



From Prediction to Design: A Paradigm Shift in AI-Driven Protein–Protein Interaction Research

Yihuang Wu, Wei Hu, and Linlin Zhao *

Department of Biochemistry, School of Medicine, Tongji University, Shanghai 200060, China

* Correspondence: lzhao@tongji.edu.cn

Received: 16 June 2026; Revised: 14 August 2026; Accepted: 17 August 2026; Published: 23 October 2026

Abstract: Protein-protein interactions (PPIs) form the molecular foundation of cellular life. Over the past two decades, computational methods for PPI analysis have primarily focused on prediction—inferring interaction properties from protein information. However, a notable conceptual shift is underway away from predictive analysis and toward rational protein interaction design. This transformation is driven by breakthroughs in protein language models, diffusion models, and equivariant neural networks. In this review, we discuss how this transition represents a notable conceptual shift, enabling exploration of sequence-structure space that evolution has never sampled. We synthesize the technical foundations, core design tasks, and challenges ahead. We address critical issues including evaluation bias, the hallucination problem, and dual-use considerations. We conclude with grand challenges for the field: mapping design ability boundaries, designing dynamic interactions, and developing interpretable methods for clinical translation. We further describe how modern AI reshapes computational approaches for biological discovery.

Keywords: protein-protein interactions; protein design; deep learning; generative models; diffusion models; protein language models

1. Introduction

Protein-protein interactions (PPIs) form the molecular foundation of cellular life [1]. The human interactome is estimated to contain 130,000–650,000 binary interactions [2], with disease-associated mutations disproportionately affecting interaction interfaces [3]. Over the past two decades, computational approaches to PPI research have primarily focused on prediction: given protein sequences or structures, infer whether and how they interact. This prediction paradigm has produced remarkable successes, including high-accuracy interaction prediction from sequence alone [4,5] and unprecedented accuracy in complex structure prediction [6,7].

Experimental approaches remain the foundation for discovering and validating PPIs. Yeast two-hybrid screening enables relatively high-throughput detection of binary interactions but can suffer from false positives, false negatives, and context dependence. Co-immunoprecipitation and affinity-purification mass spectrometry capture interactions closer to native cellular conditions but are lower-throughput, condition-specific, and often indirect. Phage or yeast display can select binders from large libraries, yet the searchable sequence space is constrained by library construction, expression bias, and selection conditions. AI-driven methods occupy a complementary position: they can generate or prioritize candidates at low marginal computational cost, but their success rates remain target-dependent and their outputs require experimental validation. This practical division of labor motivates the present review’s focus on how computational design changes candidate generation rather than replacing experimental PPI biology.

Yet a fundamental limitation constrains prediction methods: they can only describe what might happen, not create what we want to happen. For drug discovery and synthetic biology, the relevant question is not “will these proteins interact?” but “how can we design proteins that interact the way we need?” This shift from passive observation to active creation defines the emerging design paradigm.



1.1. Defining the Paradigm Shift

We argue that the transition from prediction to design represents a qualitative transformation in computational PPI research. The prediction paradigm asks: “What interactions exist in nature?” The design paradigm asks: “What interactions can we create?”

This shift is characterized by three features. First, the design paradigm enables investigation of previously inaccessible regions of sequence-structure space—configurations that evolution never explored but that may contain superior therapeutic or industrial properties. Second, the design paradigm requires fundamentally different methodological approaches: generative models rather than discriminative classifiers, inverse problems rather than forward prediction, and iterative design-test cycles rather than one-shot inference. Third, prediction and design are complementary: every design system relies on prediction capabilities for validation [8], and prediction methods benefit from novel test cases that design provides.

Figure 1 summarizes this conceptual transition while explicitly distinguishing prediction-centered analysis, earlier computational design, and modern AI-driven generative design.

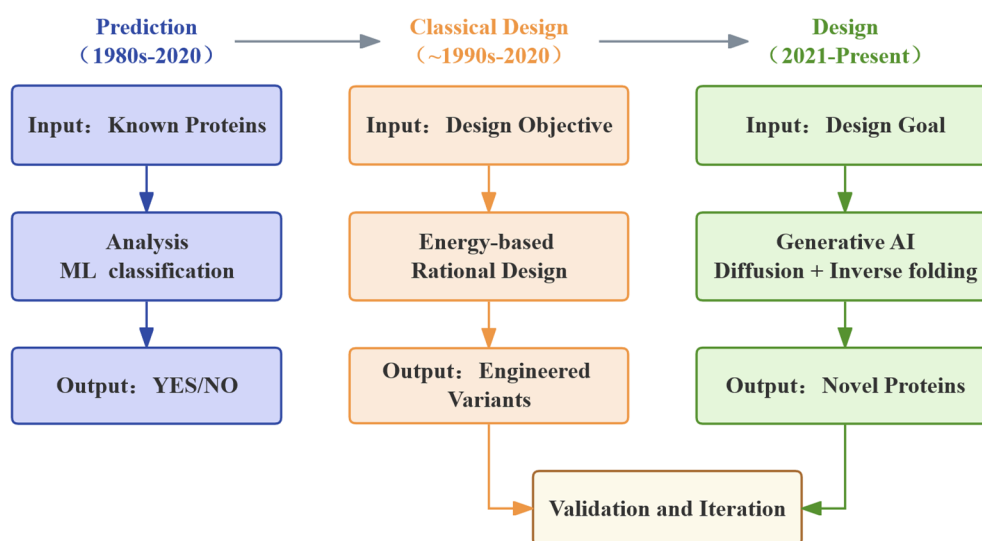


Figure 1. Three conceptual phases are color-coded: blue for prediction-centered analysis (1980s–2020), orange for classical computational protein design (~1990s–2020), and green for modern generative-AI-driven design (2021–Present). **Left panel** (blue): prediction-centered approaches infer protein interaction properties of known proteins, using sequence-based classifiers, ML (Machine Learning), docking and complex-structure prediction. **Middle panel** (orange): classical design employs energy-based and rational engineering workflows to generate engineered protein variants. **Right panel** (green): modern generative-AI-driven design leverages diffusion models and inverse folding to produce novel designed proteins. Outputs from classical and generative-AI-driven design enter the unified Validation and Iteration loop, highlighting the essential role of experimental testing for designed proteins. Grey arrows above panels denote the historical paradigm shift across three eras of protein-protein interaction study and design. These conceptual phases progressed partially in parallel and are not strictly sequential in time.

1.2. Technological Foundations and Controversies

This paradigm shift is enabled by converging technological advances: protein language models encoding evolutionary information at scale [9,10], diffusion models generating novel structures conditioned on functional constraints [8,11], and equivariant neural networks preserving geometric symmetries [12,13].

The field is not without controversy. Critics note that success rates for *de novo* design remain modest (10–50% depending on target) compared to natural protein engineering [14]. Questions persist about whether designed proteins will exhibit the stability, specificity, and lack of immunogenicity required for therapeutic use. A critical evaluation crisis has been identified: Tsishyn et al. [14] demonstrated that many reported high accuracies result from dataset-specific biases rather than genuine predictive capability, suggesting that methodological artifacts may inflate perceived progress.

1.3. Article Structure

This review provides a comprehensive analysis of this paradigm transition. Section 2 examines the trajectory from prediction to design, including methodological evolution and key milestones. Section 3 surveys core design tasks enabled by new capabilities. Section 4 details enabling technologies. Section 5 discusses open challenges including evaluation, validation, and interpretability. Section 6 examines therapeutic and industrial applications. Section 7 offers perspectives on future directions, and Section 8 synthesizes key themes and outlines grand challenges for the field.

2. The Paradigm Transition: From Prediction to Design

Figure 2 places the major technical milestones in historical context, emphasizing that recent generative models build on both earlier prediction systems and classical computational-design principles. The field of computational PPI research has evolved through distinct methodological eras, culminating in the current design-oriented approaches. This section traces this evolution, examining how prediction capabilities laid the foundation for design.

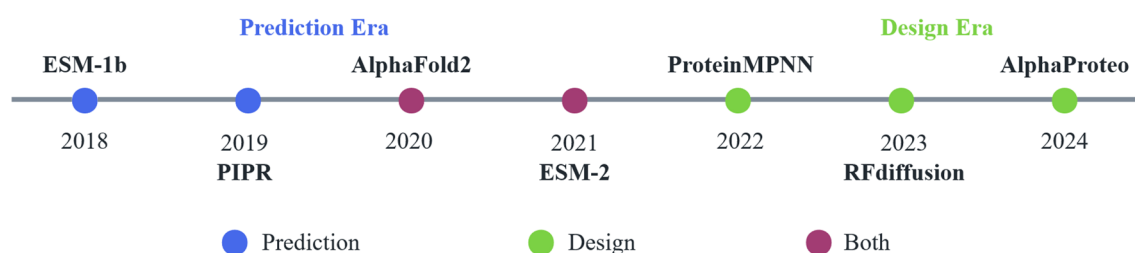


Figure 2. The timeline distinguishes earlier prediction-centered methods and protein representation learning from the recent emergence of scalable generative design systems. The transition highlighted here is therefore not the birth of computational protein design itself, but the shift from expert-guided, energy-function-based design toward data-driven, deep-learning-enabled generative design. This timeline focuses on recent deep-learning-driven advances; foundational earlier computational protein design work is acknowledged but not plotted here to preserve readability of recent milestones. The selected milestones are restricted to protein-specific or protein-adapted methods trained or evaluated on biological sequence, structure, or interaction data. Citations for each individual milestone model are provided in the main text. Color coding: blue = prediction or representation learning; green = generative design; purple = methods connecting prediction and design.

2.1. The Prediction Paradigm: Achievements and Limitations

Before the emergence of generative design, PPI research was dominated by the prediction paradigm—inferring whether, where, and how strongly proteins interact based on their sequences or structures.

Sequence-Based Prediction. Methods evolved through three generations: (1) handcrafted feature approaches using SVMs with physicochemical properties [15–17]; auto-covariance descriptors captured interactions between residues at varying sequence distances [18,19]; (2) deep learning methods enabling automatic feature learning [4,20]; and (3) pre-trained protein language models leveraging evolutionary information [5,9]. Chinese research groups contributed substantially to sequence-based PPI prediction [21]. Graph neural networks emerged as powerful approaches, with DeepRank-GNN [22] learning patterns directly from structural graphs. TUNA introduced uncertainty quantification [23], while TAGPPI incorporated predicted structural information [24]. D-SCRIPT demonstrated that structure-aware representations improve cross-species generalizability [5], and Topsy-Turvy [25] integrated sequence-based predictions with network topology.

Structure-Based Prediction. The establishment of the Protein Data Bank (PDB) in 1971 as the worldwide archive of macromolecular structures [26] provided the foundation for all structure-based prediction methods. Classical protein-protein docking methods predict bound complex structures from individual proteins: HADDOCK (version 2.4) [27] introduced information-driven docking with biochemical constraints, ZDOCK [28] uses fast Fourier transforms for shape complementarity, ClusPro [29] provides automated docking through clustering, HDock [30] combines template-based and free docking, and GalaxyTongDock [31] enables symmetric and asymmetric docking with improved energy parameters. Surface-based deep learning methods predict binding interfaces directly from molecular structures: MaSIF (version 0.1.1) [32] applied geometric deep learning to surfaces, GraphPPIS (version 1.0) [33] used graph convolutional networks, dMaSIF (version 1.0) [34] achieved 10–100× speedups with point-cloud operators, and ScanNet (version 1.0) [35] combined geometric deep learning with spatial attention for interpretable

binding site prediction. Beyond binary interaction prediction, binding site prediction identifies which residues mediate interactions—critical information for drug design and mutagenesis studies. Systematic benchmarking of AlphaFold for protein complexes revealed important determinants of prediction accuracy [36]. The most dramatic advance is end-to-end complex prediction: AlphaFold-Multimer [7] predicts complex structures using transformer architectures, while AlphaFold 3 [6] employs diffusion models to predict atomic coordinates directly, achieving superior accuracy for multimolecular complexes. Table 1 summarizes the taxonomy of PPI prediction tasks, covering task inputs, outputs, interaction types, representative methods and biological notes.

Table 1. Taxonomy of PPI prediction tasks.

Task	Input	Output	Interaction Type	Representative Methods	Notes
Binary classification	Sequence pairs	Interaction (yes/no)	Physical binary PPI	PIPR [4], D-SCRIPT [5]	Predict direct physical protein-protein contacts
Network inference	Protein set	Interaction network	Mostly functional association	STRING [1]	STRING aggregates functional/co-occurrence evidence; does not exclusively report direct physical binary PPI
Binding site prediction	Single protein	Interface residues	Physical binary PPI	MaSIF [32], GraphPPIS [33]	Predict residues participating in physical binding interfaces
Affinity prediction	Complex + mutation	$\Delta\Delta G$	Physical binary PPI	FoldX [37], SKEMPI [38]	Quantify binding energy change upon mutation
Complex structure	Sequence pairs	3D coordinates	Physical binary PPI	AlphaFold-Multimer [7], DiffDock-PP [39]	Generate 3D models for physically interacting complexes

Fundamental Limitations. Despite progress, the prediction paradigm faces inherent constraints:

1. **Data Dependence:** Models learn statistical patterns from known interactions and struggle to extrapolate to novel types [40]. Training data is heavily biased toward model organisms (human, yeast, mouse). Negative sample construction introduces systematic bias—randomly paired proteins may not represent true non-interactors [41]. The Negatome database addressed this by curating experimentally verified non-interacting pairs [42]. Prediction accuracy degrades significantly for proteins dissimilar from training examples, as demonstrated by the C1/C2/C3 difficulty classification framework.
2. **Evaluation Crisis:** A systematic analysis of binding affinity prediction methods by Tsishyn et al. [14] revealed that many reported accuracies are inflated by dataset-specific biases rather than genuine predictive capability. Eight widely-used predictors showed strong dependence on training datasets, with performance degrading substantially when evaluated on proteins from different families. Benchmark datasets suffer from incompleteness and bias, with STRING [1], BioPlex [43], and IntAct [44] each having different coverage and confidence scoring schemes [45]. This suggests that current train-test splitting protocols may be fundamentally inadequate for evaluating generalization to novel protein interactions.
3. **Passive Discovery Mode:** Prediction methods provide essential information about existing or proposed complexes, but they do not by themselves generate molecular solutions for therapeutic or synthetic-biology goals. For therapeutic applications, this provides no direct pathway from prediction to intervention.
4. **Static Assumptions:** Most methods treat interactions as binary states, ignoring tissue-specificity, condition-dependence, and conformational dynamics [46].

These limitations motivated the emergence of design-oriented approaches.

Importantly, *de novo* protein and interface design has a longer history than the current generation of AI systems. Physics-based and statistical energy-function approaches, especially Rosetta-derived methods, established that designed proteins and macromolecular interfaces could be experimentally realized before the advent of diffusion models and large protein language models. The novelty of the current phase is therefore not design per se, but the scale, automation, and conditional controllability introduced by deep learning. This distinction prevents the prediction-to-design narrative from overlooking foundational computational-design literature while preserving the central claim that modern AI has changed what is practically tractable.

2.2. Technical Drivers of the Design Paradigm

The recent transition is best understood not as the first appearance of computational protein design, but as a shift from expert-guided, energy-function-based design toward scalable AI-driven generative design, enabled by breakthroughs between 2021 and 2024.

Structure Prediction Advances. AlphaFold2 (version 2.3.2) [47] and RoseTTAFold (version 1.1.0) [48] achieved near-experimental accuracy in structure prediction. Extensions to multimeric complexes [7,49], expanded

scope to PPIs. These advances provide both validation mechanisms for designed proteins and learned representations repurposed for generation.

Generative Models. Diffusion models achieved breakthrough performance by treating protein generation as learning to reverse a noise-adding process [11]. RFdiffusion (version 1.1.0) [8] fine-tuned RoseTTAFold as a denoising model, enabling conditional generation on partial structures or target proteins. Chroma [11] incorporated physical constraints directly into the generative process, enabling generation of proteins up to 4000 amino acids. Flow-based alternatives like FoldFlow [50] offer potentially better computational efficiency.

Protein Language Models at Scale. ESM-2 [10] is a large-scale protein language model with up to 15-billion parameters; its representations power ESMFold, which enables atomic-level structure prediction directly from single sequences. These models provide learned priors over protein sequences—encoding what sequences are likely to fold into stable, functional proteins. Recent extensions like MINT [51] and PLM-interact [52] train specifically on PPI datasets.

Equivariant Neural Networks. SE(3)-Transformers [13] and Geometric Vector Perceptrons [12] process geometric data while respecting rotational and translational symmetries—critical for ensuring that generated proteins' validity does not depend on arbitrary coordinate systems. AlphaFold2's Invariant Point Attention [47] provided an influential alternative approach.

2.3. Milestone Works in the Design Era

Docking frameworks such as DiffDock-PP remain important prediction tools for evaluating or positioning existing structures, but they are not generative design methods and are therefore treated separately from the design tools summarized in Table 2. The convergence of these technologies produced breakthrough systems organized by primary function (Table 2).

Table 2. Comparison of major AI-driven protein design and inverse-design tools (2022–2024).

Method	Year	Task	Input→Output	Benchmark/Data Split	Experimental Assay	Number of Targets/Designs Tested
RFdiffusion [8]	2023	Binder design	Target → Binder	Custom held-out target set	Cryo-EM †	Hundreds of designs across multiple targets
AlphaProteo [53]	2024	Binder design	Target → Binder	Seven custom targets	SPR/BLI (nM) ‡	7 targets
ProteinMPNN [54]	2022	Inverse design	Backbone → Sequence	CATH-4.2 test set	X-ray crystallography §	Large CATH test set
ESM-IF1 [55]	2022	Inverse design	Backbone → Sequence	CATH-4.2 test set	None large-scale wet-lab validation §	Large CATH test set
Chroma [11]	2023	<i>De-Novo</i> full-protein generation	Constraints → Full protein	Internal held-out set	No wet-lab assay; only MD-simulation †	Limited public target set

† Binding success rate varies by target difficulty; data from Watson et al. [8]. ‡ Reported improvement over RFdiffusion baseline on seven targets in the original AlphaProteo (2024 release) study; independent reproduction is limited by closed-source access and unavailable model weights [53]. § Native sequence recovery on CATH 4.2 test set [54,55]. † Designability: fraction with pLDDT > 80; 70–90% depending on length [11].

Backbone Generation. RFdiffusion demonstrated successful binder design with experimental validation: cryo-EM structures confirmed atomic-level agreement with computational models (backbone RMSD: 0.5–1.5 Å). Chroma enabled generation of large proteins with explicit physical constraints.

Sequence Design. ProteinMPNN (version 1.0.1) and ESM-IF1 (version 1.0) address the inverse folding problem—given a backbone, predicting sequences that fold into it. Both achieve 50% native sequence recovery, enabling high-throughput sequence design for generated backbones.

Binder Design. AlphaProteo is a prominent recent demonstration of integrated AI-driven binder design, with the original report describing 3–300× higher success rates than RFdiffusion-based baselines across seven targets and multiple nanomolar-affinity binders. However, these results should be interpreted with appropriate caution: AlphaProteo is a proprietary system whose source code and model weights are not publicly available, limiting independent reproduction, systematic benchmarking, and method transfer by the broader academic community.

Multi-Modal Systems. AlphaFold 3 [6] and RoseTTAFold AllAtom [49] extend to complexes containing proteins, nucleic acids, and small molecules, with RFdiffusionAA enabling *de novo* design of protein binding pockets around cofactors.

2.4. The Complementarity of Prediction and Design

Prediction and design are not opposing paradigms but complementary capabilities. Every design system relies on prediction: RFDiffusion was initialized from RoseTTAFold's pretrained weights; ProteinMPNN is trained to predict foldable sequences; AlphaProteo uses structure prediction for validation. The design process can be viewed as iterative prediction: structure prediction of targets, interface prediction for binding sites, backbone generation as inverse structure prediction, and complex prediction for validation.

However, the complementarity does not negate the paradigm shift characterization. Design enables qualitatively new capabilities—exploring sequence-structure space that evolution never accessed—while prediction provides the foundation upon which design builds. The relationship is bidirectional: design also advances prediction by providing test cases that reveal limitations in current understanding.

2.5. Critical Assessment: The Design-Hallucination Boundary

A fundamental challenge in generative protein design is distinguishing genuine design from computational artifacts. Unlike language model hallucination (generating factually incorrect information), protein design “hallucination” refers to structures that score well on computational metrics but fail experimental validation [56]. This manifests as: (1) backbone incoherence violating physical constraints (Ramachandran outliers, steric clashes); (2) sequence-structure mismatch where generated sequences do not fold into designed structures; and (3) energy landscape pathology preventing spontaneous folding.

Current approaches mitigate these issues through iterative validation—structure prediction on generated sequences, molecular dynamics simulation, and experimental characterization. Garcia et al. [57] found that approximately half of designed proteins failed at various experimental stages, with many failed designs having better computational scores than successful ones, underscoring that *in silico* metrics cannot fully substitute for physical reality.

Even experimental validation has limitations: cryo-EM and X-ray crystallography capture static snapshots that may not reflect dynamic behavior in solution. Comprehensive validation requires multiple approaches: biophysical characterization for affinity, solution scattering for oligomeric state, and NMR for dynamics.

The paradigm shift is therefore not a complete replacement of traditional methods but a fundamental expansion of what is tractable. Design targets previously infeasible—novel scaffolds, *de novo* interfaces, challenging targets—are now accessible, while established methods excel in their domains of strength.

3. Core Design Tasks

The design paradigm enables fundamentally new questions. Rather than asking “what interactions exist?”, researchers can now ask “how can we create interactions with desired properties?” This section surveys major design tasks enabled by recent advances.

3.1. Protein Binder Design

Protein binder design—creating novel proteins that specifically bind target molecules—represents the most direct application of the design paradigm [8,53]. Given a target structure, the goal is to generate a binding partner with high affinity and specificity.

The problem decomposes into subtasks: epitope selection (identifying the target region), scaffold generation (creating complementary backbone), sequence design (optimizing for stability and binding), and validation (predicting binding success). A core challenge lies in the enormous sequence-space size: for a 100-residue protein there are on the order of 20^{100} distinct possible sequences, yet only a tiny fraction fold stably and bind with high affinity.

RFDiffusion [8] pioneered general-purpose binder design. Given a target structure, it generates complementary backbones via diffusion, which ProteinMPNN [54] then designs sequences for. Experimental validation demonstrated 10–50% success rates across diverse targets. AlphaProteo [53] reported an integrated generation-and-validation pipeline with improved success rates and nanomolar-affinity binders on selected therapeutic targets including PD-L1 and VEGF-A, but its closed-source implementation limits independent validation and broad accessibility. For SARS-CoV-2, designed minibinders achieved picomolar affinities (K_D : 10–100 pM), demonstrating that designed proteins can compete with evolved antibodies [8].

3.2. Antibody Design and Optimization

Antibodies present unique challenges due to their conserved framework regions and hypervariable complementarity-determining regions (CDRs). Unlike other CDRs which adopt canonical structures encoded by germline genes, CDR-H3 lacks reliable structural templates [58]. Its backbone conformation depends on the specific V(D)J recombination event, making prediction essentially a *de novo* folding problem. CDR-H3 length varies from 8 to 30+ residues (with longer loops increasingly common in therapeutic antibodies), and longer loops exhibit greater backbone conformational flexibility. Deep learning methods have made significant progress in antibody structure prediction [59–62], and recent advances have improved antibody-antigen complex modeling [63], but CDR-H3 modeling remains the primary bottleneck.

Antibody design involves: (1) designing binders against target antigens; (2) humanizing non-human antibodies to reduce immunogenicity; and (3) optimizing developability properties (stability, solubility, expression) [64]. Recent methods combine structural modeling with developability prediction to screen candidates before experimental validation. AbAgIntPre introduced sequence-only interaction prediction, while recent work applied prompt-based language models to nanobody-antigen interactions [65].

3.3. Interface Optimization

Rather than designing entirely new binding partners, interface optimization aims to improve existing interactions through targeted mutations [38].

Affinity Maturation. Increasing binding strength is central to therapeutic optimization. Physics-based methods (FoldX (version 5.0) [37], Rosetta Flex ddG (version r215) [66]) estimate affinity changes from structural calculations but require accurate structures. Machine learning methods (mCSM-PPI2 (version 1.0) [67], GeoPPI (version 1.0) [68]) learn statistical patterns from experimental data, achieving faster predictions. BeAtMuSiC (version 1.0) [69] introduced coarse-grained statistical potentials for rapid screening. SSIPe (version 1.0) [70] combines structural interface analysis with evolutionary information. DDAffinity (version 1.0) [71] uses geometric deep learning with attention for multiple simultaneous mutations, while recent work applied deep geometric representations to modeling mutation effects [68]. SSIF-Affinity (version 1.0) [72] integrates sequence, structure, and interaction features through multimodal learning.

Specificity Engineering. Beyond affinity, specificity—binding the intended target without binding similar off-targets—is critical for therapeutic applications. Designing proteins that discriminate between closely related paralogues requires subtle manipulation of interface chemistry, addressed through multi-objective optimization [67].

Stability-Developability Trade-offs. Interface mutations that improve affinity can destabilize the overall fold. Methods like ProteinMPNN perform inverse-folding-based sequence design [54]; sequences generated by this method often exhibit favorable predicted stability. In contrast, developability optimization directly targets properties including aggregation propensity, immunogenicity, and expression yield [64].

3.4. De Novo Interaction Networks

The most ambitious design task involves creating entire interaction networks—multiple proteins designed to interact in specific patterns [8]. This enables synthetic biology applications from engineered signaling pathways to self-assembling biomaterials.

Symmetric oligomer design requires simultaneously optimizing interface geometry and overall assembly structure. RFDiffusion demonstrated successful design of cyclic oligomers, with extensions enabling icosahedral cages [8]. Multi-component systems present additional challenges: each protein must fold independently while forming specific interactions with partners. Current approaches decompose the problem into pairwise interfaces, but this cannot capture global system constraints.

Three fundamental challenges remain: orthogonality (ensuring proteins interact only with intended partners), kinetics (ensuring correct assembly order), and integration (functioning within cells without disrupting essential processes). These challenges are largely unaddressed by current methods.

These tasks share a common theme: they require generative capabilities beyond prediction. The enabling technologies underlying these capabilities are detailed in Section 4.

4. Enabling Technologies

The paradigm shift is enabled by three converging technological advances: generative models, structure-aware representations, and protein language models.

4.1. Generative Models for Protein Design

Generative models learn to sample from the distribution of valid protein structures, enabling creation rather than prediction. The key insight is that protein foldability and stability can be encoded in a learned energy landscape—the model implicitly captures biophysical constraints distinguishing natural proteins from random sequences [73].

Diffusion Models. The dominant paradigm treats protein structure generation as learning to reverse a noise-adding process [74]. The forward diffusion process in Equation (1) gradually adds Gaussian noise:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon} \quad (1)$$

Here, $\mathbf{x}_0$ is the original structure, $\mathbf{x}_t$ is the noisy version at timestep t , and $\bar{\alpha}_t$ is the noise schedule. The reverse process learns to denoise by predicting the added noise commonly optimized using the objective in Equation (2):

$$\mathcal{L} = \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\epsilon}} [\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|^2] \quad (2)$$

RFdiffusion fine-tuned RoseTTAFold as a denoiser, exploiting transfer learning from pre-trained representations [8]. Chroma incorporated physical constraints (bond lengths, angles, dihedral preferences) directly into the generative process [11]. SE(3) diffusion models operate on the manifold of rigid body frames, where each residue is represented by position $\mathbf{r} \in \mathbb{R}^3$ and orientation $\mathbf{R} \in \text{SO}(3)$ [75].

Flow-Based Alternatives. Flow matching learns deterministic trajectories through structure space rather than stochastic denoising. FoldFlow demonstrated comparable quality to diffusion with potentially better computational efficiency [50], though diffusion currently leads in practical applications.

Inverse Folding. ProteinMPNN and ESM-IF1 address the complementary problem: given a backbone, what sequence will adopt it [54,55]? These methods enable sequence design for generated backbones, completing the design pipeline.

4.2. Structure-Aware Representations

Effective protein modeling requires representations capturing both discrete chemical structure (sequence) and continuous geometry (3D coordinates). Protein function depends on 3D geometry, which must be represented in a coordinate-independent manner [12,13].

Equivariant Neural Networks. SE(3)-Transformers [13] and Geometric Vector Perceptrons (GVPs) [12] process geometric data while respecting rotational and translational symmetries. SE(3) equivariance requires that for any rotation $\mathbf{R} \in \text{SO}(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$:

$$f(\mathbf{R}\mathbf{x} + \mathbf{t}) = \mathbf{R} \cdot f(\mathbf{x}) + \mathbf{t}$$

This property means network outputs transform consistently under rotation and translation, matching the transformation applied to the input structure. GVPs achieve this through scalar-vector separation: each node maintains scalar features $\mathbf{s}$ and vector features $\mathbf{v}$, with message passing operating equivariantly on vectors while aggregating scalars invariantly. AlphaFold2's Invariant Point Attention (IPA) provided an alternative through learned reference frames [47], projecting 3D coordinates onto local frames where attention becomes rotation-invariant. IPA has become a foundational architectural element in subsequent design systems.

Why Equivariance Matters for Protein Design. The equivariance constraint is not merely a mathematical convenience—it reflects fundamental physical principles. Protein structures exist in 3D space independent of coordinate system orientation. A design method producing different outputs for rotated inputs violates this physical invariance. Critically, non-equivariant methods that work well for small molecules often fail for proteins: small molecule conformations can be enumerated densely, but protein conformational space is astronomically larger ($\sim 10^{130}$ for a 100-residue protein). Equivariance provides strong inductive bias that dramatically reduces effective search space. AlphaFold2's success relied heavily on this constraint: by ensuring equivariance, the model learns identical predictions regardless of input orientation, effectively multiplying training data by the symmetry group. The practical consequence for design is substantial: equivariant models require fewer training examples to achieve equivalent performance, critical given limited experimental validation data for designed proteins.

SE(3) Diffusion vs. Euclidean Diffusion. Standard diffusion operates in Euclidean space, treating coordinates independently. However, the protein backbones are on the SE(3) manifold—each residue has position $\mathbf{r} \in \mathbb{R}^3$ and orientation $\mathbf{R} \in \text{SO}(3)$. SE(3) diffusion respects this structure, adding noise through geodesic

random walks rather than independent Gaussian noise. This ensures the noising process stays on the manifold of physically valid protein structures.

Geometric Deep Learning. MaSIF (version 0.1.1) and PeSTo (Article reproducibility, 2023) learn from molecular surfaces without relying on sequence information, demonstrating that interaction propensities are encoded in surface geometry and charge distribution [32,76]. dMaSIF replaced expensive mesh generation with point-cloud operators, achieving 10–100× speedups [34]. The biological implication is that physical determinants of protein interactions—shape complementarity, electrostatic compatibility, hydrogen bonding—can be learned from structural data and transferred to design novel interfaces.

4.3. Protein Language Models

Protein language models learn representations from evolutionary information encoded in sequence databases, providing complementary structural and functional knowledge [9,10].

Foundation Models. ESM-1b demonstrated that large-scale pre-training on protein sequences—without explicit structural supervision—yields representations capturing folding patterns, functional sites, and evolutionary constraints [9]. ESM-2, scaled to 15 billion parameters, is a protein language model whose representations underpin ESMFold (version 1.0.3). ESMFold delivers atomic-level structure prediction from single sequences without requiring multiple sequence alignments [10]. ESMFold runs 60× faster than AlphaFold2 while maintaining competitive accuracy. ProtTrans [77] established that encoder-only models outperform encoder-decoder architectures for protein understanding tasks. xTrimoPGLM scaled to 100 billion parameters for unified pre-training, representing a significant contribution from Chinese research groups to the foundation model landscape [78]. GOpoteinGNN [79] integrated protein knowledge graphs with graph neural networks leveraging Gene Ontology annotations.

Structure-Aware Language Models. SaProt introduced structure-aware tokenization where each token represents both sequence position and local structural state [80]. MINT and PLM-interact extended PLMs for interaction modeling, operating on protein sets to capture interaction propensities contextually [51,52].

Transfer Learning. The scale of PLMs enables effective fine-tuning for downstream tasks including binding site prediction, function annotation, and interaction prediction [81]. This reduces data requirements for specialized tasks while leveraging knowledge encoded in foundation models.

Computational Considerations. Training large PLMs requires enormous investment (ESM-2 required hundreds of GPU-years). Inference costs vary dramatically: ProteinMPNN runs on CPU in milliseconds, while RFdiffusion requires 16–32 GB GPU memory and minutes per design [8]. These requirements create accessibility concerns that the community must address through efficient architectures and shared infrastructure.

These technologies provide the computational substrate for design tasks. Section 5 examines the challenges limiting their practical applicability.

5. Challenges and Limitations

Despite remarkable progress, significant challenges limit practical applicability of AI-driven PPI design. This section examines major obstacles and outlines research directions.

5.1. Evaluation and Benchmarking

The evaluation crisis identified by Tsishyn et al. [14] represents a core challenge: many reported high accuracies result from data leakage—train-test overlap at the sequence similarity level. When rigorous cross-validation prevents such leakage, model performance drops substantially. This suggests that many “successful” methods primarily learn to recognize sequence similarity rather than genuine interaction features.

Proposed Evaluation Standards. Based on this analysis, we propose standardized evaluation protocols:

1. **Rigorous Sequence Identity Thresholds:** Enforce maximum 25–30% sequence identity between train and test sets, verified through all-vs-all BLAST (BLAST + 2.16.0). However, even 30% thresholds may be insufficient: Tsishyn et al. [14] demonstrated that predictors show strong dataset dependence even with standard cross-validation, suggesting that sequence-level similarity alone does not capture all sources of bias. We recommend combining sequence thresholds with structural clustering (e.g., TM-score < 0.5) and ensuring no overlapping protein families between splits.
2. **Temporal Splits:** Evaluate on proteins discovered after training data cutoff date, preventing contamination through database annotations.

3. Negative Sample Construction: Use experimentally verified non-interactors (e.g., Negatome [42]) rather than random pairing, which introduces systematic bias.
4. Multi-Dataset Validation: Require consistent performance across independent benchmarks (SKEMPI [38], DIP [82], MINT [83]) to prevent overfitting to specific dataset characteristics.
5. Negative Result Reporting: Require reporting of failure rates and unsuccessful targets to combat publication bias. This faces implementation challenges: researchers have little incentive to publish negative results. We propose three mechanisms: (1) data repository requirements mandating deposition of all experimental results (successful and failed) as a publication condition; (2) journal policy changes accepting “registered reports” where methodology is peer-reviewed before experiments begin; and (3) community recognition metrics rewarding rigor rather than just success rates. Methods achieving 20% success with complete transparency should be valued over methods claiming 50% success with selective reporting.

Design Evaluation Challenges. For design tasks, evaluation is more complex because ground truth does not exist. Success must be measured along multiple dimensions: success rate (fraction expressing, folding, binding), novelty (genuinely new vs. replicating known interactions), quality (affinities, specificities, stabilities), and practical utility (functioning in intended contexts). Reported success rates vary widely (5–50%) across studies, but direct comparison is difficult due to different experimental protocols and target difficulties.

5.2. Experimental Validation Bottleneck

Computational design outpaces experimental validation capacity. While methods generate thousands of candidates daily, validation remains time-consuming and expensive, creating a feedback bottleneck. Without systematic testing, methods cannot learn from failures [84]. Current practice validates only top-ranked designs, potentially missing high-performing designs that rank poorly computationally.

High-Throughput Solutions. Emerging platforms offer partial solutions: deep mutational scanning for parallel variant testing, yeast/phage display for library screening. However, these introduce biases—interactions functional on yeast surfaces may not work in mammalian cells. Cost structures create inequities: well-resourced labs validate more designs, accelerating their method development.

Infrastructure Needs. Addressing this bottleneck requires investment in: (1) centralized validation facilities accessible to the research community; (2) standardized protocols enabling cross-study comparison; (3) automated pipelines reducing human effort; and (4) data sharing platforms capturing both successes and failures. Developing this shared validation infrastructure will require substantial financial investment.

5.3. Exploring the Designable Space

A critical question remains unanswered: what fraction of theoretical sequence-structure space is actually designable? Not all sequences fold stably, and not all stable structures perform desired functions. The designable subset may be small [11], and designed proteins with sequences unlike training data may fail despite computational predictions [54].

Understanding designability boundaries has practical implications. If the designable space is limited, methods should focus on efficient exploration rather than random generation. If large but inefficiently sampled, improved search strategies would yield better designs. Systematic experimental characterization—deep mutational scanning combined with design—could map these boundaries, transforming design from trial-and-error to principled exploration.

5.4. Dynamic and Conditional Interactions

Most methods assume static structures and binary interaction states, ignoring biological reality: many PPIs are transient, condition-dependent, or allosterically regulated [46]. Binding often involves conformational changes, induced fit, or conformational selection. Methods designing for single static structures may fail when actual interactions involve flexible regions or multiple conformations.

AlphaFold-based predictions capture some flexibility through ensembles but do not model binding kinetics or conformational dynamics [6]. Incorporating molecular dynamics or ensemble approaches into design pipelines remains challenging. Designing proteins that bind only under specific conditions—activated by post-translational modifications or cofactors—requires modeling conditional dependencies that current methods largely ignore.

5.5. Current Approaches to Dynamic PPIs

Several computational approaches address dynamic PPIs, though none are yet integrated into generative design pipelines.

Molecular Dynamics-Enhanced Prediction. Enhanced sampling methods (metadynamics [85], replica exchange [86]) can characterize conformational ensembles of protein interfaces. However, these simulations remain computationally expensive (10^3 – 10^5 GPU-hours per complex), limiting integration into iterative design workflows. Recent work combining AlphaFold with MD simulations [87] demonstrated improved handling of flexible regions but increased computational cost 10-fold.

NMR Ensemble Integration. NMR structures capture conformational heterogeneity through ensemble refinement. Approaches like EnsembleDock [88] dock against multiple conformers, selecting poses consistent across the ensemble. However, NMR structures represent only ~5% of PDB entries, limiting applicability. For design, this suggests that generating multiple backbone conformations during diffusion and validating against all would improve robustness for flexible targets.

AlphaFold Confidence as Flexibility Indicators. AlphaFold's per-residue pLDDT scores show empirical correlation with experimentally observed flexibility—regions with pLDDT < 70 often correspond to intrinsically disordered regions [89]. Predicted aligned error (PAE) matrices can identify flexible inter-domain linkers. Importantly, pLDDT and PAE are not direct, validated readouts of molecular dynamics or binding kinetics; they reflect model prediction uncertainty rather than directly measuring protein dynamic behavior. For design, this suggests: (1) targeting high-confidence (rigid) regions reduces dynamic complexity; (2) designing for conformational ensembles requires multiple backbone generation.

Conformational Selection vs. Induced Fit. The conformational selection model posits proteins sample multiple conformations, and binding partners select the most complementary one. Induced fit involves conformational change upon binding. Both operate in nature, but current design methods implicitly assume rigid-body recognition. A promising direction is multistate design: generating binders against conformational ensembles and selecting candidates binding multiple states.

Limitations for Design. Current dynamic modeling faces three barriers: (1) computational cost (MD scales as $O(N^2)$ per timestep); (2) limited training data (few proteins have characterized dynamics); (3) evaluation challenges (measuring designed protein flexibility requires specialized experiments like hydrogen-deuterium exchange or single-molecule FRET). Near-term solutions include targeting rigid interfaces, leveraging AlphaFold ensembles with caution as heuristic indicators of flexibility, rather than validated quantitative proxies, and developing coarse-grained dynamics models.

5.6. Interpretability and Trust

Deep learning models underlying design methods are largely black boxes. When designs fail experimentally, current methods provide limited guidance for troubleshooting. Did failure result from misfolded backbone, poor sequence choice, or unpredicted interface geometry? Without interpretability, iterative design becomes trial-and-error.

For therapeutic applications, regulatory review requires mechanistic justification, validated quality attributes, and risk-based evidence rather than unsupported black-box predictions [64]. AI-generated binders that cannot be mechanistically explained or experimentally validated may therefore face translational barriers. Near-term goal: design methods providing mechanistic explanations for interface affinity and specificity.

5.7. Ethical and Dual-Use Considerations

The ability to design novel proteins binding specific targets carries dual-use potential warranting careful consideration [90]. While current applications focus on therapeutics and basic research, the same technologies could potentially be misused.

Risk Categories. Potential misuse scenarios include: (1) enhancing protein stability or affinity for harmful purposes; (2) designing novel toxins or immune modulators; and (3) creating proteins that circumvent existing countermeasures. However, several technical factors currently limit these risks: success rates remain modest (10–50%), experimental validation creates substantial bottlenecks requiring specialized expertise and equipment, and commercial DNA synthesis services implement sequence screening protocols.

Current Mitigations. The protein design community has proactively adopted responsible practices: publishing methods openly to enable broad beneficial applications while emphasizing responsible use, engaging with biosecurity experts to identify potential concerns, and supporting voluntary screening frameworks. Synthesis providers screen orders against databases of known pathogens and toxins, though AI-designed sequences may not match existing databases.

Governance Recommendations. We propose four areas for updated frameworks: (1) extending DNA synthesis screening protocols to include AI-generated sequences using machine learning classifiers for harmful properties; (2) establishing community standards for responsible publication that balance scientific openness with appropriate

safeguards; (3) engaging with regulatory bodies (NSABB, Biological Weapons Convention, WHO) to develop frameworks that address democratization of design capabilities while preserving scientific progress; and (4) developing anomaly detection systems for identifying potentially harmful design requests in computational tools.

The goal is enabling beneficial applications—therapeutics, industrial enzymes, research tools—while maintaining vigilance as capabilities advance.

5.8. Practical Guidelines for Researchers

To help researchers navigate the rapidly evolving landscape of PPI design tools, we provide brief guidance on method selection and interpretation.

Method Selection. Choice depends on task and constraints: (1) For binder design with known target structure, RFdiffusion + ProteinMPNN provides a robust starting point; AlphaProteo may offer strong reported performance for selected targets, but access restrictions, unavailable model weights, and limited independent benchmarking should be considered when selecting methods. (2) For affinity maturation, structure-based methods (FoldX, mCSM-PPI2) are fast for single mutations, while DDAffinity and GeoPPI better handle multiple mutations. (3) For binding site prediction, MaSIF and PeSTo work well when structures are available; sequence-based methods enable genome-scale screening.

Interpreting Predictions. Computational metrics provide guidance but not guarantees: pLDDT > 80 suggests reliable local structure; PAE matrices identify confident interface regions; binding energy predictions ($\Delta\Delta G$) are most reliable for single mutations at well-characterized interfaces. The Garcia et al. [57] finding—that many failed designs had good computational scores—underscores that high confidence does not guarantee experimental success.

Experimental Design. Recommended validation pipeline: (1) biophysical screening (SPR, ITC, or MST) for binding confirmation; (2) thermal stability (DSF or CD) to ensure folded structure; (3) structural validation (X-ray, cryo-EM, or NMR) for design verification. For therapeutic applications, additional developability assessments (aggregation propensity, immunogenicity, expression yield) should guide lead selection.

These challenges define the research frontier. Section 6 examines current applications despite these limitations.

6. Applications and Impact

The paradigm shift opens possibilities across therapeutic development, industrial biotechnology, and basic research (Table 3). This section examines major application areas.

Table 3. Selected applications of AI-designed PPIs.

Domain	Application	Representative Results	Status
Therapeutics	SARS-CoV-2 binders	K _D : 10–100 pM; 1000× potency [8]	Preclinical [†]
Therapeutics	PD-L1/VEGF-A binders	K _D : 1–50 nM; 3–300× success [53]	Early-stage [‡]
Immunotherapy	IL-2 selectivity	10–100× selectivity improvement [91,92]	Preclinical [†]
Enzymes	Cofactor binding pockets	Native-like geometry; catalytic activity [49]	Proof-of-concept [§]
Research	Imaging probes	High-specificity intracellular labeling [65]	Established [†]

[†] Multiple *in-vitro/in-vivo* laboratory demonstrations; no clinical trial data available. [‡] Encouraging preliminary results, but few independent reproductions, not yet widely adopted. [§] Limited laboratory demonstration in small-scale settings; no broad independent replication. [†] Repeatedly validated by multiple independent groups, widely adopted as research tool.

6.1. Therapeutic Protein Design

Designed proteins can serve as therapeutics, drug components, or research tools. The global biologics market exceeded \$350 billion in 2023, with AI-driven design offering reduced timelines and expanded target space [93].

Antibody and Binder Therapeutics. Traditional antibody development requires 12–18 months. AI-driven design accelerates this by enabling rapid candidate generation against new targets. For SARS-CoV-2, minibinders from the original LCB1 study achieved picomolar affinities (K_D: 10–100 pM) with improved stability—retaining function after weeks at 37 °C versus days for many antibodies [8]. The LCB1 minibinder showed neutralization potency (IC₅₀: 15–50 pM) exceeding many clinical-stage antibodies (IC₅₀: 100–1000 pM). Instead, select minibinder candidates can exhibit compact size, favorable thermal stability, and compatibility with bacterial expression systems in individual case-by-case examples.

AlphaProteo designed binders against cancer targets (PD-L1, VEGF-A) with nanomolar affinities without experimental optimization [53]. Multiple companies (Generate Biomedicines, Insitro, Charm Bio) are developing AI-designed therapeutics, though clinical-stage readouts for these AI-derived candidates remain to be established.

Cytokine Engineering. Natural cytokines often activate multiple receptor subtypes, causing off-target effects. Designed variants with improved selectivity could provide safer therapeutics. IL-2 variants demonstrating 10–100-fold improved selectivity between effector T cells and regulatory T cells could improve cancer immunotherapy while reducing toxicity [91,92]. Similar approaches apply to IL-4, IL-6, and interferon families.

Emerging Modalities. PROTACs (proteolysis-targeting chimeras) bring together target proteins and E3 ubiquitin ligases to trigger degradation. Over 20 PROTACs have entered clinical trials since 2019 [94]. Designing PPI interfaces between these components could improve efficiency. Bispecific antibodies—proteins binding two targets—offer immune cell redirection to tumors, with 10+ approved for oncology [95]. Designing both interfaces while maintaining stability presents complex multi-objective optimization challenges.

6.2. Industrial and Economic Impact

Cost and Timeline Reduction. Capitalized R&D costs per approved drug, including clinical failure adjustments and cost of capital, can exceed \$2.6 billion under certain estimation methods, with biologics development requiring 12–18 months for lead identification [96]. AI-driven design can substantially lower experimental burden: generating hundreds of high-quality candidates dramatically narrows the scope of experimental screening. Conventional directed evolution workflows need to screen large variant libraries (10^5 – 10^9 variants) and typically demand many months of experimental work for protein optimization. Faster development enables rapid response to emerging threats—demonstrated during COVID-19 when designed binders were generated within weeks of target availability.

Enzyme Engineering. Industrial biotechnology uses enzymes to produce chemicals, pharmaceuticals, and materials (market: \$10+ billion annually). Many transformations require multienzyme complexes channeling intermediates between active sites. RFdiffusionAA designed protein binding pockets for cofactors (heme, bilin) with native-like geometry and catalytic activity [49]. Extensions to multi-enzyme assemblies could enable metabolic pathway design for sustainable chemical production, biosynthesis of complex natural products, and bioremediation.

Agricultural Applications. Designed binding proteins could protect crops from pathogens by interfering with virulence factors, improve nitrogen fixation by engineering *Rhizobium*-plant symbiosis, or enable biosensing of environmental contaminants.

6.3. Research Tools and Basic Science

Mechanism Validation. Computational predictions can be tested through designed proteins. If a model predicts specific interface residues mediate binding, mutation in a designed binder should disrupt interaction. This enables systematic dissection of interaction mechanisms beyond natural variation [8].

Imaging and Detection. Fluorescent binders enable live-cell visualization of protein localization and dynamics [97]. Designed nanobodies fused to fluorescent proteins provide intracellular imaging capabilities impossible with traditional antibodies. This combination enables powerful imaging applications.

Synthetic Biology. Engineered signaling pathways could reprogram cellular behavior. The synNotch system combines sensing domains with transcriptional responses [98]. Deep learning methods have been developed for designing protein–protein interaction sequences, enabling precise control over interface specificity [47]. Applications include therapeutic cells sensing disease markers, industrial microbes optimizing metabolism, and biosensors detecting environmental conditions.

Structure-Function Exploration. Systematic exploration requires large variant sets. High-throughput characterization of hundreds of designs builds datasets revealing design principles. Deep mutational scanning combined with design maps fitness landscapes, identifying prediction-experiment discrepancies that point to understanding gaps [99].

6.4. Economic, Accessibility, and Sustainability Considerations

Computational divide. Training large protein language models can require hundreds of GPU-years, while inference costs vary widely: ProteinMPNN can run on CPUs in milliseconds, whereas RFdiffusion typically requires 16–32 GB GPU memory and minutes per design [8]. Access restrictions for some web-only or proprietary systems further limit large-scale use. These differences create inequities between well-resourced institutions that can screen thousands of designs and resource-limited laboratories that cannot.

Democratization Strategies. Open-source implementations such as RFdiffusion, ProteinMPNN, and ESMFold lower barriers but still require computational expertise and infrastructure. Broader access will require maintained open implementations, cloud or shared-infrastructure programs for resource-limited groups, efficient

architectures that reduce GPU requirements, and transparent benchmarks that allow researchers to select methods appropriate to their resources.

Environmental sustainability. As design campaigns scale, the carbon cost of model training and repeated inference becomes increasingly relevant. Shared pretrained models, efficient inference, model distillation, and community infrastructure can reduce duplicated computation while improving accessibility.

These applications leverage the ability to create proteins rather than search for natural variants. As methods mature, scope will expand from individual binders to multi-protein systems and programmable biological functions. Realizing this potential requires addressing challenges outlined in Section 5.

7. Future Perspectives

The paradigm shift represents not an endpoint but a transition toward deeper AI-biology integration. This section outlines emerging directions.

7.1. Prediction-Design Integration

Current workflows treat prediction and design as separate stages. Future systems will optimize this loop end-to-end through active learning—prioritizing experiments that maximally reduce uncertainty about design rules, with each result updating the model [100]. Unified models could predict interaction properties, generate designs, and evaluate quality within consistent frameworks, enabling more efficient design space exploration.

7.2. Toward Protein Intelligence

Systems operate at different capability levels: understanding (parsing structures and functions), prediction (inferring interactions), and design (creating proteins with desired functions). The emerging fourth level—creation—would generate biological systems solving problems evolution never addressed, requiring reasoning about multiple proteins within functional systems. As language models scale, few-shot and zero-shot design capabilities may expand, enabling rapid adaptation to new challenges without task-specific training [10].

7.3. Methodological Advances

Dynamic Modeling. Current methods design static interfaces, but many functional PPIs involve conformational changes and allosteric regulation. Integrating molecular dynamics with generative design could enable proteins with dynamic binding behavior. Key milestone: designing an allosterically regulated binder switching affinity in response to a small molecule or post-translational modification.

Multi-Component Systems. Extending from pairwise interactions to 3+ protein complexes would enable signaling pathways, metabolic channels, and synthetic cellular machines. This requires solving combinatorial explosion of interface configurations while maintaining assembly stability. Target capability: designing a 5-protein complex with specified architecture and function.

Enhanced Evaluation. Addressing the evaluation crisis requires: (1) community-agreed benchmark datasets with standardized splits; (2) negative result repositories capturing failures; (3) prospective evaluation requirements for published claims; and (4) computational validation standards (multiple prediction methods, confidence thresholds, structural quality metrics).

7.4. Experimental Infrastructure

Computational protein design advances far faster than experimental validation capacity. High-throughput platforms capable of testing thousands of designs per week under standardized protocols would help narrow this performance gap. Building such shared community infrastructure demands considerable capital investment. Emerging solutions include deep mutational scanning for parallel variant testing, and yeast/phage display for library screening, though these introduce system-specific biases.

7.5. Interpretability and Clinical Translation

For therapeutic applications, regulatory approval will require more than high computational confidence. AI-designed proteins must be supported by mechanistic rationale, validated analytical assays, uncertainty estimates, and evidence that model-guided changes remain within a controlled development framework. Interpretable design methods—where design decisions can be traced to physical or biological principles—are therefore essential for clinical translation. Near-term goal: methods providing mechanistic explanations for interface affinity and specificity.

Regulatory frameworks are evolving to address AI-designed therapeutics. The U.S. FDA has issued guidance for AI/ML-enabled medical devices, which emphasizes “predetermined change-control plans” for continuously learning software systems; this guidance is scoped for medical device software and does not directly apply to AI-designed therapeutic proteins as biological products. The European Medicines Agency (EMA) has highlighted the need for explainability, uncertainty characterization, and mechanistic understanding within pharmaceutical and biologic drug development. Such emerging regulatory expectations will shape methodological priorities for AI-protein design.

7.6. Physiological Context

Membrane Proteins and Post-translational Modifications. Many therapeutically important PPI targets are membrane proteins, including receptors, transporters, immune checkpoints, and viral entry proteins. Designing binders for these targets is more difficult than designing against soluble domains because native binding depends on membrane orientation, lipid composition, conformational state, oligomerization, and cell-surface accessibility. Structural models based on isolated extracellular domains may therefore overestimate epitope accessibility or miss conformations present in the cellular membrane environment.

Post-translational modifications add another layer of complexity. Glycosylation can shield epitopes, alter local electrostatics, change conformational ensembles, and introduce heterogeneous target populations that are difficult to represent in a single static structure. Phosphorylation, proteolytic processing, and cofactor-dependent states can similarly determine whether a designed binder recognizes the biologically relevant form of a target. Future design pipelines will need to integrate glycoprotein modeling, membrane-aware structural prediction, ensemble docking, and cell-based validation to improve success rates for clinically relevant targets.

7.7. Paradigm Shifts in Research Practice

Closed-loop Design Practice. Future progress will depend less on restating the philosophical contrast between prediction and design and more on building closed-loop design-build-test-learn systems [101]. In such workflows, computational models propose candidates, automated or semi-automated experiments measure binding and developability, and the resulting data are used to update the next design round [53]. This requires prospective benchmarks, standardized reporting of failures, uncertainty estimates attached to design proposals, and interoperable experimental metadata. The practical research shift is therefore toward iterative engineering systems in which prediction, generation, and validation continuously inform one another [11].

7.8. Grand Challenges for the Next Five Years

Challenge 1: Designability Mapping. What fraction of sequence-structure space is designable? Systematic experimental characterization through large-scale design and testing would transform design from trial-and-error to principled exploration. Timeline: 3–5 years with coordinated community effort.

Challenge 2: Dynamic Interaction Design. Designing proteins with conformational flexibility or conditional binding requires integrating dynamics with generation. Milestone: designing a switchable binder with $10\times$ affinity change upon ligand binding.

Challenge 3: Validation Infrastructure. Closing the computational-experimental gap requires shared high-throughput facilities accessible to the broader research community. Realizing such community-oriented infrastructure will demand substantial capital investment.

Challenge 4: Interpretability. Clinical translation requires explainable designs. Goal: methods providing atomic-level explanations linking design decisions to predicted properties.

Challenge 5: Standards and Governance. Updated frameworks for responsible development, including synthesis screening protocols and publication standards.

The prediction-to-design transition marks a shift from AI as passive observer to AI as active creator. Whether this transition fulfills its promise will be determined by collective efforts in the coming years. The technical foundations are in place; the challenges are known; the opportunity is unprecedented. The conclusion synthesizes these themes and offers final perspectives.

8. Conclusions

Computational protein-protein interaction research is undergoing a profound transformation. This review has traced the trajectory from prediction—inferring whether, where, and how proteins interact—to design—creating

proteins with desired interaction properties. This paradigm shift represents a qualitative change in how AI interfaces with biological discovery.

The technological foundations are established. Protein language models encode evolutionary and structural information at scale [9,10]. Diffusion models generate novel structures conditioned on functional constraints [8,11]. Equivariant neural networks preserve geometric symmetries essential for accurate modeling [12,13]. Together, these enable design capabilities implausible just five years ago.

Yet significant challenges remain. The evaluation crisis suggests reported accuracies may be inflated [14]. Success rates vary widely (10–50% across targets), with negative results systematically underreported. Experimental validation limits iterative improvement. Dynamic interactions, conditional specificity, and multi-protein complexes remain largely beyond current capabilities.

The implications extend beyond technical advances. Drug discovery transitions from high-throughput screening to targeted generation [53]. Synthetic biology can design interaction networks rather than repurposing natural components. Basic research gains hypothesis-driven experimental tools.

This transition highlights the expanding scope of AI tools: from analyzing naturally occurring biomolecules toward engineering proteins for predefined functional goals. This does not diminish prediction—prediction remains the foundation upon which design builds [47,54]. Rather, it positions prediction as capability to be leveraged rather than endpoint in itself.

The coming years will determine whether current successes represent a sustained revolution or incremental advance peak. Progress depends on: better evaluation, faster validation, deeper understanding of designability, dynamics modeling, and interpretable methods. Realizing the full potential of prediction-to-design workflows requires community-wide efforts. For academic researchers, this means adopting rigorous evaluation frameworks, reporting negative findings, and building coupled computational-experimental feedback loops, with priority given to unresolved problems such as dynamic interactions, conditional specificity, and multi-component assemblies. Industrial efforts should anticipate compressed discovery timelines and expanded druggable space, where robust experimental validation infrastructure and efficient design-test cycles become key competitive factors. Funding bodies would benefit from supporting fundamental methodological bottlenecks including energy-function improvement and dynamics modeling, alongside shared validation resources, rather than incremental benchmark refinements.

The ongoing prediction-to-design transition reshapes how AI tools support biological research, moving from analytical assessment of natural biological systems toward engineered functional design. The technical foundations are in place; the challenges are known; substantial new research opportunities have emerged. Whether this transition fulfills its transformative potential will be determined by the collective efforts of researchers, developers, and practitioners in the coming years.

Author Contributions: Y.W.: Conceptualization, literature collection, formal analysis, visualization, writing—original draft preparation; W.H.: Literature curation, writing—reviewing and editing; L.Z.: Conceptualization, supervision, writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable, as this article is a review and does not report new studies involving human participants, animals, or identifiable personal data.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable. No new datasets were generated or analyzed in this review article.

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

Use of AI and AI-Assisted Technologies: No AI tools were utilized for this paper.

References

1. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. <https://doi.org/10.1093/nar/gky1131>.
2. Rehman, A.U.; Khurshid, B.; Ali, Y.; Rasheed, S.; Wadood, A.; Ng, H.J.; Chen, H.Y.; Wei, Z.; Luo, R.; Zhang, J. Computational Approaches for the Design of Modulators Targeting Protein–Protein Interactions. *Expert Opin. Drug Discov.* **2023**, *18*, 315–333. <https://doi.org/10.1080/17460441.2023.2171396>.
3. Sahni, N.; Yi, S.; Taipale, M.; Fuxman Bass, J.I.; Coulombe-Huntington, J.; Yang, F.; Peng, J.; Weile, J.; Karras, G.I.; Wang, Y.; et al. Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. *Cell* **2015**, *161*, 647–660. <https://doi.org/10.1016/j.cell.2015.04.013>.

4. Chen, M.; Ju, C.J.T.; Zhou, G.; Chen, X.; Zhang, T.; Chang, K.W.; Zaniolo, C.; Wang, W. Multifaceted Protein–Protein Interaction Prediction Based on Siamese Residual RCNN. *Bioinformatics* **2019**, *35*, i305–i314. <https://doi.org/10.1093/bioinformatics/btz328>.
5. