Review Article *Open Access*

Journal of Biomedical Advancement Scientific Research

ISSN: 3069-0854

DOI: doi.org/10.63721/26JBASR0154

Artificial Intelligence in Protein Engineering: From Therapeutic Design to Precision Medicine

Ian Pranandi

Department of Biochemistry, School of Medicine and Health Sciences, Atma Jaya Catholic University of Indonesia, Jakarta 14440, Indonesia

Citation: *Ian Pranandi (2026) Artificial Intelligence in Protein Engineering: From Therapeutic Design to Precision Medicine. J. of Bio Adv Sci Research, 2(3):01-30. WMJ/JBASR-154*

Abstract

Protein engineering has become a cornerstone of modern biotechnology and therapeutic development, enabling the design and optimization of proteins for applications ranging from biologics and vaccines to precision therapeutics. Recent advances in artificial intelligence (AI) have fundamentally transformed this field by accelerating protein structure prediction, function analysis, sequence optimization, and de novo protein design. This review provides an overview of the emerging role of AI in protein engineering and its implications for therapeutic development and precision medicine. Key AI-driven technologies, including machine learning, deep learning, protein language models, and generative AI, are discussed in the context of protein structure prediction, function prediction, and rational protein design. The review further examines the application of these technologies in therapeutic protein development, antibody engineering, vaccine design, immunotherapy, and AI-assisted drug discovery. Particular attention is given to the integration of AI with multi-omics technologies, which enables the identification of patient-specific molecular characteristics and supports the development of personalized therapeutic strategies. The convergence of AI, protein engineering, and precision medicine has created new opportunities for designing biologics tailored to individual disease mechanisms and patient profiles. Despite these advances, challenges related to data quality, model interpretability, experimental validation, ethical considerations, and regulatory oversight remain important barriers to widespread clinical implementation. Emerging technologies, including biological foundation models, generative AI systems, autonomous protein engineering platforms, and digital twin technologies, are expected to further expand the capabilities of AI-driven therapeutic design. Collectively, these developments position AI as a transformative force in protein engineering, accelerating therapeutic innovation and supporting the transition toward more precise, personalized, and effective healthcare solutions.

***Corresponding author**: Ian Pranandi, Department of Biochemistry, School of Medicine and Health Sciences, Atma Jaya Catholic University of Indonesia, Jakarta 14440, Indonesia. **Email:** ian.pranandi@atmajaya.ac.id

Submitted: 07.06.2026 **Accepted**: 16.06.2026 **Published**: 30.06.2026

Keywords: Artificial Intelligence, Protein Engineering, Machine Learning, Protein Design, Therapeutic Proteins, Antibody Engineering, Precision Medicine, Multi-Omics, Generative AI, Drug Discovery

Introduction

Protein Engineering and Its Importance in Modern Medicine

Protein engineering is a multidisciplinary field that involves the design, modification, and optimization of proteins to enhance their structural properties, biological functions, and therapeutic potential. By manipulating amino acid sequences or protein structures, researchers can develop proteins with improved stability, specificity, catalytic activity, and pharmacological characteristics. Since proteins serve as the fundamental molecular machinery of living systems, advances in protein engineering have profoundly influenced modern biomedical research and clinical practice [1,2].

Over the past several decades, protein engineering has emerged as a cornerstone of biotechnology and pharmaceutical innovation. Engineered proteins are now widely utilized in numerous medical applications, including therapeutic enzymes, monoclonal antibodies, vaccines, diagnostic reagents, and gene-editing systems. Recombinant insulin, for example, revolutionized the management of diabetes mellitus, while engineered monoclonal antibodies have transformed the treatment landscape for various cancers, autoimmune disorders, and infectious diseases. Similarly, advances in vaccine engineering have enabled the rapid development of protein-based and recombinant vaccines against emerging pathogens [3,4].

Traditional protein engineering strategies have primarily relied on two complementary approaches: rational design and directed evolution. Rational design utilizes existing knowledge of protein structure-function relationships to introduce targeted modifications, whereas directed evolution employs iterative cycles of mutagenesis and screening to identify desirable protein variants. Both approaches have contributed significantly to the development of clinically relevant biomolecules, however, they are often constrained by the enormous complexity of protein sequence space. Given that even a small protein can theoretically possess an astronomical number of possible amino acid combinations, experimental exploration of all potential variants is practically impossible [5,6].

The increasing availability of biological data, coupled with advances in computational biology, has created new opportunities to overcome these limitations. High-throughput sequencing technologies, structural biology databases, and large-scale proteomic datasets have generated unprecedented amounts of information regarding protein sequences, structures, and functions. These resources provide a valuable foundation for computational methods capable of predicting protein behavior and guiding protein optimization efforts [7,8].

Consequently, protein engineering is undergoing a paradigm shift from predominantly experimental approaches toward data-driven and computationally assisted methodologies. The integration of artificial intelligence (AI) and machine learning techniques into protein engineering has enabled more efficient prediction of protein structures, identification of functional mutations, optimization of therapeutic candidates, and design of entirely novel proteins. These developments have accelerated the pace of biomedical innovation and opened new possibilities for personalized therapeutics and precision medicine [9,10].

As the demand for safer, more effective, and individualized treatments continues to grow, protein engineering is expected to play an increasingly important role in shaping the future of healthcare. The convergence of AI, biotechnology, and protein science has the potential not only to streamline therapeutic development but also to fundamentally transform how biological molecules are designed and applied in clinical medicine.

The Emergence of Artificial Intelligence in Biomedical Research

The rapid advancement of biomedical research over the past two decades has been accompanied by an unprecedented expansion in the volume, complexity, and diversity of biological data. Innovations in next-generation sequencing, high-throughput screening, structural biology, medical imaging, and multi-omics technologies have generated vast datasets that far exceed the analytical capacity of conventional statistical and computational approaches. As a result, researchers increasingly require advanced tools capable of extracting meaningful patterns, identifying hidden relationships, and generating actionable insights from complex biological information [11,12].

Artificial intelligence (AI), broadly defined as the ability of computational systems to perform tasks that typically require human intelligence, has emerged as a transformative technology in this context. Among its various branches, machine learning (ML) and deep learning (DL) have demonstrated remarkable capabilities in recognizing complex patterns within large datasets and making accurate predictions without explicit programming. These approaches have rapidly gained prominence across numerous biomedical disciplines, including genomics, transcriptomics, proteomics, drug discovery, medical imaging, and clinical decision support [13,14].

One of the primary factors driving the adoption of AI in biomedical research is the growing availability of large-scale biological databases. Public repositories containing genomic sequences, protein structures, molecular interactions, and clinical information provide extensive training datasets for AI algorithms. The combination of expanding data resources and increasing computational power has enabled the development of sophisticated predictive models capable of addressing previously intractable biological problems. Consequently, AI has become an essential component of modern biomedical research, facilitating data interpretation, hypothesis generation, and translational innovation [15,16].

In the field of molecular biology, AI has significantly accelerated the analysis of complex biological systems. Machine learning models have been applied to predict gene function, identify disease-associated biomarkers, characterize molecular interactions, and uncover regulatory networks. More recently, deep learning approaches have achieved major breakthroughs in structural biology, particularly in protein structure prediction. The development of highly accurate predictive models has substantially improved the ability of researchers to investigate protein folding, molecular recognition, and biomolecular interactions, thereby reducing reliance on time-consuming and resource-intensive experimental methods [17,18].

Beyond basic research, AI has also transformed drug discovery and therapeutic development. Traditional drug development pipelines often require substantial financial investment and extended timelines, with high rates of clinical failure. AI-driven approaches have improved target identification, virtual screening, lead optimization, and toxicity prediction, enabling more efficient prioritization of promising therapeutic candidates. These advances have contributed to a growing shift toward data-driven drug discovery strategies that integrate computational prediction with experimental validation [19,20].

The impact of AI extends further into the emerging paradigm of precision medicine. By integrating genomic, proteomic, clinical, and environmental data, AI systems can identify patient-specific disease characteristics and support the development of individualized treatment strategies. Such capabilities are particularly valuable in complex diseases characterized by significant biological heterogeneity, including cancer, neurodegenerative disorders, and autoimmune diseases. As healthcare increasingly moves toward personalized interventions, AI is expected to serve as a critical enabler of precision diagnostics and therapeutics [21,22].

Within this broader biomedical landscape, protein engineering has become one of the most promising areas for AI application. Proteins exhibit intricate relationships between sequence, structure, and function that are often difficult to decipher using traditional experimental approaches alone. AI offers powerful tools to model these relationships, predict protein behavior, and guide the rational design of novel biomolecules with desired properties. The convergence of AI and protein engineering is therefore creating new opportunities for therapeutic innovation, bridging the gap between biological discovery and precision medicine [23,24].

As AI methodologies continue to evolve, their influence on biomedical research is expected to expand further. Advances in deep learning, foundation models, and generative AI are increasingly enabling researchers not only to analyze biological data but also to design entirely new biological entities. These developments mark a transition from descriptive and predictive biology toward a more proactive and design-oriented paradigm, positioning AI as a central driver of future biomedical innovation.

Scope and Objectives of this Review

Recent advances in artificial intelligence have transformed multiple areas of biomedical research, with protein engineering emerging as one of the most promising fields for AI-driven innovation. The ability of AI algorithms to analyze vast biological datasets, predict protein structures, model sequence-function relationships, and generate novel protein designs has significantly expanded the capabilities of modern protein engineering. These developments are accelerating the discovery and optimization of therapeutic proteins while simultaneously creating new opportunities for personalized healthcare and precision medicine [25,26].

The progression of protein engineering from traditional experimental methodologies toward AI-assisted design paradigms is illustrated in Figure 1. Historically, protein engineering relied heavily on rational design and directed evolution approaches, both of which required substantial experimental effort and resources. The integration of machine learning, deep learning, and advanced structural prediction models has enabled a transition toward more efficient, data-driven strategies. More recently, generative AI and protein language models have further expanded the field by enabling the design of novel proteins with tailored biological functions. Together, these technological advances are contributing to the development of next-generation therapeutics and supporting the broader implementation of precision medicine [27,28].

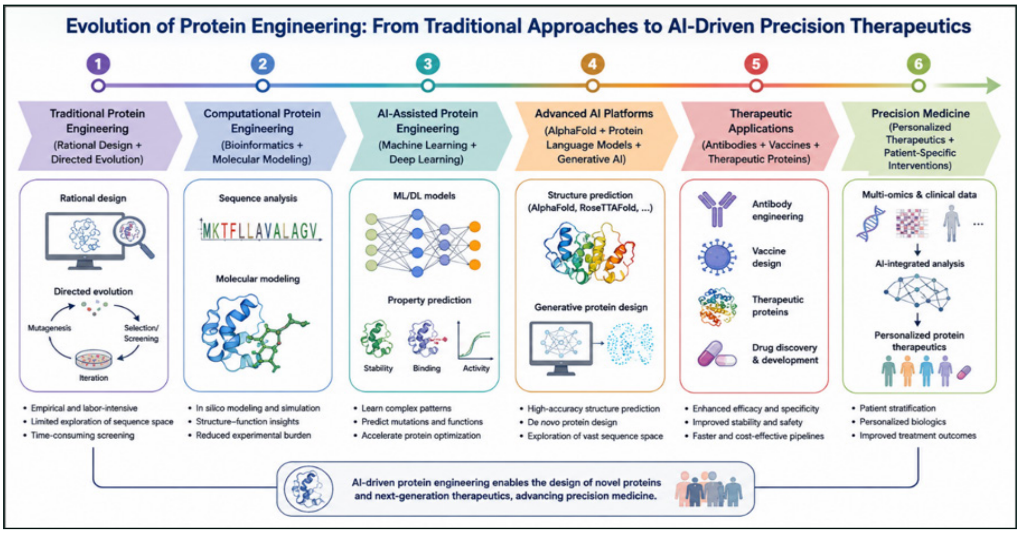

Figure 1: Evolution of Protein Engineering: From Traditional Approaches to AI-Driven Precision Therapeutics

Figure 1 illustrates the progressive evolution of protein engineering from conventional experimental methodologies to modern AI-driven approaches and their applications in precision medicine. Traditional protein engineering relied primarily on rational design and directed evolution, which were often labor-intensive and limited by the vast complexity of protein sequence space. Advances in computational biology introduced bioinformatics and molecular modeling techniques that facilitated in silico analyses of protein structure and function. The subsequent integration of machine learning and deep learning enabled the prediction of protein properties, mutation effects, and sequence-function relationships. More recently, advanced AI platforms, including structure prediction models, protein language models, and generative AI systems, have significantly expanded the ability to design and optimize proteins with desired characteristics. These innovations have accelerated the development of therapeutic proteins, antibodies, vaccines, and drug candidates, ultimately contributing to the emergence of precision medicine through the integration of multi-omics and clinical data for personalized therapeutic interventions. The figure highlights the increasing role of AI in transforming protein engineering into a data-driven discipline that supports next-generation biomedical innovation and individualized healthcare [29,30].

This review aims to provide a concise overview of the emerging role of AI in protein engineering and its implications for therapeutic development and precision medicine. Specifically, the review discusses the major AI methodologies currently employed in protein engineering, including protein structure prediction, sequence optimization, and generative protein design. Furthermore, it highlights how these technologies are being applied to therapeutic protein development, antibody engineering, vaccine design, and drug discovery. Finally, the review examines the growing integration of AI-driven protein engineering with precision medicine, as well as the challenges and future opportunities that may shape the next generation of personalized therapeutics.

Artificial Intelligence Applications in Protein Engineering

Protein Structure Prediction

The biological function of a protein is intrinsically linked to its three-dimensional structure. Protein folding determines the spatial arrangement of amino acid residues, which ultimately governs molecular interactions, enzymatic activity, stability, and biological specificity. Consequently, accurate determination of protein structures is essential for understanding protein function and for guiding the rational design of therapeutic proteins. However, experimental techniques traditionally used to determine protein structures, such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryogenic electron microscopy (cryo-EM), are often time-consuming, expensive, and technically demanding. As a result, only a small fraction of naturally occurring proteins have been structurally characterized experimentally [31,32].

For decades, predicting a protein's three-dimensional structure directly from its amino acid sequence represented one of the most challenging problems in computational biology. This challenge, commonly known as the "protein folding problem," arises from the immense number of possible conformations that a protein can theoretically adopt. Traditional computational approaches, including homology modeling and molecular dynamics simulations, provided valuable insights but were often limited in accuracy, scalability, and computational efficiency [33,34].

The emergence of artificial intelligence has fundamentally transformed protein structure prediction. Early machine learning approaches improved the prediction of secondary structures, residue contacts, and structural motifs by leveraging large biological datasets. However, the most significant breakthrough occurred with the development of deep learning-based models capable of directly learning complex relationships between protein sequences and their corresponding structures. These advances enabled researchers to achieve unprecedented levels of predictive accuracy and significantly accelerated progress in structural biology [35,36].

Among these developments, AlphaFold represented a major milestone in the field. By employing deep neural networks and attention-based architectures, AlphaFold demonstrated remarkable accuracy in predicting protein structures from amino acid sequences. The model achieved near-experimental performance in the Critical Assessment of Structure Prediction (CASP) competitions, effectively redefining expectations for computational structure prediction. The subsequent release of AlphaFold2 and the establishment of large publicly accessible structural databases have dramatically expanded the availability of high-confidence protein structures for scientific research [37,38].

The impact of AI-driven structure prediction extends far beyond structural biology. Accurate protein models provide valuable information for understanding protein function, identifying active sites, predicting molecular interactions, and evaluating the effects of genetic mutations. In protein engineering, structural predictions can guide the selection of rational mutations, improve protein stability, enhance binding affinity, and reduce immunogenicity. Consequently, AI-generated structural information has become an increasingly important component of therapeutic protein development and molecular design [39,40].

Following the success of AlphaFold, several complementary AI-based approaches have emerged.

RoseTTAFold introduced an alternative deep learning framework for protein structure prediction, while ESMFold demonstrated the potential of protein language models to predict structures directly from sequence information. More recently, AlphaFold3 expanded predictive capabilities beyond individual proteins by modeling interactions involving proteins, nucleic acids, small molecules, and other biomolecular components. These developments have further strengthened the role of AI as a foundational technology in modern structural biology and protein engineering [41,42].

Despite these remarkable advances, several challenges remain. Protein dynamics, conformational flexibility, post-translational modifications, and the influence of cellular environments are not always fully captured by current predictive models. Furthermore, computational predictions continue to require experimental validation, particularly for applications involving clinical translation and therapeutic development. Nevertheless, AI-driven protein structure prediction has substantially reduced one of the major bottlenecks in protein engineering and has established a new paradigm for the rational design of biological molecules [43,44].

The rapid progress in structure prediction illustrates how AI can convert vast amounts of biological data into actionable knowledge. By providing accurate structural information at an unprecedented scale, AI-based models have created new opportunities for protein optimization, therapeutic discovery, and precision medicine. As discussed in the following sections, these capabilities form the foundation for broader AI applications in protein function prediction, sequence optimization, and generative protein design.

Protein Function Prediction and Sequence Optimization

While protein structure provides valuable insights into molecular behavior, understanding protein function remains a central objective of protein engineering. Protein function encompasses a wide range of biological activities, including enzymatic catalysis, ligand binding, signal transduction, molecular recognition, and immune regulation. Accurate prediction of these functions is essential for designing proteins with desired properties and for accelerating the development of therapeutic molecules. However, the relationship between protein sequence and function is highly complex, often involving intricate interactions among amino acid residues that are difficult to decipher using conventional experimental approaches alone [45,46].

Artificial intelligence has emerged as a powerful tool for addressing this challenge by enabling the analysis of large-scale protein sequence datasets and the identification of patterns associated with specific functional characteristics. Traditional bioinformatics methods primarily relied on sequence alignment and homology-based inference to predict protein function. Although effective for proteins with known homologs, these approaches often struggle when analyzing novel sequences with limited evolutionary information. AI-based models overcome many of these limitations by learning sequence-function relationships directly from large biological datasets, allowing the prediction of protein properties even in the absence of closely related reference proteins [47,48].

Machine learning algorithms have been widely applied to predict various protein characteristics relevant to protein engineering. These include enzyme activity, substrate specificity, protein stability, solubility, binding affinity, and immunogenicity. By integrating sequence information with structural and biochemical features, AI models can identify determinants of protein performance and estimate the functional consequences of specific mutations. Such predictive capabilities significantly reduce the need for extensive experimental screening, enabling researchers to prioritize promising protein variants for further investigation [49,50].

One of the most significant developments in this area has been the emergence of protein language models (PLMs). Inspired by advances in natural language processing, PLMs are trained on millions of protein sequences and learn contextual relationships between amino acids in a manner analogous to how language models learn relationships between words. Models such as ESM, ProtBERT, ProtT5, and related architectures generate rich representations of protein sequences that capture structural, functional, and evolutionary information. These learned representations can subsequently be utilized for various downstream tasks, including function prediction, mutation assessment, and protein optimization [51,52].

The ability to predict the effects of amino acid substitutions is particularly important in protein engineering. Even single-point mutations can substantially influence protein stability, catalytic efficiency, binding affinity, or immunogenicity. AI-driven mutation prediction models enable researchers to evaluate large numbers of candidate mutations computationally before experimental validation. This approach facilitates rational optimization strategies by identifying variants with an increased likelihood of exhibiting desirable functional properties while minimizing costly laboratory experimentation [53,54].

Beyond prediction, AI is increasingly being used for sequence optimization. Sequence optimization involves identifying protein variants that maximize predefined objectives, such as enhanced stability, improved therapeutic efficacy, increased specificity, or reduced adverse effects. Machine learning models can explore vast regions of protein sequence space and recommend beneficial modifications that may not be apparent through conventional approaches. By iteratively combining computational predictions with experimental feedback, researchers can efficiently optimize proteins for specific biomedical applications [55,56].

These capabilities have substantial implications for therapeutic development. In antibody engineering, AI-assisted sequence optimization can improve antigen-binding affinity and reduce immunogenicity. In enzyme engineering, computational models can identify mutations that enhance catalytic performance or substrate selectivity. Similarly, in vaccine development, AI can assist in optimizing antigen sequences to improve immunogenic responses and protective efficacy. Collectively, these applications demonstrate how AI-driven sequence analysis is transforming protein engineering from a largely empirical process into a more predictive and rational discipline [57,58].

Despite considerable progress, several challenges remain. Protein function is often influenced by complex biological contexts that may not be fully represented in training datasets. Additionally, many AI models operate as "black boxes," limiting biological interpretability and mechanistic understanding. The accuracy of predictions can also vary depending on the availability and quality of experimental data used for model training. Consequently, computational predictions should be considered complementary to, rather than replacements for, experimental validation [59,60].

Nevertheless, AI-based protein function prediction and sequence optimization have become indispensable tools in modern protein engineering. By enabling rapid evaluation of protein variants and guiding rational design strategies, these technologies bridge the gap between protein sequence information and therapeutic application. As increasingly sophisticated models continue to emerge, AI is expected to further enhance our ability to engineer proteins with tailored functions, laying the groundwork for the next generation of biologics and precision therapeutics.

Generative AI for Protein Design

The recent emergence of generative artificial intelligence has introduced a new paradigm in protein engineering. While earlier AI applications primarily focused on predicting protein structures, functions, or mutation effects, generative AI extends these capabilities by enabling the creation of entirely new protein sequences with desired structural and functional properties. This transition from predictive to generative modeling represents a significant milestone in the evolution of protein engineering, offering unprecedented opportunities to explore vast regions of protein sequence space that have never been sampled by natural evolution [61,62].

Generative AI refers to a class of computational models capable of learning complex data distributions and generating novel outputs that resemble the training data while exhibiting unique characteristics. In protein engineering, these models are trained on large collections of protein sequences and structures, allowing them to capture underlying biological principles that govern protein folding, stability, and function. As a result, generative AI systems can propose novel protein variants or entirely new protein architectures tailored for specific biomedical applications [63,64].

Table 1: Major Artificial Intelligence Approaches Used in Protein Engineering and their Biomedical Applications [65].

AI Approach	Representative Models	Primary Applications	Strengths	Limitations
Machine Learning	Random Forest, SVM, XGBoost	Function prediction, mutation analysis	Interpretable, efficient	Limited representation of complex biological relationships
Deep Learning	CNNs, RNNs, Graph Neural Networks	Structure and function prediction	High predictive performance	Large training datasets required
Structure Prediction Models	AlphaFold, RoseTTAFold, ESMFold	Protein structure prediction	Near-experimental accuracy	Limited representation of protein dynamics
Protein Language Models	ESM, ProtBERT, ProtT5	Function prediction, mutation assessment, sequence representation	Capture evolutionary and contextual information	Computationally intensive
Generative AI Models	ProteinMPNN, RFdiffusion, ESM3	De novo protein design, sequence generation	Creation of novel proteins	Experimental validation remains essential

Among the most influential generative approaches are protein language models (PLMs), which have fundamentally changed how biological sequences are represented and analyzed. Similar to large language models used in natural language processing, PLMs learn statistical relationships among amino acids from millions of protein sequences. These models develop internal representations that capture structural, functional, and evolutionary information, enabling the generation of biologically plausible protein sequences. The success of PLMs has demonstrated that protein sequences possess a form of biological "language" that can be learned and manipulated computationally [51,66].

Building upon these foundations, several specialized generative frameworks have been developed for protein design. ProteinMPNN utilizes neural networks to generate amino acid sequences that are compatible with predefined protein structures, facilitating sequence optimization and inverse protein folding. Similarly, RFdiffusion applies diffusion-based generative modeling techniques to create entirely novel protein structures with desired geometries and functionalities. These methods enable researchers to design proteins that may not exist in nature while maintaining structural stability and biological relevance [29,67].

More recently, foundation models such as ESM3 have further expanded the capabilities of generative protein engineering. Unlike earlier models that focused primarily on sequence prediction, ESM3 integrates sequence, structure, and functional information within a unified framework. This multimodal approach allows the model to generate proteins while simultaneously considering multiple biological constraints, thereby improving the likelihood that generated sequences will exhibit desired properties. Such advances suggest that future AI systems may increasingly function as comprehensive biological design platforms rather than specialized predictive tools [9,25].

The implications of generative AI for therapeutic development are substantial. Novel proteins can be designed to exhibit enhanced stability, improved binding affinity, reduced immunogenicity, or optimized pharmacokinetic properties. In antibody engineering, generative models can assist in designing antibodies with higher specificity toward therapeutic targets. In vaccine development, AI-generated antigens may improve immune responses while minimizing undesirable effects. Similarly, enzyme engineering can benefit from the generation of proteins with enhanced catalytic efficiency and substrate selectivity. Collectively, these applications demonstrate the growing potential of generative AI to accelerate the development of next-generation biologics [43,61].

Despite its transformative promise, generative protein design remains associated with several challenges. The biological validity of generated proteins must be experimentally verified, as computational plausibility does not necessarily guarantee biological functionality. Furthermore, many generative models are trained on datasets that may contain biases or underrepresent rare protein families, potentially limiting their generalizability. Additional concerns include model interpretability, safety considerations, and regulatory requirements for AI-designed therapeutics intended for clinical use [35,47].

Nevertheless, generative AI is rapidly reshaping the landscape of protein engineering. By enabling the rational creation of novel proteins rather than merely modifying existing ones, these technologies are expanding the boundaries of biological design and therapeutic innovation. As generative models continue to improve in accuracy, scalability, and biological realism, they are expected to play an increasingly central role in the development of precision therapeutics and personalized medicine. The ability to design proteins with predefined characteristics represents a major step toward a future in which biological molecules can be engineered on demand to address specific clinical needs.

AI-Driven Protein Engineering for Therapeutic Development

Therapeutic Protein Design

Therapeutic proteins represent one of the most successful classes of modern biopharmaceuticals, with applications spanning oncology, autoimmune diseases, metabolic disorders, infectious diseases, and rare genetic conditions. Unlike conventional small-molecule drugs, therapeutic proteins can achieve high target specificity and complex biological functions, enabling the treatment of diseases that were previously considered difficult or impossible to address using traditional pharmacological approaches. Examples of therapeutic proteins include monoclonal antibodies, cytokines, growth factors, hormones, enzyme replacement therapies, and recombinant fusion proteins. As demand for increasingly effective and personalized biologics continues to grow, the development of optimized therapeutic proteins has become a major focus of biomedical research [23,31].

The design of therapeutic proteins is inherently challenging due to the complex relationship between protein sequence, structure, stability, immunogenicity, and biological activity. Even minor modifications to amino acid sequences can substantially influence therapeutic performance, affecting factors such as binding affinity, pharmacokinetics, solubility, manufacturability, and safety. Traditional protein engineering approaches have relied heavily on iterative cycles of mutagenesis and experimental screening, which can be labor-intensive, time-consuming, and costly. Consequently, there is a growing need for computational approaches capable of accelerating and improving the design process [11,17].

Artificial intelligence has emerged as a powerful solution to these challenges by enabling data-driven optimization of therapeutic proteins. Through the integration of large-scale biological datasets, structural information, and machine learning algorithms, AI models can identify sequence patterns associated with desirable therapeutic characteristics. These predictive capabilities allow researchers to prioritize promising protein candidates before experimental validation, thereby reducing development timelines and resource requirements [19,27].

One of the most important applications of AI in therapeutic protein design is the optimization of protein stability. Protein instability can lead to aggregation, loss of biological activity, reduced shelf life, and increased immunogenicity, all of which negatively impact clinical efficacy and manufacturability. AI-based models can predict the effects of amino acid substitutions on protein stability and identify mutations that enhance structural integrity while preserving biological function. Such approaches enable more rational and efficient optimization strategies compared with traditional trial-and-error methods [14,29].

Another major area of application involves improving target specificity and binding affinity. Therapeutic proteins often exert their effects through highly specific molecular interactions with receptors, antigens, or other biological targets. Machine learning algorithms can analyze sequence-structure relationships to predict mutations that strengthen these interactions, thereby enhancing therapeutic potency. This capability is particularly valuable in the development of biologics intended to selectively target disease-associated molecules while minimizing off-target effects [35,39].

AI is also increasingly utilized to reduce immunogenicity, a critical consideration in therapeutic protein development. Because therapeutic proteins may be recognized as foreign by the immune system, unwanted immune responses can compromise treatment efficacy and patient safety. Computational models can identify potentially immunogenic regions within protein sequences and suggest modifications that reduce immune recognition while maintaining therapeutic function. Such strategies are becoming increasingly important as biologics continue to expand across diverse clinical applications [7,35].

Recent advances in generative AI have further expanded the possibilities for therapeutic protein design. Rather than merely optimizing existing proteins, generative models can create novel protein sequences with predefined functional characteristics. These systems enable the exploration of vast regions of sequence space beyond those observed in nature, providing opportunities to discover innovative therapeutic molecules with enhanced performance. As a result, AI is increasingly shifting protein engineering from a reactive process of modifying existing proteins toward a proactive approach focused on designing therapeutics from first principles [27,63].

The impact of AI-driven therapeutic protein design is particularly evident in oncology, immunotherapy, infectious disease research, and precision medicine. In these areas, AI-assisted engineering has facilitated the development of proteins with improved efficacy, reduced toxicity, and greater patient-specific relevance. By integrating structural biology, protein engineering, and machine learning, researchers are establishing more efficient pathways for translating biological knowledge into clinically useful therapeutics [25,31].

Despite these advances, several challenges remain. Predictive models may not fully capture the complexity of biological systems, and computationally optimized proteins still require rigorous experimental validation to confirm safety and efficacy. Furthermore, regulatory frameworks for AI-assisted therapeutic design continue to evolve as increasingly sophisticated computational approaches enter the drug development pipeline. Nevertheless, the growing success of AI-driven protein optimization highlights its potential to transform therapeutic development and accelerate the emergence of next-generation biologics [29,39].

As AI technologies continue to mature, therapeutic protein design is expected to become increasingly precise, efficient, and personalized. The integration of advanced predictive models, generative AI systems, and patient-specific biological data may ultimately enable the routine development of customized protein therapeutics tailored to individual disease characteristics, bringing the vision of precision medicine closer to clinical reality.

Antibody Engineering

Monoclonal antibodies have become one of the most important classes of therapeutic proteins in modern medicine, revolutionizing the treatment of cancer, autoimmune diseases, infectious diseases, and inflammatory disorders. Their ability to recognize specific molecular targets with high affinity and selectivity has enabled the development of highly effective therapies with reduced off-target toxicity compared with many conventional drugs. Over the past several decades, advances in antibody engineering have led to the emergence of increasingly sophisticated antibody-based therapeutics, including humanized antibodies, bispecific antibodies, antibody-drug conjugates (ADCs), and engineered antibody fragments [9,33].

Despite their clinical success, antibody development remains a complex and resource-intensive process. Therapeutic efficacy depends on multiple factors, including antigen-binding affinity, specificity, stability, solubility, pharmacokinetic properties, manufacturability, and immunogenicity. Optimizing these characteristics often requires extensive experimental screening of large antibody libraries, which can be costly and time-consuming. Consequently, there is considerable interest in computational approaches that can accelerate antibody discovery and engineering while reducing development costs [13,45].

Artificial intelligence has emerged as a transformative tool in this field by enabling data-driven prediction and optimization of antibody properties. Machine learning algorithms can analyze large datasets of antibody sequences, structures, and binding interactions to identify patterns associated with favorable therapeutic characteristics. These predictive models facilitate the rapid evaluation of candidate antibodies and support rational design strategies that would be difficult to achieve using conventional experimental methods alone [9,31].

One of the most significant applications of AI in antibody engineering is affinity maturation. High binding affinity is often essential for therapeutic efficacy, particularly in oncology and immunotherapy applications where antibodies must selectively recognize disease-associated targets. AI models can predict the effects of amino acid substitutions within complementarity-determining regions (CDRs), allowing researchers to identify mutations that enhance antigen binding while preserving structural stability. By narrowing the search space for beneficial variants, AI-assisted affinity maturation substantially improves the efficiency of antibody optimization [5,9].

In addition to affinity enhancement, AI is increasingly used to improve antibody specificity. Cross-reactivity with unintended targets can compromise therapeutic efficacy and increase the risk of adverse effects. Machine learning models trained on sequence and structural data can help predict potential off-target interactions and guide the design of antibodies with greater selectivity. Such capabilities are particularly important for precision medicine applications, where highly targeted therapies are required to address specific disease mechanisms while minimizing collateral effects on healthy tissues [14,23].

Another critical application involves the reduction of antibody immunogenicity. Although modern therapeutic antibodies are often humanized or fully human, immune responses can still occur and may limit long-term treatment effectiveness. AI-based immunogenicity prediction tools can identify peptide regions likely to trigger immune recognition and recommend sequence modifications that reduce immunogenic potential while maintaining therapeutic function. This approach supports the development of safer biologics with improved clinical performance [23,27].

Recent advances in generative AI have further expanded the possibilities for antibody design. Generative models can create novel antibody sequences with predefined characteristics, enabling the exploration of sequence diversity beyond naturally occurring repertoires. These technologies facilitate the design of antibodies optimized simultaneously for affinity, specificity, stability, and developability. Furthermore, foundation models trained on large antibody datasets are increasingly capable of predicting complex sequence-structure-function relationships, providing a powerful platform for next-generation antibody discovery [55,61].

The integration of AI into antibody engineering has also accelerated the development of advanced antibody-based modalities. Bispecific antibodies, which simultaneously recognize two different targets, require careful optimization of multiple binding interfaces and structural constraints. Similarly, antibody-drug conjugates depend on the precise engineering of antibodies that can effectively deliver cytotoxic payloads to diseased cells. AI-assisted design strategies can help address these challenges by predicting structural compatibility, target interactions, and functional performance during the early stages of development [45,49].

The impact of AI-driven antibody engineering is particularly evident in oncology, where monoclonal antibodies and immune checkpoint inhibitors have become integral components of modern cancer therapy. AI-assisted optimization strategies have the potential to improve therapeutic efficacy, enhance patient stratification, and support the development of personalized immunotherapies tailored to individual disease profiles. Such advances align closely with the broader goals of precision medicine, where treatment decisions are increasingly informed by molecular and patient-specific data [11,25].

Despite these promising developments, challenges remain. Antibody behavior in vivo is influenced by numerous biological factors that may not be fully captured by computational models. Experimental validation therefore remains essential to confirm the safety, efficacy, and manufacturability of AI-designed antibodies. Additionally, the interpretability of AI models and the availability of high-quality training data continue to influence predictive performance. Nevertheless, the growing success of AI-assisted antibody engineering demonstrates its potential to fundamentally reshape biologic drug development and accelerate the discovery of next-generation therapeutic antibodies [19,39].

As AI technologies continue to evolve, antibody engineering is expected to become increasingly predictive, efficient, and personalized. The combination of structural biology, machine learning, and generative AI is creating new opportunities to develop antibody therapeutics with unprecedented precision, ultimately contributing to improved clinical outcomes and the advancement of precision medicine.

Vaccine and Immunotherapy Development

Vaccines and immunotherapies have emerged as essential components of modern medicine, providing powerful strategies for preventing and treating a wide range of diseases, including infectious diseases, cancer, and immune-mediated disorders. Unlike conventional therapeutics that directly target disease processes, vaccines and immunotherapies harness the body's immune system to generate protective or therapeutic responses. The success of these approaches depends largely on the identification and optimization of appropriate molecular targets capable of eliciting effective and durable immune responses. Consequently, protein engineering plays a central role in the development of next-generation vaccines and immunotherapeutic agents [4,13].

Traditional vaccine development has historically relied on empirical approaches involving attenuated pathogens, inactivated microorganisms, or recombinant proteins. Although these strategies have led to numerous successful vaccines, they often require extensive experimental testing and prolonged development timelines. Similarly, the design of immunotherapies, including cancer vaccines and immune-modulating biologics, involves complex processes for identifying suitable antigens, predicting immune responses, and optimizing therapeutic efficacy. These challenges have stimulated growing interest in computational approaches that can accelerate and improve the design process [49,57].

Artificial intelligence is increasingly transforming vaccine and immunotherapy development by enabling the analysis of large-scale biological and immunological datasets. Machine learning algorithms can identify patterns associated with immune recognition, predict antigenicity, and prioritize candidate targets for experimental validation. By integrating genomic, proteomic, and immunological information, AI facilitates a more rational and data-driven approach to immunotherapeutic design [25,31].

One of the most important applications of AI in this area is epitope prediction. Effective immune responses depend on the recognition of specific antigenic peptides by T cells and B cells. Computational models can predict major histocompatibility complex (MHC)-binding peptides, T-cell epitopes, and B-cell epitopes with increasing accuracy, thereby supporting the identification of promising vaccine candidates. These approaches substantially reduce the experimental burden associated with screening large numbers of potential antigens and have become a cornerstone of modern immunoinformatics [15,17].

The integration of AI with immunoinformatics has significantly accelerated the development of rational vaccine design strategies. Rather than relying solely on whole-pathogen approaches, researchers can now identify highly conserved and immunogenic protein regions suitable for targeted vaccine development. This capability is particularly valuable for rapidly evolving pathogens and diseases characterized by significant genetic variability. AI-assisted antigen selection enables more efficient prioritization of vaccine targets and supports the development of broadly protective immunogens [23,29].

In cancer immunotherapy, AI-driven protein engineering is creating new opportunities for the development of therapeutic cancer vaccines and personalized immunotherapies. Tumors often harbor unique molecular alterations that can generate neoantigens recognized by the immune system. Machine learning algorithms can analyze genomic and transcriptomic data to identify candidate neoantigens, predict their immunogenic potential, and prioritize targets for personalized vaccine development. Such approaches align closely with the principles of precision medicine by tailoring immunotherapeutic interventions to the molecular characteristics of individual patients [33,39].

AI is also contributing to the optimization of vaccine antigens and immunotherapeutic proteins. Predictive models can evaluate the impact of sequence modifications on antigen stability, immunogenicity, and structural integrity, enabling the design of proteins that generate stronger and more durable immune responses. Recent advances in generative AI further extend these capabilities by facilitating the creation of novel antigenic proteins with desirable immunological properties. These technologies provide new opportunities to engineer vaccine candidates that balance efficacy, safety, and manufacturability [51,57].

Beyond antigen design, AI has demonstrated utility in predicting vaccine efficacy and immune responses. By integrating clinical, molecular, and population-level data, machine learning models can identify factors associated with vaccine responsiveness and treatment outcomes. Such predictive capabilities may support patient stratification, optimization of immunization strategies, and the development of more personalized immunotherapeutic approaches in the future [25,35].

The impact of AI-assisted vaccine development has become increasingly evident in recent years, particularly in the context of emerging infectious diseases and cancer immunotherapy. Computational approaches have accelerated target identification, reduced development timelines, and improved the efficiency of candidate selection. These advances highlight the growing role of AI as a complementary tool that enhances both the speed and precision of vaccine and immunotherapy research [13,19].

Despite substantial progress, several challenges remain. Immune responses are highly complex and influenced by numerous genetic, environmental, and physiological factors that may not be fully captured by current computational models. Furthermore, predictions generated by AI systems require rigorous experimental and clinical validation before implementation in healthcare settings. The availability of high-quality immunological datasets and the interpretability of predictive models also remain important considerations for future development [23,29].

Nevertheless, the convergence of artificial intelligence, immunoinformatics, and protein engineering is reshaping the landscape of vaccine and immunotherapy development. By enabling more efficient target discovery, antigen optimization, and personalized therapeutic design, AI has the potential to accelerate the development of next-generation immunological interventions. As computational models continue to improve and become more deeply integrated with experimental research, AI-driven vaccine and immunotherapy development is expected to play an increasingly important role in advancing precision medicine and improving global health outcomes.

AI in Drug Discovery Pipelines

Drug discovery is a complex and resource-intensive process that traditionally requires substantial investments of time, expertise, and financial resources. The development of a new therapeutic agent typically involves multiple stages, including target identification, hit discovery, lead optimization, preclinical evaluation, and clinical testing. Despite continuous technological advances, the overall success rate of drug development remains relatively low, with many candidates failing during preclinical or clinical stages due to inadequate efficacy, toxicity, or pharmacokinetic limitations. Consequently, there is increasing interest in innovative approaches capable of improving the efficiency, accuracy, and success rate of drug discovery pipelines [19,27].

Artificial intelligence has emerged as a transformative technology across the pharmaceutical development process. By leveraging large-scale biological, chemical, and clinical datasets, AI systems can identify complex relationships that may be difficult to detect using conventional computational or experimental approaches. These capabilities enable more informed decision-making throughout the drug development lifecycle, reducing development costs and accelerating the identification of promising therapeutic candidates [23,29].

One of the earliest stages of drug discovery involves target identification, where researchers seek to identify biomolecules associated with disease pathogenesis that may serve as therapeutic intervention points.

AI algorithms can analyze multi-omics datasets, including genomic, transcriptomic, proteomic, and metabolomic information, to identify disease-associated pathways and prioritize potential therapeutic targets. Such approaches have proven particularly valuable in complex diseases characterized by multifactorial molecular mechanisms, including cancer, neurodegenerative disorders, and autoimmune diseases [39,43].

Following target identification, AI plays an increasingly important role in hit discovery and virtual screening. Traditional high-throughput screening methods require the experimental evaluation of large compound libraries, often involving significant time and resource commitments. Machine learning models can rapidly screen millions of candidate molecules in silico, predicting their likelihood of interacting with specific biological targets. This capability substantially reduces the number of compounds requiring experimental testing and improves the efficiency of early-stage drug discovery efforts [13,23].

Advances in protein engineering have further strengthened the integration of AI within drug discovery pipelines. Accurate prediction of protein structures and molecular interactions enables researchers to better understand target biology and identify potential binding sites for therapeutic intervention. AI-generated structural models can support structure-based drug design, facilitate molecular docking studies, and improve predictions of protein-ligand interactions. Consequently, the convergence of AI-driven protein engineering and computational chemistry has created powerful opportunities for rational therapeutic development [39,43].

AI is also increasingly utilized during lead optimization, where candidate molecules are refined to improve efficacy, selectivity, safety, and pharmacokinetic performance. Machine learning models can predict physicochemical properties, toxicity profiles, absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics, allowing researchers to prioritize compounds with favorable development potential. These predictive tools help reduce late-stage failures and improve the overall efficiency of pharmaceutical development [19,29].

Recent advances in generative AI have introduced new possibilities for drug discovery. Generative models can design novel molecular structures, therapeutic peptides, and protein-based therapeutics with predefined biological characteristics. By exploring vast regions of chemical and biological design space, these systems can generate candidates that may not be identified through traditional discovery approaches. The integration of generative AI with protein engineering platforms further enables the simultaneous optimization of both therapeutic targets and therapeutic molecules, creating a more comprehensive and efficient design framework [27,33].

In addition to small-molecule discovery, AI is increasingly contributing to the development of biologics, including therapeutic proteins, antibodies, and vaccines. Predictive models can assist in optimizing molecular stability, binding affinity, immunogenicity, and manufacturability, thereby improving the translational potential of biologic therapeutics. As discussed in the preceding sections, these capabilities are particularly important for the development of next-generation precision therapeutics tailored to specific disease mechanisms and patient populations [25,31].

The growing adoption of AI has also facilitated the emergence of more integrated drug discovery ecosystems. By combining biological data, structural information, clinical evidence, and real-world healthcare data, AI platforms can support decision-making across multiple stages of therapeutic development. This systems-level approach aligns closely with the objectives of precision medicine, where therapeutic strategies increasingly depend on the integration of diverse data sources to identify the most appropriate interventions for individual patients [9,14].

Despite its considerable promise, AI-driven drug discovery faces several challenges. Predictive accuracy remains dependent on the quality and diversity of training datasets, while model interpretability continues to be a concern in highly regulated healthcare environments. Furthermore, computational predictions must ultimately be validated through rigorous experimental and clinical studies. Regulatory agencies are also actively evaluating how AI-generated evidence should be incorporated into pharmaceutical development and approval processes [27,31].

Nevertheless, AI is rapidly reshaping the drug discovery landscape by improving efficiency, reducing costs, and enabling more rational therapeutic design. The integration of AI-driven protein engineering, computational biology, and pharmaceutical sciences is creating new opportunities to accelerate the translation of biological knowledge into clinically effective therapies. As AI technologies continue to mature, they are expected to play an increasingly central role in the development of innovative therapeutics and the advancement of precision medicine.

From Therapeutic Design to Precision Medicine

The Concept of Precision Medicine

Precision medicine has emerged as a transformative paradigm in modern healthcare, shifting medical practice from generalized treatment strategies toward more individualized approaches based on the unique biological characteristics of each patient. Traditionally, medical interventions have largely been developed using population-based evidence, where treatments are prescribed according to the average response observed within a broad patient population. Although this approach has led to substantial improvements in healthcare outcomes, it often fails to account for the considerable variability that exists among individuals in terms of genetics, molecular profiles, environmental exposures, and lifestyle factors [21,35].

The recognition that patients with the same clinical diagnosis may exhibit markedly different disease mechanisms, treatment responses, and prognoses has driven the development of precision medicine. Rather than relying solely on observable clinical symptoms, precision medicine seeks to integrate molecular, genetic, and environmental information to better understand disease heterogeneity and guide therapeutic decision-making. This approach aims to deliver the right treatment to the right patient at the right time, thereby maximizing therapeutic efficacy while minimizing adverse effects [13,25].

Advances in high-throughput technologies have been instrumental in enabling the precision medicine paradigm. Innovations in genomics, transcriptomics, proteomics, metabolomics, and other omics disciplines have generated unprecedented insights into the molecular basis of human health and disease. These technologies allow researchers and clinicians to identify disease-associated biomarkers, characterize molecular subtypes of disease, and predict individual treatment responses. As a result, precision medicine has become increasingly relevant across a wide range of clinical areas, including oncology, infectious diseases, cardiovascular disorders, and rare genetic conditions [19,45].

Among these applications, oncology has emerged as one of the most prominent examples of precision medicine in practice. Molecular profiling of tumors has revealed substantial genetic and biological heterogeneity even among patients with the same cancer type. The identification of specific molecular alterations has enabled the development of targeted therapies designed to selectively inhibit disease-driving pathways. Consequently, treatment decisions are increasingly informed by genomic and molecular information rather than solely by histopathological classification. This shift has contributed to improved patient outcomes and has demonstrated the clinical value of individualized therapeutic strategies [9,21].

The implementation of precision medicine extends beyond genomics alone. While genetic information provides valuable insights into disease susceptibility and therapeutic response, additional biological factors also contribute to clinical variability. Protein expression patterns, immune responses, metabolic profiles, microbiome composition, and environmental influences all play important roles in shaping disease progression and treatment outcomes. Therefore, modern precision medicine increasingly relies on the integration of diverse data sources to develop a more comprehensive understanding of individual patient biology [17,23].

Artificial intelligence has become a critical enabler of this data-intensive healthcare paradigm. The vast volume and complexity of multi-omics, clinical, and real-world health data present significant analytical challenges that often exceed the capabilities of conventional statistical approaches. Machine learning and deep learning algorithms provide powerful tools for identifying clinically relevant patterns, predicting disease trajectories, and supporting personalized treatment recommendations. By transforming large-scale biological datasets into actionable knowledge, AI facilitates the practical implementation of precision medicine in both research and clinical settings [5,17].

The growing convergence of AI, multi-omics technologies, and protein engineering is creating new opportunities for the development of personalized therapeutics. As discussed in previous sections, advances in AI-driven protein engineering have accelerated the design of therapeutic proteins, antibodies, vaccines, and other biologics. When combined with patient-specific molecular information, these technologies offer the potential to tailor therapeutic interventions to individual biological profiles, moving beyond population-based treatments toward truly personalized healthcare solutions [23,39].

Despite its considerable promise, precision medicine faces several challenges. The integration of heterogeneous data sources, concerns regarding data privacy and security, unequal access to advanced technologies, and the need for robust clinical validation remain significant barriers to widespread implementation. Furthermore, translating complex molecular information into actionable clinical decisions requires interdisciplinary collaboration among clinicians, biologists, bioinformaticians, and data scientists [21,43].

Nevertheless, precision medicine represents one of the most important developments in contemporary healthcare. By recognizing and addressing biological variability among patients, this approach has the potential to improve diagnostic accuracy, optimize therapeutic efficacy, and reduce treatment-related adverse effects. As AI-driven protein engineering and multi-omics technologies continue to advance, precision medicine is expected to play an increasingly central role in shaping the future of biomedical research and personalized therapeutic development.

Integrating AI, Multi-Omics, and Protein Engineering

The successful implementation of precision medicine depends on the ability to comprehensively characterize individual biological variability and translate this information into actionable therapeutic strategies. While traditional clinical parameters provide valuable diagnostic insights, they often fail to capture the underlying molecular complexity of disease. Advances in high-throughput technologies have therefore led to the emergence of multi-omics approaches, which integrate information from multiple biological layers, including genomics, transcriptomics, proteomics, metabolomics, and epigenomics, to generate a more comprehensive understanding of human health and disease [5,11].

Each omics platform provides distinct but complementary biological information. Genomics identifies inherited and acquired genetic variations that influence disease susceptibility and therapeutic response. Transcriptomics reveals patterns of gene expression and cellular activity, while proteomics provides direct insights into protein abundance, structure, interactions, and function. Metabolomics further captures the downstream biochemical consequences of cellular processes, reflecting the dynamic physiological state of an individual. Collectively, these datasets offer an unprecedented opportunity to characterize disease mechanisms at a systems level [11,15].

Despite their potential, multi-omics datasets present substantial analytical challenges due to their scale, complexity, and heterogeneity. Individual datasets often contain thousands of variables, complex interactions, and non-linear relationships that are difficult to interpret using conventional analytical methods. Artificial intelligence has emerged as a critical tool for addressing these challenges by enabling the integration, analysis, and interpretation of large-scale biological data. Machine learning and deep learning algorithms can identify hidden patterns, uncover molecular signatures, and generate predictive models that support both biological discovery and clinical decision-making [35,55].

The integration of AI and multi-omics has already demonstrated significant value in disease stratification and biomarker discovery. By analyzing molecular profiles across multiple biological layers, AI models can identify patient subgroups with distinct disease mechanisms, therapeutic vulnerabilities, or prognostic outcomes. Such capabilities are particularly important in heterogeneous diseases such as cancer, where molecularly distinct tumors may exhibit markedly different responses to the same treatment. These insights provide a foundation for more personalized therapeutic strategies and support the broader goals of precision medicine [13,21].

Protein engineering represents a critical translational component within this framework. While multi-omics analyses can identify disease-associated molecular targets and patient-specific biological characteristics, protein engineering provides the tools necessary to develop therapeutic interventions that address these findings. AI-driven protein engineering can utilize molecular information derived from multi-omics analyses to guide the design of therapeutic proteins, antibodies, vaccines, enzymes, and other biologics optimized for specific disease contexts. This integration effectively bridges the gap between molecular diagnosis and therapeutic development [29,61].

One promising application involves the development of targeted biologics based on patient-specific molecular profiles. For example, genomic and transcriptomic analyses may identify disease-driving mutations or dysregulated signaling pathways that serve as therapeutic targets. AI-assisted protein engineering can subsequently be employed to design proteins or antibodies with enhanced specificity toward these targets. Such approaches enable the creation of therapeutics that are tailored not only to a disease category but also to the underlying molecular characteristics of individual patients [15,27].

Similarly, advances in immunogenomics and immunoproteomics are creating new opportunities for personalized immunotherapy. AI algorithms can analyze tumor-specific mutations, predict neoantigen candidates, and prioritize immunogenic targets for therapeutic intervention. Protein engineering platforms can then be utilized to develop customized vaccines, engineered antibodies, or other immunotherapeutic agents designed to target these patient-specific antigens. This strategy represents one of the most compelling examples of how AI, multi-omics, and protein engineering can converge to support precision medicine [21,31].

Recent advances in foundation models and generative AI further enhance these capabilities. Modern AI systems are increasingly capable of integrating diverse biological data types within unified computational frameworks, enabling simultaneous analysis of sequence, structure, function, and clinical information. Such multimodal approaches may facilitate the design of therapeutic proteins that account for multiple biological constraints and patient-specific variables. As these technologies mature, they are expected to play an increasingly important role in the development of individualized therapeutics [29,55].

The conceptual relationship among AI, multi-omics, protein engineering, and precision medicine is illustrated in Figure 2. Multi-omics technologies generate comprehensive molecular data that are analyzed using AI-driven approaches to identify actionable biological insights. These insights subsequently guide protein engineering strategies aimed at developing optimized therapeutics, ultimately supporting personalized treatment decisions and improved clinical outcomes [21,39].

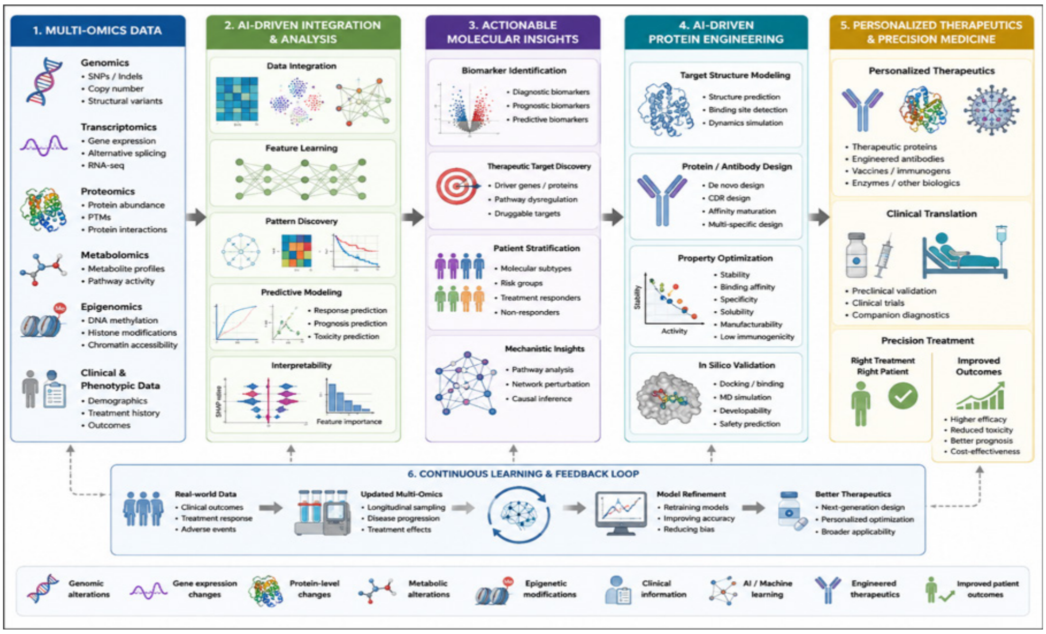

Figure 2: Integration of Artificial Intelligence, Multi-Omics, and Protein Engineering for Precision Medicine

Figure 2 illustrates a conceptual framework demonstrating how the integration of multi-omics technologies, artificial intelligence (AI), and protein engineering supports the development of precision medicine. Comprehensive biological data, including genomics, transcriptomics, proteomics, metabolomics, epigenomics, and clinical information, serve as the foundation for molecular characterization of individual patients. AI-driven analytical platforms integrate these heterogeneous datasets to identify biomarkers, discover therapeutic targets, stratify patients, and generate mechanistic insights into disease biology. The resulting actionable molecular information guides AI-assisted protein engineering processes, including target structure modeling, therapeutic protein and antibody design, sequence optimization, and in silico validation. These engineered biologics can then be translated into personalized therapeutic interventions, such as therapeutic proteins, engineered antibodies, vaccines, and other precision biologics. Clinical implementation supports the selection of the most appropriate treatment for individual patients, leading to improved therapeutic efficacy, reduced toxicity, enhanced patient stratification, and better clinical outcomes. The continuous feedback loop between clinical outcomes, real-world data, multi-omics profiling, and AI model refinement further enables iterative optimization of therapeutic development, creating a dynamic ecosystem that accelerates the realization of precision medicine [21,39].

Despite substantial progress, several challenges remain. Integrating diverse omics datasets requires standardized methodologies, robust computational infrastructure, and high-quality biological data. Furthermore, translating AI-generated insights into clinically effective therapeutics requires extensive experimental validation and interdisciplinary collaboration. Ethical considerations related to data privacy, algorithmic bias, and equitable access to precision medicine technologies must also be addressed to ensure responsible implementation [29,47].

Nevertheless, the convergence of AI, multi-omics, and protein engineering represents a powerful framework for advancing personalized healthcare. By transforming complex biological data into tailored therapeutic solutions, these technologies are accelerating the transition from population-based treatment strategies toward truly individualized medicine. As computational methods, biological datasets, and protein engineering technologies continue to evolve, their integration is expected to become a cornerstone of future precision therapeutic development.

Future Potential of Personalized Protein Therapeutics

The convergence of artificial intelligence, multi-omics technologies, and protein engineering is laying the foundation for a new generation of personalized protein therapeutics. While most current biologic therapies are designed for broad patient populations, future therapeutic strategies may increasingly incorporate individual molecular characteristics to optimize efficacy, safety, and clinical outcomes. Such an approach represents a significant evolution from traditional precision medicine toward a more comprehensive framework in which therapeutic proteins can be tailored to the unique biological profiles of individual patients [29,47].

One of the primary drivers of this transition is the growing availability of patient-specific molecular data. Advances in genomic sequencing, transcriptomic profiling, proteomics, and other omics technologies are generating increasingly detailed insights into disease mechanisms at the individual level. These datasets provide valuable information regarding genetic variants, dysregulated signaling pathways, immune responses, and molecular biomarkers that may influence therapeutic effectiveness. By integrating such information with AI-driven analytical platforms, researchers can identify individualized therapeutic targets and develop more precise intervention strategies [14,21].

Artificial intelligence is expected to play a central role in translating molecular information into personalized therapeutic designs. Machine learning and generative AI models can analyze complex biological datasets to predict optimal protein sequences, structural modifications, and functional characteristics for specific clinical contexts. Rather than relying solely on standardized therapeutic products, future AI platforms may be capable of generating customized protein therapeutics designed to address the unique molecular features of individual patients or patient subgroups [14,27].

Personalized cancer immunotherapy provides one of the most compelling examples of this emerging paradigm. Tumors often harbor unique mutational landscapes that generate patient-specific neoantigens. AI-assisted analysis of genomic and transcriptomic data can identify these neoantigens and predict their immunogenic potential, enabling the design of individualized therapeutic vaccines or engineered antibodies. Such approaches have the potential to improve treatment specificity while reducing unintended effects on healthy tissues. As computational methods continue to advance, personalized immunotherapeutic strategies may become increasingly feasible in routine clinical practice [21,39].

Similarly, future developments in antibody engineering may enable the creation of antibodies optimized for individual disease profiles. AI-driven design platforms could incorporate patient-specific molecular information, such as target expression levels, mutational status, or immune signatures, to generate antibodies with enhanced affinity, specificity, and therapeutic performance. This level of customization may improve treatment responses in diseases characterized by significant biological heterogeneity, including cancer, autoimmune disorders, and infectious diseases [17,45].

Recent advances in generative AI and biological foundation models further expand the possibilities for personalized therapeutics. Models trained on large-scale biological datasets are increasingly capable of integrating sequence, structure, function, and clinical information within unified computational frameworks. Such systems may eventually support the rapid design of therapeutic proteins tailored to specific molecular conditions, accelerating the development of individualized biologics while reducing reliance on extensive experimental screening. The ability to generate novel proteins on demand represents a potentially transformative shift in therapeutic development [39,63].

Another emerging concept is the application of digital twins in personalized medicine. Digital twins are computational representations of individual patients that integrate biological, clinical, and environmental data to simulate disease progression and therapeutic responses. In the future, AI-driven digital twin platforms may enable virtual testing of engineered protein therapeutics before clinical administration. By predicting efficacy, safety, and treatment outcomes in silico, these systems could support more informed therapeutic decision-making and improve the efficiency of personalized treatment development [23,45].

Despite these exciting prospects, several scientific and practical challenges must be addressed before personalized protein therapeutics can be widely implemented. The generation and interpretation of large-scale patient-specific datasets require substantial computational infrastructure and robust data governance frameworks. Manufacturing customized biologics may also present logistical and economic challenges compared with traditional large-scale production models. Furthermore, regulatory agencies will need to establish appropriate evaluation frameworks for AI-designed and potentially patient-specific therapeutics to ensure safety, efficacy, and quality [33,59].

Ethical considerations are equally important. Issues related to patient privacy, algorithmic transparency, healthcare accessibility, and potential disparities in access to advanced technologies must be carefully addressed to ensure equitable implementation of personalized medicine. As therapeutic design becomes increasingly dependent on AI systems, maintaining transparency, accountability, and clinical oversight will remain essential for responsible innovation [5,11].

Nevertheless, the future of personalized protein therapeutics appears highly promising. The integration of AI-driven protein engineering with multi-omics technologies, precision diagnostics, and advanced computational modeling is creating unprecedented opportunities to develop biologics tailored to individual biological characteristics. As these technologies continue to mature, personalized protein therapeutics may evolve from a conceptual aspiration into a practical clinical reality, fundamentally transforming how diseases are diagnosed, treated, and managed. Ultimately, this convergence of disciplines has the potential to establish a new era of precision medicine in which therapeutic interventions are designed not only for specific diseases but also for the unique molecular identities of individual patients.

Challenges and Future Perspectives

Data Availability and Quality

The remarkable progress of artificial intelligence in protein engineering has been driven largely by the increasing availability of biological data. Large-scale repositories containing protein sequences, structural information, molecular interactions, and clinical datasets have provided the foundation for the development of sophisticated machine learning and deep learning models. The success of many contemporary AI systems, including protein structure prediction models, protein language models, and generative design platforms, depends heavily on the quantity, diversity, and quality of the data used during training. Consequently, data availability and quality represent some of the most important factors influencing the performance and reliability of AI-driven protein engineering [7,15].

Despite substantial advances in biological data generation, significant challenges remain regarding data completeness and representativeness. Current databases are heavily enriched for certain protein families, model organisms, and well-studied biological systems, whereas many proteins from less-characterized organisms remain underrepresented. This imbalance may introduce biases into AI models, limiting their ability to generalize across diverse biological contexts. As a result, predictive performance may vary substantially depending on the availability of relevant training data for specific protein classes or therapeutic applications [1,13].

Data quality is equally important. Biological datasets frequently contain experimental variability, annotation errors, incomplete records, and inconsistencies arising from differences in experimental methodologies. Such issues can adversely affect model training and lead to inaccurate predictions or misleading conclusions. In protein engineering applications, even minor inaccuracies in sequence annotations, structural information, or functional characterization can influence downstream predictions related to stability, activity, binding affinity, or therapeutic performance. Therefore, robust data curation and quality control procedures are essential for ensuring the reliability of AI-driven analyses [11,19].

Another important challenge involves the integration of heterogeneous data types. Modern precision medicine increasingly relies on the combination of genomic, transcriptomic, proteomic, metabolomic, structural, and clinical information. While these datasets provide complementary insights into biological systems, they often differ substantially in format, scale, dimensionality, and quality. Integrating such diverse data sources into unified computational frameworks remains a complex task that requires advanced analytical methodologies and standardized data processing pipelines. Inadequate integration can reduce model performance and limit the clinical utility of AI-generated insights [21,25].

The availability of high-quality clinical data presents additional difficulties. Many AI applications in precision medicine require access to patient-specific molecular profiles, treatment histories, and clinical outcomes. However, clinical datasets are often fragmented across institutions, subject to privacy regulations, and limited by incomplete documentation or inconsistent data collection practices. These factors may restrict the development of robust predictive models and hinder the translation of AI-driven protein engineering into clinical applications [14,21].

Data diversity is also critical for ensuring equitable and generalizable AI systems. Numerous biological and clinical datasets disproportionately represent specific populations, geographic regions, or healthcare systems. Consequently, AI models trained on these datasets may exhibit reduced performance when applied to underrepresented populations. Addressing such disparities is particularly important in precision medicine, where individualized therapeutic strategies must be applicable across diverse patient groups to avoid exacerbating existing healthcare inequalities [25,35].

Recent advances in collaborative data-sharing initiatives have begun to address some of these limitations. Public repositories, structural databases, multi-omics resources, and international research consortia have significantly expanded the availability of biological information. The growing adoption of standardized data formats and FAIR (Findable, Accessible, Interoperable, and Reusable) data principles further supports the development of more robust and reproducible AI models. Such efforts are expected to play an increasingly important role in enhancing the quality and accessibility of biological datasets for future research [43,49].

Emerging technologies may also help mitigate data-related challenges. Transfer learning, self-supervised learning, foundation models, and synthetic data generation are increasingly being utilized to overcome limitations associated with scarce or incomplete datasets. Protein language models, for example, can learn meaningful biological representations from large collections of unlabeled protein sequences, reducing dependence on extensive experimentally annotated datasets. Similarly, generative AI approaches may contribute to the creation of synthetic biological data that can supplement existing resources and improve model robustness [39,51].

Nevertheless, data availability and quality remain fundamental constraints on the continued advancement of AI-driven protein engineering. The predictive power of AI systems ultimately depends on the accuracy and representativeness of the information from which they learn. Addressing challenges related to data completeness, standardization, diversity, and accessibility will therefore be essential for maximizing the reliability of AI-based predictions and ensuring the successful translation of protein engineering innovations into therapeutic and precision medicine applications.

Model Interpretability and Reliability

The increasing adoption of artificial intelligence in protein engineering has enabled remarkable advances in protein structure prediction, function prediction, therapeutic design, and drug discovery. However, alongside these achievements, concerns regarding model interpretability and reliability have emerged as significant challenges for both research and clinical translation. While many AI systems demonstrate impressive predictive performance, understanding how these models arrive at their predictions remains difficult, particularly for complex deep learning architectures. Consequently, improving model transparency and ensuring prediction reliability have become important priorities for the future development of AI-driven protein engineering [25,37].

Model interpretability refers to the ability to understand the reasoning processes underlying AI-generated predictions. Traditional statistical models often provide relatively transparent relationships between input variables and outcomes, allowing researchers to identify factors contributing to specific predictions. In contrast, many contemporary AI systems, particularly deep neural networks and foundation models, operate as highly complex computational architectures containing millions or even billions of parameters. Although these models can capture intricate biological relationships, their internal decision-making processes are often difficult to interpret, leading to the characterization of such systems as "black boxes" [23,31].

The lack of interpretability presents several challenges in protein engineering. Researchers may obtain highly accurate predictions regarding protein structure, stability, binding affinity, or functional properties without fully understanding the biological features driving these outcomes. This limitation can reduce confidence in model predictions and hinder the generation of mechanistic insights that are essential for scientific discovery. In therapeutic development, the inability to explain why a model recommends specific protein modifications may complicate decision-making processes and limit acceptance among researchers, clinicians, and regulatory authorities [19,47].

Reliability is closely related to interpretability but represents a distinct concern. A reliable AI model should generate accurate and consistent predictions across different datasets, biological contexts, and experimental conditions. However, predictive performance often varies depending on the quality and characteristics of the training data. Models trained on specific protein families, organisms, or experimental datasets may not generalize effectively to previously unseen biological systems. Such limitations can lead to prediction errors when AI systems are applied outside their original training domains [33,39].

Another challenge involves uncertainty estimation. Many AI models generate predictions without adequately communicating the level of confidence associated with their outputs. In protein engineering applications, this can create difficulties when prioritizing candidate proteins for experimental validation. For example, two predicted protein variants may appear equally promising, yet one prediction may be associated with substantially greater uncertainty. Incorporating uncertainty quantification into AI frameworks can help researchers make more informed decisions and allocate experimental resources more effectively [9,27].

The growing complexity of modern AI systems further amplifies these concerns. Foundation models and generative AI platforms are increasingly capable of integrating diverse biological information and generating novel protein designs. While these capabilities offer tremendous opportunities for innovation, they also introduce additional challenges related to validation and trustworthiness. Generated proteins may appear biologically plausible according to computational metrics while exhibiting unexpected behavior in experimental or clinical settings. Consequently, robust evaluation frameworks are essential to ensure that AI-generated outputs are both scientifically valid and therapeutically relevant [19,29].

Several approaches have been proposed to improve interpretability and reliability in biomedical AI. Explainable artificial intelligence (XAI) techniques aim to provide insights into the features, variables, or biological patterns that contribute to specific predictions. Methods such as attention visualization, feature importance analysis, saliency mapping, and SHAP (Shapley Additive Explanations) have been applied to identify influential sequence regions, structural features, or molecular determinants associated with model outputs. These approaches can improve transparency and facilitate biological interpretation of AI-generated predictions [7,31].

Model robustness can also be enhanced through rigorous validation strategies. Independent external validation datasets, prospective evaluations, benchmark competitions, and reproducibility assessments are increasingly recognized as essential components of AI model development. In protein engineering, experimental verification remains the gold standard for assessing predictive accuracy and biological relevance. Combining computational predictions with laboratory validation helps establish confidence in model performance and supports the responsible application of AI technologies [11,39].

The issue of interpretability is particularly important in precision medicine, where AI-generated recommendations may directly influence therapeutic decisions. Clinicians and regulatory agencies often require clear explanations regarding the rationale behind treatment recommendations before incorporating AI systems into clinical workflows. Therefore, future AI platforms will likely need to balance predictive performance with transparency, ensuring that generated insights are both accurate and understandable to end users [17,29].

Despite these challenges, significant progress is being made toward developing more interpretable and reliable AI systems. Advances in explainable AI, uncertainty estimation, model auditing, and validation methodologies are improving confidence in AI-driven predictions across biomedical applications. As these approaches continue to evolve, they are expected to enhance the trustworthiness of AI-assisted protein engineering and facilitate broader adoption in therapeutic development and precision medicine [9,23].

Ultimately, the long-term success of AI in protein engineering will depend not only on predictive accuracy but also on the ability to generate reliable, transparent, and biologically meaningful insights. Addressing challenges related to interpretability and reliability will therefore be critical for ensuring the responsible integration of AI technologies into future biomedical research, therapeutic innovation, and clinical practice.

Experimental Validation Requirements

Artificial intelligence has dramatically accelerated protein engineering by enabling rapid prediction of protein structures, functions, molecular interactions, and therapeutic properties. Modern AI systems can analyze vast biological datasets and generate hypotheses at a scale that would be impractical using conventional experimental approaches alone. Despite these advances, computational predictions remain fundamentally dependent on experimental validation. The successful translation of AI-generated insights into clinically relevant applications requires rigorous laboratory confirmation to ensure biological accuracy, functional performance, and therapeutic safety [7,19].

One of the primary limitations of AI-based protein engineering is that computational models provide predictions rather than direct evidence of biological function. Although many contemporary models achieve impressive predictive accuracy, biological systems are inherently complex and influenced by numerous factors that may not be fully represented within computational frameworks. Protein folding dynamics, post-translational modifications, cellular environments, molecular interactions, and physiological conditions can all affect protein behavior in ways that are difficult to capture computationally. Consequently, experimental studies remain essential for verifying whether predicted properties are realized under real biological conditions [17,31].

Validation typically begins with in vitro experiments designed to evaluate the structural and functional characteristics of engineered proteins. Recombinant expression systems can be used to produce candidate proteins, followed by assessments of stability, solubility, enzymatic activity, binding affinity, and structural integrity. These experiments provide critical information regarding the accuracy of AI-generated predictions and help identify discrepancies between computational expectations and observed biological behavior. Such evaluations are particularly important when novel protein sequences are generated using generative AI approaches, as these proteins may lack natural biological precedents [14,35].

Structural validation represents another important component of the experimental workflow. Although AI-driven structure prediction models have achieved near-experimental accuracy in many cases, confirmation using techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, or cryogenic electron microscopy (cryo-EM) remains valuable, particularly for therapeutic applications. Experimental structural data can verify predicted conformations, identify unexpected structural features, and provide insights into molecular interactions that may influence therapeutic performance [27,43].

Functional validation is especially critical in therapeutic protein development. AI-generated proteins may exhibit favorable computational characteristics while failing to demonstrate adequate biological activity under experimental conditions. Therefore, in vitro assays are often complemented by cell-based studies that evaluate efficacy, specificity, toxicity, immunogenicity, and mechanism of action. These experiments help determine whether engineered proteins can achieve the desired biological effects while maintaining acceptable safety profiles [13,29].

For therapeutic candidates intended for clinical applications, preclinical validation represents an additional essential step. Animal models are commonly used to assess pharmacokinetics, biodistribution, efficacy, immunogenicity, and toxicity before progression to human studies. Such evaluations provide important information regarding therapeutic performance within complex physiological environments and help identify potential safety concerns that may not be apparent from computational or in vitro analyses alone [39,47].

Experimental validation is particularly important in the context of generative AI. Modern generative models are capable of producing entirely novel protein sequences and structures that may not exist in nature. While these systems offer unprecedented opportunities for innovation, they also increase the risk of generating proteins with unforeseen properties. Comprehensive experimental evaluation is therefore necessary to ensure that generated proteins exhibit the intended biological functions and do not introduce unintended safety risks. As generative AI becomes increasingly integrated into therapeutic development pipelines, robust validation frameworks will become even more important [55,61].

The relationship between AI and experimental biology should be viewed as complementary rather than competitive. AI excels at rapidly exploring vast regions of sequence and structural space, identifying promising candidates, and prioritizing experimental efforts. Experimental research, in turn, provides the empirical evidence necessary to confirm predictions and refine computational models. This iterative cycle of prediction, validation, and model improvement creates a powerful feedback loop that can accelerate protein engineering while maintaining scientific rigor [45,47].

Recent developments have also highlighted the emergence of active learning and closed-loop protein engineering platforms. In these systems, AI models generate predictions that are experimentally tested, and the resulting data are subsequently used to retrain and improve the models. Such approaches enable continuous optimization of predictive performance while reducing the number of experiments required to achieve desired outcomes. This integration of computational and experimental methodologies is increasingly recognized as a promising strategy for future protein engineering workflows [15,31].

Despite substantial technological progress, experimental validation remains the definitive standard for establishing the credibility and translational potential of AI-generated discoveries. Regulatory agencies, clinicians, and researchers all require robust empirical evidence before novel proteins can be considered for therapeutic applications. Consequently, the future success of AI-driven protein engineering will depend not only on advances in computational methodologies but also on the development of efficient validation frameworks capable of translating computational predictions into reliable biological and clinical outcomes [49,63].

As AI technologies continue to evolve, the role of experimental validation is unlikely to diminish. Instead, validation strategies will become increasingly integrated with AI-driven workflows, creating synergistic systems that combine the speed of computational prediction with the reliability of empirical evidence. Such integration will be essential for realizing the full potential of AI-assisted protein engineering in therapeutic development and precision medicine.

Ethical and Regulatory Considerations

The rapid advancement of artificial intelligence in protein engineering offers unprecedented opportunities for therapeutic innovation and precision medicine. However, alongside these scientific and technological developments, important ethical and regulatory challenges have emerged. As AI systems become increasingly involved in the design of therapeutic proteins, antibodies, vaccines, and other biologics, ensuring their safe, transparent, and responsible use has become a critical priority for researchers, healthcare providers, industry stakeholders, and regulatory agencies [47,59].

One of the primary ethical concerns relates to transparency and accountability in AI-driven decision-making. Many advanced machine learning and deep learning models operate as highly complex systems whose internal reasoning processes may be difficult to interpret. In therapeutic development, AI-generated recommendations may influence critical decisions regarding target selection, protein design, candidate prioritization, and clinical development strategies. The inability to fully explain how these recommendations are generated can complicate scientific evaluation and may reduce confidence among researchers, clinicians, patients, and regulators. Consequently, there is growing interest in developing explainable and trustworthy AI systems that provide both accurate predictions and meaningful insights into their underlying decision-making processes [33,43].

Data privacy and security represent another major consideration, particularly in the context of precision medicine. AI-driven therapeutic design increasingly relies on large-scale genomic, proteomic, clinical, and multi-omics datasets derived from individual patients. While these datasets provide valuable information for personalized healthcare, they also contain highly sensitive biological and medical information. Unauthorized access, data breaches, or inappropriate use of personal health data could compromise patient privacy and undermine public trust in AI-enabled healthcare systems. Therefore, robust data governance frameworks, secure data-sharing practices, and compliance with applicable privacy regulations are essential components of responsible AI implementation [11,27].

The use of AI-generated biological designs also raises important questions regarding safety and risk management. Generative AI systems are increasingly capable of producing novel protein sequences and molecular structures that may not exist in nature. Although these technologies offer substantial opportunities for therapeutic innovation, they may also introduce unforeseen biological properties or safety concerns that are difficult to predict computationally. Comprehensive experimental validation, rigorous risk assessment, and appropriate regulatory oversight are therefore necessary to ensure that AI-designed therapeutics meet established standards of safety, efficacy, and quality before clinical application [27,35].

Regulatory agencies face the challenge of adapting existing frameworks to accommodate rapidly evolving AI technologies. Traditional regulatory pathways were developed primarily for therapeutics generated through conventional research and development processes. In contrast, AI-assisted and generative approaches may involve dynamic computational models, continuously evolving datasets, and increasingly autonomous design systems. These developments require regulatory frameworks capable of evaluating not only the final therapeutic product but also the computational processes used during its development. Consequently, regulatory authorities worldwide are actively exploring approaches for assessing the reliability, transparency, and validation of AI-enabled biomedical innovations [14,39].

Another important consideration involves fairness and equitable access. Advanced AI technologies, large-scale computational infrastructure, and personalized therapeutic platforms may initially be concentrated within well-funded institutions and high-resource healthcare systems. If access to these innovations remains uneven, existing healthcare disparities could be further exacerbated. Ensuring equitable access to AI-driven diagnostics, biologics, and precision medicine interventions will therefore be essential for maximizing the societal benefits of these technologies. International collaboration, data-sharing initiatives, and policies that promote accessibility may help address some of these concerns [15,19].

Intellectual property considerations are also becoming increasingly relevant as AI assumes a larger role in therapeutic design. Questions regarding ownership of AI-generated protein sequences, computationally designed biologics, and machine-generated innovations remain subjects of ongoing legal and regulatory debate. Clear guidelines will be necessary to balance incentives for innovation with broader goals related to scientific collaboration and public health [17,35].

The ethical implementation of AI in protein engineering additionally requires careful consideration of algorithmic bias. AI models trained on incomplete or unrepresentative datasets may generate predictions that perform unevenly across different populations, disease subtypes, or biological contexts. Such biases could influence therapeutic development and ultimately affect patient outcomes. Continuous evaluation of model performance across diverse datasets is therefore necessary to ensure fairness, robustness, and generalizability [11,34].

Despite these challenges, ethical and regulatory considerations should not be viewed as barriers to innovation. Rather, they provide essential safeguards that support the responsible development and deployment of AI-enabled technologies. Transparent governance, rigorous validation, interdisciplinary collaboration, and adaptive regulatory frameworks can help ensure that advances in AI-driven protein engineering translate into safe and effective therapeutic solutions [15,29].

As AI becomes increasingly integrated into biomedical research and therapeutic development, ethical and regulatory oversight will play a pivotal role in shaping its future impact. Establishing appropriate frameworks for transparency, accountability, privacy, safety, and equity will be critical for maintaining public trust and enabling the successful translation of AI-assisted protein engineering into clinical practice. Ultimately, responsible innovation will be essential for realizing the full potential of AI-driven therapeutics and precision medicine.

Future Directions

Artificial intelligence has already transformed multiple aspects of protein engineering, from structure prediction and function analysis to therapeutic design and drug discovery. However, current achievements likely represent only the early stages of a broader technological transformation. Continued advances in computational biology, multi-omics technologies, and machine learning are expected to further expand the capabilities of AI-driven protein engineering, enabling increasingly sophisticated approaches to therapeutic development and precision medicine [7,27].

One of the most promising areas of future development involves biological foundation models. Inspired by large language models in natural language processing, foundation models are trained on massive biological datasets encompassing protein sequences, structures, molecular interactions, and functional annotations. These systems are increasingly capable of learning generalizable biological representations that can be adapted to a wide range of downstream tasks. Future foundation models may integrate multiple biological modalities simultaneously, allowing comprehensive analysis of sequence, structure, function, and clinical data within unified computational frameworks. Such capabilities could significantly accelerate biological discovery and therapeutic design [33,51].

Generative AI is also expected to play an increasingly important role in the creation of novel proteins and biologics. Current generative models already demonstrate the ability to design proteins with predefined structural and functional characteristics. As these technologies continue to mature, researchers may be able to generate highly optimized therapeutic proteins, antibodies, enzymes, and vaccines tailored to specific clinical applications. The ability to explore vast regions of protein sequence space beyond those sampled by natural evolution may unlock entirely new classes of therapeutic molecules with unprecedented capabilities [43,61].

Another emerging direction involves autonomous protein engineering platforms. Traditional protein engineering relies on iterative cycles of computational prediction, experimental validation, and optimization. Future systems may increasingly integrate AI-driven design, robotic experimentation, and automated data analysis within closed-loop workflows. Such platforms could continuously generate hypotheses, conduct experiments, analyze results, and refine models with minimal human intervention. By reducing development timelines and increasing experimental efficiency, autonomous systems may substantially accelerate therapeutic innovation [13,25].

The integration of AI with multi-omics technologies is expected to further strengthen precision medicine initiatives. Future therapeutic development may increasingly rely on comprehensive molecular characterization of individual patients, enabling the design of personalized interventions based on specific genetic, transcriptomic, proteomic, and immunological profiles. AI-driven protein engineering could facilitate the creation of customized biologics optimized for individual disease mechanisms, moving healthcare beyond population-based treatment strategies toward truly personalized therapeutics [29,43].

Digital twin technologies represent another exciting frontier. Digital twins are computational representations of individual patients that incorporate biological, clinical, and environmental data to simulate disease progression and therapeutic responses. As computational models become increasingly sophisticated, digital twins may support virtual testing of therapeutic proteins and biologics before clinical administration. Such approaches could improve treatment selection, reduce development risks, and enhance the efficiency of personalized healthcare [7,15].

The future of protein engineering will likely be characterized by increasing convergence among artificial intelligence, systems biology, synthetic biology, and precision medicine. Rather than functioning as isolated disciplines, these fields are expected to become integrated components of a broader biomedical innovation ecosystem. Advances in one area will increasingly inform and accelerate developments in others, creating synergistic opportunities for scientific discovery and therapeutic development [19,25].

Despite these promising prospects, continued progress will require addressing several key challenges related to data quality, interpretability, validation, regulation, and accessibility. Future research efforts should therefore focus not only on improving predictive performance but also on enhancing transparency, reproducibility, and clinical applicability. Interdisciplinary collaboration among computational scientists, biologists, clinicians, engineers, and regulatory experts will be essential for ensuring that technological advances translate into meaningful healthcare benefits [15,23].

Table 2: Current Challenges and Future Opportunities in AI-Driven Protein Engineering [15,19].

Challenge	Impact on Protein Engineering	Potential Solutions	Future Outlook
Limited data quality and availability	Reduced predictive accuracy and generalizability	Improved data curation, standardized databases, FAIR principles	More comprehensive and diverse biological datasets
Model interpretability	Reduced trust and regulatory acceptance	Explainable AI, uncertainty quantification, transparent architectures	More interpretable and clinically acceptable AI systems
Experimental validation requirements	Increased development time and cost	Active learning, automated validation pipelines, closed-loop systems	Faster integration of computational and experimental workflows
Regulatory uncertainty	Delayed clinical translation	Adaptive regulatory frameworks and international guidelines	Clearer pathways for AI-assisted therapeutic development
Data privacy and security concerns	Limited access to clinical and multi-omics data	Secure data-sharing platforms and privacy-preserving AI	Broader access to high-quality patient data with enhanced protections

Limited personalization of current therapeutics	Variable treatment responses among patients	Multi-omics integration and AI-guided therapeutic design	Development of individualized protein therapeutics
Fragmented biomedical workflows	Inefficient translation from discovery to application	Integrated AI-biology-clinical platforms	End-to-end AI-enabled therapeutic development ecosystems

In summary, the future of AI-driven protein engineering extends far beyond the prediction and optimization of existing proteins. Emerging technologies are increasingly enabling the design of novel biological molecules, the integration of complex molecular data, and the development of personalized therapeutic strategies tailored to individual patients. As advances in artificial intelligence continue to converge with innovations in biotechnology and precision medicine, protein engineering is poised to evolve into a highly predictive and design-oriented discipline capable of transforming the future of healthcare. The realization of this vision will depend on the responsible development, validation, and implementation of AI technologies, ensuring that scientific innovation ultimately translates into safer, more effective, and more personalized therapeutic solutions.

Conclusion

Artificial intelligence is fundamentally transforming protein engineering by enabling unprecedented advances in protein structure prediction, function analysis, sequence optimization, and de novo protein design. The integration of machine learning, deep learning, protein language models, and generative AI has shifted protein engineering from a predominantly experimental discipline toward a more predictive and data-driven science. These developments have substantially accelerated the discovery and optimization of therapeutic proteins, antibodies, vaccines, and other biologics, creating new opportunities for innovation across the biomedical landscape [35,68].

The impact of AI extends beyond improving existing protein engineering workflows. Recent advances in computational modeling and generative design are increasingly enabling the creation of novel proteins with tailored structural and functional characteristics. Such capabilities have significant implications for therapeutic development, where AI-assisted approaches can improve target specificity, stability, efficacy, and safety while reducing development timelines and resource requirements. Consequently, AI-driven protein engineering is emerging as a powerful platform for the development of next-generation therapeutics [55,69].

At the same time, the convergence of AI with multi-omics technologies is strengthening the foundation of precision medicine. By integrating genomic, transcriptomic, proteomic, and clinical information, AI systems can identify patient-specific molecular characteristics and support the development of more individualized therapeutic strategies. Protein engineering serves as a critical translational bridge within this framework, enabling the design of biologics that address disease mechanisms at increasingly precise molecular levels. This integration represents an important step toward the realization of personalized therapeutics tailored to the unique biological profiles of individual patients [49,70].

Despite remarkable progress, several challenges remain. Issues related to data quality and availability, model interpretability, experimental validation, regulatory oversight, and ethical governance continue to influence the reliability and clinical translation of AI-driven innovations. Addressing these challenges will require interdisciplinary collaboration among computational scientists, biologists, clinicians, engineers, industry stakeholders, and regulatory authorities. The responsible development of AI technologies, coupled with rigorous validation and transparent governance, will be essential for ensuring their safe and effective implementation in healthcare [15,19].

Looking ahead, advances in biological foundation models, generative AI, autonomous protein engineering platforms, and digital twin technologies are expected to further expand the capabilities of AI-assisted therapeutic design. As these technologies continue to mature, protein engineering may evolve beyond the optimization of existing biomolecules toward the routine creation of customized therapeutics designed for specific diseases, molecular targets, and individual patients. Such developments have the potential to redefine the future of biomedical research and therapeutic development [9,31].

In conclusion, the integration of artificial intelligence and protein engineering represents one of the most significant technological advances in contemporary biomedical science. By accelerating therapeutic discovery, enabling rational molecular design, and supporting the emergence of precision medicine, AI is reshaping the way proteins are engineered and applied in healthcare. Continued innovation at the intersection of artificial intelligence, biotechnology, and medicine will be instrumental in realizing the next generation of personalized and highly effective therapeutic solutions.

References

1. Gonlepa MK, Osotuyi TB, Ofuonye CG, Durojaye OA (2025) Protein engineering as a driver of innovation in therapeutics biotechnology and the global bioeconomy. Discover Chemistry 2: 226.
2. Kaliaperumal K, Govindarajan S, Ravi K, Pranandi I, Kumar PV, et al (2025). AI-assisted design and biochemical optimization of protein structures for enhanced drug delivery in chemotherapy. J Neonatal Surg 14: 147-151.
3. Mbanya JC, Sandow J, Landgraf W, Owens DR (2017) Recombinant human insulin in global diabetes management – focus on clinical efficacy. Eur Endocrinol 13: 21-25.
4. Vishweshwaraiah YL, Dokholyan NV (2022) Toward rational vaccine engineering. Adv Drug Deliv Rev 183: 114142.
5. Mohammed A, Ibrahim NA, Basher NS (2025) Protein engineering and drug discovery: importance, methodologies, challenges, and prospects. Biomolecules 15: 1628.
6. Pranandi I (2025) Integrative biochemical diagnostics: from prenatal genomics to environmental and behavioral biomarkers. J Nat Sci Res Rev 1: 123-130.
7. Fan M, Zou Q (2026) From protein structure to drug discovery: bioinformatics breakthroughs in 2024–2025. Curr Issues Mol Biol 48: 33.
8. Pranandi I (2025) Bioinformatics exploration of biochemical traits associated with culturally distinct populations: between genetics and identity. J Bio Adv Sci Research 1: 1-19.
9. Jin S, Wu Q, Fu G, Lu D, Wang F, et al. (2025) Breaking evolution's ceiling: AI-powered protein engineering. Catalysts. 2025,15(9):842.
10. Pranandi I, Tjhay F (2025) Artificial intelligence and machine learning in biochemical and molecular diagnostics: a transformative review of current applications and future prospects. Int J Comput Exp Sci Eng 11: 4138-4147.
11. Abu-El-Ruz R, Hasan A, Hijazi D, Masoud O, Abdallah AM, et al. (2025) Artificial intelligence in biomedical sciences: a scoping review. Br J Biomed Sci 82: 14362.
12. William W, Sudiyono N, Pranandi I (2025) Artificial intelligence in circadian physiology: predicting biochemical and hormonal rhythms in health and disease. J Bio Adv Sci Research 1: 1-14.
13. Cesarelli G, Ponsiglione AM, Sansone M, Amato F, Donisi L, (2024) Machine learning for biomedical applications. Bioengineering 11: 790.
14. Wang Y, Liu L, Wang C (2023) Trends in using deep learning algorithms in biomedical prediction systems. Front Neurosci 17: 1256351.
15. Aviles M, Tovar-Arriaga S, Pérez-Soto GI, Camarillo-Gómez KA, Rodríguez-Reséndiz J (2025) Artificial intelligence in biomedical technology: advances and challenges. Technologies 13: 208.
16. Clarina S, Siswanto FM, Pranandi I, Handayani MDN, Dewi R (2025) Identification of mir-103a/PLEKHA1 pair as candidate biomarkers and therapeutic targets for skin aging by bioinformatics analysis. Front Health Inform 14: 2245-2254.
17. Sahu MK (2021) AI-based systems biology: modeling complex biological systems to understand disease mechanisms. J Sci Technol (JST) 2: 181-227.
18. Yerramsetty T, Moinuddin MM, Ranjan K, Pranandi I, Gobinath VM, (2025) Harnessing big data analytics and deep learning for predictive bug finding and test automation in complex embedded software systems. J Inf Syst Eng Manag 10: 432-442.
19. Blanco-González A, Cabezón A, Seco-González A, Conde-Torres D, Antelo-Riveiro P, et al. (2023) The role of AI in drug discovery: challenges, opportunities, and strategies. Pharmaceuticals 16: 891.
20. Pranandi I, Rosari BP, Siswanto FM, Surja SS (2025) Molecular Crosstalk between Blastocystis hominis Infection and Colorectal Carcinogenesis: An in Silico Investigation of Shared Pathways and Biomarkers. J Bio Adv Sci Research 1: 1-10.
21. Ahmed Z, Alam MN (2026) Artificial intelligence and precision medicine. Sci Rep 16: 2186.

22. Pranandi I, Arieselia Z (2026) Integrative transcriptomic profiling of human neural tissues reveals core molecular signatures of neurodegeneration. J Nat Sci Res Rev 2: 62-66.
23. Singh N, Lane S, Yu T, Lu J, Ramos A, et al. (2025) A generalized platform for artificial intelligence-powered autonomous enzyme engineering. Nat Commun 16: 5648.
24. Pranandi I, Tanjung JR, Handayani MDN (2025) Snail (Achatina fulica) slime extract effectively increases the speed of incised wound healing process. Front Health Inform 14: 2119-2126.
25. Esmaeilpour D, Hamblin MR, Cheng J, Khosravi A, Liu J, et al. (2026) Artificial intelligence driven protein design and sustainable nanomedicine for advanced theranostics. Bioact Mater 60 :425-455.
26. Pranandi I (2025) Biochemical pathways in neonatal intestinal atresia: vascular and genetic perspectives. J Neonatal Surg 14: 713-719.
27. Das U (2025) Generative AI for drug discovery and protein design: the next frontier in AI-driven molecular science. Med Drug Discov 27:100213.
28. Pranandi I (2025) Fetal-maternal cell-free DNA and RNA in plasma: biochemical insights into non-invasive prenatal testing. Int J Environ Sci 11: 28-48.
29. Rosenkranz AA, Slastnikova TA (2023) Prospects of using protein engineering for selective drug delivery into a specific compartment of target cells. Pharmaceutics 15: 987.
30. Pranandi I (2025) Chronic urticaria as the sole clinical manifestation of autoimmune thyroid disease: a case report. J Clin Case Rep Med Imag Health Sci 12: 1-2.
31. AlQuraishi M (2021) Machine learning in protein structure prediction. Curr Opin Chem Biol 65: 1-8.
32. Pranandi I (2025) Recurrent elbow pain in an adolescent: an atypical presentation of missed radial head subluxation in outpatient practice. EAS J Orthop Physiother 7: 103-105.
33. Grandori R (2023) Protein structure and dynamics in the era of integrative structural biology. Front Biophys 1: 1219843.
34. Pranandi I (2025) Chronic posterior ankle pain in a recreational runner: a rare case of os trigonum syndrome diagnosed in an outpatient setting. J Ortho Physio 3: 1-3.
35. Alanazi W, Meng D, Pollastri G (2025) Advancements in one-dimensional protein structure prediction using machine learning and deep learning. Comput Struct Biotechnol J 27: 1416-1430.
36. Pranandi I (2025) Development of a low-cost kit for microplastic detection in household tap water: a public health perspective. J Bio Adv Sci Research 1: 1-16.
37. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 59: 583-589.
38. Pranandi I, Hananta L, Arieselia Z, Kurniawan SV, Lonah L, et al. (2025) Computational prediction of miRNA targets in cardiovascular disease pathways: an in silico study. J Bio Adv Sci Research 1: 1-9.
39. Qiu X, Li H, Ver Steeg G, Godzik A (2024) Advances in AI for protein structure prediction: implications for cancer drug discovery and development. Biomolecules 2024,14: 339.
40. Pranandi I (2025) In silico prediction of salivary biomarkers associated with nicotine metabolism. J Carcinog 24: 195-210.
41. Lisanza SL, Gershon JM, Tipps SWK, Sims JN, Arnoldt L, et al. (2023) Multistate and functional protein design using RoseTTAFold sequence space diffusion. Nat Biotechnol 43: 1288-1298.
42. Pranandi I (2025) Herpes zoster of the trigeminal nerve presenting as dental pain: a diagnostic challenge in outpatient practice. J Clin Med Images Case Rep 5: 1807.
43. Niazi SK (2025) Critical assessment of AI-based protein structure prediction: Fundamental challenges and future directions in drug discovery. Comput Struct Biotechnol Rep 2: 100064.
44. Goel AK, Chumjinda UC, Pranandi I, Senthur NS (2025) Evidence-based pain management strategies in post-operative patients: advancing SDG 3. Vasc Endovasc Rev 8: 298-305.
45. Park Y, Metzger BPH, Thornton JW (2025) The simplicity of protein sequence-function relationships. Nat Commun 15: 7953.
46. Hananta L, Pranandi I, Handayani MDN, Harjanti DA, Kurniawan F, et al. (2025) Chronic low back pain in a young female farmer from a remote coastal island in Indonesia: a case from a community health charity event. J Ortho Physio 3: 1-5.
47. Hein ZM, Guruparan D, Okunsai B, Che Mohd Nassir CMN, Ramli MDC, et al. (2025) AI and machine learning in biology: from genes to proteins. Biology (Basel) 14: 1453.
48. Handayani MDN, Pakpahan HA, Siswanto FM, Dewi R, Pranandi I, et al. (2025) Edukasi pentingnya

menjaga kebugaran fisik untuk mencegah sindrom metabolik pada karyawan Universitas Katolik Indonesia Atma Jaya. J Panrita Abdi 9: 977-986.
49. Horne J, Shukla D (2022) Recent advances in machine learning variant effect prediction tools for protein engineering. Ind Eng Chem Res 61: 6235-6245.
50. Pranandi I (2025) Toward a universal cancer vaccine: telomerase reverse transcriptase (hTERT) as the core antigen for immunological cancer prevention. J Bio Adv Sci Research 1: 1-24.
51. Leclercq M, Droit A (2026) Protein language models: applications and perspectives. J Proteome Res 25: 507-524.
52. Pranandi I, Lumbanraja MP, Andani BN, Urbinas MI, Jessica M (2025) Sembilang fish spine injury leading to systemic hypersensitivity: a rural coastal case from Muaragembong, Indonesia. J Nat Sci Res Rev 1: 187-189.
53. Mushtaq H, Shah SS, Zarlashat Y, Iqbal M, Abbas W (2024) Cell culture adaptive amino acid substitutions in FMDV structural proteins: a key mechanism for altered receptor tropism. Viruses 16: 512.
54. Pranandi I (2026) Immunoinformatic prediction of broadly reactive T-cell epitopes for universal cancer vaccines. J Bio Adv Sci Research 2: 1-10.
55. Chen L, Li Q, Nasif KFA, Xie Y, Deng B, et al. (2024) AI-driven deep learning techniques in protein structure prediction. Int J Mol Sci 25: 8426.
56. Pranandi I, Dewi R, Margaret AL (2026) Transcriptomic signatures of music perception in the human cortex. J Bio Adv Sci Research 2: 1-11.
57. El Arab RA, Alkhunaizi M, Alhashem YN, Al Khatib A, Bubsheet M, et al. (2025) Artificial intelligence in vaccine research and development: an umbrella review. Front Immunol 16: 1567116.
58. Pranandi I (2026) Differential expression analysis identifies novel regulatory genes in endometriosis. Journal Sehat Indonesia (JUSINDO) 8: 267-275.
59. Lin B, Luo X, Liu Y, Jin X (2025) A comprehensive review and comparison of existing computational methods for protein function prediction. Brief Bioinform 25: bbae289.
60. Jessica M, Urbinas MI, Hartanto KL, Rahelia D, Andani BN, et al. (2026) Edukasi perilaku hidup bersih dan sehat pada siswa sekolah dasar menggunakan pendekatan asset-based community development. Jurnal Pembelajaran Pemberdayaan Masyarakat (JP2M) 7: 542-555.
61. Brackmann M, Reiners S, Hoogendoorn M, Moser M (2026) Protein design, generative AI and biological security. Front Microbiol 17: 1817535.
62. Katuuk SHH, Pranandi I (2026) Circulating miRNA signatures as non-invasive biomarkers for MASLD: a bioinformatics approach. J Bio Adv Sci Research 2: 1-8.
63. Ille AM, Markosian C, Burley SK, Mathews MB, Pasqualini R, et al. (2024) Generative artificial intelligence performs rudimentary structural biology modeling. Sci Rep 14: 19372.
64. Sudiyono N, William W, Pranandi I (2026) Identifikasi gen pengatur sentral pada stroke iskemik menggunakan pendekatan berbasis jaringan. J. Sains. Kes 7: 309-317.
65. Villalobos-Alva J, Ochoa-Toledo L, Villalobos-Alva MJ, Aliseda A, Pérez-Escamirosa F, et al. (2022) Protein science meets artificial intelligence: a systematic review and a biochemical meta-analysis of an inter-field. Front Bioeng Biotechnol 10: 788300.
66. Lamb KD, Hughes J, Lytras S, Young F, Koci O, et al. From single-sequences to evolutionary trajectories: protein language models capture the evolutionary potential of SARS-CoV-2. Nat Commun 17: 2938.
67. Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, et al. (2022) Robust deep learning based protein sequence design using ProteinMPNN. Science 7378: 49-56.
68. Winnifrith A, Outeiral C, Hie BL (2024) Generative artificial intelligence for de novo protein design. Curr Opin Struct Biol 86: 102794.
69. Zhang Y, Wu Y, Wang K, Zhang Z, Liu Y, et al. (2025) Revolutionizing anesthesia: AI-driven protein structure prediction and its clinical impact. Anesthesiol Perioper Sci 3: 38.
70. Liu Y, Zhu K, Peng W, Liu Z, Mao X (2026) Multi-omics and artificial intelligence for precision drug discovery and potential clinical applications. Signal Transduct Target Ther 11: 210.