identifies two aspartic acid residues involved in the pH-dependent activity of OG-OXIDASE 1. *Plant Physiol. Biochem.* **2021**, *169*, 171–182. [\[CrossRef\]](#) [\[PubMed\]](#)
94. Yin, J.; Fenley, A.T.; Henriksen, N.M.; Gilson, M.K. Toward Improved Force-Field Accuracy through Sensitivity Analysis of Host-Guest Binding Thermodynamics. *J. Phys. Chem. B* **2015**, *119*, 10145–10155. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Meng, X.Y.; Zhang, H.X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput.-Aid Drug* **2011**, *7*, 146–157. [\[CrossRef\]](#)
96. Agu, P.C.; Afiukwa, C.A.; Orji, O.U.; Ezech, E.M.; Ofoke, I.H.; Ogbu, C.O.; Ugwuja, E.I.; Aja, P.M. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. *Sci. Rep.* **2023**, *13*, 13398. [\[CrossRef\]](#) [\[PubMed\]](#)
97. Paggi, J.M.; Pandit, A.; Dror, R.O. The Art and Science of Molecular Docking. *Annu. Rev. Biochem.* **2024**, *93*, 389–410. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Li, Y.; Yi, J.C.; Li, H.; Li, K.; Kang, F.H.; Deng, Y.C.; Wu, C.K.; Fu, X.Z.; Jiang, D.J.; Cao, D.S. Decoding the limits of deep learning in molecular docking for drug discovery. *Chem. Sci.* **2025**, *16*, 17374–17390. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Mukherjee, S.; Balias, T.E.; Rizzo, R.C. Docking Validation Resources: Protein Family and Ligand Flexibility Experiments. *J. Chem. Inf. Model.* **2010**, *50*, 1986–2000. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Sciortino, G.; Garribba, E.; Maréchal, J.D. Validation and Applications of Protein-Ligand Docking Approaches Improved for Metalloligands with Multiple Vacant Sites. *Inorg. Chem.* **2019**, *58*, 294–306. [\[PubMed\]](#)
101. Bender, B.J.; Gahbauer, S.; Lutten, A.; Lyu, J.K.; Webb, C.M.; Stein, R.M.; Fink, E.A.; Balias, T.E.; Carlsson, J.; Irwin, J.J.; et al. A practical guide to large-scale docking. *Nat. Protoc.* **2021**, *16*, 4799–4832. [\[CrossRef\]](#) [\[PubMed\]](#)
102. Martins, F.G.; Santos, H.A.; Sousa, S.F. A Review of Current Computational Tools for Peptide-Protein Docking. *J. Comput. Chem.* **2026**, *47*, e70328. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Vennelakanti, V.; Nazemi, A.; Mehmood, R.; Steeves, A.H.; Kulik, H.J. Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 9–17. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Magalhaes, R.P.; Fernandes, H.S.; Sousa, S.F. Modelling Enzymatic Mechanisms with QM/MM Approaches: Current Status and Future Challenges. *Isr. J. Chem.* **2020**, *60*, 655–666. [\[CrossRef\]](#)

105. Murphy, A.R.; Hix, M.A.; Walker, A.R. Exploring the Effects of Mutagenesis on FusionRed by Using Excited-State QM/MM Dynamics and Classical Force Field Simulations. *Chembiochem* **2023**, *24*, e202200799. [[CrossRef](#)] [[PubMed](#)]
106. Bím, D.; Navrátil, M.; Gutten, O.; Konvalinka, J.; Kutil, Z.; Culka, M.; Navrátil, V.; Alexandrova, A.N.; Barinka, C.; Rulísek, L. Predicting Effects of Site-Directed Mutagenesis on Enzyme Kinetics by QM/MM and QM Calculations: A Case of Glutamate Carboxypeptidase II. *J. Phys. Chem. B* **2022**, *126*, 132–143. [[CrossRef](#)] [[PubMed](#)]
107. Birch-Price, Z.; Hardy, F.J.; Lister, T.M.; Kohn, A.R.; Green, A.P. Noncanonical Amino Acids in Biocatalysis. *Chem. Rev.* **2024**, *124*, 8740–8786. [[CrossRef](#)] [[PubMed](#)]
108. Brouwer, B.; Della-Felice, F.; Illies, J.H.; Iglesias-Moncayo, E.; Roelfes, G.; Drienovska, I. Noncanonical Amino Acids: Bringing New-to-Nature Functionalities to Biocatalysis. *Chem. Rev.* **2024**, *124*, 10877–10923. [[CrossRef](#)] [[PubMed](#)]
109. Hickey, J.L.; Sindhikara, D.; Zultanski, S.L.; Schultz, D.M. Beyond 20 in the 21st Century: Prospects and Challenges of Non-canonical Amino Acids in Peptide Drug Discovery. *ACS Med. Chem. Lett.* **2023**, *14*, 557–565. [[CrossRef](#)] [[PubMed](#)]
110. Cui, X.C.; Zheng, Y.; Liu, Y.; Yuchi, Z.; Yuan, Y.J. AI-driven de novo enzyme design: Strategies, applications, and future prospects. *Biotechnol. Adv.* **2025**, *82*, 108603. [[CrossRef](#)] [[PubMed](#)]
111. Kortemme, T. De novo protein design—From new structures to programmable functions. *Cell* **2024**, *187*, 526–544. [[CrossRef](#)] [[PubMed](#)]
112. Wang, X.; Xu, K.; Tan, Y.; Liu, S.; Zhou, J. Possibilities of Using De Novo Design for Generating Diverse Functional Food Enzymes. *Int. J. Mol. Sci.* **2023**, *24*, 3827. [[CrossRef](#)] [[PubMed](#)]
113. Sellés Vidal, L.; Isalan, M.; Heap, J.T.; Ledesma-Amaro, R. A primer to directed evolution: Current methodologies and future directions. *RSC Chem. Biol.* **2023**, *4*, 271–291. [[CrossRef](#)] [[PubMed](#)]
114. Tang, M.; Ge, F.; Li, A.; Hu, L.; Wang, C.; Tang, J.; Song, X.; Liu, X.; Shi, H.; Tan, Z. Artificial Intelligence-Driven de Novo Design of Robust Enzymes to Enhance Their Performance. *ACS Synth. Biol.* **2025**, *14*, 4178–4201. [[CrossRef](#)] [[PubMed](#)]
115. Braun, M.; Tripp, A.; Chakatok, M.; Kaltenbrunner, S.; Fischer, C.; Stoll, D.; Bijelic, A.; Elailly, W.; Totaro, M.G.; Moser, M.; et al. Computational enzyme design by catalytic motif scaffolding. *Nature* **2026**, *649*, 237–245. [[PubMed](#)]
116. Kortemme, T. De novo protein design: From new structures to programmable molecular and cellular functions. *Eur. Biophys. J. Biophys.* **2025**, *54*, S44.
117. Anishchenko, I.; Pellock, S.J.; Chidyausiku, T.M.; Ramelot, T.A.; Ovchinnikov, S.; Hao, J.; Bafna, K.; Norn, C.; Kang, A.; Bera, A.K.; et al. De novo protein design by deep network hallucination. *Nature* **2021**, *600*, 547–552. [[CrossRef](#)] [[PubMed](#)]
118. 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.; et al. De novo design of luciferases using deep learning. *Nature* **2023**, *614*, 774–780. [[CrossRef](#)] [[PubMed](#)]
119. Zou, Z.; Kalvet, I.; Lozhkin, B.; Morris, E.; Zhang, K.L.; Chen, D.P.; Ernst, M.L.; Zhang, X.; Baker, D.; Ward, T.R. De novo design and evolution of an artificial metathase for cytoplasmic olefin metathesis. *Nat. Catal.* **2025**, *8*, 1208–1219. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.