Acta Pharmaceutica Sinica B 2026;16(2):788−801

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

REVIEW

Harnessing deep learning to accelerate the development of antibodies and aptamers

Pan Tan[a,c,†], Song Li[b,†], Jin Huang[e,†], Ziyi Zhou[d], Liang Hong[a,b,c,d,e,*]

[a]*Chongqing Artificial Intelligence Research Institute of Shanghai Jiao Tong University, Chongqing 401329, China*
[b]*Shanghai Matwings Technology Co., Ltd., Shanghai 200240, China*
[c]*Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China*
[d]*Zhangjiang Institute for Advanced Study, Shanghai Jiao Tong University, Shanghai 201203, China*
[e]*School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China*

Received 22 December 2024; received in revised form 11 November 2025; accepted 18 November 2025

KEY WORDS

Antibody;
RNA aptamer;
Deep learning;
Antibody structure prediction;
RNA structure prediction;
Protein language models;
Protein engineering;
De novo design

Abstract Artificial intelligence (AI) has revolutionized the design of antibodies and RNA aptamers, driving significant advancements in molecular therapeutics. In antibody design, AI enables accurate structure prediction and optimization of binding affinity, specificity, and stability, thereby accelerating the development of therapies targeting challenging antigens, such as those associated with viral infections and cancer. By integrating sequence and structural data, AI significantly reduces experimental costs and development timelines, streamlining the creation of next-generation antibody-based therapeutics. Similarly, AI has transformed RNA aptamer design, addressing long-standing challenges in structure prediction and binding optimization. AI-driven approaches allow for the rapid generation of aptamers with enhanced specificity, stability, and functional properties, expanding their potential applications in both therapeutics and diagnostics. These advancements offer scalable, cost-effective, and highly customizable solutions for precision medicine. As AI systems continue to evolve and integrate with experimental validation, they hold immense promise for developing more effective treatments for complex diseases, including cancer, autoimmune disorders, and viral infections. This marks the beginning of a new era in therapeutic innovation, where AI plays a pivotal role in addressing the challenges of modern medicine.

*Corresponding author.
E-mail address: hongl3liang@sjtu.edu.cn (Liang Hong).
[†]These authors made equal contributions to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2025.12.017

1. Introduction

Macromolecular drugs, such as recombinant proteins, peptides, and nucleic acid therapeutics, have emerged as a cornerstone of modern biomedicine due to their high specificity, strong targeting capabilities, and favorable safety profiles. Over the past four decades (1980−2022), the U.S. Food and Drug Administration (FDA) has approved 235 therapeutic biologics and 17 gene and cell therapies[1]. These innovative therapeutics have provided transformative and promising solutions for patients suffering from cancer, autoimmune diseases, and neurological disorders, where conventional treatments often fall short. Among macromolecular drugs, monoclonal antibodies (mAbs) represent the most critical and rapidly growing class. Remarkably, more than 50% of FDA-approved biologics are antibodies or antibody-derived therapeutics[2], underscoring their pivotal role in clinical medicine. Monoclonal antibodies offer unparalleled specificity in targeting disease-associated antigens, making them highly effective for treating complex and difficult-to-target conditions[3-5].

However, the traditional approaches to antibody development are fraught with significant challenges. These methods rely heavily on *in vivo* immunization or high-throughput screening of *in vitro* antibody libraries, which are both time-consuming and resource-intensive[6]. The conventional antibody development pipeline—from initial discovery to final clinical approval—typically spans several decades and demands investments of billions of dollars. Such long timelines and high costs have hindered the efficiency and accessibility of novel antibody therapies. In this context, the field is increasingly turning to innovative technologies, such as artificial intelligence (AI) and computational biology, to overcome these challenges. AI-driven methods have the potential to accelerate antibody discovery, optimize molecular design, and reduce development costs, thereby revolutionizing the landscape of macromolecular drug development[7]. AI is transforming the research and development paradigm, enabling faster and more cost-effective approaches to antibody discovery and optimization.

Nucleic acid aptamers are single-stranded oligonucleotides that specifically bind to target molecules[8]. Due to their targeting properties, which are similar to monoclonal antibodies, they are often referred to as "chemical antibodies"[9]. Aptamers achieve this specificity through a combination of molecular interactions, including nucleotide base-pairing, hydrogen bonding, $\pi-\pi$ stacking, and electrostatic forces, which drive adaptive folding into specific three-dimensional structures[10]. These structures enable high-affinity binding to target molecules, with dissociation constants reaching nanomolar or picomolar levels, comparable to monoclonal antibodies. Compared to traditional monoclonal antibodies, nucleic acid aptamers offer several advantages, including a broader range of target molecules, superior thermal stability, smaller molecular size, lower immunogenicity, chemical synthesize-ability, minimal batch-to-batch variation, and ease of modification[11]. Aptamers have demonstrated broad utility in biosensing applications. When specifically binding to target substances, aptamers undergo conformational changes. Exploiting this property, researchers have developed electrochemical aptamer-based sensors[12], which are characterized by portability, simplicity, and low cost. Beyond biosensing, aptamers have found applications in disease diagnosis, therapeutics[13], and clinical diagnostics, with advancements in their stability in blood circulation[14]. In nuclear medicine, aptamers have been employed in Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET) imaging[15]. Since now, aptamers have found widespread applications in biosensing, drug delivery[16], bioimaging, disease diagnosis, and therapeutics.

However, the traditional preparation of aptamers relies on the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) technique[17]. This method involves multiple rounds of experiments, making the process cumbersome and costly, thus limiting its broader applicability. Despite significant advances in aptamer discovery, challenges remain in optimizing their affinity, stability, and specificity. The introduction of AI and computational biology has provided new solutions for aptamer design. By simulating the interactions between aptamers and their targets, AI can predict their structures and functions, significantly shortening the screening and optimization process. Additionally, AI can guide aptamer design, improving their affinity and stability, and advancing their potential for clinical applications.

2. Applications of AI in antibody drug development

2.1. AI-driven precision in predicting antibody structures

AI has revolutionized antibody development across four critical domains: structure prediction, affinity maturation, developability assessment, and *de novo* design (Fig. 1). Among these, accurate prediction of antibody structures forms the cornerstone of AI-driven antibody design, enabling precise engineering of therapeutic molecules.

2.1.1. Importance of predicting antibody structures

As crucial immune molecules in living organisms, antibodies play a pivotal role in disease diagnosis and therapy. Accurately predicting antibody structures is essential for understanding their functions in depth. On the one hand, the structure of an antibody determines its mode of binding and specificity to antigens. By predicting antibody structures, researchers can better understand how antibodies recognize and bind to specific antigens, providing critical theoretical insights for the design of antibody-based therapeutics. For instance, computational protein modeling and in-depth analysis of the relationship between antibody sequences and structures enable predictive modeling of antibody structures, facilitating optimized design to enhance binding affinity and stability. On the other hand, advances in AI technology have unlocked significant potential for predicting antibody structures in drug development. The 2024 Nobel Prize in Chemistry, awarded to three scientists for groundbreaking contributions to protein structure prediction and *de novo* design, underscores the transformative impact of AI on scientific research. Algorithms such as AlphaFold2[18] have resolved the decades-long challenge of accurately predicting protein structures, providing a powerful tool for research in biology, medicine, and drug design. In the realm of antibody drug development, precise antibody structure prediction can accelerate the drug discovery process and reduce development costs.

Additionally, predicting antibody structures is instrumental in addressing certain limitations of antibody therapeutics. Although antibody drugs offer advantages over small-molecule drugs in terms of specificity, safety, and half-life, they face challenges such as limited intracellular access, difficulty crossing the blood−brain barrier, and potential immunogenicity. Accurate structure prediction can inform innovative solutions to these challenges. For example, designing antibodies with structures tailored to specific intracellular targets can enhance their therapeutic efficacy within cells. Alternatively, developing antibodies capable of traversing

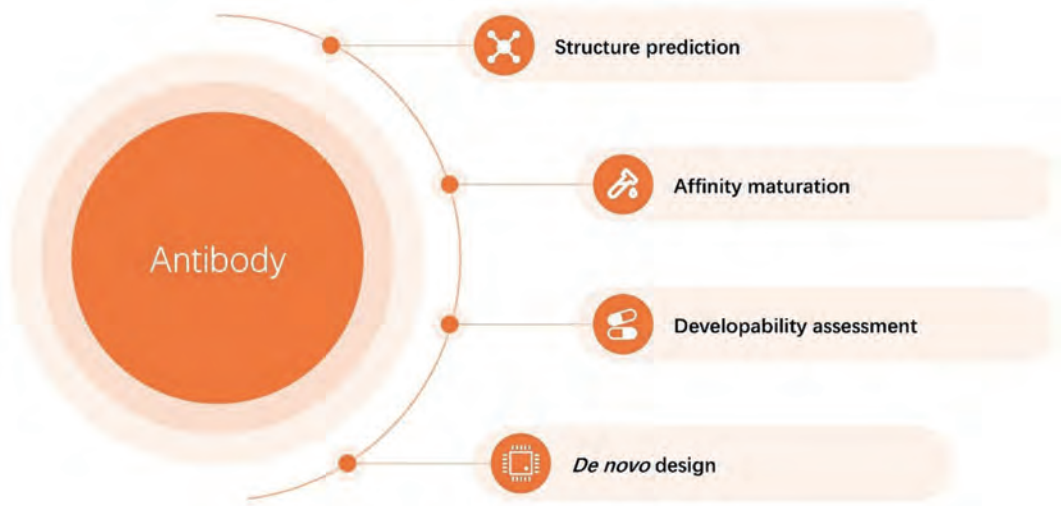

Figure 1 The application of artificial intelligence in antibody development. This schematic illustrates the multifaceted applications of AI in antibody development, encompassing four key areas: structure prediction, affinity maturation, developability assessment, and *de novo* design.

the blood−brain barrier could expand treatment options for central nervous system disorders.

