

Review

Not peer-reviewed version

Revolutionizing Molecular Design for Innovative Therapeutic Applications through Artificial Intelligence

[Ahrum Son](#) , [Jongham Park](#) , [Woojin Kim](#) , Yoonki Yoon , [Sangwoon Lee](#) , [Yongho Park](#) , [Hyunsoo Kim](#) *

Posted Date: 2 September 2024

doi: 10.20944/preprints202409.0013.v1

Keywords: Computational Biology; Protein Engineering; Artificial Intelligence; Molecular Design; De Novo Protein Design; Therapeutic Proteins; Synthetic Biology



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Revolutionizing Molecular Design for Innovative Therapeutic Applications through Artificial Intelligence

Ahrum Son ¹, Jongham Park ², Woojin Kim ², Yoonki Yoon ², Sangwoon Lee ², Yongho Park ² and Hyunsoo Kim ^{2,3,4,5*}

¹ Department of Molecular Medicine, Scripps Research, La Jolla, CA 92037, USA; ahson@scripps.edu (A.S.)

² Department of Bio-AI convergence, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon, 34134, Republic of Korea; 975pjh@gmail.com (J.P.); woojin1544@gmail.com (W.K.); dbsrl0218@gmail.com (Y.Y.); sanguni088@gmail.com (S.L.); kmalrpkr13@gmail.com (Y.P.)

³ Department of Convergent Bioscience and Informatics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea

⁴ Protein AI Design Institute, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea

⁵ SCICS, Sciences for Panomics, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea

* Correspondence: kimlab@cnu.ac.kr (H.K.)

Abstract: The field of computational protein engineering has been transformed by recent advancements in machine learning, artificial intelligence, and molecular modeling, enabling the design of proteins with unprecedented precision and functionality. Computational methods now play a crucial role in enhancing the stability, activity, and specificity of proteins for diverse applications in biotechnology and medicine. Techniques such as deep learning, reinforcement learning, and transfer learning have dramatically improved protein structure prediction, optimization of binding affinities, and enzyme design. These innovations have streamlined the process of protein engineering by allowing the rapid generation of targeted libraries, reducing experimental sampling, and enabling the rational design of proteins with tailored properties. Furthermore, the integration of computational approaches with high-throughput experimental techniques has facilitated the development of multifunctional proteins and novel therapeutics. However, challenges remain in bridging the gap between computational predictions and experimental validation, and in addressing ethical concerns related to AI-driven protein design. This review provides a comprehensive overview of the current state and future directions of computational methods in protein engineering, emphasizing their transformative potential in creating next-generation biologics and advancing synthetic biology.

Keywords: Computational Biology; Protein Engineering; Artificial Intelligence; Molecular Design; *De Novo* Protein Design; Therapeutic Proteins; Synthetic Biology

1. Introduction

In recent years, the subject of computational biology has experienced rapid and significant expansion, leading to a fundamental shift in how we comprehend and manipulate biological systems. The impact of computational approaches on protein engineering and molecular design is especially noticeable, as they have completely transformed the capacity to create and enhance proteins with new and unique capabilities. The incorporation of computational methodologies alongside conventional biological methods has created new opportunities for advancement in biotechnology, medicines, and related disciplines. This collaboration has resulted in improved and focused approaches for manipulating proteins, finding new drugs, and creating innovative biomolecules with improved capabilities.

Computational methods are becoming essential for customizing proteins for different biotechnological uses. Each year, a variety of tools and methodologies are being created and improved to keep up with the growing needs and difficulties of protein engineering [1]. The progress in machine learning and artificial intelligence has greatly improved the precision of protein structure predictions and the detection of functional regions, enabling more accurate manipulation of protein activities [2]. The use of computational approaches has greatly influenced the field of enzyme design. These approaches have allowed for the development of proteins that have enhanced catalytic efficiencies and new functionality [3]. For example, the utilization of machine learning models to forecast protein stability and interactions has simplified the design procedure, enabling the quick creation and manufacture of proteins without the limitations of living cells.

The combination of computational and experimental methods has expedited the design process by allowing the development of targeted libraries for laboratory evolution. This has resulted in a reduction of the extensive sequence space that requires sampling [4]. Platforms such as Mutexa demonstrate attempts to develop intelligent ecosystems that integrate fast computation with bioinformatics and quantum chemistry, making the process of identifying potential protein variants more efficient [5]. However, there are still obstacles to overcome in expanding the use of these technologies and making them available to a wider group of academics. This is crucial in order to fully utilize their potential in addressing global issues like sustainable development and healthcare [6].

Computational methods have gained significance in the field of drug development, thanks to recent progress in deep learning and artificial intelligence. These advancements have made it easier to quickly identify a wide range of powerful and specific ligands. These advancements have the capacity to make the drug discovery process more accessible to the general public, offering new possibilities for the efficient creation of safer and more efficient small-molecule medicines. The advancement of computational tools and their integration with experimental approaches is paving the way for remarkable innovation and application in protein design within the field of synthetic biology.

The continuous progress in computational biology is paving the way for a forthcoming period of protein engineering and molecular design, marked by enhanced accuracy, efficiency, and creativity. In order to overcome current hurdles and fully utilize the promise of biotechnology and pharmaceuticals, it is imperative to integrate computational and experimental approaches as the area continues to develop. This study seeks to present a thorough summary of the most recent developments in computational approaches used in protein engineering and molecular design. It emphasizes the significant influence of these technologies on the field.

2. Machine Learning and AI Applications in Protein Design

2.1. Deep Learning Approaches

2.1.1. Convolutional Neural Networks (CNNs) for Structure Prediction

Convolutional Neural Networks (CNNs) have greatly enhanced the field of structure prediction in computational biology, specifically for proteins and RNA. CNNs are utilized for their capacity to do hierarchical feature extraction, rendering them well-suited for jobs that involve identifying intricate patterns in biological sequences and structures. CNNs have been utilized in protein structure prediction to forecast inter-residue distances and contact maps. This approach is exemplified in AlphaFold, which incorporates ResNets to improve prediction accuracy by incorporating translational invariance in the data [7,8]. In addition, CNNs have been modified for the purpose of predicting RNA secondary structure. Models such as CDPFold and E2Efold utilize convolutional layers to estimate the probability of base-pairing, and then employ dynamic programming to extract the structure [9]. Recent progress has involved combining CNNs with other deep learning architectures, such as transformers, to enhance the accuracy of predicting protein secondary structures. This approach capitalizes on the benefits of both convolutional and attention mechanisms [10]. In addition, 3D Convolutional Neural Networks (CNNs) have been used to forecast the local

fitness landscapes of protein structures. This helps in recognizing the wild-type and consensus amino acids based on their structural contexts [11]. The applications mentioned highlight the flexibility and effectiveness of CNNs in solving various and intricate problems in structural bioinformatics. This makes them a fundamental component in the continuous development of computational biology [7,8,12] (Figure 1A).

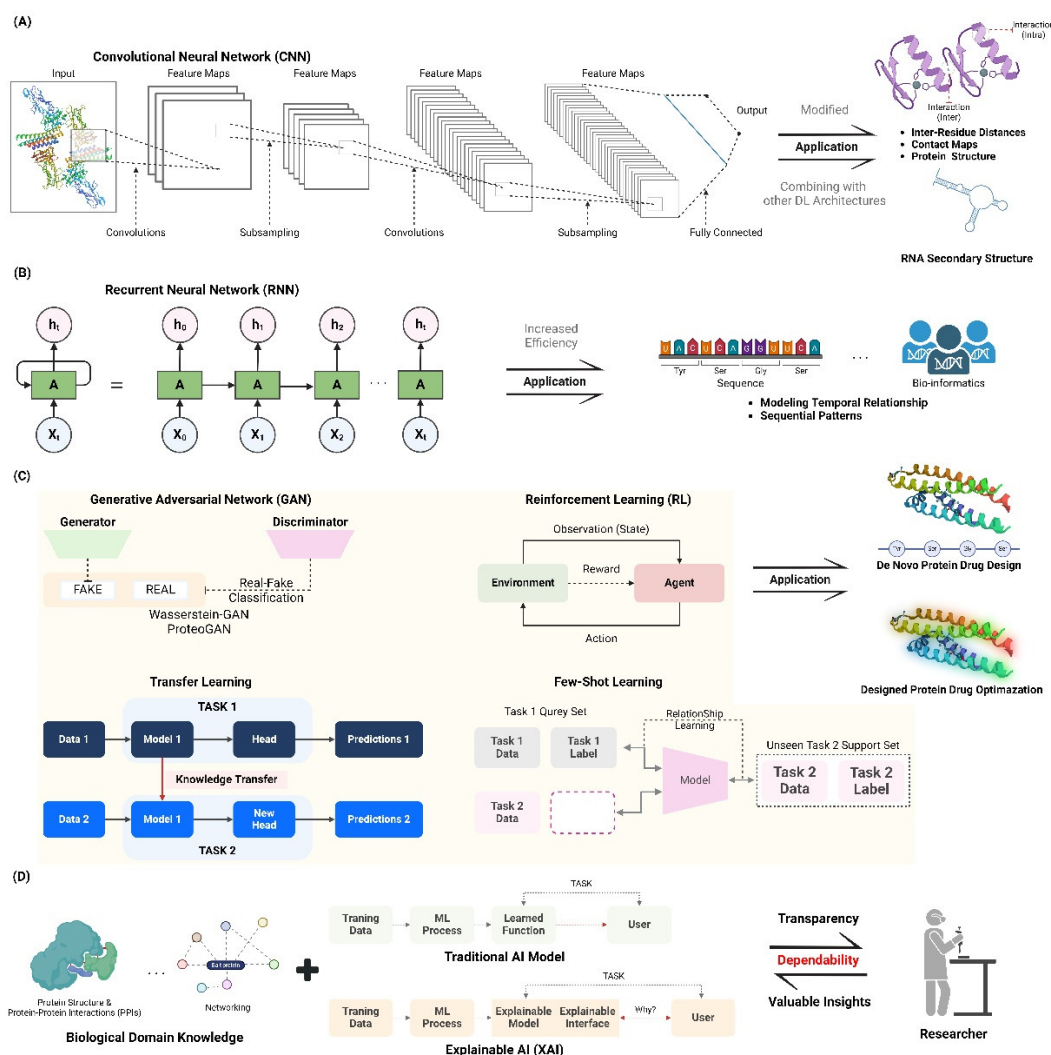


Figure 1. Development and application of AI algorithms in biotechnology. (A), (B) Various AI algorithms significantly contribute to the development of biotechnology. Representatively, CNN (Convolutional Neural Network) are utilized for protein structure prediction through the prediction of distances and contact maps between residues. Additionally, RNN (Recurrent Neural Network) play a crucial role in sequence optimization through temporal relationship and sequential pattern modeling. (C) Recently, algorithm such as GAN (Generative Adversarial Network), RL (Reinforcement Learning), Transfer Learning and Few-Shot Learning have demonstrated their efficiency in modeling protein structures and interactions. These advanced algorithms are being utilized to overcome limitations in data collection required for model training as well as limitations in designing new proteins. (D) Explainable AI (XAI) provides transparency and insight into modeling results by elucidating the decision-making process behind the vague “black box” judgment criteria of existing AI-based predictive models. Advances in AI algorithms have significantly progressed protein engineering. however, they still require experimental validation. The integration of domain expertise and AI-based methodologies, also known as informed AI, can potentially enhance model efficiency, reliability, and to provide more accurate insights consistent with validated domain knowledge.

2.1.2. Recurrent Neural Networks (RNNs) for Sequence Optimization

Recurrent Neural Networks (RNNs) are a potent tool for optimizing sequences, demonstrating their effectiveness in modeling temporal relationships and sequential patterns. RNNs, specifically LSTM and GRU architectures, are commonly used due to their ability to address the vanishing gradient problem and capture long-range dependencies in sequential data [13]. Current studies have concentrated on enhancing Recurrent Neural Networks (RNNs) for many purposes, such as predicting future values in time series data, understanding and generating human language, and analyzing biological information [14]. The convergence and performance of RNNs across many tasks have been greatly enhanced by the development of weight initialization schemes, such as Xavier/Glorot and He initialization [15]. Moreover, the utilization of optimization techniques such as adaptive learning rate approaches and gradient descent-based algorithms has played a vital role in improving the training efficiency and generalization performance of RNN models. Research has also investigated the combination of Recurrent Neural Networks (RNNs) with other neural network structures, like Convolutional Neural Networks (CNNs), to utilize their complementary advantages for sequence modeling and feature extraction [16]. The adaptability and robustness of RNNs in sequence optimization are emphasized by these achievements, establishing them as essential components in the continuous progress of machine learning and artificial intelligence [14] (Figure 1B).

2.1.3. Generative Adversarial Networks (GANs) In De Novo Protein Design

GANs have significantly transformed the field of de novo protein design by allowing the creation of new protein sequences that possess specific desirable characteristics. Generative Adversarial Networks (GANs), including a generator and a discriminator network, have demonstrated remarkable efficacy in modeling the intricate interactions between sequence, structure, and function that are inherent in proteins. Recent research has shown that GANs can be used to create proteins with specific structures and functions. This was achieved by using a Wasserstein-GAN with gradient penalty to design proteins with unique folds [17]. In addition, I created ProteoGAN, a conditional GAN that produces protein sequences using hierarchical functional labels from the Gene Ontology. This model outperformed other deep learning baselines in generating protein sequences [18]. The ability to produce proteins with precise enzymatic activity and solubility profiles has been improved by advancements in conditional generative modeling. This is exemplified by the hierarchical conditional GAN framework outlined. In addition, a comprehensive analysis was conducted on several deep generative models, emphasizing the crucial contribution of GANs in suggesting innovative proteins that closely mimic natural equivalents in terms of stability and expression [19]. The advancements highlight the profound capacity of GANs in creating new proteins with specific characteristics for various biotechnological and medicinal uses, demonstrating their ability to rapidly and effectively design proteins (Figure 1C).

2.2. Reinforcement Learning in Protein Engineering

2.2.1. Optimization of Protein Properties

Reinforcement Learning (RL) has demonstrated significant potential in the domain of protein engineering, namely in the enhancement of protein characteristics. RL techniques, like those used in ProteinRL, utilize generative protein language models to optimize protein sequences for specific structural and functional properties. This allows for the creation of new proteins with high charge content or diverse sequences that have high solubility and structural confidence [20]. Self-play RL is a new tool that helps optimize protein sequences to achieve desired features. This has a substantial impact on drug discovery and other biotechnological applications [21]. Moreover, the integration of reinforcement learning (RL) with fitness landscape modeling, exemplified by the microFormer framework, enables the efficient exploration of the extensive mutant space. This integration facilitates the design of protein variants that exhibit improved activity and stability [22]. One recent development involves using protein language models as reward functions in RL frameworks to create biologically realistic sequences. These sequences are then optimized using smaller proxy models to efficiently handle computational expenses [21]. Model-based reinforcement learning (RL)

methods, like the ones that use AlphaZero, have shown success in protein backbone design. They outperform standard Monte Carlo tree search methods by adding secondary objectives and introducing new reward structures [23]. These discoveries demonstrate the profound impact of RL on protein engineering, enabling the development of proteins with customized characteristics for a wide range of uses in medicine, biotechnology, and synthetic biology.

2.2.2. Design of Protein-Protein Interactions

Reinforcement Learning (RL) has demonstrated considerable promise in the development of protein-protein interactions by facilitating the enhancement of binding affinities and the refinement of interaction specificities. Advancements in recent RL methods have resulted in the creation of advanced models capable of predicting and improving protein-protein interactions. An example of this is the RL pipeline that was created to find communities in weighted protein-protein interaction networks. This pipeline showed enhanced accuracy and speed in detecting new protein complexes, which emphasizes the scalability and efficiency of RL in this specific field [24]. Another significant contribution is the research that introduced the PPI-former model. This model utilized a large-scale dataset and SE(3)-equivariant representations to predict the effects of mutations on protein-protein interactions. The model achieved state-of-the-art performance in practical case studies, including SARS-CoV-2 antibody design [25]. In addition, the UniBind framework was introduced. It uses deep learning to examine protein-protein interactions at the residue and atom levels. This framework has been successful in accurately predicting the impact of mutations on binding affinities. Furthermore, it offers valuable insights into viral infectivity and variant evolution. This information is based on a study cited as [26]. These works highlight the significant influence of reinforcement learning (RL) and deep learning in the field of protein engineering. This enables the creation of proteins with customized interaction features, which can be used in various fields such as medicine, biotechnology, and synthetic biology (Figure 1C).

2.3. *Transfer Learning and Few-Shot Learning*

2.3.1. Leveraging Pre-Trained Models for Protein Design

Transfer Learning and Few-Shot Learning are innovative methods in protein design that utilize pre-trained models to enhance protein properties with limited experimental data. These strategies facilitate the adjustment of models that have been trained on huge and varied datasets to specific protein engineering activities, thereby greatly minimizing the requirement for additional data gathering. For example, the effectiveness of pre-trained protein language models (PLMs) such as ESM-2 and ProGen in predicting protein fitness landscapes using few-shot learning was shown, thus improving the accuracy of protein design with little wet-lab data [27]. Furthermore, it was demonstrated how transfer learning may be utilized to optimize deep learning models for the purpose of predicting protein expression based on 5'UTR sequences in various situations. This approach enhances the ability of these models to generalize and be applied to varied genetic backgrounds [28]. A different significant work examined the combination of deep learning and transfer learning in protein design, emphasizing the potential of both techniques to create functional sites and develop new protein interactions with great accuracy [29]. The progress made in transfer learning and few-shot learning highlights the ability to transform protein engineering by facilitating the efficient and economical creation of proteins with specific properties for use in medicine, biotechnology, and synthetic biology (Figure 1C).

2.3.2. Addressing the Challenge of Limited Data in Protein Engineering

The integration of powerful computational approaches and machine learning techniques has made it increasingly practical to tackle the obstacle of limited data in protein engineering. Efficient algorithms are necessary to navigate and optimize protein attributes due to the wide sequence space and combinatorial complexity of protein creation [30]. Machine learning models, namely those utilizing semi-supervised and transfer learning methods, have played a crucial role in estimating

protein fitness landscapes with a small amount of experimental data. As a result, they have been able to guide protein engineering campaigns more efficiently [31]. In addition, data-driven methods have utilized high-throughput experimental data to enhance the catalytic activity and selectivity of enzymes, demonstrating the promise of machine learning in dealing with limited data availability [32]. By using a variety of training datasets, such as those obtained from X-ray crystallography, NMR, and cryo-EM, the performance of the model has been improved. This is achieved by reducing biases and enhancing the ability to apply the model to varied protein structures [33]. In addition, the utilization of evolutionary probability and stacking regression models has been employed to enhance protein characteristics, emphasizing the significance of computational techniques in addressing the constraints imposed by limited experimental data [34]. The progress made in computational and machine learning techniques highlights their crucial role in tackling the difficulties posed by limited data in protein engineering. This progress also paves the path for more effective and creative strategies for designing proteins.

2.4. Interpretable AI for Protein Design

2.4.1. Explainable AI Models for Rational Protein Engineering

Interpretable AI, also known as XAI, is gaining recognition as an essential element in protein design. It provides transparency and valuable insights into the decision-making processes of machine learning models used for rational protein engineering. The incorporation of Explainable Artificial Intelligence (XAI) techniques tackles the issue of the “black box” phenomenon that arises in intricate AI models, hence improving the credibility and dependability of forecasts [35]. For example, researchers have used feature attribution approaches and instance-based analysis to clarify the underlying mechanisms of protein-protein interactions. This has led to an improvement in the interpretability of prediction models [36]. The latest progress has shown the practical use of XAI in detecting DNA-binding proteins and enhancing the brightness of Green Fluorescent Proteins. This highlights the effectiveness of explainable models in real-world protein engineering activities. In addition, the advancement of self-explaining models and uncertainty assessment methods has made it easier to create proteins with specific features by offering clear justifications for model predictions [37]. These methods not only improve the clarity of the model but also provide guidance for experimental verification, guaranteeing that protein designs guided by AI are both dependable and efficient [38]. The integration of XAI into protein engineering pipelines is expected to transform the design and optimization of proteins, leading to more efficient and interpretable AI-driven solutions in biotechnology and synthetic biology [20] (Figure 1D).

2.4.2. Integration of Domain Knowledge with AI-Driven Approaches

The fusion of domain expertise with AI-driven methodologies is an emerging field of study that seeks to improve the effectiveness, comprehensibility, and dependability of machine learning models. This approach, also known as informed AI, utilizes human experience to direct the development and improvement of AI systems, thus overcoming some limits that exist in solely data-driven methodologies. Embedding domain knowledge into AI models can greatly enhance their interpretability and resilience, as demonstrated by recent research in diverse domains like health, engineering, and environmental science [39]. Integrating clinical guidelines and expert knowledge into machine learning pipelines in the medical field has been proven to improve the accuracy, interpretability, and adherence to clinical standards of models, especially in situations where data is scarce [40]. Similarly, the utilization of many artificial intelligence agents that are specialized in different domains has shown to have greater capacities in discovering knowledge across other domains. This, in turn, enables the generation of more complete and precise insights. In addition, domain expertise can be included at different points in the AI pipeline, including data preprocessing, model training, and evaluation, to guarantee that the models are not only precise but also consistent with recognized principles particular to the domain [41,42]. This strategy, which combines data-driven and knowledge-driven techniques, tackles important difficulties such as expensive data

collection and the risk of overfitting. As a result, it leads to the development of more generalizable and dependable AI systems [43]. Incorporating domain expertise is vital for the development of explainable AI systems, which are necessary for establishing confidence and enabling the ethical implementation of AI technologies in sensitive sectors such as healthcare and finance. In general, combining domain knowledge with AI-driven methods has great potential for enhancing the capabilities of AI systems, making them more efficient, dependable, and in line with human expertise and ethical standards [44].

3. Computational Methods in Enzyme Engineering

3.1. Structure-Based Design Strategies

3.1.1. Homology Modeling and Threading Techniques

Homology modeling and threading are essential tools in structure-based protein design, enabling the prediction of protein structures in the absence of experimental data [45,46]. Homology modeling, also known as comparative modeling, is based on the assumption that proteins with comparable sequences would have similar structures. This makes it the preferred method when a homologous structure is present in the Protein Data Bank (PDB) [47]. This method has played a crucial role in the process of finding new therapeutics. It enables researchers to create accurate three-dimensional models of certain proteins, which helps them gain insights into how these proteins interact with drug molecules and aids in the development of novel medications. Advancements in homology modeling, including superior sequence alignment methods and loop modeling techniques, have greatly improved the accuracy of these models, even for proteins that have a low sequence identity to their templates. Alternatively, threading, which is sometimes referred to as fold recognition, is used in cases where no homologous structures are present [48]. This method involves aligning the desired sequence with a database of established protein folds. A score system is then used to assess the compatibility between the sequence and each template structure [47,49]. Threading methods have advanced to include advanced algorithms, such as probabilistic graphical models and dynamic programming, in order to enhance alignment precision and model quality. Both techniques are essential components of contemporary drug discovery processes, facilitating the identification of potential targets for drug development and the creation of new therapeutic treatments using virtual screening and molecular docking. The combination of AI and machine learning has advanced these techniques, increasing their ability to forecast and operate efficiently. This integration also enables the management of extensive datasets produced by genomic and proteomic research [46]. In summary, the combination of homology modeling and threading approaches, supported by computational progress, remains a key driver of breakthroughs in predicting protein structures and designing drugs [45,47] (Figure 2A).

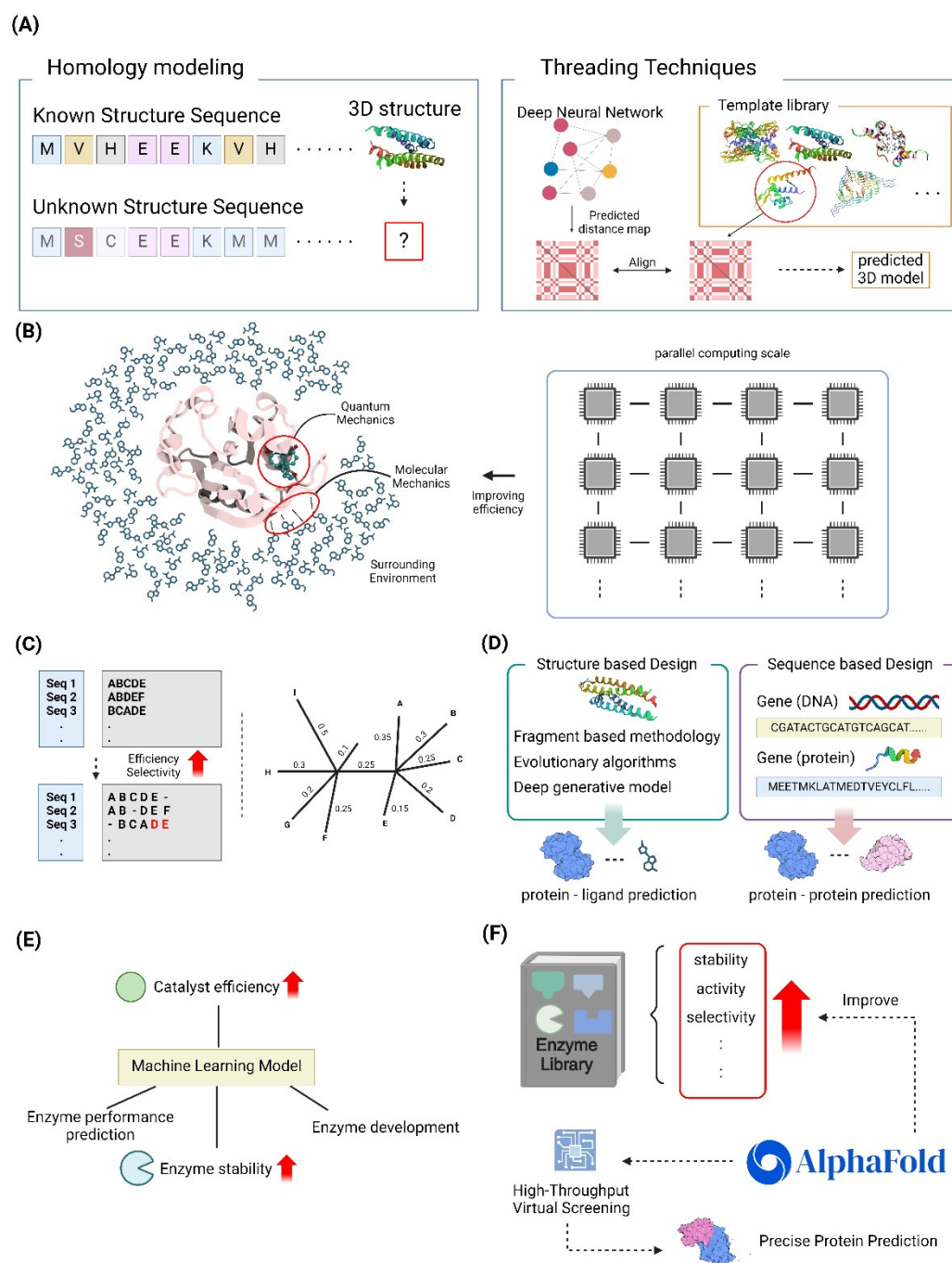


Figure 2. This figure illustrates the advanced computational techniques used in protein structure prediction, ligand-protein interaction modeling, and enzyme engineering. (A) Homology modeling (left image) infers the structure of a protein with an unknown structure by using the structure of a related sequence, based on the observation that proteins with similar sequences tend to have similar structures, while threading techniques (right image) predict a new structure by scoring the alignment of the target sequence against a template library with protein fold information when no structurally similar sequences are available; both methods are utilized for protein structure prediction in the absence of experimental data. (B) Quantum mechanics is used to predict the interactions between a ligand and a protein, while molecular mechanics is applied to model the interactions between a protein and its surrounding environment. The combined use of these two approaches, known as a hybrid method, has been enhanced by recent advancements in parallel computing technologies, overcoming previous limitations and contributing to the development of high-success-rate drugs. (C) The diagram on the left illustrates the process of aligning various protein sequences, enabling researchers to extract information more efficiently from refined sequences. Phylogenetic analysis

allows for the determination of relative distances between elements, and by integrating MSA (Multiple Sequence Alignment) with phylogenetic approaches, information can be analyzed more effectively. (D) Structure-based design methods (left) are used for protein-ligand binding and provide examples of various underlying analytical techniques. Sequence-based design methods (right) are primarily applied to protein-protein interactions and can be broadly categorized into gene and protein sequence analysis. (E) Applying machine learning to enzyme engineering allows for predicting enzyme activity based on library data, improving enzyme stability, and facilitating enzyme development. It also helps explore methods to enhance the efficiency of catalysts or assists in selecting the appropriate catalyst. (F) The development of deep learning software such as AlphaFold has enabled rapid results in high-throughput virtual screening without the need for experimental procedures. Additionally, such software can significantly contribute to understanding enzyme-protein interactions within enzyme libraries, particularly in terms of stability, activity, and selectivity.

3.1.2. Quantum Mechanics/Molecular Mechanics (QM/MM) Approaches

QM/MM techniques have become indispensable in structure-based design methodologies, especially in drug development, because of their precise modeling of intricate biomolecular systems. Hybrid approaches integrate the accuracy of quantum mechanics (QM) in modeling the active site with the efficiency of molecular mechanics (MM) in representing the surrounding environment. This enables detailed simulations of enzyme reactions and interactions with ligands. Recent progress has been made in enhancing the scalability and efficiency of QM/MM simulations by utilizing exascale computing. This allows for the handling of huge biological systems and extended simulation timelines, which were previously difficult due to computational constraints [50,51]. The emergence of interfaces such as the MiMiC framework has showcased substantial parallel efficiency, facilitating the precise examination of thermodynamics and kinetics in drug targets with a high level of precision [50]. In addition, the use of machine learning techniques has increased the accuracy of QM/MM methodologies, making it easier to study energy transfer processes in biomolecular machines. The advancements discussed here demonstrate the potential of QM/MM techniques to significantly transform drug design. These approaches offer chemically precise insights into molecular interactions, leading to an enhanced success rate in drug development initiatives [52]. With the continuous expansion of computer resources, QM/MM approaches are in a position to make even more significant advancements in the field. These methods can tackle more intricate biological inquiries and facilitate more accurate therapeutic interventions [53,54] (Figure 2B).

3.2. Sequence-Based Design Methods

3.2.1. Multiple Sequence Alignments and Phylogenetic Analysis

Multiple sequence alignment (MSA) and phylogenetic analysis are essential techniques for designing sequences based on their alignment and evolutionary relationships. These technologies have made substantial progress in recent years. The utilization of MSA is essential for a range of biological investigations, such as the estimation of phylogeny and the prediction of RNA structure. The scalability and accuracy of MSA algorithms, such as the EMMA technique, have been enhanced by recent advancements. These improvements are particularly beneficial for large datasets. The EMMA approach does this by efficiently managing computational resources through a divide-and-conquer strategy [55]. Researchers have also investigated bioinspired algorithms, which provide innovative methods to improve the precision and speed of alignment [56]. Phylogenetic analysis, which utilizes Multiple Sequence Alignments (MSAs) to deduce evolutionary connections, has been enhanced by advanced computer techniques such as maximum likelihood and Bayesian inference. These methods provide reliable frameworks for generating phylogenetic trees [57]. Recent research has shown that DNA sequences can be just as successful as protein sequences in determining deep phylogenies. This challenges long-held notions and broadens the range of phylogenetic approaches that can be used [58]. The integration of advanced computational tools and methods has supported these improvements, leading to better resolution and reliability of phylogenetic trees. As a result, our understanding of evolutionary processes has been enhanced [59]. As sequencing technology progresses,

it is crucial to continue developing and improving Multiple Sequence Alignment (MSA) and phylogenetic approaches. These advancements are essential for tackling intricate biological inquiries and pushing forward the discipline of bioinformatics [60] (Figure 2C).

3.2.2. Coevolution-Based Approaches for Enzyme Design

Coevolution-based methodologies have become a potent instrument in the field of enzyme design. These methodologies utilize the evolutionary information included in protein sequences to pinpoint crucial interactions and mutations that can improve the activity of enzymes. These techniques employ numerous sequence alignments to identify coevolving residues, which are pairings of amino acids that have evolved together to preserve structural integrity and function. Notable progress in this area involves the creation of methods such as SCANEER, which use sequence coevolution analysis to forecast enzyme performance. This enables the identification of specific mutations that can enhance enzyme efficiency and substrate selectivity [61]. These methods have effectively been used on several enzymes, such as beta-lactamase and aminoglycoside phosphotransferase, to show their ability to enhance enzyme activity for industrial and pharmacological purposes. In addition, the investigation of coevolution has played a key role in the identification of allosteric sites. These sites are essential for controlling enzyme activity and can be specifically targeted for the design of drugs [62]. The combination of computational tools and machine learning has increased the effectiveness of coevolution-based techniques, allowing for the creation of enzymes with new catalytic characteristics and enhanced stability [63,64]. As research progresses, coevolution-based methods are expected to have a crucial impact on the deliberate development of enzymes, providing valuable insights that connect natural evolution with synthetic biology.

3.3. Hybrid Methods

3.3.1. Integration of Structure and Sequence Information

Hybrid approaches in drug and protein design combine both structure-based and sequence-based tactics to enhance the optimization of novel therapies. Structure-based design utilizes the three-dimensional structures of target proteins to uncover and enhance therapeutic candidates. This approach involves techniques such as fragment-based methodologies, evolutionary algorithms, and deep generative models, as demonstrated in recent works [65,66]. This method takes advantage of improvements in computational capacity and machine learning, which improve the ability to anticipate interactions between proteins and ligands and explore the field of chemistry [67]. Conversely, sequence-based design prioritizes the analysis of genetic and amino acid sequences in order to forecast protein activities and interactions. Direct coupling analysis and statistical modeling are employed to deduce co-evolutionary characteristics, which are essential for the advancement of hybrid proteins and genetic sensors [68,69]. By integrating the characteristics of both approaches, the integration of these methodologies in hybrid modeling provides a more thorough understanding of protein dynamics and function. This facilitates the design of more effective medications and proteins, as observed in the field of protein research [67]. Recent studies highlight the possibility of merging these tactics to overcome the inherent constraints of each method when employed separately, hence facilitating the development of inventive solutions in drug discovery and protein engineering [70] (Figure 2D).

3.3.2. Machine Learning-Assisted Enzyme Engineering

Machine learning-assisted enzyme engineering is an advancing discipline that integrates computational and experimental methods to improve enzyme characteristics for many uses. Recent progress has shown that machine learning (ML) models can be used to forecast enzyme performance and stability, enhance catalytic efficiency, and assist in the logical development of enzymes. ML models can effectively explore the extensive protein sequence space to discover potential enzyme variations. This study focuses on the use of ML in predicting protein architectures and substrate

specificity [71]. Moreover, the combination of machine learning (ML) with directed evolution has been demonstrated to expedite the process of enzyme optimization by lessening the workload of experiments. This highlights the significance of ML in providing guidance for directed evolution in the field of protein engineering [72]. In addition, the advancement of innovative machine learning algorithms, such as MODIFY, has made it possible to simultaneously optimize both the effectiveness and variety of enzymes. This has greatly facilitated the identification of enzyme activities that are unique to the natural world [73]. The progress made in ML in enzyme engineering highlights the significant and profound influence it has, providing new opportunities for developing biocatalysts that have improved performance and unique capabilities (Figure 2E).

3.4. High-Throughput Virtual Screening

3.4.1. In-Silico Directed Evolution

High-throughput virtual screening (HTVS) and in-silico directed evolution are innovative methods used in drug discovery and protein engineering. These methods utilize computing capacity to efficiently explore large chemical and protein spaces. HTVS employs computational models to efficiently assess extensive collections of compounds, discovering potential bioactive molecules without the necessity of physical synthesis. This approach overcomes the constraints of traditional high-throughput screening (HTS), which relies on pre-existing compounds [74,75]. Recent progress in machine learning, specifically convolutional neural networks such as AtomNet, has shown great success in identifying new drug-like structures in different medical fields. This suggests that computational methods can effectively replace high-throughput screening (HTS) in the early stages of drug discovery [75]. In-silico directed evolution utilizes computational algorithms to model the process of evolution, enhancing protein functionalities through repeated cycles of mutation and selection. The utilization of deep learning models, such as AlphaFold2, has improved this method. These models are capable of accurately predicting protein structures, thereby enabling the creation of proteins with specific binding capabilities [76]. EvoPro is a new pipeline that combines deep learning to predict protein structure and optimize protein sequences. It demonstrates the effectiveness of in-silico approaches in evolving protein binders. These computational methodologies not only speed up the process of discovery but also increase the range of chemicals and proteins that researchers may access, thereby enabling the development of unique therapeutic solutions [77,78] (Figure 2F).

3.4.2. Computational Library Design for Enzyme Engineering

Computational library design for enzyme engineering is an innovative method that use sophisticated computational techniques to enhance enzyme characteristics, including stability, activity, and substrate selectivity. This approach entails the generation of extensive and varied collections of enzyme variations, which can be computationally analyzed to pinpoint potential candidates possessing specific characteristics. The effectiveness of this technique has been greatly improved by recent breakthroughs in machine learning and structural bioinformatics. For example, advanced tools such as AlphaFold have brought about a significant transformation in the field of protein structure prediction. These tools enable researchers to precisely model enzyme structures and forecast the impact of mutations on enzyme activity [79,80]. Machine learning methods are being more and more utilized to analyze large datasets produced from high-throughput sequencing and screening. This allows for the detection of advantageous mutations and the forecasting of enzyme performance in different circumstances [81,82]. Computational approaches not only decrease the time and expense of traditional experimental methods, but also broaden the range of enzyme engineering by exploring a wider sequence space. Computational library design is positioned to have a vital impact on the development of new biocatalysts for industrial and pharmacological purposes [3,79] (Figure 2F).

4. Molecular Dynamics Simulation Studies of Biomolecular Systems

4.1. *Advanced Sampling Techniques*

4.1.1. Replica Exchange Molecular Dynamics

Replica Exchange Molecular Dynamics (REMD) is a powerful enhanced sampling technique widely utilized in molecular dynamics simulations to overcome the limitations of traditional MD methods, particularly in exploring rugged energy landscapes of biomolecular systems. REMD involves simulating multiple copies, or replicas, of a system at different temperatures, allowing for the efficient sampling of conformational space by periodically exchanging configurations between replicas based on a Metropolis criterion. This method is particularly effective in studying systems with high energy barriers, such as protein folding, aggregation, and receptor-ligand interactions. Recent studies have demonstrated the utility of REMD in elucidating the mechanisms of protein aggregation associated with diseases like Alzheimer's and Parkinson's, as well as in the structural prediction of transmembrane proteins using implicit solvent models to reduce computational costs [83–85]. The method's adaptability to parallel computing environments further enhances its efficiency, making it suitable for large-scale simulations on supercomputers [84]. Moreover, advancements such as the multicanonical replica-exchange method (MUCAREM) and the integration of implicit solvent models have been developed to improve sampling efficiency and reduce computational demands [84]. Overall, REMD continues to be a vital tool in biomolecular research, providing detailed insights into the dynamic behavior of complex systems at an atomic level (Figure 3A).

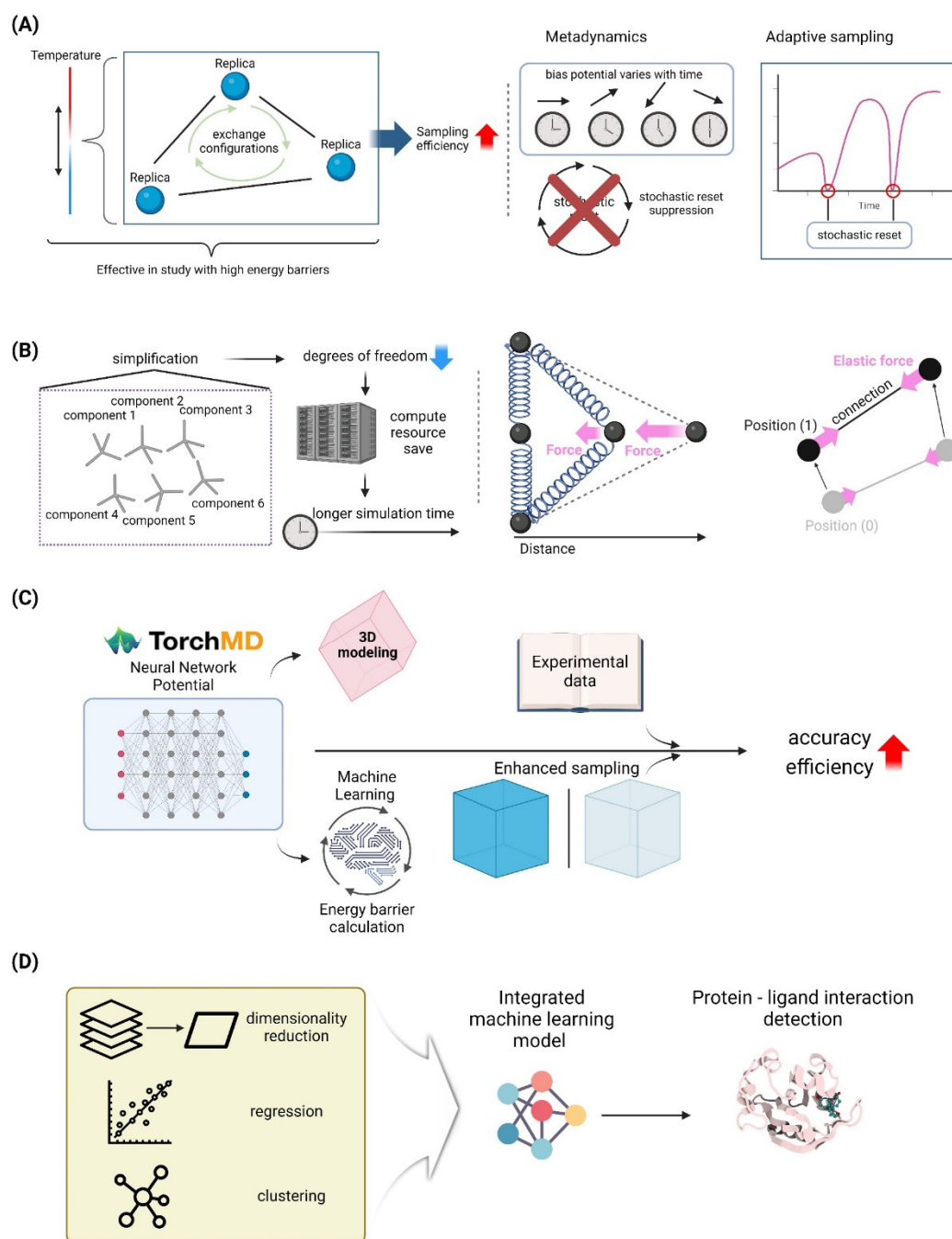


Figure 3. This figure illustrates various computational techniques used to enhance sampling efficiency and reduce computational resources in biomolecular simulations, highlighting their distinct approaches and applications. (A) Diagram of replica exchange molecular dynamics (left). This method forms multiple replicas and allows efficient simulation sampling through periodic exchanges of components between these replicas. It is particularly suitable for scenarios involving high-energy barriers in biomolecular interactions and can be conducted at different temperatures. Diagram illustrating the difference between metadynamics and adaptive sampling methods in terms of stochastic reset (right). Stochastic reset refers to the model probabilistically reverting to a previous state; metadynamics prevents this by introducing a bias potential, while adaptive sampling intentionally restarts the model at specific locations to enhance the sampling method. (B) Diagram of the MARTINI model and its advantages (left). The MARTINI model simplifies molecular systems by grouping multiple elements (primarily atoms) into larger entities called beads, rather than treating each element individually. This simplification reduces the degrees of freedom, significantly lowering computational resources required and enabling longer simulations with limited resources. Schematic

of Elastic Network Models (ENMs) (right). ENMs represent the forces between biomolecules in large simulation environments using a spring model, where each node typically represents an alpha carbon. The longer the distance, the stronger the pulling force, allowing the possible conformations of biomolecules upon deformation to be inferred through this model. (C) Neural network potentials, such as Torch MD, enable 3D modeling and high-energy barrier calculations through machine learning. When combined with enhanced sampling techniques or experimental data, neural network potentials can achieve greater accuracy and efficiency. (D) An integrated model utilizing machine learning tools such as dimensionality reduction, regression, and clustering enables the modeling of complex biomolecular systems, such as detecting protein-ligand interactions.

4.1.2. Metadynamics and Adaptive Sampling Methods

Metadynamics and adaptive sampling approaches are essential tools in molecular dynamics (MD) simulations, specifically for investigating the intricate energy landscapes of biomolecular systems. Metadynamics improves the efficiency of sampling by introducing a bias potential that varies with time. This potential discourages the system from returning to states that have already been examined, enabling it to overcome energy barriers and explore novel conformations. The effectiveness of metadynamics relies heavily on the choice of collective variables (CVs), which must precisely reflect the sluggish phases of the system's dynamics [86]. Recent advancements, such as the combination of stochastic resetting and metadynamics, have demonstrated potential in speeding up simulations even when less than ideal variables are utilized. This approach offers a substantial increase in speed without incurring any extra computing expenses [86]. However, adaptive sampling methods, such as adaptive path sampling and machine learning-enhanced sampling, maintain the thermodynamic ensemble while improving sampling by selectively restarting MD trajectories at specific locations. By employing deep learning, these techniques have proven to be highly successful in capturing protein conformational changes. They achieve this by accurately predicting the most favorable areas of the conformational space to investigate [87]. Ongoing research is dedicated to enhancing the efficiency and applicability of both metadynamics and adaptive sampling approaches. This study aims to broaden their scope to encompass a wider spectrum of biomolecular systems. By doing so, it will provide a more comprehensive understanding of protein dynamics and facilitate drug development efforts [87,88] (Figure 3A).

4.2. Coarse-Grained Models

4.2.1. MARTINI force Field and Its Applications

The MARTINI force field is a well-established coarse-grained model employed in molecular dynamics simulations for the investigation of biomolecular systems. It provides a favorable trade-off between computational efficiency and accuracy. The MARTINI model, created by Marrink et al., simplifies molecular structures by combining several atoms into larger "beads." This simplification reduces the complexity of the system and enables simulations of massive biomolecular complexes over extended periods of time. This method has proven to be especially successful in replicating lipid membranes, protein folding, and interactions within intricate biological settings. The model MARTINI 3 has increased its application through recent advances. These advancements have improved the depiction of small molecules and increased the accuracy of lipid and protein simulations. This has been demonstrated in studies that have explored drug delivery systems and protein-protein interactions [89,90]. The integration of both top-down and bottom-up parameterization methodologies has enabled these improvements, resulting in a force field that accurately reproduces experimental partitioning free energies [91]. The MARTINI force field's adaptability is emphasized by its successful integration into several simulation platforms, such as OpenMM, allowing for its extensive application in both academic and industrial research environments [92]. Continuing work in the field are focused on improving the model's parameters and broadening its application range, namely in drug development and the examination of membrane proteins and cryptic pockets [90] (Figure 3B).

4.2.2. Elastic Network Models for Large-Scale Simulations

Elastic Network Models (ENMs) are a widely used method in molecular dynamics simulations that are particularly useful for studying the overall movements of biomolecular systems. Elastic network models (ENMs) describe biomolecules as networks of nodes connected by springs, with the nodes commonly representing the C α atoms of proteins. This representation enables the rapid calculation of normal modes and the study of slow, large-scale conformational changes. This approach is beneficial for investigating computationally challenging processes, such as protein folding, allosteric transitions, and massive biomolecular assemblies, which cannot be effectively studied using all-atom models. Recent progress has been made in improving the precision and usefulness of ENMs by combining them with other computational methods, such as molecular dynamics simulations and perturbation response scanning. This integration allows for the study of intricate systems, such as ubiquitin-specific protease 7 (USP7) and its mechanisms of allosteric regulation [93,94]. In addition, ENMs have been modified to different resolutions and parameterizations in order to accurately represent the dynamics of diverse biomolecular systems. This adaptation has shown resilience across numerous formalisms and applications [95]. These models are continuously improved to enhance their ability to make accurate predictions and to integrate them into multiscale modeling frameworks. This expansion increases their usefulness in the fields of structural biology and drug development [93,95] (Figure 3B).

4.3. Long-Timescale Simulations

4.3.1. Specialized Hardware for MD Simulations

Advanced hardware has transformed long-term molecular dynamics (MD) simulations, allowing researchers to investigate biomolecular systems with exceptional precision and effectiveness. Notable progress has been made through the utilization of Graphics Processing Units (GPUs), Field-Programmable Gate Arrays (FPGAs), and Application-Specific Integrated Circuits (ASICs), each providing unique benefits in terms of velocity and computational capability. Originally intended for parallel processing in graphics, GPUs have been adapted to expedite MD simulations by effectively managing non-bonded interactions, resulting in a substantial decrease in computation time and cost [96,97]. FPGAs have the advantage of flexibility and efficiency, enabling the customization and optimization of MD algorithms. This customization can result in significant improvements in the speed of specific computational workloads [98,99]. ASICs, like the ones seen in Anton supercomputers, are designed exclusively for MD simulations. They provide impressive performance improvements by optimizing every component of the simulation process [97,100]. The hardware developments have increased the possible duration of simulations to the millisecond range and made MD simulations more accessible to a wider group of researchers. This has led to significant progress in drug discovery and structural biology [97]. The continuous advancement of technology is anticipated to boost the capabilities of MD simulations by integrating machine learning with specialized hardware. This integration will enable more detailed and precise examinations of complicated biomolecular processes.

4.3.2. Enhanced Sampling Techniques for Accessing Biologically Relevant Timescales

Enhanced sampling approaches play a crucial role in expanding the time span of molecular dynamics (MD) simulations, allowing us to explore biologically significant time scales that would otherwise be impossible due to computational limitations. These methods, including metadynamics, replica-exchange molecular dynamics (REMD), and stochastic resetting, aim to tackle the difficulty of surpassing high-energy obstacles and investigating the complex energy patterns commonly found in biomolecular systems. Metadynamics is a method that improves sampling by introducing a bias potential that changes over time along specific collective variables. This helps to explore unusual events and calculate differences in free energy [86]. REMD, in contrast, utilizes the simulation of numerous duplicates of the system at various temperatures to enable effective sampling of diverse conformations by promoting transitions over energy barriers. Recent advancements, such as the

integration of metadynamics with stochastic resetting, have shown substantial improvement in sampling efficiency. This improvement is observed even when suboptimal collective variables are employed, hence expanding the range of applications for these methods [86]. These advanced sampling techniques not only enhance the precision of molecular dynamics (MD) simulations, but also broaden their applicability in investigating intricate biological processes such as protein folding, ligand binding, and allosteric regulation. As a result, they contribute to the advancement of our comprehension of molecular mechanisms and assist in the discovery of new drugs [101] (Figure 3A).

4.4. Machine Learning-Enhanced MD Simulations

4.4.1. Neural Network Potentials for Accurate and Efficient Simulations

Neural network potentials (NNPs) are a revolutionary method in molecular dynamics (MD) simulations that offer both precision and efficiency in modeling intricate biomolecular systems. Natural language processing (NNPs) utilize machine learning techniques to estimate potential energy surfaces, providing a computationally efficient alternative to conventional quantum mechanical calculations. This is especially advantageous for simulating extensive systems over extended durations. Recent technological developments, exemplified by TorchMD and its successor TorchMD-Net 2.0, have shown that neural network potentials (NNPs) may reliably simulate molecules that were not part of their training data. This demonstrates the ability of NNPs to generalize and perform well in diverse scenarios, indicating their robustness and versatility [102,103]. The models are trained utilizing data from accurate simulations or experimental observations, as demonstrated in the Differentiable Trajectory Reweighting approach. This method incorporates experimental data to improve Neural Network Potentials (NNPs) without the need to differentiate through extensive Molecular Dynamics (MD) simulations [104]. Moreover, incorporating active learning procedures, as explored in recent research, improves the capacity of NNPs to forecast infrequent occurrences, like bond breaking, by continuously updating the model with fresh data obtained through increased sampling approaches [105]. The inclusion of equivariance in neural networks, which acknowledges the spatial symmetries of molecular systems, has enhanced the precision and dependability of NNPs, rendering them a potent tool in both academic research and industrial applications [106]. These advancements highlight the capacity of NNPs to greatly enhance our comprehension of molecular dynamics, enabling major progress in fields like drug discovery and materials science (Figure 3C).

4.4.2. AI-Driven Analysis of MD Trajectories

The utilization of artificial intelligence (AI) to analyze molecular dynamics (MD) trajectories has emerged as a revolutionary method for comprehending intricate biomolecular systems. This strategy harnesses machine learning (ML) to derive valuable insights from extensive datasets. By incorporating machine learning techniques, including as dimensionality reduction, clustering, regression, and classification, it becomes possible to analyze and interpret MD simulation data more efficiently. This overcomes the limitations of traditional methods that mainly rely on manual inspection and intuition [107]. Unsupervised deep learning techniques, such as graph neural networks, have shown promise in detecting complex patterns in MD data with many dimensions. They can capture the dynamics of protein-ligand interactions that are often difficult to analyze using traditional methods [108]. In addition, trajectory-based machine learning methods such as TrajML enable the development of precise force fields by training on ab initio molecular dynamics data. This improves the accuracy of MD simulations without the computational complexity associated with conventional techniques [109]. AI-enhanced techniques enhance the accuracy and efficiency of MD simulations and offer new opportunities to study protein dynamics, ligand-binding affinities, and other important biological processes. This ultimately contributes to the progress of drug discovery and materials science in fields such as [104,110]. The integration of AI with MD simulations is anticipated to better the modeling of intricate biomolecular systems, leading to greater understanding and allowing the development of innovative therapeutic approaches (Figure 3D).

5. Advances in Computational Docking and Drug Design

5.1. Protein-Ligand Docking

5.1.1. Flexible Docking Algorithms

Flexible docking methods have greatly improved the field of protein-ligand docking by enabling the dynamic modeling of ligands and protein targets. This has resulted in more accurate predictions of binding modes and has made drug development easier. Flexible docking is a docking method that allows for conformational changes in both the protein and ligand. This is important for accurately mimicking biological interactions, unlike typical rigid docking methods. Methods like as global optimization, step-by-step building, and multi-conformer docking have been created to investigate a broad spectrum of conformations, as observed in software applications like AutoDock Vina, DOCK, and MDOCK. Although these methods require significant computer resources, they have demonstrated higher success rates in predicting the position of flexible ligands. However, they do not consistently beat rigid docking in virtual screening due to difficulties in accurately scoring the results [111]. Recent research highlights the importance of improved scoring methods that can precisely consider the energetic effects of ligand flexibility, including internal strain and changes in entropy [112,113]. Machine learning methods are getting more and more incorporated to improve the accuracy of scoring and decrease the computational expenses, which shows potential for breakthroughs in flexible docking approaches [112,114] (Figure 4A).

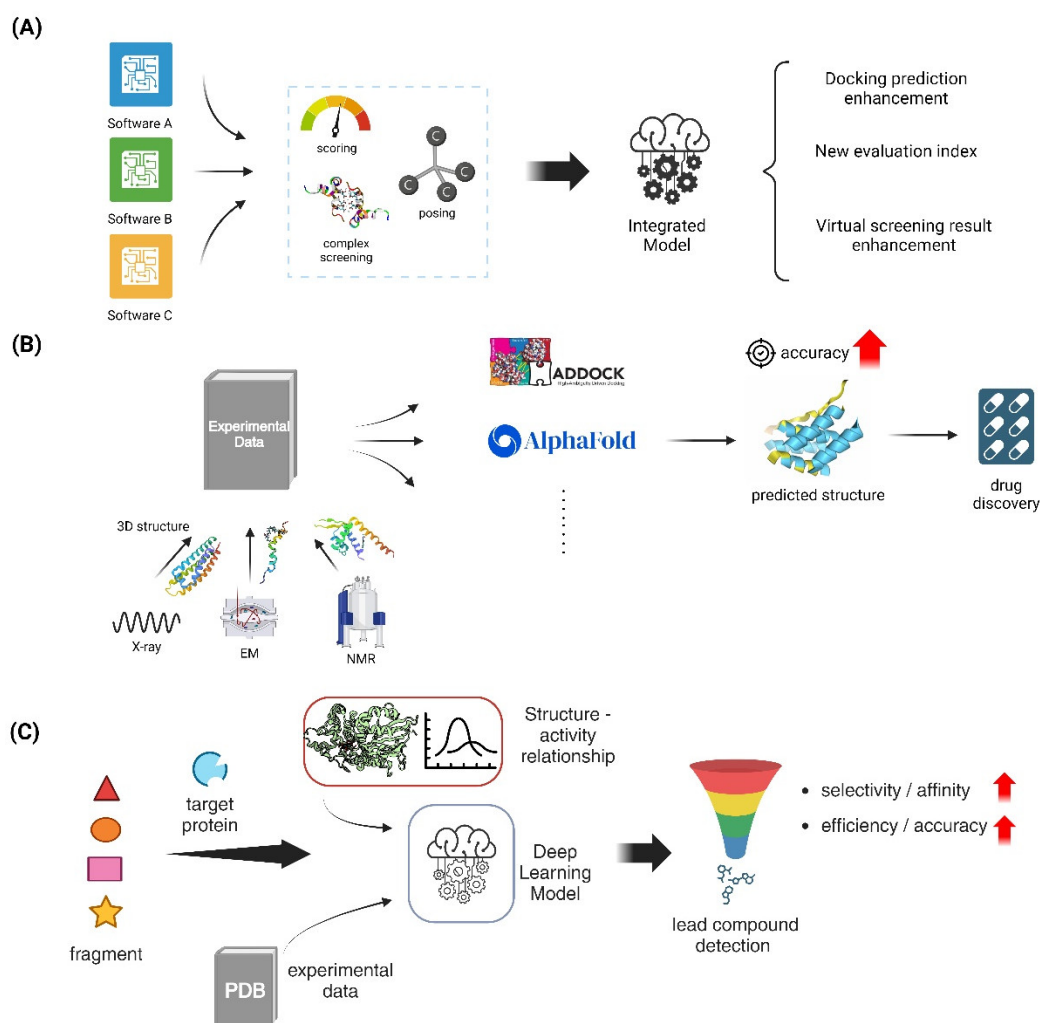


Figure 4. This figure highlights various approaches that enhance the accuracy and reliability of drug discovery processes by integrating computational models, experimental data, and deep learning methods. It showcases how combining these elements can improve prediction performance, structural accuracy, and lead compound optimization. (A) A model integrating output data from

various software improves prediction performance, generates new evaluation metrics, and provides more reliable information during the virtual screening stage. Input parameters include docking scores, molecular (or component) poses, and representations of complexes. (B) Experimental data-based libraries enable the use of various software tools. These libraries compile 3D structures obtained through methods such as X-ray crystallography, electron microscopy (EM), and NMR spectroscopy. By leveraging actual data, software like AlphaFold and HADDOCK can achieve highly accurate structural predictions, ultimately contributing to the drug development process. (C) A deep learning model for simulating the binding of lead compound candidates to target proteins can achieve superior performance by integrating structure-activity relationship data with experimental data. Experimental data can be sourced from databases like PDB, which mainly include data obtained from X-ray crystallography, electron microscopy (EM), and NMR spectroscopy. Ultimately, the integrated deep learning model enhances selectivity and affinity during the lead compound optimization stage, improving efficiency and accuracy at every step.

5.1.2. Consensus Docking Approaches

The significance of consensus docking approaches in protein-ligand interactions has been emphasized by recent advancements in computer docking and drug design. These approaches have greatly enhanced the accuracy and dependability of predictions. Consensus docking approaches, which merge the outcomes of several docking programs, have been demonstrated to improve the results of virtual screening by averaging the scores or ranks of individual molecules. This approach overcomes the restrictions of using a single docking algorithm [115,116]. An example of this is the MetaDOCK method, which combines the data from Auto-Dock4.2, LeDock, and rDOCK. It has been shown to outperform individual programs in terms of scoring, posing, and screening protein-ligand complexes [117]. Furthermore, new consensus measures such as the Exponential Consensus Rank (ECR) have been created to overcome the drawbacks of conventional approaches. These metrics provide enhancements by employing rank-based techniques instead of score-based strategies, which are not influenced by score units and scales [115]. The integration of machine learning approaches enhances the prediction capacities of consensus docking, complementing these improvements. Consensus docking is anticipated to have a vital role in the rational development of therapies as the science advances. It will offer a thorough comprehension of molecular interactions and aid in the identification of new drugs [116] (Figure 4A).

5.2. Protein-Protein Docking

5.2.1. Template-Based Docking Methods

Advancements in computational docking have greatly enhanced protein-protein docking techniques, with template-based docking emerging as a highly efficient method. Template-based docking utilizes the structural information obtained from known protein complexes to forecast the interaction surfaces of novel protein pairings. This method provides a more precise alternative to classic *ab initio* methods, but it requires the availability of suitable templates [118]. This method has been improved through the creation of extensive template libraries, such as those produced from the Protein Data Bank (PDB), which consist of several protein complexes that are used as benchmarks for docking predictions [119]. Recent research has shown that template-based approaches are useful in capturing the conformational dynamics of protein-protein interactions, which is crucial for accurately modeling these complexes. For instance, the combination of AlphaFold2 and template-based docking has demonstrated potential in accurately predicting protein complexes. This is achieved by employing deep learning algorithms to generate structural templates [120]. Furthermore, the utilization of paired interfacial residue restraints has been demonstrated to enhance docking predictions, particularly in situations requiring moderate to substantial conformational alterations [118]. With the continuous expansion of computer resources and structural databases, template-based docking is anticipated to have a growing significance in predicting protein-protein interactions. This will aid in advancing medication design and enhancing our comprehension of intricate biological processes (Figure 4B).

5.2.2. Integration of Experimental Data in Docking Protocols

Computational docking has made substantial progress in improving protein-protein docking methods. This progress has been achieved by integrating experimental data, resulting in greater accuracy and dependability of docking predictions. Integrative methodologies that merge computational docking with experimental techniques, such as small-angle X-ray scattering (SAXS), electron microscopy (EM), and nuclear magnetic resonance (NMR), have demonstrated the ability to enhance docking success rates by offering supplementary structural constraints and filtering capabilities [121,122]. The integrative docking method, as reported by Trinh et al., employs simulated experimental data to enhance the accuracy of docking. This approach showcases the possibility of integrating different experimental methodologies to enhance the quality of docking models. In addition, techniques such as pyDockSAXS and HADDOCK have integrated SAXS data to improve and optimize docked models. This integration allows for better prediction of protein-protein interactions by utilizing low-resolution shape information [122]. By including evolutionary data, such as sequence conservation and coevolution, the accuracy of docking predictions is improved. This is achieved by gaining valuable information about the interface residues that are highly important for the interaction [122]. The incorporation of various experimental datasets into docking protocols is anticipated to have a significant impact on the advancement of the field. This integration, made possible by the continuous development of computational and experimental techniques, will enhance the accuracy of protein-protein interaction modeling and facilitate drug discovery endeavors (Figure 4B).

5.3. *Fragment-Based Drug Design*

5.3.1. In Silico Fragment Growing and Linking Strategies

Advancements in fragment-based drug design (FBDD) have greatly improved the methods of in silico fragment growing and linking. These strategies are crucial in converting first fragment hits into powerful lead compounds. In silico methods, as reported by Moira et al., utilize computational tools to aid in the process of optimizing fragments into lead compounds. These methods integrate techniques such as hot spot analysis and structure-activity relationship (SAR) predictions to guide the expansion of fragments [123]. ACFIS 2.0 incorporates dynamic fragment growth techniques, which facilitate the comprehensive sampling of protein conformations. This enhances the precision of fragment binding predictions and enables the creation of a wide range of compound libraries [124]. Moreover, recent studies have emphasized the effectiveness of employing deep learning models in fragment optimization to expedite the discovery of synthesizable molecules. These models can predict bioactivity and pharmacokinetic features, thereby making the drug discovery process more efficient [123]. By combining computational tactics with experimental data from techniques like X-ray crystallography and NMR, the fragment growth and linking processes can be improved. This ensures that the final compounds have the best possible binding affinities and drug-like features [65]. With the increasing growth of computer power and algorithm sophistication, in silico tactics are anticipated to have a progressively vital part in the efficient development of new therapeutic medicines (Figure 4C).

5.3.2. Machine Learning in Fragment-Based Approaches

We utilized machine learning techniques to augment the in silico fragment growing and linking tactics, resulting in a substantial improvement in the efficiency and accuracy of drug discovery operations. Recent studies in de novo drug design have demonstrated the successful application of machine learning models, namely those applying deep reinforcement learning (DRL), to optimize molecular structures. These algorithms learn how to change existing molecules in order to enhance their attributes. [125]. By incorporating geometric deep learning frameworks such as FRAME, FBDD has been enhanced by properly determining the optimal locations for adding fragments to a ligand and assessing the geometric properties of these additions. This has resulted in improved predictions of the affinity and selectivity of the resultant molecules [126]. Moreover, the utilization of graph-

based deep generative models in conjunction with evolutionary learning procedures has been utilized to optimize several objectives, including binding affinity and pharmacokinetic features, in the creation of innovative compounds [127]. These machine learning-based methods not only simplify the process of designing drugs based on fragments, but also have the ability to efficiently explore large chemical regions, thereby enabling the rapid synthesis of new therapeutic agents. With the increasing computer power and advancement in algorithms, the incorporation of machine learning in FBDD (Fragment-Based Drug Discovery) is expected to have a significant impact on the future of drug discovery. This integration will allow for more accurate and efficient development of drug candidates.

5.4. *Structure-Based Virtual Screening*

5.4.1. Pharmacophore Modeling and Shape-Based Screening

The merging of pharmacophore modeling with shape-based screening has greatly improved structure-based virtual screening, leading to substantial breakthroughs in the drug discovery process. Pharmacophore modeling is a technique that determines the specific arrangement of features required for molecules to interact with each other. It has been very useful in narrowing down large compound libraries to find potential matches. This has been demonstrated in several studies that have used databases like ZINCPharmer for efficient screening. [128,129]. Shape-based screening enhances the analysis by emphasizing the compatibility of the ligand and the target protein in terms of their three-dimensional shapes. This approach has been improved with advanced algorithms like O-LAP, which enhances docking enrichment by comparing shape similarities with inverted binding cavities [130]. By utilizing these methods, it becomes possible to identify a wide range of compounds that have different structures but yet fulfill the requirements of pharmacophoric and form criteria. This enables the exploration of various molecular scaffolds and the finding of new potential drugs. Recent studies have emphasized the significance of machine learning in speeding up pharmacophore-based virtual screening. This allows for the effective management of large chemical spaces and enhances the identification of potential ligand candidates [131]. The advancement of computational tools and databases is likely to have a significant impact on drug design and development. The synergy between pharmacophore modeling and shape-based screening is anticipated to play a crucial part in this advancement [128,131] (Figure 4C).

5.4.2. AI-Driven Virtual Screening Pipelines

The drug development process has been greatly improved by AI-driven virtual screening pipelines, which have transformed structure-based virtual screening. These advancements have led to increased efficiency and accuracy. AI-driven techniques utilize advanced algorithms to assess the intricate three-dimensional structures of target proteins and accurately forecast their interactions with prospective therapeutic molecules. This process greatly simplifies the discovery of highly promising candidates from extensive chemical libraries [132]. These technologies employ machine learning methods, namely graph neural networks (GNNs), to forecast chemical features and enhance drug design by properly simulating intricate molecular interactions [132]. AI has been successfully incorporated into virtual screening, resulting in faster drug discovery processes. One example is ZairaChem, a platform that utilizes AI/ML models to conduct quantitative structure-activity/property relationship modeling. This approach has significantly reduced attrition rates in experimental pipelines, as evidenced by research [133]. In addition, the use of AI-driven methods has allowed for the creation of prediction models that may estimate binding affinities without requiring substantial molecular docking. This has been demonstrated in studies where machine learning has expedited pharmacophore-based virtual screening [131]. These advancements not only expedite the quick detection of lead compounds but also make strong computational tools more accessible, thus enhancing the efficiency and cost-effectiveness of drug development efforts [6]. The incorporation of AI technologies into virtual screening pipelines is anticipated to boost the precision and speed of drug discovery, ultimately resulting in the development of safer and more effective treatments [134].

6. Design and Development of Novel Proteins with Enhanced Functionalities

6.1. De Novo Protein Design

6.1.1. Computational Design of Protein Backbones

The field of de novo protein design has been greatly advanced by recent developments in computational techniques, namely in the design of protein backbones. These advancements have enabled the production of new proteins with improved capabilities. The advancement of complex algorithms, as described by MacDonald and Freemont, has enabled the integration of backbone plasticity into design processes. This overcomes the constraints of using rigid backbone templates and broadens the range of potential protein structures [135]. The ability to be flexible is extremely important for exploring a larger range of sequences and obtaining more intricate functionality. This has been emphasized by recent attempts to create new folds and functional sites using the extensive structural data found in the Protein Data Bank (PDB) [136]. RFdiffusion, an advanced technique, utilizes deep learning to generate novel protein backbones. This is achieved by repeatedly refining random residue frames. The results of this approach show substantial enhancements in the design of proteins with specific structural and functional needs [137]. In addition, the use of machine learning models, such as AlphaFold2 and ProteinMPNN, has significantly enhanced the effectiveness and achievement rates of de novo protein design. These models effectively forecast and optimize both the backbone structures and their related sequences, leading to improved efficiency [138]. These advancements not only improve our capacity to create proteins with specific functions, but also open up possibilities for future use in biomedicine and synthetic biology, where precise manipulation of protein structure and function is crucial [136,139] (Figure 5A).

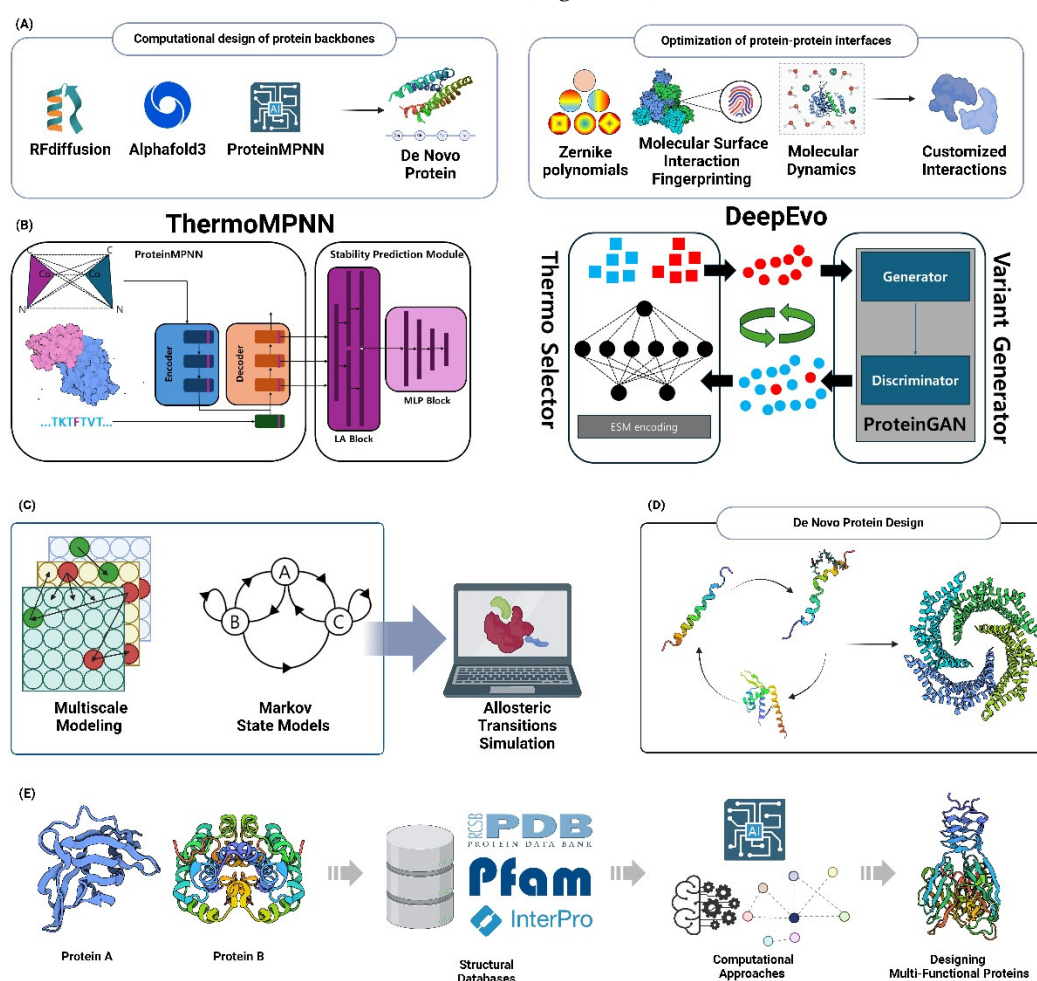


Figure 5. Enhanced functionalities of proteins through computational protein design and development. (A) Advancements in computational techniques, including deep learning models like

RFdiffusion, AlphaFold2, and ProteinMPNN, have significantly improved de novo protein design. Zernike polynomials, Molecular Surface Interaction Fingerprinting (MaSIF), and molecular dynamics techniques help optimize protein-protein interactions. (B) ThermoMPNN is a computational tool that uses a deep neural network trained to predict stability changes in point mutations of a given protein with an initial structure. DeepEvo is an AI-based protein engineering strategy using a protein language model that can predict thermostability variants. (C) Allosteric transition simulations using multiscale modeling and Markov state models can predict protein functions, enabling the creation of customized allosteric regulatory proteins and the development of new protein functions. (D) Deep learning-based computational tools like Rosetta precisely modify protein structures to enhance binding capabilities, enabling the de novo protein design with customized binding properties. (E) Computational Design for domain fusion and chimeric proteins uses structural databases and computer technologies such as machine learning to generate multifunctional proteins.

6.1.2. Optimization of Protein-Protein Interfaces

Computational approaches have greatly improved the optimization of protein-protein interfaces through de novo protein design. These methods allow for exact engineering of molecular interactions, leading to greater functioning. Methods, such as the use of Zernike polynomials, have been created to represent the shape and electrical characteristics of binding sites. These methods enable the improvement of the compatibility of protein surfaces that interact with each other [140]. This method has effectively been used to create protein mutants that have stronger binding affinities. This has been proved in research that focused on the interaction between Ferritin and the Transferrin Receptor [140]. In addition, the incorporation of deep learning frameworks, such as Molecular Surface Interaction Fingerprinting (MaSIF), has introduced a new approach for capturing the essential geometric and chemical characteristics involved in protein-protein interactions. This method has greatly aided in the development of novel protein binders with high specificity and affinity [141]. The use of Monte Carlo simulations and molecular dynamics helps validate and improve interface designs, ensuring that altered proteins attain the expected functional outcomes [140]. As these approaches progress, they provide significant potential for use in synthetic biology and biomedicine. This is because they allow for the creation of proteins with customized interactions, which can lead to the development of new therapies and biomaterials [138,141] (Figure 5A).

6.2. Protein Stability Engineering

6.2.1. Computational Prediction of Stabilizing Mutations

The latest developments in computational methods for predicting stabilizing mutations have greatly improved the field of protein stability engineering. However, the scarcity of these mutations still poses hurdles. ThermoMPNN, a type of computational tool, has demonstrated potential by obtaining a precision rate of 68% in predicting stabilizing mutations for proteins like the bacterial toxin CcdB. However, it has only shown small increases in thermal stability, with an increase of approximately 1°C in the melting temperature [142]. Nevertheless, these methods frequently encounter difficulties when dealing with more intricate targets, such as influenza neuraminidase, underscoring the necessity for enhanced predictive precision [142]. Research has highlighted the drawbacks of existing techniques, pointing out that whereas several computational tools successfully forecast changes that cause destabilization, they struggle to reliably detect variants that promote stabilization [143]. Current endeavors have concentrated on amalgamating empirical data with computational forecasts to augment precision, as exemplified by logistic regression models that were trained on yeast surface display libraries. These models achieved a precision rate of 90% and a 3°C elevation in thermal stability for CcdB [142]. In addition, RaSP, a type of deep learning model, has been created to quickly forecast changes in stability. This provides a scalable approach for analyzing protein variants on a wide scale. However, there are still difficulties in reliably predicting mutations that enhance stability [144]. The progress made in merging computational and experimental methods highlights the potential for improving the accuracy of predicting stabilizing mutations. This is

essential for protein engineering and the creation of new proteins with improved functions [142–144] (Figure 5B).

6.2.2. Design of Thermostable Proteins

Computational techniques have played a significant role in driving recent improvements in the design of thermostable proteins. These approaches have made it possible to engineer proteins with improved stability, which is beneficial for a range of industrial and biological uses. FireProt and its updated version, FireProt 2.0, are tools that have played a crucial role in automating the process of designing thermostable proteins. They achieve this by combining energy- and evolution-based methods to predict mutations that enhance stability. As a result, it becomes possible to create multiple-point mutants that exhibit improved thermal stability [145,146]. These platforms utilize both sequence and structural data, applying advanced algorithms to reduce antagonistic effects caused by mutations and improve stability without compromising function [145,146]. In addition, the utilization of deep learning models, such as DeepEvo, has made it possible to forecast thermostable variations by simulating evolutionary processes. This offers a new method for protein engineering that avoids the time-consuming old techniques. Molecular dynamics simulations have been important in comprehending the stability and dynamics of engineered proteins, providing valuable knowledge about the structural foundation of thermostability and driving the improvement of protein interfaces to promote functionality [147]. In addition, ancestral sequence reconstruction has become a promising approach that utilizes phylogenetic analysis to revive ancient proteins with naturally stable structures. This expands the range of tools that may be used to build strong proteins for commercial and medicinal purposes [148,149]. These computational advancements enhance the effectiveness of designing proteins that can withstand high temperatures and also create opportunities for their use in demanding conditions, thus progressing the area of protein engineering (Figure 5B).

6.3. Protein Functionalization

6.3.1. Computational Design of Allosteric Regulation

The latest progress in the computational design of allosteric regulation has greatly improved the capacity to manipulate proteins and create new functions. This research has specifically concentrated on optimizing allosteric sites to achieve precise control over protein activity. The utilization of computational tools, as described by Duan et al., has played a crucial role in understanding the routes of allosteric communication. These methods have allowed for the identification and creation of allosteric sites that can be specifically targeted for the purpose of discovering new drugs [150]. These approaches employ bioinformatics and machine learning to simulate the dynamic and network-based characteristics of allosteric control. They offer valuable insights into the structural alterations that enable allosteric signaling [151,152]. Recent research has utilized multiscale modeling and Markov state models to simulate allosteric transitions. This approach provides a quantitative framework for predicting how mutations or ligand binding can affect protein function [151]. The combination of computational and experimental methods has improved these models, enabling the creation of proteins with improved allosteric properties. This has been demonstrated through the manipulation of allosteric networks to enhance enzyme activity and biosensor performance [153]. As these computational tools progress, they offer the potential to enhance the range of methods for creating proteins with customized allosteric regulation. This, in turn, will contribute to the advancement of synthetic biology and therapeutic development (Figure 5C).

6.3.2. Engineering Proteins with Novel Binding Properties

The development of proteins with new binding properties has been greatly influenced by the use of computational and experimental methods to improve the specificity and strength of protein interactions. Computational tools like Rosetta have played a crucial role in the development of proteins with novel binding sites. These tools enable precise modifications to protein structure,

resulting in improved binding capacities. Recent research on de novo protein design have emphasized the significance of these advancements [3,136]. These technologies employ algorithms that forecast the most effective interactions between proteins and their targets, enabling the development of proteins with customized binding properties for particular applications, such as therapeutic targets or biosensors [154]. Furthermore, machine learning techniques have been included into protein design in order to forecast and enhance binding interactions. This is achieved by utilizing extensive datasets from the Protein Data Bank to guide design choices and enhance precision [136]. Directed evolution is an experimental technique that complements computational methods. It involves iteratively refining protein sequences to acquire specific binding qualities. This process enhances the functionalization of proteins for various biomedical purposes [154]. The integration of these computational and experimental methods not only speeds up the progress of proteins with unique binding characteristics but also broadens their potential for use in areas such as pharmaceutical development and synthetic biology [154]. As these approaches progress, they provide the potential to improve the accuracy and effectiveness of protein engineering, leading to new and creative solutions in the fields of health and biotechnology [136,155] (Figure 5D).

6.4. *Designing Multi-Functional Proteins*

6.4.1. Computational Approaches for Domain Fusion

Advancements in computational methodologies for domain fusion have greatly improved the design and creation of multi-functional proteins with new binding characteristics and capabilities. The fusion of protein domains enables the formation of chimeric proteins possessing distinctive combinations of functionalities. This process largely depends on precise predictions of both structure and function, as demonstrated in recent research utilizing AlphaFold II and other modeling techniques [156]. Computational approaches encounter difficulties in accurately anticipating the spatial orientation and interactions of fused domains, but they provide a structure for investigating new protein structures that do not exist in nature. Relational algebra is suggested as a potent technique for detecting functionally connected proteins in domain fusion analysis. This approach utilizes extensive domain databases like Pfam and InterPro to anticipate domain fusions and their potential functional associations [157]. Furthermore, the design of inter-domain linkers plays a vital role in preserving the structural integrity and functionality of fused proteins. Recent investigations have identified the ideal features of linkers that prevent undesirable interactions and improve protein stability [158]. Deep learning techniques, like those used in DeepAssembly, enhance the prediction of multi-domain protein structures by properly simulating inter-domain interactions and boosting the accuracy of domain assembly [159]. These computational breakthroughs not only make it easier to design proteins with improved functions, but also broaden the range of possible uses for modified proteins in areas like drug discovery and synthetic biology (Figure 5E).

6.4.2. Rational Design of Chimeric Proteins

Computational techniques have greatly advanced the rational design of chimeric proteins, which entails strategically fusing different protein domains to form multifunctional proteins. These methods utilize knowledge about the structure and function of proteins to direct the merging of protein domains, with the goal of improving or introducing new functions. For instance, the utilization of computational tools such as Protlego simplifies the process of designing and analyzing chimeric proteins by automating the selection and combining of protein fragments. This is done by considering evolutionary conservation and structural compatibility [160]. This strategy has been confirmed by effective applications in producing proteins with enhanced stability and catalytic capabilities, as shown in studies that focus on chimeric enzymes combining domains to boost biocatalytic efficiency [161]. In addition, the combination of machine learning and structural databases, including the Protein Data Bank, enables precise forecasting of domain interfaces and the enhancement of linker regions. These regions are essential for preserving the structural integrity and functionality of the chimeras [156]. These developments not only simplify the design process but also

broaden the possible uses of chimeric proteins in the creation of therapies, synthetic biology, and industrial biotechnology. With the ongoing advancement of computational tools, there is a potential for significant improvement in the accuracy and effectiveness of chimeric protein design. This progress opens up opportunities for groundbreaking solutions in diverse scientific disciplines (Figure 5E).

7. Case Studies and Applications in Biotechnology and Pharmaceuticals

7.1. Engineered Antibodies and Immunotherapeutics

7.1.1. Computational Design of Antibody-Antigen Interfaces

The use of advanced algorithms in computational design has greatly improved the production of modified antibodies and immuno-therapeutics by enhancing the prediction and optimization of binding interactions in antibody-antigen interfaces. The utilization of computational approaches, as exemplified by Norman et al., involves the use of structural modeling to discover crucial residues in antibody-antigen interactions. This process aids in the development of antibodies with enhanced specificity and affinity [162]. Machine learning techniques, such as Parapred, which is a deep learning algorithm, have been used to forecast paratope areas. This has resulted in enhanced precision in antibody design by specifically targeting important binding sites [70]. By combining computational methodologies with high-throughput sequencing data, it has been possible to create more potent therapeutic antibodies. This approach allows for the quick evaluation and enhancement of potential anti-body candidates [163]. Moreover, the application of geometric deep learning has enhanced the ability to forecast protein interaction surfaces, offering valuable knowledge about the structural factors that influence antibody-antigen binding and assisting in the development of innovative antibody forms [70]. The computational breakthroughs not only simplify the process of designing antibodies, but also broaden their potential for use in treating many diseases. This is evident from the growing number of computationally produced antibodies that are being tested in clinical studies [164]. As the field progresses, these methods hold the potential to improve the accuracy and effectiveness of antibody-based treatments, aiding in the advancement of advanced immunotherapies (Figure 6A).

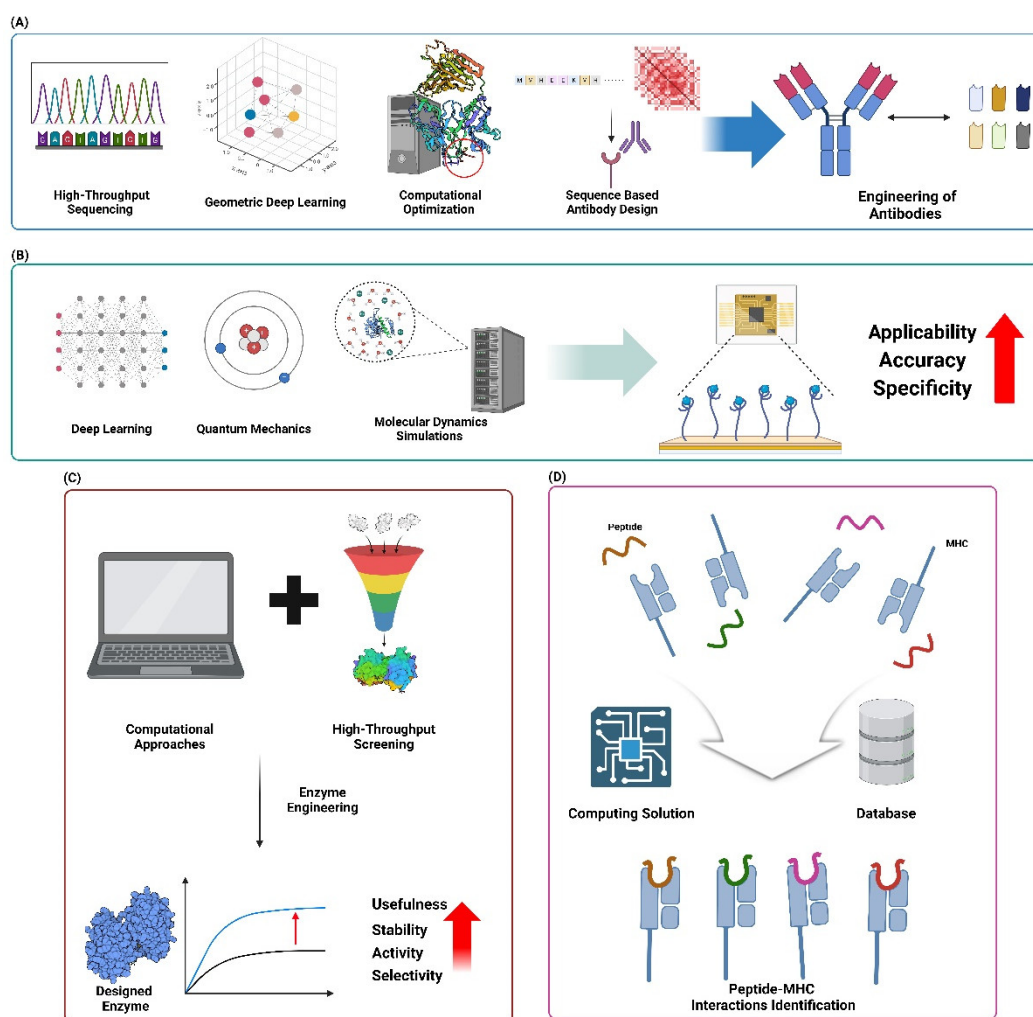


Figure 6. Protein engineering applications using computational approaches in biotechnology and pharmaceuticals. (A) High-throughput sequencing data and geometric deep learning can enhance antibody binding prediction capabilities. Computational technologies such as deep learning enable sequence-based antibody design, providing advanced approaches to antibody engineering. (B) Computational and structural methods, such as deep learning and quantum mechanical molecular dynamics simulations, have enabled the prediction of atomic-level movements of biomolecules, leading to improvements in the applicability, accuracy, and specificity of protein-based biosensors. (C) Advancements in computational technologies such as machine learning, combined with high-throughput screening, have enabled improved enzyme engineering with enhanced catalytic properties, leading to increased stability, activity, and selectivity of enzymes. (D) Computational technologies play a crucial role in therapeutic protein design, particularly in predicting peptide-MHC binding affinity. These methods not only advance personalized medicine but also accelerate the clinical application of protein therapeutics.

7.1.2. In Silico Optimization of Antibody Stability and Specificity

The latest progress in the computational optimization of antibody stability and specificity has greatly improved the creation of engineered antibodies and immunotherapeutics. This is achieved by using computational approaches to simplify and increase the process of designing antibodies. The computational approach, as outlined by Norman et al., employ structural modeling to forecast and improve the stability and specificity of antibodies. The main focus is on optimizing specific residues at the interface between the antibody and antigen to enhance binding strength and decrease the likelihood of immune response [162]. Deep learning algorithms, such as DeepAb, have been utilized to directly forecast the structures of antibody Fv based on their sequences. This allows for the creation

of improved variants with higher thermostability and affinity, eliminating the requirement for considerable experimental data [165]. These models combine high-throughput sequencing data and machine learning to quickly evaluate and improve antibody candidates, resulting in a significant reduction in the time and cost required by traditional experimental methods [163]. In addition, the incorporation of artificial intelligence in the process of creating antibodies has made it possible to anticipate the specificity of antigens based on antibody sequences. This has enabled the production of synthetic antibodies that have enhanced binding properties [163]. As these computational techniques advance, they provide the potential to improve the accuracy and effectiveness of antibody optimization. This progress will facilitate the creation of next-generation immunotherapeutics with enhanced therapeutic characteristics (Figure 6A).

7.2. Biosensors and Diagnostics

7.2.1. Rational Design of Protein-Based Biosensors

The latest progress in the logical development of protein-based biosensors has greatly improved their use in biotechnology and diagnostics. This has been achieved by utilizing computational and structural methods to boost the binding specificity and sensitivity. Computational techniques, as described by Kaczmarek et al., employ knowledge about the structure and evolution of biosensors to design sensors that have enhanced ability to bind to specific molecules and exhibit improved fluorescence properties. This allows for accurate identification of small molecules in complicated biological settings [166]. The study published in *Nature* showcases the potential of de novo designed protein switches in the development of modular and tunable biosensor platforms. These protein switches can sense a wide range of targets by linking conformational changes to sensitive outputs, thereby enhancing the versatility of biosensor applications [167]. Moreover, the incorporation of synthetic biology methods has enabled the development of genetically engineered biosensors that can actively control metabolic pathways, providing the ability to monitor and manipulate cellular processes in real-time. This has been demonstrated in research involving biosensors based on transcription factors [168]. These improvements enhance the functionality and adaptability of protein-based biosensors, making them suitable for various applications like environmental monitoring, healthcare diagnostics, and industrial biotechnology. The advancement of computational tools and synthetic biology is anticipated to boost the precision and efficiency of protein-based biosensors, facilitating the development of creative solutions for intricate analytical problems.

7.2.2. Computational Approaches for Enhancing Sensor Sensitivity and Specificity

Advancements in computational techniques have greatly enhanced the sensitivity and specificity of biosensors, leading to their increased use in biotechnology and pharmaceutical industries. The enhancements are primarily propelled by the incorporation of sophisticated algorithms and simulations that enhance the efficiency of sensor functionality. The use of molecular dynamics simulations and quantum mechanics computations has played a crucial role in accurately predicting the behavior of biomolecules at the atomic level. This enables the precise adjustment of biosensor components to achieve certain performance characteristics [169,170]. Computational fluid dynamics has been used to improve the advancement of microfluidic devices, which are important for enhancing the sensitivity and specificity of biosensors by regulating fluid dynamics and analyte transport. In addition, researchers have used hybrid computational methods that combine molecular docking and virtual screening to discover new sensing components that have both high specificity and affinity. This has enabled the creation of biosensors that can detect low levels of target substances in complex biological samples [171]. Machine learning and artificial intelligence have improved biosensor design, providing new opportunities to enhance the predictive capability and precision of computational models, hence facilitating the creation of more advanced biosensing devices [170]. As these computational tools progress, they hold the potential to enhance the field of biosensors, making

them more efficient for use in healthcare diagnostics, environmental monitoring, and food safety (Figure 6B).

7.3. Industrial Enzymes

7.3.1. Computational Engineering of Enzymes for Biocatalysis

Computational engineering of industrial enzymes for biocatalysis is an advanced field in biotechnology and pharmaceuticals that aims to improve enzyme functioning for industrial use. Improvements in machine learning have had a substantial impact on enzyme engineering. These improvements provide tools to predict interactions between enzymes and substrates, which is essential for designing enzymes with improved catalytic characteristics [172]. By combining computational approaches with high-throughput screening, researchers may effectively explore large enzyme design spaces. This enables the synthesis of stable and selective biocatalysts that are essential for cost-effective bio-based processes [79]. In addition, the combination of molecular dynamics simulations and ML models allows for a detailed understanding of enzyme processes at the atomic level. This enables precise adjustments that improve enzyme stability and activity in industrial settings. The combination of computational and experimental methods has resulted in the successful modification of enzymes to perform new tasks, increasing their usefulness in drug production and environmental cleanup [173]. These advancements highlight the significant impact of using computational enzyme engineering to develop environmentally-friendly and effective biocatalytic processes. This, in turn, enhances the capacities of biotechnology and pharmaceutical industries (Figure 6C).

7.3.2. Design of enzymes for Biodegradation and Environmental Applications

Enzyme design for biodegradation and environmental applications is a rapidly growing area in biotechnology, propelled by breakthroughs in protein engineering and computational techniques. Recent study emphasizes the utilization of directed evolution and rational design to augment the enzymatic capacity to break down persistent pollutants, including plastics and other synthetic substances, so aiding in environmental preservation [174]. Enzymes that have been specifically designed have been enhanced to break down polyethylene terephthalate (PET), a commonly used plastic. This has been achieved by improving their ability to speed up chemical reactions and their ability to remain stable over time. This demonstrates the promise of using biological catalysts in recycling and managing garbage [175]. In addition, the combination of computational modeling and experimental methods has made it possible to create enzymes that can work under harsh environmental circumstances, thereby expanding their usefulness in various industrial processes [176]. These advancements highlight the significant impact of enzyme engineering in tackling environmental issues, providing sustainable methods for managing pollutants and recovering resources (Figure 6C).

7.4. Therapeutic Protein Design

7.4.1. Computational Approaches for Improving Protein Drug Properties

The field of therapeutic protein design has experienced notable progress, especially with the incorporation of computational methods that improve the feasibility of developing protein-based therapeutics. Computational methods, such as molecular dynamics and artificial intelligence, play a crucial role in tackling important aspects of therapeutic proteins, such as affinity, selectivity, stability, and solubility. These factors are essential for the successful application of these proteins in clinical settings [177]. These techniques allow for the anticipation and enhancement of protein structures, making it easier to create proteins with enhanced therapeutic characteristics. For example, deep learning algorithms have been used to forecast protein interactions and improve sequences to decrease immunogenicity and increase stability. These computational solutions not only make the medication development process more efficient but also save expenses by reducing the necessity for large experimental trials [6]. The collaboration between computational scientists and pharmaceutical

developers is essential for closing the divide between theoretical models and real applications, guaranteeing the appropriate utilization of computational tools in drug discovery [177]. As these technologies continue to advance, they hold the potential to greatly transform the process of designing therapeutic proteins. They offer more accurate and effective methods for building new protein-based therapeutics [139] (Figure 6D).

7.4.2. In Silico Prediction of Immunogenicity and Optimization of Protein Therapeutics

The topic of in silico prediction of immunogenicity and optimization of protein therapeutics is fast advancing, utilizing computational technologies to improve the safety and effectiveness of biologic medications. These methods are crucial for detecting possible immune-stimulating regions in protein-based treatments, enabling their alteration or removal prior to use in clinical settings. Machine learning algorithms have been recently combined with classical bioinformatics methods to identify T-cell epitopes. This is done by analyzing peptide-MHC binding affinities, which is important for evaluating immunogenic potential [178,179]. The utilization of extensive databases such as the Immune Epitope Database (IEDB) has enabled the refinement of these algorithms, enhancing their precision and suitability across various HLA haplotypes [178]. In addition, computational techniques are used to enhance protein sequences by minimizing their immunogenicity while yet ensuring their therapeutic effectiveness. This approach tackles obstacles such as MHC polymorphism and the intricate nature of peptide-MHC interactions [178,179]. In silico methodologies not only optimize the drug development process by minimizing the requirement for extensive in vitro and in vivo testing, but also facilitate the tailoring of protein treatments to unique patient profiles, hence boosting personalized medicine [179]. As these technologies progress, they have the potential to greatly decrease the failure rates of protein therapies due to immunogenicity, therefore speeding up their journey towards clinical application [179] (Figure 6D).

8. Challenges and Future Perspectives

8.1. Integration of Multi-Scale Modeling Approaches

The incorporation of multi-scale modeling methods in computational protein engineering poses obstacles and offers future prospects for enhancing molecular design. Multi-scale modeling is crucial for understanding the intricate dynamics of protein systems at several levels, ranging from electronic to macroscopic, by integrating atomistic, coarse-grained, and continuum models. This methodology overcomes the constraints of conventional methods that face difficulties in dealing with the extensive range of protein conformations and the lengthy simulation times needed for in-depth protein investigations [180]. Machine learning has recently made significant progress in enhancing multi-scale modeling. This progress has resulted in improved prediction accuracy and the ability to efficiently explore protein design spaces [181]. These computational tools aid in the discovery of protein structures and interactions, which are essential for the development of proteins with new activities and enhanced stability. Nevertheless, there are still obstacles to overcome when it comes to merging data from various scales and guaranteeing that models precisely depict biological phenomena. Future prospects involve the creation of hybrid models that effortlessly combine different scales, aided by advancements in processing power and algorithms [6]. As these models advance in complexity, they have the capacity to transform protein engineering by offering comprehensive understanding of protein activity, thereby expediting the creation of new medicines and biomaterials (Figure 7B).

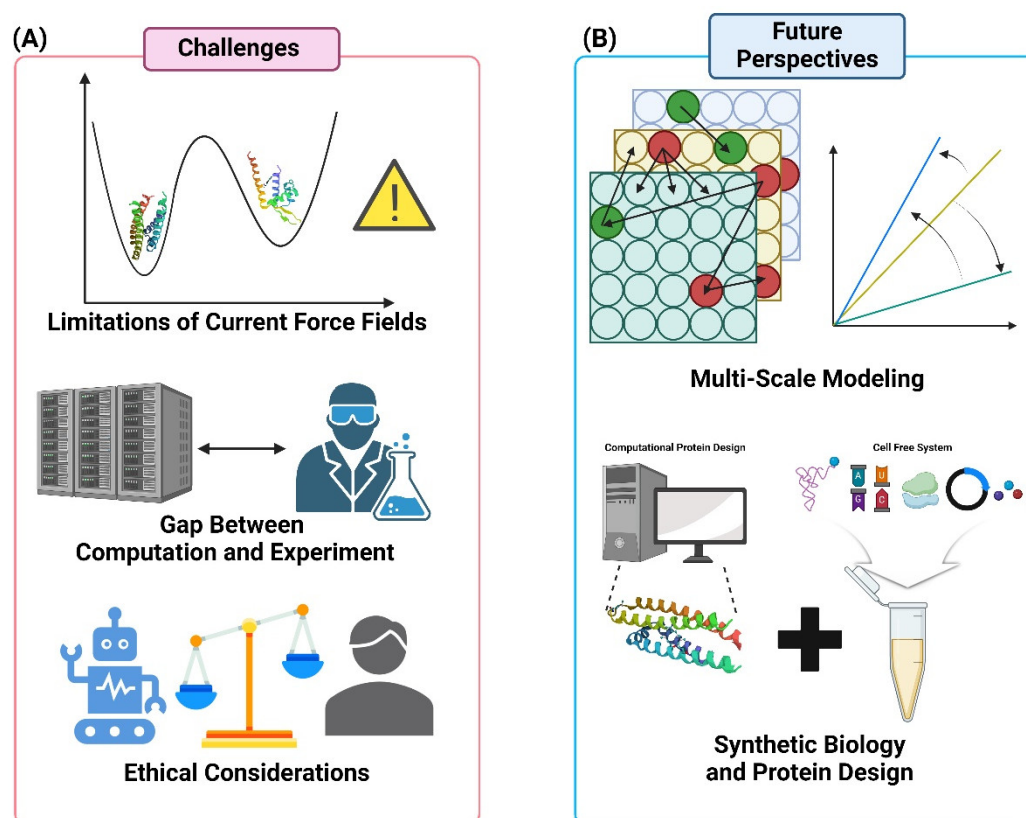


Figure 7. Challenges and future perspectives in computational approaches to protein engineering applications. (A) Current force fields have limitations in accurately capturing changes in electrostatic interactions, which impacts the accuracy and reliability of simulations. Integrating computational tools with experimental validation is essential for enhancing the accuracy and efficiency of protein design. Ethical issues related to bias, transparency, and accountability arise in the application of AI in protein engineering. (B) The integration of multi-scale modeling approaches is essential for understanding the complex dynamics of protein systems and developing proteins with new functions, and the advancement of these models holds great potential in the field of computational protein design. The combination of computational protein design and synthetic biology enables the development of innovative proteins.

8.2. Addressing the Limitations of Current Force Fields

Overcoming the constraints of existing force fields in computational protein engineering and molecular design is a crucial task that greatly affects the precision and dependability of molecular simulations. Conventional force fields commonly utilize stationary charges located at the atoms, which may not accurately capture the changing behavior of electrostatic interactions. As a result, this can lead to mistakes when simulating protein folding and interactions [182,183]. Polarizable force fields, such as the Drude and AMOEBA models, have been developed to incorporate electronic polarization effects. These improvements aim to enhance the accuracy of representing molecular interactions and energy landscapes [182,184]. Nevertheless, these models require significant computational resources and can be very responsive to initial conditions, which presents obstacles to their extensive implementation [183,184]. Integrating both polarizable and non-polarizable elements in hybrid models is a potential strategy to achieve a compromise between accuracy and computational efficiency [182,183]. Furthermore, the application of machine learning and automated fitting techniques has demonstrated promise in improving force field parameters by utilizing extensive datasets of experimental and simulation data [182]. The increasing computer capacity allows for the integration of advanced force fields with multi-scale modeling techniques. This

integration is expected to improve the accuracy of simulations, making it easier to design proteins with new functionalities and better stability [177] (Figure 7A).

8.3. Bridging the Gap Between Computation and Experiment

The integration of modern computational tools with empirical validation is crucial for bridging the gap between computational and experimental approaches in protein engineering and molecular design. This integration aims to enhance the design and functionality of proteins. Advancements in computational technologies, including machine learning and artificial intelligence, have greatly enhanced the accuracy of predicting protein structures and identifying functional areas. This has made it easier to tailor protein functionalities with more precision [3,185]. Nevertheless, due to the intricate nature of biological systems and the constraints of computer models, it is essential to conduct experimental verification in order to guarantee the dependability of these forecasts [186]. The emergence of platforms such as Mutexa showcases the endeavor to establish intelligent protein engineering ecosystems that integrate high-throughput computation with bioinformatics and quantum chemistry. This integration aims to simplify the process of identifying potential protein variants that show promise [4]. Furthermore, the combination of computational and experimental methods might expedite the design process by enabling the development of targeted libraries for laboratory evolution, thus minimizing the extensive sequence space that requires sampling [187]. With the increasing computer power and advancement of algorithms, the combination of computation and experimentation has the potential to greatly impact protein engineering. This could result in the creation of new proteins that have improved stability, activity, and therapeutic properties [139] (Figure 7A).

8.4. Ethical Considerations in AI-Driven Protein Engineering

The incorporation of artificial intelligence (AI) into protein engineering and molecular design gives rise to noteworthy ethical concerns that warrant attention in order to guarantee responsible and advantageous progress in the domain. The utilization of AI in protein engineering has significant promise for the creation of innovative medicines and biomaterials. However, it also presents concerns of bias, transparency, and accountability. The main ethical concerns with AI systems are centered around their ability to perpetuate pre-existing biases present in the training data, resulting in unfair outcomes in healthcare applications [188,189]. Furthermore, the capacity to provide clear explanations for AI models is essential in order to uphold trust and guarantee that AI-driven decisions in protein design are visible and comprehensible to stakeholders [190]. Researchers and developers are encouraged to actively participate in ethical frameworks and principles that prioritize fairness, the prevention of harm, and the respect for human autonomy in the implementation of AI applications [189,191]. Additionally, it is imperative for scientists, ethicists, and legislators to work together in order to establish strong governance systems that effectively tackle ethical dilemmas and encourage the conscientious application of AI in protein engineering. In order to maintain a balance between innovation and societal values and to prevent the misuse of AI technology, it is crucial for the field to engage in ongoing debate and adjust ethical standards as it evolves [192] (Figure 7A).

8.5. Emerging Opportunities in Synthetic Biology and Protein Design

The integration of modern computational tools is driving emerging opportunities in synthetic biology and protein design, which have transformational potential in the fields of biotechnology and molecular design. Synthetic biology, a field that focuses on creating new biological components and systems, is using machine learning more and more to improve protein engineering. This allows for the development of proteins with new functions and better performance in industrial and medical applications [193]. Cell-free protein synthesis (CFPS) is a promising technique that enables the quick prototyping and manufacturing of proteins without the limitations of living cells. This method facilitates the investigation of novel protein designs and functionalities [194]. Moreover, the merging of synthetic biology and metagenomics is creating opportunities to construct intricate biological

systems, hence improving our capacity to control and exploit microbial populations for biotechnological purposes [195]. However, there are still difficulties in expanding the use of these technologies and making sure that they are available to a wider group of academics. This is crucial in order to fully utilize their potential in addressing global issues like sustainable development and healthcare. [185,196]. The advancement of computational tools and their integration with experimental methodologies is paving the way for groundbreaking innovation and application of protein design in synthetic biology across several domains (Figure 7B).

9. Conclusion

The domain of computational protein engineering and molecular design is swiftly progressing, propelled by improvements in machine learning, molecular modeling techniques, and high-performance computing. This study has emphasized the wide range of applications and creative methods in this rapidly evolving subject, including AI-powered protein design, molecular dynamics research, and computational drug discovery. In the future, it will be essential to combine these computational methods with experimental validation in order to fully realize their promise. The ongoing advancement of increasingly precise and effective algorithms, together with the growing accessibility of biological data, holds the potential to expedite the identification and creation of new proteins and molecules with improved capabilities. The research showcased in this Special Issue of *Molecules* highlights the significant influence of computational methods on protein engineering and molecular design. As these methodologies progress and develop further, they will surely have a growing impact on our comprehension of biological systems and the creation of inventive solutions to urgent difficulties in biotechnology, medicine, and other fields.

Author Contributions: Conceptualization, investigation, writing, and original draft preparation, A.S. (Ahrum Son); H.K. (Hyunsoo Kim) – Visualization, and proofreading, J.P. (Jongham Park); W.K. (Woojin Kim); Y.Y. (Yoonki Yoon); S.L. (Sangwoon Lee); Y.P. (Yongho Park) – Supervision, Project Administration, Funding Acquisition, Review and Editing, H.K. (Hyunsoo Kim). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00209456). This work was supported by the Korea Basic Science Institute (National research Facilities and Equipment Center) grant funded by the Korea government (MSIT) (No. RS-2024-00402298). This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2022-00155857, Artificial Intelligence Convergence Innovation Human Resources Development (Chungnam National University)).

Acknowledgments: This manuscript was proofread and edited with the assistance of ChatGPT-4, a language model by OpenAI. Figures were created using BioRender (BioRender.com).

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

References

1. Sequeiros-Borja, C.E.; Surpeta, B.; Brezovsky, J. Recent advances in user-friendly computational tools to engineer protein function. *Brief Bioinform* **2021**, *22*, doi:10.1093/bib/bbaa150.
2. Xu, B.; Chen, Y.; Xue, W. Computational Protein Design - Where it goes? *Curr Med Chem* **2024**, *31*, 2841-2854, doi:10.2174/0929867330666230602143700.
3. Derat, E.; Kamerlin, S.C.L. Computational Advances in Protein Engineering and Enzyme Design. *J Phys Chem B* **2022**, *126*, 2449-2451, doi:10.1021/acs.jpcc.2c01198.
4. Yang, Z.J.; Shao, Q.; Jiang, Y.; Jurich, C.; Ran, X.; Juarez, R.J.; Yan, B.; Stull, S.L.; Gollu, A.; Ding, N. Mutexa: A Computational Ecosystem for Intelligent Protein Engineering. *J Chem Theory Comput* **2023**, *19*, 7459-7477, doi:10.1021/acs.jctc.3c00602.
5. Wang, S.; Ben-Sasson, A.J. Precision materials: Computational design methods of accurate protein materials. *Curr Opin Struct Biol* **2022**, *74*, 102367, doi:10.1016/j.sbi.2022.102367.
6. Sadybekov, A.V.; Katritch, V. Computational approaches streamlining drug discovery. *Nature* **2023**, *616*, 673-685, doi:10.1038/s41586-023-05905-z.

7. Gligorijevic, V.; Renfrew, P.D.; Kosciolk, T.; Leman, J.K.; Berenberg, D.; Vatanen, T.; Chandler, C.; Taylor, B.C.; Fisk, I.M.; Vlamakis, H.; et al. Structure-based protein function prediction using graph convolutional networks. *Nat Commun* **2021**, *12*, 3168, doi:10.1038/s41467-021-23303-9.
8. Gao, W.; Mahajan, S.P.; Sulam, J.; Gray, J.J. Deep Learning in Protein Structural Modeling and Design. *Patterns (N Y)* **2020**, *1*, 100142, doi:10.1016/j.patter.2020.100142.
9. Saman Booy, M.; Ilin, A.; Orponen, P. RNA secondary structure prediction with convolutional neural networks. *BMC Bioinformatics* **2022**, *23*, 58, doi:10.1186/s12859-021-04540-7.
10. Gao, M.; Zhang, D.; Chen, Y.; Zhang, Y.; Wang, Z.; Wang, X.; Li, S.; Guo, Y.; Webb, G.I.; Nguyen, A.T.N.; et al. GraphormerDTI: A graph transformer-based approach for drug-target interaction prediction. *Comput Biol Med* **2024**, *173*, 108339, doi:10.1016/j.combiomed.2024.108339.
11. Kulikova, A.V.; Diaz, D.J.; Loy, J.M.; Ellington, A.D.; Wilke, C.O. Learning the local landscape of protein structures with convolutional neural networks. *J Biol Phys* **2021**, *47*, 435-454, doi:10.1007/s10867-021-09593-6.
12. Kumar, N.; Srivastava, R. Deep learning in structural bioinformatics: current applications and future perspectives. *Brief Bioinform* **2024**, *25*, doi:10.1093/bib/bbae042.
13. Asabuki, T.; Kokate, P.; Fukai, T. Neural circuit mechanisms of hierarchical sequence learning tested on large-scale recording data. *PLoS Comput Biol* **2022**, *18*, e1010214, doi:10.1371/journal.pcbi.1010214.
14. Jain, R.; Jain, A.; Mauro, E.; LeShane, K.; Densmore, D. ICOR: improving codon optimization with recurrent neural networks. *BMC Bioinformatics* **2023**, *24*, 132, doi:10.1186/s12859-023-05246-8.
15. Tang, F.; Wang, Z.; Sun, Y.; Fan, L.; Yang, Y.; Guo, X.; Wang, Y.; Yan, S.; Qiao, Z.; Li, Y.; et al. Recurrent neural network for predicting absence of heterozygosity from low pass WGS with ultra-low depth. *BMC Genomics* **2024**, *25*, 470, doi:10.1186/s12864-024-10400-4.
16. Das, S.; Tariq, A.; Santos, T.; Kantareddy, S.S.; Banerjee, I. Recurrent Neural Networks (RNNs): Architectures, Training Tricks, and Introduction to Influential Research. In *Machine Learning for Brain Disorders*, Colliot, O., Ed.; New York, NY, 2023; pp. 117-138.
17. Lin, E.; Lin, C.H.; Lane, H.Y. De Novo Peptide and Protein Design Using Generative Adversarial Networks: An Update. *J Chem Inf Model* **2022**, *62*, 761-774, doi:10.1021/acs.jcim.1c01361.
18. Kucera, T.; Togninalli, M.; Meng-Papaxanthos, L. Conditional generative modeling for de novo protein design with hierarchical functions. *Bioinformatics* **2022**, *38*, 3454-3461, doi:10.1093/bioinformatics/btac353.
19. Strokach, A.; Kim, P.M. Deep generative modeling for protein design. *Curr Opin Struct Biol* **2022**, *72*, 226-236, doi:10.1016/j.sbi.2021.11.008.
20. Atz, K.; Cotos, L.; Isert, C.; Hakansson, M.; Focht, D.; Hilleke, M.; Nippa, D.F.; Iff, M.; Ledergerber, J.; Schiebroek, C.C.G.; et al. Prospective de novo drug design with deep interactome learning. *Nat Commun* **2024**, *15*, 3408, doi:10.1038/s41467-024-47613-w.
21. Kim, H.; Choi, H.; Kang, D.; Lee, W.B.; Na, J. Materials discovery with extreme properties via reinforcement learning-guided combinatorial chemistry. *Chem Sci* **2024**, *15*, 7908-7925, doi:10.1039/d3sc05281h.
22. Zhou, J.; Huang, M. Navigating the landscape of enzyme design: from molecular simulations to machine learning. *Chem Soc Rev* **2024**, doi:10.1039/d4cs00196f.
23. Zhang, Y.; Li, S.; Xing, M.; Yuan, Q.; He, H.; Sun, S. Universal Approach to De Novo Drug Design for Target Proteins Using Deep Reinforcement Learning. *ACS Omega* **2023**, *8*, 5464-5474, doi:10.1021/acsomega.2c06653.
24. Palukuri, M.V.; Patil, R.S.; Marcotte, E.M. Molecular complex detection in protein interaction networks through reinforcement learning. *BMC Bioinformatics* **2023**, *24*, 306, doi:10.1186/s12859-023-05425-7.
25. Dietrich, L.; Rathmer, B.; Ewan, K.; Bange, T.; Heinrichs, S.; Dale, T.C.; Schade, D.; Grossmann, T.N. Cell Permeable Stapled Peptide Inhibitor of Wnt Signaling that Targets beta-Catenin Protein-Protein Interactions. *Cell Chem Biol* **2017**, *24*, 958-968 e955, doi:10.1016/j.chembiol.2017.06.013.
26. Wang, G.; Liu, X.; Wang, K.; Gao, Y.; Li, G.; Baptista-Hon, D.T.; Yang, X.H.; Xue, K.; Tai, W.H.; Jiang, Z.; et al. Deep-learning-enabled protein-protein interaction analysis for prediction of SARS-CoV-2 infectivity and variant evolution. *Nat Med* **2023**, *29*, 2007-2018, doi:10.1038/s41591-023-02483-5.
27. Zhou, Z.; Zhang, L.; Yu, Y.; Wu, B.; Li, M.; Hong, L.; Tan, P. Enhancing efficiency of protein language models with minimal wet-lab data through few-shot learning. *Nat Commun* **2024**, *15*, 5566, doi:10.1038/s41467-024-49798-6.
28. Chowdhury, R.; Bouatta, N.; Biswas, S.; Floristean, C.; Kharkar, A.; Roy, K.; Rochereau, C.; Ahdritz, G.; Zhang, J.; Church, G.M.; et al. Single-sequence protein structure prediction using a language model and deep learning. *Nat Biotechnol* **2022**, *40*, 1617-1623, doi:10.1038/s41587-022-01432-w.
29. Khakzad, H.; Igashov, I.; Schneuing, A.; Goverde, C.; Bronstein, M.; Correia, B. A new age in protein design empowered by deep learning. *Cell Syst* **2023**, *14*, 925-939, doi:10.1016/j.cels.2023.10.006.
30. 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, doi:10.1038/s41580-024-00718-y.
31. Kafri, M.; Metzl-Raz, E.; Jona, G.; Barkai, N. The Cost of Protein Production. *Cell Rep* **2016**, *14*, 22-31, doi:10.1016/j.celrep.2015.12.015.

32. Ao, Y.F.; Dorr, M.; Menke, M.J.; Born, S.; Heuson, E.; Bornscheuer, U.T. Data-Driven Protein Engineering for Improving Catalytic Activity and Selectivity. *Chembiochem* **2024**, *25*, e202300754, doi:10.1002/cbic.202300754.
33. 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.
34. Illig, A.M.; Siedhoff, N.E.; Davari, M.D.; Schwaneberg, U. Evolutionary Probability and Stacked Regressions Enable Data-Driven Protein Engineering with Minimized Experimental Effort. *J Chem Inf Model* **2024**, doi:10.1021/acs.jcim.4c00704.
35. Medl, M.; Leisch, F.; Durauer, A.; Scharl, T. Explainable deep learning enhances robust and reliable real-time monitoring of a chromatographic protein A capture step. *Biotechnol J* **2024**, *19*, e2300554, doi:10.1002/biot.202300554.
36. Lee, M. Recent Advances in Deep Learning for Protein-Protein Interaction Analysis: A Comprehensive Review. *Molecules* **2023**, *28*, doi:10.3390/molecules28135169.
37. Kim, M.; Kang, D.; Kim, M.S.; Choe, J.C.; Lee, S.H.; Ahn, J.H.; Oh, J.H.; Choi, J.H.; Lee, H.C.; Cha, K.S.; et al. Acute myocardial infarction prognosis prediction with reliable and interpretable artificial intelligence system. *J Am Med Inform Assoc* **2024**, *31*, 1540-1550, doi:10.1093/jamia/ocae114.
38. Malinverno, L.; Barros, V.; Ghisoni, F.; Visona, G.; Kern, R.; Nickel, P.J.; Ventura, B.E.; Simic, I.; Stryeck, S.; Manni, F.; et al. A historical perspective of biomedical explainable AI research. *Patterns (N Y)* **2023**, *4*, 100830, doi:10.1016/j.patter.2023.100830.
39. Dash, T.; Chitlangia, S.; Ahuja, A.; Srinivasan, A. A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Sci Rep* **2022**, *12*, 1040, doi:10.1038/s41598-021-04590-0.
40. Sirocchi, C.; Bogliolo, A.; Montagna, S. Medical-informed machine learning: integrating prior knowledge into medical decision systems. *BMC Med Inform Decis Mak* **2024**, *24*, 186, doi:10.1186/s12911-024-02582-4.
41. Laxmi, B.; Devi, P.U.M.; Thanjavur, N.; Buddolla, V. The Applications of Artificial Intelligence (AI)-Driven Tools in Virus-Like Particles (VLPs) Research. *Curr Microbiol* **2024**, *81*, 234, doi:10.1007/s00284-024-03750-5.
42. Khlaif, Z.N.; Mousa, A.; Hattab, M.K.; Itmazi, J.; Hassan, A.A.; Sanmugam, M.; Ayyoub, A. The Potential and Concerns of Using AI in Scientific Research: ChatGPT Performance Evaluation. *JMIR Med Educ* **2023**, *9*, e47049, doi:10.2196/47049.
43. Musa, N.; Gital, A.Y.; Aljojo, N.; Chiroma, H.; Adewole, K.S.; Mojeed, H.A.; Faruk, N.; Abdulkarim, A.; Emmanuel, I.; Folawiyo, Y.Y.; et al. A systematic review and Meta-data analysis on the applications of Deep Learning in Electrocardiogram. *J Ambient Intell Humaniz Comput* **2023**, *14*, 9677-9750, doi:10.1007/s12652-022-03868-z.
44. Dikmen, M.; Burns, C. The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human-Computer Studies* **2022**, *162*, doi:10.1016/j.ijhcs.2022.102792.
45. Wodak, S.J.; Vajda, S.; Lensink, M.F.; Kozakov, D.; Bates, P.A. Critical Assessment of Methods for Predicting the 3D Structure of Proteins and Protein Complexes. *Annu Rev Biophys* **2023**, *52*, 183-206, doi:10.1146/annurev-biophys-102622-084607.
46. Wuyun, Q.; Chen, Y.; Shen, Y.; Cao, Y.; Hu, G.; Cui, W.; Gao, J.; Zheng, W. Recent Progress of Protein Tertiary Structure Prediction. *Molecules* **2024**, *29*, doi:10.3390/molecules29040832.
47. Bertoline, L.M.F.; Lima, A.N.; Krieger, J.E.; Teixeira, S.K. Before and after AlphaFold2: An overview of protein structure prediction. *Front Bioinform* **2023**, *3*, 1120370, doi:10.3389/fbinf.2023.1120370.
48. Bhattacharya, S.; Roche, R.; Shuvo, M.H.; Bhattacharya, D. Recent Advances in Protein Homology Detection Propelled by Inter-Residue Interaction Map Threading. *Front Mol Biosci* **2021**, *8*, 643752, doi:10.3389/fmolb.2021.643752.
49. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493-500, doi:10.1038/s41586-024-07487-w.
50. Raghavan, B.; Paulikat, M.; Ahmad, K.; Callea, L.; Rizzi, A.; Ippoliti, E.; Mandelli, D.; Bonati, L.; De Vivo, M.; Carloni, P. Drug Design in the Exascale Era: A Perspective from Massively Parallel QM/MM Simulations. *J Chem Inf Model* **2023**, *63*, 3647-3658, doi:10.1021/acs.jcim.3c00557.
51. Rossetti, G.; Mandelli, D. How exascale computing can shape drug design: A perspective from multiscale QM/MM molecular dynamics simulations and machine learning-aided enhanced sampling algorithms. *Curr Opin Struct Biol* **2024**, *86*, 102814, doi:10.1016/j.sbi.2024.102814.
52. Ginex, T.; Vazquez, J.; Estarellas, C.; Luque, F.J. Quantum mechanical-based strategies in drug discovery: Finding the pace to new challenges in drug design. *Curr Opin Struct Biol* **2024**, *87*, 102870, doi:10.1016/j.sbi.2024.102870.
53. Kubar, T.; Elstner, M.; Cui, Q. Hybrid Quantum Mechanical/Molecular Mechanical Methods For Studying Energy Transduction in Biomolecular Machines. *Annu Rev Biophys* **2023**, *52*, 525-551, doi:10.1146/annurev-biophys-111622-091140.

54. Giese, T.J.; Zeng, J.; Lerew, L.; McCarthy, E.; Tao, Y.; Ekesan, S.; York, D.M. Software Infrastructure for Next-Generation QM/MM-DeltaMLP Force Fields. *J Phys Chem B* **2024**, *128*, 6257-6271, doi:10.1021/acs.jpcc.4c01466.
55. Shen, C.; Liu, B.; Williams, K.P.; Warnow, T. EMMA: a new method for computing multiple sequence alignments given a constraint subset alignment. *Algorithms Mol Biol* **2023**, *18*, 21, doi:10.1186/s13015-023-00247-x.
56. Ibrahim, M.K.; Yusof, U.K.; Eisa, T.A.E.; Nasser, M. Bioinspired Algorithms for Multiple Sequence Alignment: A Systematic Review and Roadmap. *Applied Sciences* **2024**, *14*, doi:10.3390/app14062433.
57. Zou, Y.; Zhang, Z.; Zeng, Y.; Hu, H.; Hao, Y.; Huang, S.; Li, B. Common Methods for Phylogenetic Tree Construction and Their Implementation in R. *Bioengineering (Basel)* **2024**, *11*, doi:10.3390/bioengineering11050480.
58. Kapli, P.; Kotari, I.; Telford, M.J.; Goldman, N.; Yang, Z. DNA Sequences Are as Useful as Protein Sequences for Inferring Deep Phylogenies. *Syst Biol* **2023**, *72*, 1119-1135, doi:10.1093/sysbio/syad036.
59. Cao, W.; Wu, L.Y.; Xia, X.Y.; Chen, X.; Wang, Z.X.; Pan, X.M. A sequence-based evolutionary distance method for Phylogenetic analysis of highly divergent proteins. *Sci Rep* **2023**, *13*, 20304, doi:10.1038/s41598-023-47496-9.
60. Chao, J.; Tang, F.; Xu, L. Developments in Algorithms for Sequence Alignment: A Review. *Biomolecules* **2022**, *12*, doi:10.3390/biom12040546.
61. Kim, D.; Noh, M.H.; Park, M.; Kim, I.; Ahn, H.; Ye, D.Y.; Jung, G.Y.; Kim, S. Enzyme activity engineering based on sequence co-evolution analysis. *Metab Eng* **2022**, *74*, 49-60, doi:10.1016/j.ymben.2022.09.001.
62. Xie, J.; Zhang, W.; Zhu, X.; Deng, M.; Lai, L. Coevolution-based prediction of key allosteric residues for protein function regulation. *Elife* **2023**, *12*, doi:10.7554/eLife.81850.
63. Hossack, E.J.; Hardy, F.J.; Green, A.P. Building Enzymes through Design and Evolution. *ACS Catalysis* **2023**, *13*, 12436-12444, doi:10.1021/acscatal.3c02746.
64. Pinto, G.P.; Corbella, M.; Demkiv, A.O.; Kamerlin, S.C.L. Exploiting enzyme evolution for computational protein design. *Trends Biochem Sci* **2022**, *47*, 375-389, doi:10.1016/j.tibs.2021.08.008.
65. Tang, Y.; Moretti, R.; Meiler, J. Recent Advances in Automated Structure-Based De Novo Drug Design. *J Chem Inf Model* **2024**, *64*, 1794-1805, doi:10.1021/acs.jcim.4c00247.
66. Isert, C.; Atz, K.; Schneider, G. Structure-based drug design with geometric deep learning. *Curr Opin Struct Biol* **2023**, *79*, 102548, doi:10.1016/j.sbi.2023.102548.
67. Banerjee, A.; Saha, S.; Tvedt, N.C.; Yang, L.W.; Bahar, I. Mutually beneficial confluence of structure-based modeling of protein dynamics and machine learning methods. *Curr Opin Struct Biol* **2023**, *78*, 102517, doi:10.1016/j.sbi.2022.102517.
68. Kinshuk, S.; Li, L.; Meckes, B.; Chan, C.T.Y. Sequence-Based Protein Design: A Review of Using Statistical Models to Characterize Coevolutionary Traits for Developing Hybrid Proteins as Genetic Sensors. *Int J Mol Sci* **2024**, *25*, doi:10.3390/ijms25158320.
69. Zhang, J.; Basu, S.; Kurgan, L. HybridDBRpred: improved sequence-based prediction of DNA-binding amino acids using annotations from structured complexes and disordered proteins. *Nucleic Acids Res* **2024**, *52*, e10, doi:10.1093/nar/gkad1131.
70. Hummer, A.M.; Abanades, B.; Deane, C.M. Advances in computational structure-based antibody design. *Curr Opin Struct Biol* **2022**, *74*, 102379, doi:10.1016/j.sbi.2022.102379.
71. Siedhoff, N.E.; Schwaneberg, U.; Davari, M.D. Machine learning-assisted enzyme engineering. *Methods Enzymol* **2020**, *643*, 281-315, doi:10.1016/bs.mie.2020.05.005.
72. Gantz, M.; Neun, S.; Medcalf, E.J.; van Vliet, L.D.; Hollfelder, F. Ultrahigh-Throughput Enzyme Engineering and Discovery in In Vitro Compartments. *Chem Rev* **2023**, *123*, 5571-5611, doi:10.1021/acs.chemrev.2c00910.
73. Ding, K.; Chin, M.; Zhao, Y.; Huang, W.; Mai, B.K.; Wang, H.; Liu, P.; Yang, Y.; Luo, Y. Machine learning-guided co-optimization of fitness and diversity facilitates combinatorial library design in enzyme engineering. *Nat Commun* **2024**, *15*, 6392, doi:10.1038/s41467-024-50698-y.
74. Atomwise, A.P. AI is a viable alternative to high throughput screening: a 318-target study. *Sci Rep* **2024**, *14*, 7526, doi:10.1038/s41598-024-54655-z.
75. Carlsson, J.; Lutten, A. Structure-based virtual screening of vast chemical space as a starting point for drug discovery. *Curr Opin Struct Biol* **2024**, *87*, 102829, doi:10.1016/j.sbi.2024.102829.
76. Goudy, O.J.; Nallathambi, A.; Kinjo, T.; Randolph, N.Z.; Kuhlman, B. In silico evolution of autoinhibitory domains for a PD-L1 antagonist using deep learning models. *Proc Natl Acad Sci U S A* **2023**, *120*, e2307371120, doi:10.1073/pnas.2307371120.
77. McLure, R.J.; Radford, S.E.; Brockwell, D.J. High-throughput directed evolution: a golden era for protein science. *Trends in Chemistry* **2022**, *4*, 378-391, doi:10.1016/j.trechm.2022.02.004.
78. Shao, Q.; Jiang, Y.; Yang, Z.J. EnzyHTP Computational Directed Evolution with Adaptive Resource Allocation. *J Chem Inf Model* **2023**, *63*, 5650-5659, doi:10.1021/acs.jcim.3c00618.

79. Orsi, E.; Schada von Borzyskowski, L.; Noack, S.; Nikel, P.I.; Lindner, S.N. Automated in vivo enzyme engineering accelerates biocatalyst optimization. *Nat Commun* **2024**, *15*, 3447, doi:10.1038/s41467-024-46574-4.
80. Scherer, M.; Fleishman, S.J.; Jones, P.R.; Dandekar, T.; Bencurova, E. Computational Enzyme Engineering Pipelines for Optimized Production of Renewable Chemicals. *Front Bioeng Biotechnol* **2021**, *9*, 673005, doi:10.3389/fbioe.2021.673005.
81. Vanella, R.; Kovacevic, G.; Doffini, V.; Fernandez de Santaella, J.; Nash, M.A. High-throughput screening, next generation sequencing and machine learning: advanced methods in enzyme engineering. *Chem Commun (Camb)* **2022**, *58*, 2455-2467, doi:10.1039/d1cc04635g.
82. Zhou, L.; Tao, C.; Shen, X.; Sun, X.; Wang, J.; Yuan, Q. Unlocking the potential of enzyme engineering via rational computational design strategies. *Biotechnol Adv* **2024**, *73*, 108376, doi:10.1016/j.biotechadv.2024.108376.
83. Bernardi, R.C.; Melo, M.C.R.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim Biophys Acta* **2015**, *1850*, 872-877, doi:10.1016/j.bbagen.2014.10.019.
84. Gong, X.; Zhang, Y.; Chen, J. Advanced Sampling Methods for Multiscale Simulation of Disordered Proteins and Dynamic Interactions. *Biomolecules* **2021**, *11*, doi:10.3390/biom11101416.
85. Qi, R.; Wei, G.; Ma, B.; Nussinov, R. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. *Methods Mol Biol* **2018**, 1777, 101-119, doi:10.1007/978-1-4939-7811-3_5.
86. Blumer, O.; Reuveni, S.; Hirshberg, B. Combining stochastic resetting with Metadynamics to speed-up molecular dynamics simulations. *Nat Commun* **2024**, *15*, 240, doi:10.1038/s41467-023-44528-w.
87. Kleiman, D.E.; Nadeem, H.; Shukla, D. Adaptive Sampling Methods for Molecular Dynamics in the Era of Machine Learning. *J Phys Chem B* **2023**, *127*, 10669-10681, doi:10.1021/acs.jpcc.3c04843.
88. Brooks, C.L., 3rd; MacKerell, A.D., Jr.; Post, C.B.; Nilsson, L. Biomolecular dynamics in the 21st century. *Biochim Biophys Acta Gen Subj* **2024**, 1868, 130534, doi:10.1016/j.bbagen.2023.130534.
89. Marrink, S.J.; Monticelli, L.; Melo, M.N.; Alessandri, R.; Tieleman, D.P.; Souza, P.C.T. Two decades of Martini: Better beads, broader scope. *WIREs Computational Molecular Science* **2022**, *13*, doi:10.1002/wcms.1620.
90. Kjolbye, L.R.; Pereira, G.P.; Bartocci, A.; Pannuzzo, M.; Albani, S.; Marchetto, A.; Jimenez-Garcia, B.; Martin, J.; Rossetti, G.; Cecchini, M.; et al. Towards design of drugs and delivery systems with the Martini coarse-grained model. *QRB Discov* **2022**, *3*, e19, doi:10.1017/qrd.2022.16.
91. Periole, X.; Marrink, S.J. The Martini coarse-grained force field. *Methods Mol Biol* **2013**, *924*, 533-565, doi:10.1007/978-1-62703-017-5_20.
92. MacCallum, J.L.; Hu, S.; Lenz, S.; Souza, P.C.T.; Corradi, V.; Tieleman, D.P. An implementation of the Martini coarse-grained force field in OpenMM. *Biophys J* **2023**, *122*, 2864-2870, doi:10.1016/j.bpj.2023.04.007.
93. Togashi, Y.; Flechsig, H. Coarse-Grained Protein Dynamics Studies Using Elastic Network Models. *Int J Mol Sci* **2018**, *19*, doi:10.3390/ijms19123899.
94. Xu, J.; Wang, Y.; Zhang, J.; Abdelmoneim, A.A.; Liang, Z.; Wang, L.; Jin, J.; Dai, Q.; Ye, F. Elastic network models and molecular dynamic simulations reveal the molecular basis of allosteric regulation in ubiquitin-specific protease 7 (USP7). *Comput Biol Med* **2023**, *162*, 107068, doi:10.1016/j.combiomed.2023.107068.
95. Leioatts, N.; Romo, T.D.; Grossfield, A. Elastic Network Models are Robust to Variations in Formalism. *J Chem Theory Comput* **2012**, *8*, 2424-2434, doi:10.1021/ct3000316.
96. Stone, J.E.; Hynninen, A.P.; Phillips, J.C.; Schulten, K. Early Experiences Porting the NAMD and VMD Molecular Simulation and Analysis Software to GPU-Accelerated OpenPOWER Platforms. *High Perform Comput (2016)* **2016**, 9945, 188-206, doi:10.1007/978-3-319-46079-6_14.
97. Ahmed, M.; Maldonado, A.M.; Durrant, J.D. From byte to bench to bedside: molecular dynamics simulations and drug discovery. *BMC Biol* **2023**, *21*, 299, doi:10.1186/s12915-023-01791-z.
98. Chiu, M.; Herbordt, M.C. Molecular Dynamics Simulations on High-Performance Reconfigurable Computing Systems. *ACM Trans Reconfigurable Technol Syst* **2010**, *3*, doi:10.1145/1862648.1862653.
99. Jones, D.; Allen, J.E.; Yang, Y.; Drew Bennett, W.F.; Gokhale, M.; Moshiri, N.; Rosing, T.S. Accelerators for Classical Molecular Dynamics Simulations of Biomolecules. *J Chem Theory Comput* **2022**, *18*, 4047-4069, doi:10.1021/acs.jctc.1c01214.
100. Hollingsworth, S.A.; Dror, R.O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129-1143, doi:10.1016/j.neuron.2018.08.011.
101. Rizzi, V.; Aureli, S.; Ansari, N.; Gervasio, F.L. OneOPES, a Combined Enhanced Sampling Method to Rule Them All. *J Chem Theory Comput* **2023**, *19*, 5731-5742, doi:10.1021/acs.jctc.3c00254.
102. Doerr, S.; Majewski, M.; Perez, A.; Kramer, A.; Clementi, C.; Noe, F.; Giorgino, T.; De Fabritiis, G. TorchMD: A Deep Learning Framework for Molecular Simulations. *J Chem Theory Comput* **2021**, *17*, 2355-2363, doi:10.1021/acs.jctc.0c01343.

103. Pelaez, R.P.; Simeon, G.; Galvelis, R.; Mirarchi, A.; Eastman, P.; Doerr, S.; Tholke, P.; Markland, T.E.; De Fabritiis, G. TorchMD-Net 2.0: Fast Neural Network Potentials for Molecular Simulations. *J Chem Theory Comput* **2024**, *20*, 4076–4087, doi:10.1021/acs.jctc.4c00253.
104. Thaler, S.; Zavadlav, J. Learning neural network potentials from experimental data via Differentiable Trajectory Reweighting. *Nat Commun* **2021**, *12*, 6884, doi:10.1038/s41467-021-27241-4.
105. Jung, G.S.; Choi, J.Y.; Lee, S.M. Active learning of neural network potentials for rare events. *Digital Discovery* **2024**, *3*, 514–527, doi:10.1039/d3dd00216k.
106. Duignan, T.T. The Potential of Neural Network Potentials. *ACS Phys Chem Au* **2024**, *4*, 232–241, doi:10.1021/acsphyschemau.4c00004.
107. Kaptan, S.; Vattulainen, I. Machine learning in the analysis of biomolecular simulations. *Advances in Physics: X* **2022**, *7*, doi:10.1080/23746149.2021.2006080.
108. Mustali, J.; Yasuda, I.; Hirano, Y.; Yasuoka, K.; Gautieri, A.; Arai, N. Unsupervised deep learning for molecular dynamics simulations: a novel analysis of protein-ligand interactions in SARS-CoV-2 M(pro). *RSC Adv* **2023**, *13*, 34249–34261, doi:10.1039/d3ra06375e.
109. Han, R.; Lubner, S. Trajectory-based machine learning method and its application to molecular dynamics. *Molecular Physics* **2020**, *118*, doi:10.1080/00268976.2020.1788189.
110. Prašnikar, E.; Ljubič, M.; Perdih, A.; Borišek, J. Machine learning heralding a new development phase in molecular dynamics simulations. *Artificial Intelligence Review* **2024**, *57*, doi:10.1007/s10462-024-10731-4.
111. Huang, S.Y. Comprehensive assessment of flexible-ligand docking algorithms: current effectiveness and challenges. *Brief Bioinform* **2018**, *19*, 982–994, doi:10.1093/bib/bbx030.
112. Yang, C.; Chen, E.A.; Zhang, Y. Protein-Ligand Docking in the Machine-Learning Era. *Molecules* **2022**, *27*, doi:10.3390/molecules27144568.
113. Huang, S.Y.; Zou, X. Advances and challenges in protein-ligand docking. *Int J Mol Sci* **2010**, *11*, 3016–3034, doi:10.3390/ijms11083016.
114. Torres, P.H.M.; Sodero, A.C.R.; Jofily, P.; Silva-Jr, F.P. Key Topics in Molecular Docking for Drug Design. *Int J Mol Sci* **2019**, *20*, doi:10.3390/ijms20184574.
115. Palacio-Rodriguez, K.; Lans, I.; Cavasotto, C.N.; Cossio, P. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Sci Rep* **2019**, *9*, 5142, doi:10.1038/s41598-019-41594-3.
116. Blanes-Mira, C.; Fernandez-Aguado, P.; de Andres-Lopez, J.; Fernandez-Carvajal, A.; Ferrer-Montiel, A.; Fernandez-Ballester, G. Comprehensive Survey of Consensus Docking for High-Throughput Virtual Screening. *Molecules* **2022**, *28*, doi:10.3390/molecules28010175.
117. Kamal, I.M.; Chakrabarti, S. MetaDOCK: A Combinatorial Molecular Docking Approach. *ACS Omega* **2023**, *8*, 5850–5860, doi:10.1021/acsomega.2c07619.
118. Xue, L.C.; Rodrigues, J.; Dobbs, D.; Honavar, V.; Bonvin, A. Template-based protein-protein docking exploiting pairwise interfacial residue restraints. *Brief Bioinform* **2017**, *18*, 458–466, doi:10.1093/bib/bbw027.
119. Meng, Q.; Guo, F.; Wang, E.; Tang, J. ComDock: A novel approach for protein-protein docking with an efficient fusing strategy. *Comput Biol Med* **2023**, *167*, 107660, doi:10.1016/j.compbiomed.2023.107660.
120. Bryant, P.; Pozzati, G.; Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* **2022**, *13*, 1265, doi:10.1038/s41467-022-28865-w.
121. Schneidman-Duhovny, D.; Rossi, A.; Avila-Sakar, A.; Kim, S.J.; Velazquez-Muriel, J.; Strop, P.; Liang, H.; Krukenberg, K.A.; Liao, M.; Kim, H.M.; et al. A method for integrative structure determination of protein-protein complexes. *Bioinformatics* **2012**, *28*, 3282–3289, doi:10.1093/bioinformatics/bts628.
122. Tsuchiya, Y.; Yamamori, Y.; Tomii, K. Protein-protein interaction prediction methods: from docking-based to AI-based approaches. *Biophys Rev* **2022**, *14*, 1341–1348, doi:10.1007/s12551-022-01032-7.
123. de Souza Neto, L.R.; Moreira-Filho, J.T.; Neves, B.J.; Maidana, R.; Guimaraes, A.C.R.; Furnham, N.; Andrade, C.H.; Silva, F.P., Jr. In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery. *Front Chem* **2020**, *8*, 93, doi:10.3389/fchem.2020.00093.
124. Shi, X.X.; Wang, Z.Z.; Wang, F.; Hao, G.F.; Yang, G.F. ACFIS 2.0: an improved web-server for fragment-based drug discovery via a dynamic screening strategy. *Nucleic Acids Res* **2023**, *51*, W25–W32, doi:10.1093/nar/gkad348.
125. Mouchlis, V.D.; Afantitis, A.; Serra, A.; Fratello, M.; Papadiamantis, A.G.; Aidinis, V.; Lynch, I.; Greco, D.; Melagraki, G. Advances in de Novo Drug Design: From Conventional to Machine Learning Methods. *Int J Mol Sci* **2021**, *22*, doi:10.3390/ijms22041676.
126. Powers, A.S.; Yu, H.H.; Suriana, P.; Koodli, R.V.; Lu, T.; Paggi, J.M.; Dror, R.O. Geometric Deep Learning for Structure-Based Ligand Design. *ACS Cent Sci* **2023**, *9*, 2257–2267, doi:10.1021/acscentsci.3c00572.
127. Mukaidaisi, M.; Vu, A.; Grantham, K.; Tchagang, A.; Li, Y. Multi-Objective Drug Design Based on Graph-Fragment Molecular Representation and Deep Evolutionary Learning. *Front Pharmacol* **2022**, *13*, 920747, doi:10.3389/fphar.2022.920747.
128. Opo, F.; Rahman, M.M.; Ahammad, F.; Ahmed, I.; Bhuiyan, M.A.; Asiri, A.M. Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification

- of natural anti-cancer agents targeting XIAP protein. *Sci Rep* **2021**, *11*, 4049, doi:10.1038/s41598-021-83626-x.
129. Giordano, D.; Biancanello, C.; Argenio, M.A.; Facchiano, A. Drug Design by Pharmacophore and Virtual Screening Approach. *Pharmaceuticals (Basel)* **2022**, *15*, doi:10.3390/ph15050646.
 130. Moyano-Gomez, P.; Lehtonen, J.V.; Pentikainen, O.T.; Postila, P.A. Building shape-focused pharmacophore models for effective docking screening. *J Cheminform* **2024**, *16*, 97, doi:10.1186/s13321-024-00857-6.
 131. Cieslak, M.; Danel, T.; Krzysztynska-Kuleta, O.; Kalinowska-Tluscik, J. Machine learning accelerates pharmacophore-based virtual screening of MAO inhibitors. *Sci Rep* **2024**, *14*, 8228, doi:10.1038/s41598-024-58122-7.
 132. Visan, A.I.; Negut, I. Integrating Artificial Intelligence for Drug Discovery in the Context of Revolutionizing Drug Delivery. *Life (Basel)* **2024**, *14*, doi:10.3390/life14020233.
 133. Turon, G.; Hlozek, J.; Woodland, J.G.; Kumar, A.; Chibale, K.; Duran-Frigola, M. First fully-automated AI/ML virtual screening cascade implemented at a drug discovery centre in Africa. *Nat Commun* **2023**, *14*, 5736, doi:10.1038/s41467-023-41512-2.
 134. Qureshi, R.; Irfan, M.; Gondal, T.M.; Khan, S.; Wu, J.; Hadi, M.U.; Heymach, J.; Le, X.; Yan, H.; Alam, T. AI in drug discovery and its clinical relevance. *Heliyon* **2023**, *9*, e17575, doi:10.1016/j.heliyon.2023.e17575.
 135. MacDonald, J.T.; Freemont, P.S. Computational protein design with backbone plasticity. *Biochem Soc Trans* **2016**, *44*, 1523-1529, doi:10.1042/BST20160155.
 136. Pan, X.; Kortemme, T. Recent advances in de novo protein design: Principles, methods, and applications. *J Biol Chem* **2021**, *296*, 100558, doi:10.1016/j.jbc.2021.100558.
 137. 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.; et al. De novo design of protein structure and function with RFdiffusion. *Nature* **2023**, *620*, 1089-1100, doi:10.1038/s41586-023-06415-8.
 138. Bennett, N.R.; Coventry, B.; Goreschnik, I.; Huang, B.; Allen, A.; Vafeados, D.; Peng, Y.P.; Dauparas, J.; Baek, M.; Stewart, L.; et al. Improving de novo protein binder design with deep learning. *Nat Commun* **2023**, *14*, 2625, doi:10.1038/s41467-023-38328-5.
 139. Kortemme, T. De novo protein design-From new structures to programmable functions. *Cell* **2024**, *187*, 526-544, doi:10.1016/j.cell.2023.12.028.
 140. Di Rienzo, L.; Milanetti, E.; Testi, C.; Montemiglio, L.C.; Baiocco, P.; Boffi, A.; Ruocco, G. A novel strategy for molecular interfaces optimization: The case of Ferritin-Transferrin receptor interaction. *Comput Struct Biotechnol J* **2020**, *18*, 2678-2686, doi:10.1016/j.csbj.2020.09.020.
 141. Gainza, P.; Wehrle, S.; Van Hall-Beauvais, A.; Marchand, A.; Scheck, A.; Harteveld, Z.; Buckley, S.; Ni, D.; Tan, S.; Sverrisson, F.; et al. De novo design of protein interactions with learned surface fingerprints. *Nature* **2023**, *617*, 176-184, doi:10.1038/s41586-023-05993-x.
 142. Ganesan, S.; Mittal, N.; Bhat, A.; Adiga, R.S.; Ganesan, A.; Nagarajan, D.; Varadarajan, R. Improved Prediction of Stabilizing Mutations in Proteins by Incorporation of Mutational Effects on Ligand Binding. *Proteins* **2024**, doi:10.1002/prot.26738.
 143. Zheng, F.; Liu, Y.; Yang, Y.; Wen, Y.; Li, M. Assessing computational tools for predicting protein stability changes upon missense mutations using a new dataset. *Protein Sci* **2024**, *33*, e4861, doi:10.1002/pro.4861.
 144. Blaabjerg, L.M.; Kassem, M.M.; Good, L.L.; Jonsson, N.; Cagiada, M.; Johansson, K.E.; Boomsma, W.; Stein, A.; Lindorff-Larsen, K. Rapid protein stability prediction using deep learning representations. *Elife* **2023**, *12*, doi:10.7554/eLife.82593.
 145. Musil, M.; Stourac, J.; Bendl, J.; Brezovsky, J.; Prokop, Z.; Zendulka, J.; Martinek, T.; Bednar, D.; Damborsky, J. FireProt: web server for automated design of thermostable proteins. *Nucleic Acids Res* **2017**, *45*, W393-W399, doi:10.1093/nar/gkx285.
 146. Musil, M.; Jezik, A.; Horackova, J.; Borko, S.; Kabourek, P.; Damborsky, J.; Bednar, D. FireProt 2.0: web-based platform for the fully automated design of thermostable proteins. *Brief Bioinform* **2023**, *25*, doi:10.1093/bib/bbad425.
 147. Gonzalez, N.A.; Li, B.A.; McCully, M.E. The stability and dynamics of computationally designed proteins. *Protein Eng Des Sel* **2022**, *35*, doi:10.1093/protein/gzac001.
 148. Thomson, R.E.S.; Carrera-Pacheco, S.E.; Gillam, E.M.J. Engineering functional thermostable proteins using ancestral sequence reconstruction. *J Biol Chem* **2022**, *298*, 102435, doi:10.1016/j.jbc.2022.102435.
 149. Sumida, K.H.; Nunez-Franco, R.; Kalvet, I.; Pellock, S.J.; Wicky, B.I.M.; Milles, L.F.; Dauparas, J.; Wang, J.; Kipnis, Y.; Jameson, N.; et al. Improving Protein Expression, Stability, and Function with ProteinMPNN. *J Am Chem Soc* **2024**, *146*, 2054-2061, doi:10.1021/jacs.3c10941.
 150. Ni, D.; Liu, Y.; Kong, R.; Yu, Z.; Lu, S.; Zhang, J. Computational elucidation of allosteric communication in proteins for allosteric drug design. *Drug Discov Today* **2022**, *27*, 2226-2234, doi:10.1016/j.drudis.2022.03.012.
 151. Verkhivker, G.M.; Agajanian, S.; Hu, G.; Tao, P. Allosteric Regulation at the Crossroads of New Technologies: Multiscale Modeling, Networks, and Machine Learning. *Front Mol Biosci* **2020**, *7*, 136, doi:10.3389/fmolb.2020.00136.

152. Sheik Amamuddy, O.; Veldman, W.; Manyumwa, C.; Khairallah, A.; Agajanian, S.; Oluyemi, O.; Verkhivker, G.; Tastan Bishop, O. Integrated Computational Approaches and Tools for Allosteric Drug Discovery. *Int J Mol Sci* **2020**, *21*, doi:10.3390/ijms21030847.
153. Chen, J.; Vishweshwaraiah, Y.L.; Dokholyan, N.V. Design and engineering of allosteric communications in proteins. *Curr Opin Struct Biol* **2022**, *73*, 102334, doi:10.1016/j.sbi.2022.102334.
154. Ebrahimi, S.B.; Samanta, D. Engineering protein-based therapeutics through structural and chemical design. *Nat Commun* **2023**, *14*, 2411, doi:10.1038/s41467-023-38039-x.
155. Alvisi, N.; de Vries, R. Biomedical applications of solid-binding peptides and proteins. *Mater Today Bio* **2023**, *19*, 100580, doi:10.1016/j.mtbio.2023.100580.
156. Vymetal, J.; Mertova, K.; Bousova, K.; Sulc, J.; Tripsianes, K.; Vondrasek, J. Fusion of two unrelated protein domains in a chimera protein and its 3D prediction: Justification of the x-ray reference structures as a prediction benchmark. *Proteins* **2022**, *90*, 2067-2079, doi:10.1002/prot.26398.
157. Truong, K.; Ikura, M. Domain fusion analysis by applying relational algebra to protein sequence and domain databases. *BMC Bioinformatics* **2003**, *4*, 16, doi:10.1186/1471-2105-4-16.
158. Chen, X.; Zaro, J.L.; Shen, W.C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* **2013**, *65*, 1357-1369, doi:10.1016/j.addr.2012.09.039.
159. Xia, Y.; Zhao, K.; Liu, D.; Zhou, X.; Zhang, G. Multi-domain and complex protein structure prediction using inter-domain interactions from deep learning. *Commun Biol* **2023**, *6*, 1221, doi:10.1038/s42003-023-05610-7.
160. Ferruz, N.; Noske, J.; Hocker, B. Protlego: a Python package for the analysis and design of chimeric proteins. *Bioinformatics* **2021**, *37*, 3182-3189, doi:10.1093/bioinformatics/btab253.
161. Garcia-Paz, F.M.; Del Moral, S.; Morales-Arrieta, S.; Ayala, M.; Trevino-Quintanilla, L.G.; Olvera-Carranza, C. Multidomain chimeric enzymes as a promising alternative for biocatalysts improvement: a minireview. *Mol Biol Rep* **2024**, *51*, 410, doi:10.1007/s11033-024-09332-9.
162. Norman, R.A.; Ambrosetti, F.; Bonvin, A.; Colwell, L.J.; Kelm, S.; Kumar, S.; Krawczyk, K. Computational approaches to therapeutic antibody design: established methods and emerging trends. *Brief Bioinform* **2020**, *21*, 1549-1567, doi:10.1093/bib/bbz095.
163. Kim, J.; McFee, M.; Fang, Q.; Abidin, O.; Kim, P.M. Computational and artificial intelligence-based methods for antibody development. *Trends Pharmacol Sci* **2023**, *44*, 175-189, doi:10.1016/j.tips.2022.12.005.
164. Madsen, A.V.; Mejias-Gomez, O.; Pedersen, L.E.; Preben Morth, J.; Kristensen, P.; Jenkins, T.P.; Goletz, S. Structural trends in antibody-antigen binding interfaces: a computational analysis of 1833 experimentally determined 3D structures. *Comput Struct Biotechnol J* **2024**, *23*, 199-211, doi:10.1016/j.csbj.2023.11.056.
165. Hutchinson, M.; Ruffolo, J.A.; Haskins, N.; Iannotti, M.; Vozza, G.; Pham, T.; Mehzaheen, N.; Shandilya, H.; Rickert, K.; Croasdale-Wood, R.; et al. Toward enhancement of antibody thermostability and affinity by computational design in the absence of antigen. *MAbs* **2024**, *16*, 2362775, doi:10.1080/19420862.2024.2362775.
166. Kaczmarzski, J.A.; Mitchell, J.A.; Spence, M.A.; Vongsouthi, V.; Jackson, C.J. Structural and evolutionary approaches to the design and optimization of fluorescence-based small molecule biosensors. *Curr Opin Struct Biol* **2019**, *57*, 31-38, doi:10.1016/j.sbi.2019.01.013.
167. Quijano-Rubio, A.; Yeh, H.W.; Park, J.; Lee, H.; Langan, R.A.; Boyken, S.E.; Lajoie, M.J.; Cao, L.; Chow, C.M.; Miranda, M.C.; et al. De novo design of modular and tunable protein biosensors. *Nature* **2021**, *591*, 482-487, doi:10.1038/s41586-021-03258-z.
168. Liu, Y.; Liu, Y.; Wang, M. Design, Optimization and Application of Small Molecule Biosensor in Metabolic Engineering. *Front Microbiol* **2017**, *8*, 2012, doi:10.3389/fmicb.2017.02012.
169. Singh, A.; Sharma, A.; Ahmed, A.; Sundramoorthy, A.K.; Furukawa, H.; Arya, S.; Khosla, A. Recent Advances in Electrochemical Biosensors: Applications, Challenges, and Future Scope. *Biosensors (Basel)* **2021**, *11*, doi:10.3390/bios11090336.
170. Naresh, V.; Lee, N. A Review on Biosensors and Recent Development of Nanostructured Materials-Enabled Biosensors. *Sensors (Basel)* **2021**, *21*, doi:10.3390/s21041109.
171. Pham, C.; Stogios, P.J.; Savchenko, A.; Mahadevan, R. Computation-guided transcription factor biosensor specificity engineering for adipic acid detection. *Comput Struct Biotechnol J* **2024**, *23*, 2211-2219, doi:10.1016/j.csbj.2024.05.002.
172. Markus, B.; C, G.C.; Andreas, K.; Arkadij, K.; Stefan, L.; Gustav, O.; Elina, S.; Radka, S. Accelerating Biocatalysis Discovery with Machine Learning: A Paradigm Shift in Enzyme Engineering, Discovery, and Design. *ACS Catal* **2023**, *13*, 14454-14469, doi:10.1021/acscatal.3c03417.
173. 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. *Nature Reviews Methods Primers* **2021**, *1*, doi:10.1038/s43586-021-00044-z.
174. Radley, E.; Davidson, J.; Foster, J.; Obexer, R.; Bell, E.L.; Green, A.P. Engineering Enzymes for Environmental Sustainability. *Angew Chem Weinheim Bergstr Ger* **2023**, *135*, e202309305, doi:10.1002/ange.202309305.

175. Qiu, J.; Chen, Y.; Zhang, L.; Wu, J.; Zeng, X.; Shi, X.; Liu, L.; Chen, J. A comprehensive review on enzymatic biodegradation of polyethylene terephthalate. *Environ Res* **2024**, *240*, 117427, doi:10.1016/j.envres.2023.117427.
176. Mesbah, N.M. Industrial Biotechnology Based on Enzymes From Extreme Environments. *Front Bioeng Biotechnol* **2022**, *10*, 870083, doi:10.3389/fbioe.2022.870083.
177. Chen, Z.; Wang, X.; Chen, X.; Huang, J.; Wang, C.; Wang, J.; Wang, Z. Accelerating therapeutic protein design with computational approaches toward the clinical stage. *Comput Struct Biotechnol J* **2023**, *21*, 2909-2926, doi:10.1016/j.csbj.2023.04.027.
178. Ewaisha, R.; Anderson, K.S. Immunogenicity of CRISPR therapeutics-Critical considerations for clinical translation. *Front Bioeng Biotechnol* **2023**, *11*, 1138596, doi:10.3389/fbioe.2023.1138596.
179. Harris, C.T.; Cohen, S. Reducing Immunogenicity by Design: Approaches to Minimize Immunogenicity of Monoclonal Antibodies. *BioDrugs* **2024**, *38*, 205-226, doi:10.1007/s40259-023-00641-2.
180. Ingolfsson, H.I.; Bhatia, H.; Aydin, F.; Oppelstrup, T.; Lopez, C.A.; Stanton, L.G.; Carpenter, T.S.; Wong, S.; Di Natale, F.; Zhang, X.; et al. Machine Learning-Driven Multiscale Modeling: Bridging the Scales with a Next-Generation Simulation Infrastructure. *J Chem Theory Comput* **2023**, *19*, 2658-2675, doi:10.1021/acs.jctc.2c01018.
181. Qiu, Y.; Wei, G.W. Artificial intelligence-aided protein engineering: from topological data analysis to deep protein language models. *Brief Bioinform* **2023**, *24*, doi:10.1093/bib/bbad289.
182. Poletto, M.D.; Lemkul, J.A. Integration of Experimental Data and Use of Automated Fitting Methods in Developing Protein Force Fields. *Commun Chem* **2022**, *5*, doi:10.1038/s42004-022-00653-z.
183. Kamenik, A.S.; Handle, P.H.; Hofer, F.; Kahler, U.; Kraml, J.; Liedl, K.R. Polarizable and non-polarizable force fields: Protein folding, unfolding, and misfolding. *J Chem Phys* **2020**, *153*, 185102, doi:10.1063/5.0022135.
184. Lopes, P.E.; Guvench, O.; MacKerell, A.D., Jr. Current status of protein force fields for molecular dynamics simulations. *Methods Mol Biol* **2015**, *1215*, 47-71, doi:10.1007/978-1-4939-1465-4_3.
185. Bamezai, S.; Maresca di Serracapriola, G.; Morris, F.; Hildebrandt, R.; Amil, M.A.S.; Sporadic, G.E.M.T.; Ledesma-Amaro, R. Protein engineering in the computational age: An open source framework for exploring mutational landscapes in silico. *Eng Biol* **2023**, *7*, 29-38, doi:10.1049/enb2.12028.
186. Barrozo, A.; Borstnar, R.; Marloie, G.; Kamerlin, S.C. Computational protein engineering: bridging the gap between rational design and laboratory evolution. *Int J Mol Sci* **2012**, *13*, 12428-12460, doi:10.3390/ijms131012428.
187. Verma, R.; Schwaneberg, U.; Roccatano, D. Computer-Aided Protein Directed Evolution: a Review of Web Servers, Databases and other Computational Tools for Protein Engineering. *Comput Struct Biotechnol J* **2012**, *2*, e201209008, doi:10.5936/csbj.201209008.
188. Carobene, A.; Padoan, A.; Cabitza, F.; Banfi, G.; Plebani, M. Rising adoption of artificial intelligence in scientific publishing: evaluating the role, risks, and ethical implications in paper drafting and review process. *Clin Chem Lab Med* **2024**, *62*, 835-843, doi:10.1515/cclm-2023-1136.
189. Kargl, M.; Plass, M.; Muller, H. A Literature Review on Ethics for AI in Biomedical Research and Biobanking. *Yearb Med Inform* **2022**, *31*, 152-160, doi:10.1055/s-0042-1742516.
190. Holzinger, A.; Keiblinger, K.; Holub, P.; Zatloukal, K.; Muller, H. AI for life: Trends in artificial intelligence for biotechnology. *N Biotechnol* **2023**, *74*, 16-24, doi:10.1016/j.nbt.2023.02.001.
191. Resnik, D.B.; Hosseini, M. The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI and Ethics* **2024**, doi:10.1007/s43681-024-00493-8.
192. Maccaro, A.; Stokes, K.; Statham, L.; He, L.; Williams, A.; Pecchia, L.; Piaggio, D. Clearing the Fog: A Scoping Literature Review on the Ethical Issues Surrounding Artificial Intelligence-Based Medical Devices. *J Pers Med* **2024**, *14*, doi:10.3390/jpm14050443.
193. Kohyama, S.; Frohn, B.P.; Babl, L.; Schwill, P. Machine learning-aided design and screening of an emergent protein function in synthetic cells. *Nat Commun* **2024**, *15*, 2010, doi:10.1038/s41467-024-46203-0.
194. Yue, K.; Chen, J.; Li, Y.; Kai, L. Advancing synthetic biology through cell-free protein synthesis. *Comput Struct Biotechnol J* **2023**, *21*, 2899-2908, doi:10.1016/j.csbj.2023.05.003.
195. Levin, D.B.; Budisa, N. Synthetic biology encompasses metagenomics, ecosystems, and biodiversity sustainability within its scope. *Frontiers in Synthetic Biology* **2023**, *1*, doi:10.3389/fsybi.2023.1255472.
196. Yamagata, M. SynBio: A Journal for Advancing Solutions to Global Challenges. *SynBio* **2023**, *1*, 190-193, doi:10.3390/synbio1030013.

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.