In summary, AI-driven structural predictions not only deepen our understanding of antibody function but also pave the way for innovative strategies in antibody drug development, ultimately advancing the therapeutic potential of these critical biologics.

2.1.2. Methods for antibody structure prediction

2.1.2.1. Breakthroughs in protein structure prediction with AlphaFold2. The 2024 Nobel Prize in Chemistry was awarded to three pioneering scientists for their groundbreaking contributions to protein structure prediction and design, underscoring the transformative role of AI in scientific research. AlphaFold2[18], an AI-based algorithm developed by DeepMind, represents a monumental leap in this field by accurately predicting the complex three-dimensional structures of proteins. This breakthrough addresses a longstanding challenge in protein structure prediction, leveraging extensive databases of protein structures and sequences combined with deep learning techniques. AlphaFold2 has proven invaluable for advancing research in biology, medicine, and drug discovery.

In 2024, a new iteration, AlphaFold3[19], was introduced, offering an enhanced architecture and significantly improved performance. While AlphaFold3 achieves similar accuracy to AlphaFold2 in single-protein structure prediction, its primary strength lies in more precise predictions of protein complexes and its expanded applicability to other biomolecules, encompassing nearly all entities within the Protein Data Bank (PDB). For example, AlphaFold3 outperformed traditional docking tools like the latest version of Vina and state-of-the-art deep learning models such as RoseTTAFold All-Atom[20] in predicting protein−ligand interfaces during the PoseBusters benchmark. This benchmark includes 428 protein−ligand structures published in the PDB before 2021. Additionally, AlphaFold3 demonstrated superior accuracy compared to RoseTTAFold2NA[21] and CASP15's top AI algorithm, Alchemy_RNA[22], in predicting protein-nucleic acid complexes and RNA structures, based on CASP15 examples and the PDB protein-nucleic acid dataset.

AlphaFold3 also accurately predicts the effects of covalent modifications on proteins, RNA, or DNA, such as bonded ligands, glycosylation, and modified residues or nucleic acid bases. However, these functionalities have yet to be benchmarked against other tools. The AlphaFold3 architecture retains similarities to AlphaFold2 but incorporates multiple refinements for enhanced complex structure prediction. Like its predecessor, AlphaFold3 employs template structure searches and multiple sequence alignment (MSA) searches. However, its MSA module processes fewer alignments, resulting in faster computations. The Pairformer module replaces AlphaFold2's Evoformer module, and the structural modeling component transitions to a diffusion-based model. This generative approach produces structural distributions rather than a single structure with uncertainties, leading to more accurate predictions without parameterization. Notably, AlphaFold3 eliminates the physics-based minimization of side-chain atom positions previously handled by AMBER in AlphaFold2. To address unstructured regions, AlphaFold3 uses training data from AlphaFold-Multimer v2.3, which incorporates loops from such regions through cross-distillation. A confidence module evaluates prediction accuracy at both atomic and pairwise levels. The model significantly improves predictions of protein−protein complexes, with notable advancements in antibody−protein interfaces compared to AlphaFold-Multimer v2.3. Leveraging diffusion-based generation, AlphaFold3 processes large-scale inputs to directly predict the structures of biomolecules such as proteins, DNA, and RNA. The enhanced Evoformer module streamlines processing molecular sequences through diffusion networks, generating highly accurate molecular structures. Unlike AlphaFold2, AlphaFold3 abandons rigid stereochemical invariance, allowing greater flexibility in model design by removing constraints imposed by AlphaFold2's additional architectural features.

2.1.2.2. Other antibody-specific structure prediction models. Several antibody-optimized structure prediction models have been developed alongside AlphaFold. ABodyBuilder2[23], based on the AlphaFold-Multimer[24] architecture, specializes in predicting antibody structures and efficiently determines CDRH3 loop conformations. The tFold-Ab[25] model integrates protein language modeling with AlphaFold-Multimer's structure prediction framework, improving both accuracy and efficiency in antibody modeling. ColabDock[26] integrates AlphaFold2 for protein−protein structure prediction, enhancing accuracy without

extensive retraining. It employs AlphaFold2 as its core structural prediction model, outperforming other methods by incorporating various experimental constraints. ABlooper[27] employs E(n)-Equivariant graph neural networks to directly predict the backbone atom positions of CDR loops without relying on external tools or MSA, offering rapid predictions albeit with occasional non-physical results. IgFold[28] combines AntiBERTy[29] embeddings with a graph-based Transformer to rapidly predict antibody and nanobody structures. By utilizing pretrained language models for sequence representation, IgFold directly predicts backbone atomic coordinates, achieving accuracy comparable to AlphaFold2 with predictions completed in under 1 min. DeepAb[30] leverages a bidirectional LSTM pretrained model combined with Rosetta for structure optimization. This model operates in two stages: a deep residual convolutional network and a Rosetta-based fast structure design phase. DeepAb demonstrates improved performance for diverse antibody prediction tasks, delivering lower RMSD values for heavy and light chain frameworks compared to alternative methods. ESMFold[31], powered by ESM-2 embeddings, predicts protein structures directly from sequences without requiring MSAs. While it offers rapid folding capabilities, the high model dimensionality poses computational challenges. Similarly, EquiFold[32] adopts coarse-grained structural representations and avoids MSAs or protein language models, achieving faster prediction speeds for small proteins but occasionally producing non-physical results. OmegaFold[33] leverages protein sequence embeddings from large-scale language models to predict antibody and nanobody structures without relying on MSAs. And HelixFold-Single[34] is optimized for single-chain predictions, employing a streamlined architecture to predict structures efficiently and without the complexity of MSA-based methods.

Fig. 2 presents a comparative analysis of various AI methods in predicting the structure of antibodies and antigen−antibody complexes[25]. The upper panel evaluates the performance of these models in predicting antigen−antibody (IgG) complex structures, using DockQ and TM-Score metrics. AlphaFold-Multimer, RoseTTAFold2, tFold-Ag, and AlphaFold3 are assessed for their success rates, defined by DockQ scores exceeding 0.23. Among these, AlphaFold3 demonstrated the highest success rate (36.7%), outperforming AlphaFold-Multimer (18.2%) and tFold-Ag (28.3%). Notably, RoseTTAFold2 did not achieve any successful predictions in this benchmark. AlphaFold3's superior performance highlights its effectiveness in modeling antigen−antibody binding interfaces with improved precision. The lower panel compares AI models' performance in predicting monomeric antibody (IgG) and nanobody (VHH) structures, using root mean square deviation (RMSD) values for HCDR3 and CDR3 as indicators of structural accuracy. AlphaFold2, OmegaFold, and tFold-Ab achieved relatively low RMSD values, indicating accurate modeling of both antibody and nanobody structures. Conversely, DeepAb exhibited significantly higher RMSD values for the prediction of CDR3 of VHH, suggesting suboptimal accuracy in the structure prediction of highly variable region of the nanobody.

2.2. *Enhancing antigen−antibody binding with AI*

2.2.1. *Advantages of AI in affinity maturation compared to traditional methods*

Affinity maturation is a critical step in antibody drug development to enhance the binding strength between antibodies and antigens. Traditional methods, such as random mutagenesis, and directed evolution, rely on constructing mutant antibody libraries and screening for high-affinity variants. However, these approaches are time-consuming, costly, and have a relatively low success rate. In contrast, AI-driven affinity maturation offers several significant advantages:

(1) Efficiency

AI-based methods operate at inference speeds hundreds to thousands of times faster than traditional kinetic methods, enabling rapid evaluation of large-scale antibody mutants in virtual screening and multi-target antibody design. This drastically shortens the time required for affinity maturation. For example, the "Pro" series of protein language models developed by Hong Liang's research group[35] at Shanghai Jiao Tong University can efficiently design antibody candidate sequences with high affinity.

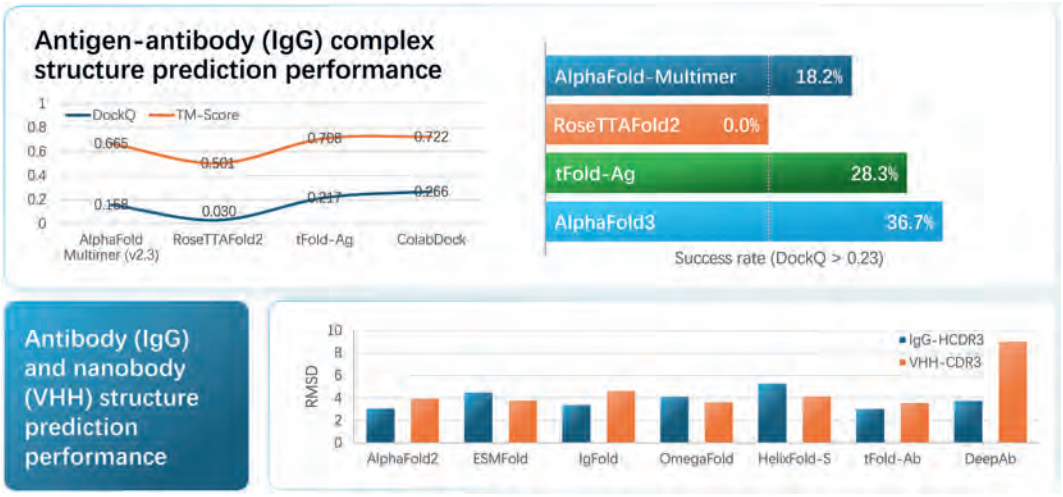

Figure 2 The performances comparison of AI methods on the task of predicting antibody structure. The upper panel assesses the predicted structure of antigen−antibody (IgG) complexes, while the lower panel compares the predictions for antibody monomers (IgG and VHH). For antigen-antibody complexes, DockQ and TM-Score were used to evaluate the predicted precision of binding interface. For antibody and nanobody, RMSD of HCDR3 or CDR3 was used to characterize the modelling performance.

In the case of VHH domains, Pro-PRIME identified high-affinity and highly alkali-resistant single-point mutations without requiring any prior experimental data.

(2) Accuracy

While general structure prediction models like AlphaFold2 are less sensitive to mutations, particularly for side chains at binding interfaces, innovative AI methods simulate detailed protein−protein interactions using multi-scale geometric information transmission. These methods construct multi-relation graphs for all heavy atoms at the interface and employ hierarchical message-passing schemes to capture the complexity of protein interaction interfaces. For instance, GearBind[36] leverages contrastive pretraining on large protein structure datasets to integrate critical structural insights into the model, enhancing the accuracy of antibody affinity prediction. Sequence-based affinity maturation models, though less reliable for ranking protein complex binding affinities, benefit significantly from incorporating structural data. This approach plays a pivotal role in constructing precise algorithms for modeling protein−protein interactions.

(3) Cost and time reduction

Traditional methods, reliant on mutagenesis, display, and screening experiments, often require 2−3 months or longer to complete, incurring substantial costs. AI-driven methods can economically and efficiently identify optimized mutants within a shorter time frame by leveraging algorithms to explore antibody−antigen structures and interactions. This reduces the number of required experiments and associated costs. Given the vast combination space of mutations in antibody complementarity-determining regions, purely experimental approaches are insufficient to cover all possibilities. AI-based simulations efficiently screen promising antibody mutants, minimizing experimental efforts.

For instance, in July 2024, researchers at Shanghai Jiao Tong University and Shanghai AI Laboratory developed Pro-FSFP[37], a few-shot learning algorithm integrating meta-learning, ranking learning, and parameter-efficient fine-tuning. Using just 20 experimental data points, Pro-FSFP successfully trained protein pre-trained models and significantly improved protein mutation-property prediction during the second experimental iteration. Similarly, researchers from Harvard Medical School introduced EVOLVEpro[38], which combines sequence embeddings from ESM2-15B with a random forest regression head and active learning strategies. Despite requiring over five rounds of wet-lab iterations, EVOLVEpro demonstrated effectiveness in directed protein evolution under low-data regimes. However, no direct benchmarking between EVOLVEpro and FSFP was conducted in their study. To address this gap, we compared these two few-shot learning methods under identical conditions in the present paper. Using just two rounds of active learning iterations, we found that FSFP outperformed EVOLVEpro in top-*k* positive rates for single-point mutations, the critical metric. This benchmark was conducted on the ProteinGym DMS dataset, where fitness values for top 15%, 10%, and 5% of mutations were used as wild-type cutoffs, FSFP consistently achieved higher positive mutation rates.

Table 1 provides a comparative evaluation of two few-shot protein engineering methodologies, FSFP and EVOLVEpro, in the context of three different initial single-point mutation selection strategies:

1. Random selection of 40 single-point mutations.
2. Top-40 mutations predicted by the ESM-1v model.
3. A combination of the top-20 ESM-1v-predicted mutations and 20 randomly selected mutations.

The results reveal that the highest positive rates were achieved when initial mutations were selected based on the ESM-1v top-40 predictions. Training on experimental data from the first round led to higher positive rates for both FSFP and EVOLVEpro in the second round, particularly when the initial data were randomly selected. This highlights the interplay between computational predictions and experimental data in optimizing antibody design pipelines.

2.2.2. Deep learning-based methods for antibody affinity maturation

2.2.2.1. Sequence based protein language models. Pretrained protein language models like ESM-1b[39] and ESM-1v[40] have made significant advancements in the field of antibody affinity

Table 1 Comparative analysis of two few-shot protein engineering methods, FSFP and EVOLVEpro, across three initial single-point mutation selection strategies.

Type	FSFP						
	First round				Second round		
Strategy	Top-40 (P = 15%)	Top-40 (P = 10%)	Top-40 (P = 5%)	Spearman	Top-40 (P = 15%)	Top-40 (P = 10%)	Top-40 (P = 5%)
Random 40	0.167	0.117	0.06	0.501	0.383	0.292	0.173
ESM-1v top-40	0.316	0.231	0.13	0.44	0.354	0.272	0.167
ESM-1v top-20 + random 20	0.248	0.178	0.102	0.468	0.36	0.273	0.166
Type	EvolvePro						
	First round				Second round		
Strategy	Top-40 (P = 15%)	Top-40 (P = 10%)	Top-40 (P = 5%)	Spearman	Top-40 (P = 15%)	Top-40 (P = 10%)	Top-40 (P = 5%)
Random 40	0.167	0.117	0.06	0.252	0.313	0.231	0.133
ESM-1v top-40	0.316	0.231	0.13	0.064	0.236	0.172	0.096
ESM-1v top-20 + random 20	0.248	0.178	0.102	0.197	0.288	0.216	0.123

maturation. These models are trained on large-scale, non-redundant protein sequence datasets, enabling them to learn complex patterns and evolutionary rules within protein sequences. By predicting possible amino acid substitutions that may occur in natural proteins, these models offer guidance for laboratory-based protein evolution. In practical applications, such models not only improve antibody binding affinity but also provide efficient evolutionary guidance under various protein family contexts and selective pressures, such as antibiotic resistance and enzyme activity. For example, using the ESM-1v model, researchers achieved a sevenfold increase in binding affinity for four clinically relevant, highly evolved antibodies, and a 160-fold improvement for three underdeveloped antibodies, all with just 20 or fewer variants selected in two rounds of laboratory evolution. Protein generation models such as ProGen[41] and AbGPT[42] offer a distinct paradigm for *de novo* antibody sequence design. ProGen, a transformer-based framework developed by Salesforce Research, employs autoregressive pretraining on diverse protein families to enable conditional generation of functional sequences. Unlike masked language modeling (*e.g.*, ESM), ProGen iteratively predicts residues in a left-to-right manner, making it particularly suited for generating novel antibody scaffolds while preserving structural viability. These results highlight that even without specific task-related training data or prior antigen knowledge, AI models can effectively guide antibody evolution and enhance binding affinity[43].

2.2.2.2. Models incorporating structural information. The ESM-IF1[44] represents an extension of AI applications in antibody drug development by integrating three-dimensional structural information. This model enables more accurate predictions regarding how amino acid substitutions affect protein function, activity, and evolvability. Although trained solely on single-chain structures, it can be extended to engineered protein complexes, including antibody−antigen interactions. In practical use, researchers employed ESM-IF1 to mature the affinity of two therapeutic clinical antibodies targeting SARS-CoV-2. The results demonstrated up to a 25-fold enhancement in neutralizing capacity and a 37-fold improvement in binding affinity against escape variants BQ.1.1 and XBB.1.5[45].

The GeoPPI[46] integrates graph attention networks and gradient-boosted trees to learn features from large-scale protein structure and sequence datasets. It predicts the impact of amino acid substitutions on binding free energy (ΔG), offering an effective strategy for antibody affinity maturation. By analyzing how mutations influence binding free energy, GeoPPI assists researchers in selecting mutations with higher potential for affinity improvement, making it a valuable tool for optimizing antibody−antigen interactions.

Shan and colleagues[47] utilized a Transformer architecture to learn key residue pair information from large datasets of SARS-CoV-2 variants. This method predicts how mutations impact the binding affinity between SARS-CoV-2 variants and antibodies. It provides insights into the evolutionary trends of SARS-CoV-2 variants and aids in the development of antibodies targeting new variants. This approach proves particularly valuable in the rapidly evolving landscape of infectious disease, offering a framework for adapting antibody therapeutics. GearBind[36] is a pre-trained geometric graph neural network that constructs multi-relation graphs, utilizes geometric message-passing techniques, and pre-trains on large-scale unlabeled protein data. By simulating fine-grained protein−protein interactions, GearBind predicts binding free energy changes. The model builds multi-relation graphs for all heavy atoms at protein interfaces and uses a multi-level message-passing approach. It incorporates a contrastive learning pretraining algorithm to enhance model performance. In antibody affinity maturation projects, GearBind has demonstrated exceptional performance, effectively designing antibodies with enhanced binding affinity. This makes it a powerful tool for optimizing antibody−antigen interactions and guiding the design of high-affinity antibodies. These deep learning-based methods, with their ability to integrate structural and sequence data, significantly enhance the precision and efficiency of antibody affinity maturation, offering powerful tools for the design of next-generation antibody therapeutics.

2.3. AI in assessing antibody developability

2.3.1. Antibody developability

As biological drugs, the stability of antibodies is crucial. Instability in antibodies can lead to loss of activity during storage and transport, ultimately impacting drug efficacy. For example, in long-term stability studies, we observed that antibody stability varied under different temperatures. Under accelerated conditions at 40 °C, some antibodies showed significant molecular aggregation and fragmentation, while full-length monoclonal antibodies (mAbs) performed relatively better.

Aggregation propensity is also a critical issue. Antibody aggregation can lead to precipitation, shorten shelf life, and increase immunogenicity when aggregates form *in vivo*. In biophysical stability studies, we assessed the colloidal stability of antibodies using various methods and found that different antibody forms exhibited varying self-interaction parameters (K_d), with substantial variability observed within the same antibody species.

Immunogenicity can reduce therapeutic efficacy and pose safety risks by triggering adverse drug reactions. Since therapeutic antibodies often originate from non-human species, they may be recognized as foreign antigens by the immune system, eliciting immunogenic responses. Predicting immunogenicity is therefore essential for assessing the clinical safety and effectiveness of antibody drugs. Chemical degradation, such as deamidation and oxidation, may affect the structure and function of antibodies. These chemical processes can lead to a decrease in antibody activity, making them critical factors in assessing antibody stability and efficacy.

2.3.2. AI platforms for antibody developability assessment

Antibody developability can be predicted based on the physicochemical properties of its amino acid sequence, including hydrophobicity, electrostatic charge, and topological interactions. Advanced platforms and models have been developed to assess and optimize the developability of therapeutic antibodies, mainly focusing on five key aspects: expression, thermal stability, aggregation, solubility, and immunogenicity (Fig. 3). For expression and thermal stability, the methods such as ProSST[48], SaProt[49], ESM-IF1[44] and ProtSSN[50] in ProteinGym[51] benchmark show superior performance on zero-shot prediction of protein expression and stability. Tools such as ProteinMPNN[52] analyze sequence properties and structural features to predict and enhance the manufacturability and stability of therapeutic antibodies.

Aggregation propensity is closely related to hydrophobicity, solubility, viscosity, self-interaction, and protein stability. Aggregation can lead to precipitation, shorten storage life, and increase the immunogenicity of drugs. To predict protein solubility and

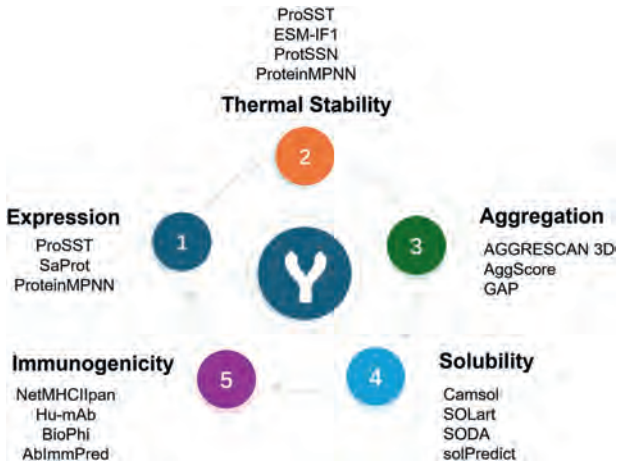

Figure 3 The main indicators for assessing antibody developability, as well as examples of related AI prediction methods.

aggregation propensity, various tools and machine learning models have been developed, such as Camsol[53], SOLart[54], and Aggrescan3D (A3D)[55], AggScore[56], and GAP[57]. These models utilize antibody sequence or structural information to predict aggregation tendencies, guiding the engineering optimization of antibody drug candidates. Antibody solubility plays a vital role in drug formulation and delivery. Predictive tools like Camsol[53], SOLart[54], SODA[58], and solPredict[59] estimate solubility by leveraging sequence or structural features, providing valuable insights for modifications that reduce precipitation risks and enhance solubility. Finally, immunogenicity prediction is crucial for therapeutic safety and efficacy, as antibodies derived from non-human species can trigger immune responses. AI-based tools such as NetMHCIIpan[60], Hu-mAb[61], BioPhi[62], and AbImmPred[63] identify B and T cell epitopes and classify sequences for humanization, reducing immunogenicity risks. More prediction tools are provided in IEDB analysis resource[64] for immunogenicity prediction. Together, these computational approaches provide a comprehensive framework for optimizing antibody properties, enabling the development of safe, effective, and manufacturable therapeutic candidates.

In addition to these methods for handling individual tasks, integrative AI platforms may be more attractive for users. One notable example is the therapeutic antibody profile (TAP) model, which evaluates developability by assessing critical features such as CDR lengths, surface hydrophobicity, charge distribution in CDRs, and heavy-light chain charge asymmetry[65]. TAP assigns risk levels to these attributes, helping to filter out candidates with poor developability. Genscript Biotech has developed an AI-based antibody optimization platform that supports *in vitro* developability assessments[66]. By using the TAP tool, researchers can calculate and compare biophysical properties to clinically validated molecules, facilitating the selection of candidates with higher stability and manufacturability potential. Similarly, Chen et al.[67] utilized the SAbDab database to train machine learning models for predicting antibody developability, showcasing the utility of data-driven approaches in optimizing antibody design.

2.4. *AI-driven de novo antibody design*

David Baker's lab has made significant advancements in AI-driven *de novo* antibody design. They have utilized a fine-tuned version of RFdiffusion[68], a neural network-based technique akin to the models used in image generation AI. The network was fine-tuned by training on thousands of experimentally determined antibody structures, as well as other real examples of antibody interactions. The fine-tuned RFdiffusion can design antibody structures that closely match the input framework and are capable of aligning and binding to the provided molecular targets.

Using this technology, the lab has designed thousands of antibodies targeting specific regions of bacterial and viral proteins, including those used by SARS-CoV-2 and the flu virus to invade cells, as well as cancer drug targets. Some of the designed nanobodies were synthesized and tested in the lab, with approximately 1% of the antibody designs meeting the desired performance, exhibiting medium affinity. Cryo-electron microscopy (cryo-EM) was used to assess the accuracy of the designs, and the structures of VHHs (nanobodies) designed to bind influenza hemagglutinin were found to closely match the designed models, with nearly identical configurations of the CDR loops and overall binding sites[69].

In addition to the work of David Baker's team, other research groups have also made significant progress in AI-driven *de novo* antibody design. For example, research teams have utilized PALM-H3[70] to generate antibodies with the ability to bind to SARS-CoV-2 antigens, including emerging XBB variants. This was validated through both *in-silico* analysis and *in-vitro* experiments, demonstrating the potential of AI-driven antibody design to target new viral strains effectively. Google DeepMind's AlphaProteo [71] algorithm is capable of generating high-affinity binding molecules without the need for multiple rounds of experimental testing. AlphaProteo includes a generative model and a filtering mechanism that can predict designs likely to succeed in experiments. And Bruno E. Correia et al.[72] developed BindCraft, an automated pipeline that uses AlphaFold2's deep learning weights to generate nanomolar-level binders. Although the results of antibody *de novo* design have not been mentioned, its binder generation technology should be extended to antibody binder design tasks. And this approach eliminates the need for high-throughput screening or experimental optimization, streamlining the antibody design process. For example, recently Nabla Bio released a technical report representing its ability to *de novo* design epitope-specific antibodies against soluble and multi-pass membrane proteins with high specificity, developability, and function[73]. They present JAM, a generative protein design system that enables fully computational design of antibodies with therapeutic-grade properties for the first time. Fig. 4 shows the success rate reported in their studies (except the result of IL-9Rα was obtained from Luis Serrano's research[74]. It is a significant breakthrough that the success rate of *de novo* design for VHHs has approached that of protein binders. However, the success rate for *de novo* design of complete antibodies, such as single-chain variable fragments (ScFv), remains exceptionally low (0.0017%). This highlights a critical limitation in current technologies, suggesting that substantial advancements in computational methods, structural modeling, or experimental validation will be necessary to overcome these challenges. Future innovations in AI-driven antibody design, combined with enhanced understanding of antibody structures and interactions, are expected to play a pivotal role in improving the success rate of full antibody design.

2.5. *Challenges of AI-driven antibody design*

The significance of AI-driven antibody design lies in its ability to rapidly and efficiently optimize and generate antibodies with specific functions, offering enormous potential for the treatment of

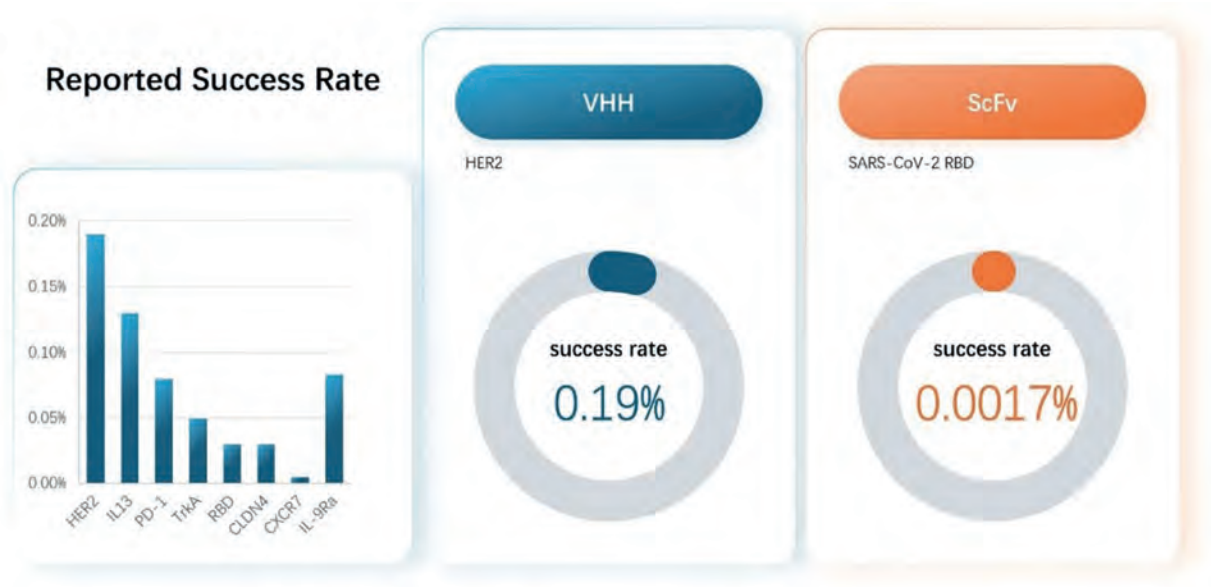

Figure 4 Reported success rate for the task of *de novo* design antibody, including VHH and ScFvs. The bar chart on the left shows the experimental success rate of VHH binder designed, and the chart on the right shows the highest reported success rate of VHH and ScFv designed respectively.

various diseases. This approach can dramatically accelerate the development of antibody therapeutics, providing more targeted and personalized treatments. However, there are still several challenges in this field.

2.5.1. Dynamic property screening limitations

AI-based design approaches have not yet established a systematic screening strategy for the dynamic properties of antibodies. A study have shown that the CDRH3 region of antibodies may adopt different conformational states under varying pH conditions, which can significantly impact their function[75]. For example, in the highly acidic tumor microenvironment[76], pH-dependent conformational changes can serve as a "switch" mechanism, enhancing antibody targeting specificity and reducing off-target effects. However, the current datasets lack sufficient dynamic structural information, making it challenging for AI tools to effectively learn and accurately predict these dynamic properties, thereby limiting their application in optimizing antibody function.

2.5.2. CDRH3 conformational modeling

Accurate modeling of the CDRH3 loop remains a significant challenge. The CDRH3 loop exhibits high flexibility and can adopt multiple stable conformational states[77]. This conformational variability is not solely sequence-dependent but is also influenced by external factors such as experimental conditions, making precise structural prediction difficult[78]. Furthermore, existing structural data are predominantly derived from antibody−antigen complexes rather than free antibody structures, leading to a lack of comprehensive characterization of the unbound state. This data bias further increases the complexity of CDRH3 modeling.

2.5.3. Post-translational modification (PTM) effects

Another key challenge for current AI models in antibody design is the accurate modeling of post-translational modifications (PTMs), which play a crucial role in antibody structure, stability, immune interactions, and antigen-binding properties[79]. Common PTMs, such as oxidation, deamidation, acetylation, cyclization, and glycosylation, can induce significant molecular changes. For instance, methionine oxidation may alter the overall conformation and solution stability of an antibody[79], asparagine (Asn) deamidation can lead to the formation of charge isomers[80], and tryptophan (Trp) oxidation may modify antibody hydrophobicity or hydrophilicity[81]. Isomerization within the Fab region can compromise conformational stability[82]. Additionally, glycosylation at specific sites within the CDR loops may directly decrease antigen-binding affinity, thereby impacting antibody function[83]. However, current AI prediction models are primarily trained on structural data from unmodified antibodies, lacking a systematic approach to account for the effects of PTMs. Moreover, since structural characterization of antibody−antigen complexes often does not fully consider PTMs, AI models are further limited in their ability to predict how these modifications influence antibody function.

2.5.4. Genetic diversity and SNP-driven variability

The impact of single nucleotide polymorphisms (SNPs) introduces an additional layer of complexity. Population-level amino acid variations in target proteins (*e.g.*, viral envelope proteins, cytokine receptors) may alter epitope conformations, potentially reducing antibody efficacy across genetically diverse cohorts. Current AI frameworks often prioritize binding to reference sequences, neglecting clinically prevalent SNP-derived variants. For example, rsID-documented mutations with $>1\%$ allele frequency in key ethnic groups (European, African, East Asian) could destabilize paratope-epitope interactions. While emerging strategies—such as integrating gnomAD[84] variant frequencies during epitope prioritization and multi-objective optimization against SNP libraries—show promise, most models remain insufficiently trained on population-scale variant data.

2.5.5. Spatiotemporal complexities of molecular interactions

The reductionist assumption of static binding interfaces fundamentally limits current AI approaches. Two critical dimensions demand urgent attention:

- Temporal dynamics: transient interaction states—such as GPCR-arrestin complex conformations persisting for ~50 ms—may govern antibody accessibility to cryptic epitopes[85]. Time-resolved cryo-EM studies reveal that many of therapeutic targets exhibit functionally relevant conformational fluctuations at microsecond timescales[86], yet AI models predominantly utilize static crystallographic snapshots.
- Spatial compartmentalization: subcellular localization imposes physical constraints on antibody-target engagement. Nuclear membrane-restricted mRNA−protein interactions, for instance, remain inaccessible to conventional IgG antibodies, necessitating engineered delivery systems absent from current AI pipelines.

Emerging solutions employ multiscale graph temporal networks that integrate:

a) Molecular dynamics-derived B-factors to model residue flexibility
b) Spatial transcriptomics data to predict target accessibility across cellular compartments

Such frameworks enable probabilistic predictions of epitope availability across spatiotemporal contexts, though systematic validation in physiologically relevant environments remains challenging.

2.6. *Future theoretical directions*

As AI technology continues to evolve, it is expected that these challenges will gradually be overcome. With improvements in machine learning models, structural prediction algorithms, and molecular simulation techniques, AI will play an increasingly prominent role in the design and development of antibody drugs. Specifically, the integration of population genetics databases (*e.g.*, TOPMed), SNP-aware docking protocols, and ethnic group-specific validation pipelines could enhance the robustness of AI-designed antibodies in genetically diverse populations. This will lead to more efficient and targeted treatments for a wide range of diseases, expanding the potential of antibody therapeutics and revolutionizing the drug development process.

3. Aptamers

Transcription factors play a crucial role in regulating cellular fate, and their dysregulation can lead to various diseases such as cancer, infectious diseases, and neurodegenerative disorders[87,88]. However, the development of small molecule chemical inhibitors targeting transcription factors has faced significant challenges due to the lack of ideal binding pockets for most transcription factor interactions[89]. Consequently, exploring novel regulators that can target intracellular transcription factor interactions has become a hot research topic.

Aptamers, which are single-stranded DNA (deoxyribonucleic acid) or RNA (ribonucleic acid) molecules, fold into three-dimensional structures to bind specifically to target proteins, earning them the nickname "chemical antibodies"[9]. The discovery that nucleic acids (DNA or RNA) can act as ligands to regulate protein activity originated from viral research. Specifically, molecular biological studies on the human immunodeficiency virus (HIV) and adenoviruses revealed that these viruses encode small structured RNAs that bind to host proteins[90]. These RNA ligands can either promote viral replication or alleviate the host's antiviral responses. In 1990, Sullenger et al.[91] published a study demonstrating that an RNA aptamer designed to bind viral proteins could prevent the binding of viral RNA to proteins, thereby inhibiting viral replication. The TAR RNA (*trans*-activation response element, a short RNA sequence synthesized by the HIV virus itself) derived aptamer, expressed as a decoy, binds to the Tat and Cyclin T1 proteins in $CD4^+$ T cells, conferring resistance to viral replication in cells that express the TAR aptamer, suggesting the therapeutic potential of RNA aptamers[92]. RNA aptamers, due to their significant specificity and high affinity for a wide range of targets—including proteins, sugars, lipids, and cellular metabolites—have shown enormous potential in therapeutic, diagnostic, imaging, and drug delivery applications. Due to their large molecular size, broad binding capabilities on protein surfaces (without the need for a "classic drug pocket"), and effective cellular internalization, aptamers have become important tools for modulating transcription factor interactions. Since the invention of the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) method[93], numerous aptamers have been selected to target ions, small molecules, proteins, tissues, whole cells, and even bacteria, viruses, and parasites. The high affinity and selectivity of aptamers can rival the characteristics of antibodies. Moreover, aptamers offer distinct advantages over antibodies in terms of thermal and chemical stability, cost-effectiveness, reproducibility, ease of generation, and modification. These unique characteristics have led to the growing potential of aptamer-based technologies in medical diagnostics and treatments, particularly in the field of infectious diseases, over the past few decades.

3.1. *RNA structure prediction*

An essential challenge in the computational design of RNA aptamers is the prediction of RNA's three-dimensional (3D) structure. Determining the 3D shape of biomolecules is one of the most difficult problems in modern biology and medical discoveries. Many companies and research institutions spend millions of dollars trying to determine molecular structures, often with limited success. Recently, the use of AI to predict molecular structures has become a hot research topic, demonstrated by the success of DeepMind's AlphaFold2 and AlphaFold3 in protein structure prediction. However, achieving accurate molecular structure prediction requires vast datasets.

For RNA molecules, their 3D structure is crucial for their biological function, and this is especially significant in the discovery of RNA aptamer drugs. However, known RNA structures are still limited, and predicting RNA structure computationally remains a challenging task. In 2021, a research team from Stanford University addressed this problem using machine learning techniques. Remarkably, the researchers used only 18 known RNA structures for training and were able to predict accurate structural models without any prior knowledge of the specific properties of those structures.

The key method used in this breakthrough was a machine learning approach that developed a scoring function known as Atomic Rotationally Equivariant Scorer (ARES), which significantly outperformed previous methods[94]. In order to train ARES, the team used 18 RNA molecules published between 1994 and 2006, and applied the Rosetta FARFAR2 sampling method to generate 1000 structural models for each RNA molecule without

using any known structure. The novelty of this work lies in the fact that the authors proposed an effective neural network architecture for training scoring functions for large RNA molecular systems and applied it to RNA structure evaluation. By combining the high-precision scoring function with specific RNA structural generation or sampling methods, they were able to achieve accurate predictions of RNA's 3D structure.

The training goal was to assess the difference between a given RNA structure and its optimal structure in terms of RMSD values. This was achieved by generating thousands of RNA structures from a small set of known RNA structures, creating corresponding RMSD values for training purposes. As a result, the model training became feasible. Fig. 5 illustrates the architecture of ARES, where the input consists of atomic coordinates and element types. ARES ensures the translation and rotation equivariance of the input molecules through an equivariant geometric neural network. Additionally, the feature encoding of each atom includes information about its neighboring atoms. The network ultimately predicts the RMSD between the predicted structure and the actual structure.

This work represents a significant advancement in RNA 3D structure prediction, opening new avenues for the computational design of RNA-based therapeutics, including aptamers.

The ARES method, developed by Stanford University, achieved state-of-the-art (SOTA) results in the RNA Puzzles challenge[95], a blind RNA structure prediction competition. One of the method's key strengths is its ability to effectively learn from a limited dataset, overcoming the typical limitations of standard deep neural networks. Additionally, ARES uses only atomic coordinates as input, without relying on specific RNA sequence information, making it applicable to a wide range of problems in structural biology, chemistry, and materials science.

In recent developments, a collaborative effort from teams at The Chinese University of Hong Kong, Fudan University, Harvard University, and Zelixir Biotech introduced RhoFold+[96], a deep learning method for rapid and accurate *de novo* RNA 3D structure prediction. Trained on approximately 23.7 million RNA sequences, this method addresses the challenge of limited data availability and offers a fully automated, end-to-end pipeline for RNA structure prediction. RhoFold + demonstrates high accuracy in predicting single-stranded RNA structures and excels in generalizing across different RNA families and types. It captures local features such as inter-helical angles and secondary structures, improving the overall quality of structure predictions. Retrospective evaluations on RNA-Puzzles[97] and CASP15 RNA targets show that RhoFold + outperforms existing methods, including ARES and human expert groups.

The training data for RhoFold + consists of RNA 3D structures curated from the BGSU Representative RNA Structure Set (2022-04-13), sourced from the Protein Data Bank (PDB). Redundancy was reduced by clustering RNA sequences based on an 80% sequence similarity threshold using CD-HIT, resulting in 782 unique RNA sequence clusters. The pipeline incorporates the RNA-FM model, a large RNA language model, to transform RNA sequences into embeddings that capture evolutionary and structural knowledge. These embeddings, along with multiple sequence alignment (MSA) features, are fed into the Rhoformer transformer network, which iteratively refines the structure. The method uses a geometric-aware attention mechanism, similar to that in AlphaFold2, combined with an Invariant Point Attention (IPA) module, to optimize the local backbone coordinates and torsion angles of key RNA atoms. During the reconstruction of full-atom coordinates, secondary structure and base-pairing constraints ensure that the generated structures are biologically and physically realistic.

RhoFold + offers a fully automated and highly accurate approach for RNA 3D structure prediction, eliminating the need for manual interventions or high-throughput experimental optimizations. With its advanced integration of RNA sequence data, evolutionary knowledge, and geometric constraints, RhoFold + sets a new benchmark in the field of RNA structural biology and holds promise for applications in RNA-based drug discovery and therapeutic development.

3.2. *De novo design of RNA aptamers*

AI has emerged as a powerful tool for the *de novo* design of RNA aptamers, which are short, single-stranded nucleic acids that can bind to specific molecular targets. Leveraging advanced RNA structure prediction algorithms, researchers have developed AI-driven platforms that can generate novel RNA aptamers with desired structural and functional properties.

One notable example is the RhoDesign platform developed by a team from MIT[98]. This platform integrates AI techniques with RNA structure prediction to design RNA aptamers from scratch. RhoDesign builds upon the RhoFold model, which has shown promising performance in predicting RNA structures and outperforms other models in various benchmarks. The platform utilizes a deep learning approach that can design RNA aptamers with structurally similar but sequence-different counterparts, capable of fluorescing in the presence of small molecules, thereby enabling high-throughput screening.

To develop RhoDesign, the researchers first curated a dataset of 369,499 RNA 3D structures predicted by RhoFold and 3435

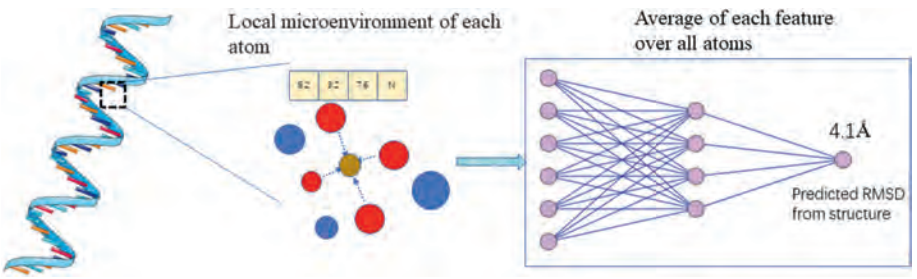

Figure 5 The architecture of ARES. The ARES model processes atomic coordinates and element types as input, ensuring translation and rotation equivariance through an equivariant geometric neural network. Each atom's features incorporate information from neighboring atoms. The model predicts the root mean square deviation (RMSD) between the predicted and actual molecular structures.

experimentally determined RNA structures from the Protein Data Bank (PDB). These structures, along with their corresponding sequences, were used to train the deep learning model. Additional secondary structure predictions, either computationally generated or experimentally determined, were incorporated to improve the model's accuracy.

After training, the model was evaluated using metrics such as recovery rate, template modeling (TM) score, RMSD, and perplexity. In benchmark comparisons, RhoDesign outperformed several existing RNA aptamer design platforms, including LEARNA, *Meta*-LEARNA, RiboLogic, gRNAde, RDesign, and eM2dRNAs.

The development of RhoDesign represents a significant step forward in RNA aptamer design. It provides a powerful computational tool for designing RNA sequences with specific structural characteristics, offering advantages over traditional methods. The integration of RhoDesign with other RNA design platforms is expected to further enhance the accuracy and efficiency of RNA aptamer development. Moving forward, the platform will benefit from the use of more reliable RNA 3D structure prediction methods and additional experimental validation with diverse RNA targets.

3.3. Challenges of AI-driven aptamer design

The function of RNA aptamers relies on their precise structure. However, accurately predicting RNA secondary and tertiary structures based solely on sequences remains a significant challenge. A major limitation is the scarcity of available RNA structural data for training AI models. Currently, the Protein Data Bank (PDB) contains 227,222 registered protein structures, whereas only 4602 nucleic acid structures are available (February 2025). This data insufficiency severely constrains the ability of AI to effectively model RNA structures.

Furthermore, natural RNA aptamers are prone to degradation by nucleases *in vivo*, necessitating specific chemical modifications (such as 2′-*O*-methylation and phosphorothioate modifications) to enhance their stability[99]. However, existing AI training datasets are predominantly based on unmodified RNA, lacking comprehensive information on how different modifications impact aptamer structure and function[100]. This limitation hinders AI-driven optimization of RNA aptamer stability and its broader applicability in therapeutic and diagnostic contexts.

4. Conclusions

The application of AI in both antibody and aptamer design has marked a transformative shift in the landscape of molecular therapeutics. AI technologies, particularly deep learning and neural networks, have revolutionized the way we approach the design of therapeutic molecules, enabling more efficient, precise, and scalable solutions than ever before.

In the realm of antibody design, AI-driven methods, such as deep sequence-structure mapping and generative models, have significantly advanced the ability to predict and engineer antibodies with optimized binding affinity, specificity, and stability. Tools like ESM-1v, GeoPPI, and RFdiffusion have demonstrated their potential to guide antibody maturation and predict the effects of mutations on binding properties. These technologies have not only accelerated the development of antibodies for existing targets but also enabled the design of antibodies for challenging new targets, including viral infections (such as SARS-CoV-2) and cancer-related antigens. The integration of structural information into AI models has further improved the precision of antibody design, offering a promising path for next-generation therapeutics.

Similarly, the field of RNA aptamer design has been transformed by AI, with platforms like RhoDesign showing great promise in the *de novo* design of functional RNA molecules. AI methods have overcome traditional challenges in RNA design, such as predicting 3D structures from sequence and optimizing binding affinities. By leveraging large datasets and advanced deep learning algorithms, AI has enabled the rapid generation of aptamers with desired properties, including high specificity, stability, and in some cases, fluorescence for detection purposes. This has vast implications for drug development, diagnostics, and therapeutic applications, particularly in targeting complex biological molecules or pathways where traditional small molecules and antibodies face limitations.

Together, AI-driven approaches to antibody and aptamer design are setting the stage for a new era of precision medicine. These technologies not only offer faster, more cost-effective solutions for therapeutic development but also enhance the capacity to design molecules that are more tailored to specific disease targets. As AI models continue to evolve, their integration with experimental validation will further refine their accuracy, leading to the discovery of highly effective, novel drugs and diagnostic tools.

In conclusion, AI is poised to significantly accelerate the pace of innovation in antibody and aptamer design, with the potential to revolutionize the development of next-generation therapeutics. By enabling the creation of molecules with improved efficacy, specificity, and safety profiles, AI is becoming an indispensable tool in the fight against a wide range of diseases, including cancer, viral infections, and autoimmune disorders. As research and technology advance, the synergy between AI and molecular biology will continue to unlock new possibilities for the treatment and prevention of disease.

Acknowledgments

This work was supported by Science and Technology Innovation Key R&D Program of Chongqing (CSTB2022TIAD-STX0017, CSTB2024TIAD-STX0032, China), the Computational Biology Key Program of Shanghai Science and Technology Commission (23JS1400600, China), Shanghai Jiao Tong University Scientific and Technological Innovation Funds (21X010200843, China), and the Student Innovation Center at Shanghai Jiao Tong University, and Shanghai Artificial Intelligence Laboratory.

Author contributions

Pan Tan: Writing-review & editing, Writing-original draft, Visualization, Investigation, Data curation, Conceptualization. Song Li: Writing-review & editing, Writing-original draft, Visualization, Investigation. Jin Huang: Writing-review & editing, Writing-original draft. Ziyi Zhou: Writing-review & editing. Liang Hong: Supervision.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Torchilin V. Tumor delivery of macromolecular drugs based on the EPR effect. *Adv Drug Deliv Rev* 2011;**63**:131−5.
2. Li J, Zhu Z. Research and development of next generation of antibody-based therapeutics. *Acta Pharmacol Sin* 2010;**31**:1198−207.
3. Singh S, Tank NK, Dwiwedi P, Charan J, Kaur R, Sidhu P, et al. Monoclonal antibodies: a review. *Curr Clin Pharmacol* 2018;**13**: 85−99.
4. Su Z, Xiao D, Xie F, Liu L, Wang Y, Fan S, et al. Antibody−drug conjugates: recent advances in linker chemistry. *Acta Pharm Sin B* 2021;**11**:3889−907.
5. Wang L, Yin H, Jiang J, Li Q, Gao C, Li W, et al. A rationally designed CD19 monoclonal antibody−triptolide conjugate for the treatment of systemic lupus erythematosus. *Acta Pharm Sin B* 2024; **14**:4560−76.
6. Tiller KE, Tessier PM. Advances in antibody design. *Annu Rev Biomed Eng* 2015;**17**:191−216.
7. He XH, Li JR, Xu J, Shan H, Shen SY, Gao SH, et al. AI-driven antibody design with generative diffusion models: current insights and future directions. *Acta Pharmacol Sin* 2025;**46**:565−74.
8. Rimmele M. Nucleic acid aptamers as tools and drugs: recent developments. *Chembiochem* 2003;**4**:963−71.
9. Zhou G, Wilson G, Hebbard L, Duan W, Liddle C, George J, et al. Aptamers: a promising chemical antibody for cancer therapy. *Oncotarget* 2016;**7**:13446−63.
10. Patel DJ. Structural analysis of nucleic acid aptamers. *Curr Opin Chem Biol* 1997;**1**:32−46.
11. Byun J. Recent progress and opportunities for nucleic acid aptamers. *Life* 2021;**11**:193.
12. Schoukroun-Barnes LR, Macazo FC, Gutierrez B, Lottermoser J, Liu J, White RJ. Reagentless, structure-switching, electrochemical aptamer-based sensors. *Annu Rev Anal Chem* 2016;**9**:163−81.
13. He J, Duan Q, Ran C, Fu T, Liu Y, Tan W. Recent progress of aptamer−drug conjugates in cancer therapy. *Acta Pharm Sin B* 2023; **13**:1358−70.
14. Wu L, Wang Y, Xu X, Liu Y, Lin B, Zhang M, et al. Aptamer-based detection of circulating targets for precision medicine. *Chem Rev* 2021;**121**:12035−105.
15. Röthlisberger P, Gasse C, Hollenstein M. Nucleic acid aptamers: emerging applications in medical imaging, nanotechnology, neurosciences, and drug delivery. *Int J Mol Sci* 2017;**18**:2430.
16. Li B, Wei J, Di C, Lu Z, Qi F, Zhang Y, et al. Molecularly engineered truncated tissue factor with therapeutic aptamers for tumor-targeted delivery and vascular infarction. *Acta Pharm Sin B* 2021;**11**:2059−69.
17. Lorenz C, von Pelchrzim F, Schroeder R. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc* 2006;**1**:2204−12.
18. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583−9.
19. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;**630**:493−500.
20. Krishna R, Wang J, Ahern W, Sturmfels P, Venkatesh P, Kalvet I, et al. Generalized biomolecular modeling and design with RoseTTAFold all-atom. *Science* 2024;**384**:eadl2528.
21. Baek M, McHugh R, Anishchenko I, Jiang H, Baker D, DiMaio F. Accurate prediction of protein−nucleic acid complexes using RoseTTAFoldNA. *Nat Methods* 2024;**21**:117−21.
22. Das R, Kretsch RC, Simpkin AJ, Mulvaney T, Pham P, Rangan R, et al. Assessment of three-dimensional RNA structure prediction in CASP15. *bioRxiv* 2023;**91**:1747−70.
23. Abanades B, Wong WK, Boyles F, Georges G, Bujotzek A, Deane CM. ImmuneBuilder: deep-learning models for predicting the structures of immune proteins. *Commun Biol* 2023;**6**:575.
24. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021. 2021.10.04.463034.
25. Wu J, Wu F, Jiang B, Liu W, Zhao P. tFold-Ab: fast and accurate antibody structure prediction without sequence homologs. *bioRxiv* 2022. 2022.11.10.515918.
26. Feng S, Chen Z, Zhang C, Xie Y, Ovchinnikov S, Gao YQ, et al. Integrated structure prediction of protein−protein docking with experimental restraints using ColabDock. *Nat Mach Intell* 2024;**6**: 924−35.
27. Abanades B, Georges G, Bujotzek A, Deane CM. ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation. *Bioinformatics* 2022;**38**:1877−80.
28. Ruffolo JA, Chu LS, Mahajan SP, Gray JJ. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nat Commun* 2023;**14**:2389.
29. Ruffolo JA, Gray JJ, Sulam J. Deciphering antibody affinity maturation with language models and weakly supervised learning. *arXiv preprint arXiv:211207782* 2021. Available from: https://doi.org/10.48550/arXiv.2112.07782.
30. Ruffolo JA, Sulam J, Gray JJ. Antibody structure prediction using interpretable deep learning. *Patterns* 2022;**3**:100406.
31. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;**379**:1123−30.
32. Lee JH, Yadollahpour P, Watkins A, Frey NC, Leaver-Fay A, Ra S, et al. EquiFold: protein structure prediction with a novel coarse-grained structure representation. *bioRxiv* 2022. 2022.10.07.511322.
33. Wu R, Ding F, Wang R, Shen R, Zhang X, Luo S, et al. High-resolution *de novo* structure prediction from primary sequence. *bioRxiv* 2022. Available from: https://doi.org/10.1101/2022.07.21.500999.
34. Fang X, Wang F, Liu L, He J, Lin D, Xiang Y, et al. A method for multiple-sequence-alignment-free protein structure prediction using a protein language model. *Nat Mach Intell* 2023;**5**:1087−96.
35. Jiang F, Li M, Dong J, Yu Y, Sun X, Wu B, et al. A general temperature-guided language model to design proteins of enhanced stability and activity. *Sci Adv* 2024;**10**:eadr2641.
36. Cai H, Zhang Z, Wang M, Zhong B, Li Q, Zhong Y, et al. Pre-trainable geometric graph neural network for antibody affinity maturation. *Nat Commun* 2024;**15**:7785.
37. Zhou Z, Zhang L, Yu Y, Wu B, Li M, Hong L, et al. Enhancing efficiency of protein language models with minimal wet-lab data through few-shot learning. *Nat Commun* 2024;**15**:5566.
38. Jiang K, Yan Z, Di Bernardo M, Sgrizzi SR, Villiger L, Kayabolen A, et al. Rapid *in silico* directed evolution by a protein language model with EVOLVEpro. *Science* 2024;**387**:eadr6006.
39. Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc Natl Acda Sci* 2021;**118**: e2016239118.
40. Meier J, Rao R, Verkuil R, Liu J, Sercu T, Rives A. Language models enable zero-shot prediction of the effects of mutations on protein function. *bioRxiv* 2021. 2021.07.09.450648.
41. Madani A, McCann B, Naik N, Keskar NS, Anand N, Eguchi RR, et al. ProGen: language modeling for protein generation. *bioRxiv* 2020. Available from: https://doi.org/10.48550/arXiv.2004.03497.
42. Kuan D, Farimani AB. AbGPT: *de novo* antibody design *via* generative language modeling. *arXiv* 2024. Available from: https://doi.org/10.48550/arXiv.2409.06090.
43. Hie BL, Shanker VR, Xu D, Bruun TUJ, Weidenbacher PA, Tang S, et al. Efficient evolution of human antibodies from general protein language models. *Nat Biotechnol* 2024;**42**:275−83.
44. Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, et al. Learning inverse folding from millions of predicted structures. *bioRxiv* 2022. 2022.04.10.487779.
45. Shanker VR, Bruun TUJ, Hie BL, Kim PS. Unsupervised evolution of protein and antibody complexes with a structure-informed language model. *Science* 2024;**385**:46−53.

46. Liu X, Luo Y, Li P, Song S, Peng J. Deep geometric representations for modeling effects of mutations on protein−protein binding affinity. *PLoS Comput Biol* 2021;**17**:e1009284.
47. Shan S, Luo S, Yang Z, Hong J, Su Y, Ding F, et al. Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization. *Proc Natl Acda Sci* 2022;**119**:e2122954119.
48. Li M, Tan Y, Ma X, Zhong B, Yu H, Zhou Z, et al. ProSST: protein language modeling with quantized structure and disentangled attention. *NeurIPS* 2024;**37**:35700−26.
49. Su J, Han C, Zhou Y, Shan J, Zhou X, Yuan F. Saprot: protein language modeling with structure-aware vocabulary. *bioRxiv* 2023. Available from: https://doi.org/10.1101/2023.10.01.560349.
50. Tan Y, Zhou B, Zheng L, Fan G, Hong L. Semantical and geometrical protein encoding toward enhanced bioactivity and thermostability. *bioRxiv* 2023. Available from: https://doi.org/10.1101/2023.12.01.569522.
51. Notin P, Dias M, Frazer J, Hurtado JM, Gomez AN, Marks D, et al. Tranception: protein fitness prediction with autoregressive transformers and inference-time retrieval. *39th ICML* 2022;**162**:16990−7017.
52. Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, et al. Robust deep learning−based protein sequence design using ProteinMPNN. *Science* 2022;**378**:49−56.
53. Sormanni P, Amery L, Ekizoglou S, Vendruscolo M, Popovic B. Rapid and accurate *in silico* solubility screening of a monoclonal antibody library. *Sci Rep* 2017;**7**:8200.
54. Hou Q, Kwasigroch JM, Rooman M, Pucci F. SOLart: a structure-based method to predict protein solubility and aggregation. *Bioinformatics* 2020;**36**:1445−52.
55. Kuriata A, Iglesias V, Pujols J, Kurcinski M, Kmiecik S, Ventura S. Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility. *Nucleic Acids Res* 2019;**47**:W300−7.
56. Sankar K, Krystek Jr SR, Carl SM, Day T, Maier JKX. AggScore: prediction of aggregation-prone regions in proteins based on the distribution of surface patches. *Proteins* 2018;**86**:1147−56.
57. Thangakani AM, Kumar S, Nagarajan R, Velmurugan D, Gromiha MM. GAP: towards almost 100 percent prediction for β-strand-mediated aggregating peptides with distinct morphologies. *Bioinformatics* 2014;**30**:1983−90.
58. Paladin L, Piovesan D, Tosatto Silvio CE. SODA: prediction of protein solubility from disorder and aggregation propensity. *Nucleic Acids Res* 2017;**45**:W236−40.
59. Feng J, Jiang M, Shih J, Chai Q. Antibody apparent solubility prediction from sequence by transfer learning. *iScience* 2022;**25**:105173.
60. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 2020;**48**:W449−54.
61. Marks C, Hummer AM, Chin M, Deane CM. Humanization of antibodies using a machine learning approach on large-scale repertoire data. *Bioinformatics* 2021;**37**:4041−7.
62. Prihoda D, Maamary J, Waight A, Juan V, Fayadat Dilman L, Svozil D, et al. BioPhi: a platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning. *mAbs* 2022;**14**:2020203.
63. Wang H, Hao X, He Y, Fan L. AbImmPred: an immunogenicity prediction method for therapeutic antibodies using AntiBERTy-based sequence features. *PLoS One* 2024;**19**:e0296737.
64. Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, et al. IEDB-AR: immune epitope database—analysis resource in 2019. *Nucleic Acids Res* 2019;**47**:W502−6.
65. Raybould MIJ, Marks C, Krawczyk K, Taddese B, Nowak J, Lewis AP, et al. Five computational developability guidelines for therapeutic antibody profiling. *Proc Natl Acad Sci U S A* 2019;**116**:4025−30.
66. Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, Rooman M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* 2009;**25**:2537−43.
67. Chen X, Dougherty T, Hong C, Schibler R, Zhao YC, Sadeghi R, et al. Predicting antibody developability from sequence using machine learning. *bioRxiv* 2020. 2020.06.18.159798.
68. Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. *De novo* design of protein structure and function with RFdiffusion. *Nature* 2023;**620**:1089−100.
69. Bennett NR, Watson JL, Ragotte RJ, Borst AJ, See DL, Weidle C, et al. Atomically accurate *de novo* design of antibodies with RFdiffusion. *bioRxiv* 2025. 2024.03.14.585103.
70. He H, He B, Guan L, Zhao Y, Jiang F, Chen G, et al. *De novo* generation of SARS-CoV-2 antibody CDRH3 with a pre-trained generative large language model. *Nat Commun* 2024;**15**:6867.
71. Zambaldi V, La D, Chu AE, Patani H, Danson AE, Kwan TOC, et al. *De novo* design of high-affinity protein binders with AlphaProteo. *arXiv preprint arXiv 240908022* 2024. Available from: https://doi.org/10.48550/arXiv.2409.08022.
72. Pacesa M, Nickel L, Schellhaas C, Schmidt J, Pyatova E, Kissling L, et al. One-shot design of functional protein binders with BindCraft. *Nature* 2025;**646**:483−92.
73. Bio N. *De novo* design of epitope-specific antibodies against soluble and multipass membrane proteins with high specificity, developability, and function. *Nabla Bio* 2024;**23**. Available from: https://doi.org/10.1101/2025.01.21.633066.
74. Zhang Z, van der Kant R, Marković I, Vizarraga D, Garcia T, Maragkou K, et al. *In silico* design of stable single-domain antibodies with high affinity. *bioRxiv* 2024. 2024.04.22.589762.
75. Lan W, Valente JJ, Ilott A, Chennamsetty N, Liu Z, Rizzo JM, et al. Investigation of anomalous charge variant profile reveals discrete pH-dependent conformations and conformation-dependent charge states within the CDR3 loop of a therapeutic mAb. *mAbs* 2020;**12**:1763138.
76. Estrella V, Chen T, Lloyd M, Wojtkowiak J, Cornnell HH, Ibrahim-Hashim A, et al. Acidity generated by the tumor microenvironment drives local invasion. *Cancer Res* 2013;**73**:1524−35.
77. Greenshields-Watson A, Vavourakis O, Spoendlin FC, Cagiada M, Deane CM. Challenges and compromises: predicting unbound antibody structures with deep learning. *Curr Opin Struct Biol* 2025;**90**:102983.
78. North B, Lehmann A, Dunbrack Jr RL. A new clustering of antibody CDR loop conformations. *J Mol Biol* 2011;**406**:228−56.
79. Houde D, Peng Y, Berkowitz SA, Engen JR. Post-translational modifications differentially affect IgG1 conformation and receptor binding. *Mol Cell Proteomics* 2010;**9**:1716−28.
80. Vatsa S. *In silico* prediction of post-translational modifications in therapeutic antibodies. *mAbs* 2022;**14**:2023938.
81. Pavon JA, Xiao L, Li X, Zhao J, Aldredge D, Dank E, et al. Selective tryptophan oxidation of monoclonal antibodies: oxidative stress and modeling prediction. *Anal Chem* 2019;**91**:2192−200.
82. Yan Y, Wei H, Fu Y, Jusuf S, Zeng M, Ludwig R, et al. Isomerization and oxidation in the complementarity-determining regions of a monoclonal antibody: a study of the modification-structure-function correlations by hydrogen-deuterium exchange mass spectrometry. *Anal Chem* 2016;**88**:2041−50.
83. Haberger M, Bomans K, Diepold K, Hook M, Gassner J, Schlothauer T, et al. Assessment of chemical modifications of sites in the CDRs of recombinant antibodies: susceptibility *vs.* functionality of critical quality attributes. *mAbs* 2014;**6**:327−39.
84. Karczewski KJ, Francioli LC, Tiao G, Cummings B, BAlföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;**581**:434−43.
85. Draper-Joyce C, Furness SGB. Conformational transitions and the activation of heterotrimeric G proteins by G protein-coupled receptors. *ACS Pharmacol Transl Sci* 2019;**2**:285−90.
86. Son A, Kim W, Park J, Lee W, Lee Y, Choi S, et al. Utilizing molecular dynamics simulations, machine learning, Cryo-EM, and NMR spectroscopy to oredict and validate protein dynamics. *Int J Mol Sci* 2024;**25**:9725.
87. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human transcription factors. *Cell* 2018;**172**:650−65.

88. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell* 2013;**152**:1237−51.
89. Lu H, Zhou Q, He J, Jiang Z, Peng C, Tong R, et al. Recent advances in the development of protein−protein interactions modulators: mechanisms and clinical trials. *Sig Trans Targ Thera* 2020;**5**: 213.
90. Aparicio O, Razquin N, Zaratiegui M, Narvaiza I, Fortes P. Adenovirus virus-associated RNA is processed to functional Interfering RNAs involved in virus production. *J Virol* 2006;**80**: 1376−84.
91. Sullenger BA, Gallardo HF, Ungers GE, Gilboa E. Overexpression of TAR sequences renders cells resistant to human immunodeficiency virus replication. *Cell* 1990;**63**:601−8.
92. Breaker RR. Natural and engineered nucleic acids as tools to explore biology. *Nature* 2004;**432**:838−45.
93. Klug SJ, Famulok M. All you wanted to know about SELEX. *Mol Biol Rep* 1994;**20**:97−107.
94. Townshend RJL, Eismann S, Watkins AM, Rangan R, Karelina M, Das R, et al. Geometric deep learning of RNA structure. *Science* 2021;**373**:1047−51.
95. Miao Z, Adamiak RW, Antczak M, Boniecki MJ, Bujnicki J, Chen S-J, et al. RNA-puzzles round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA* 2020;**26**:982−95.
96. Shen T, Hu Z, Sun S, Liu D, Wong F, Wang J, et al. Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nat Methods* 2024;**21**:2287−98.
97. Magnus M, Antczak M, Zok T, Wiedemann J, Lukasiak P, Cao Y, et al. RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. *Nucleic Acids Res* 2020;**48**:576−88.
98. Wong F, He D, Krishnan A, Hong L, Wang AZ, Wang J, et al. Deep generative design of RNA aptamers using structural predictions. *Nat Comput Sci* 2024;**4**:829−39.
99. Incarnato D, Oliviero S. The RNA epistructurome: uncovering RNA function by studying structure and post-transcriptional modifications. *Trends Biotechnol* 2017;**35**:318−33.
100. Sato K, Hamada M. Recent trends in RNA informatics: a review of machine learning and deep learning for RNA secondary structure prediction and RNA drug discovery. *Briefings Bioinf* 2023;**24**: bbad186.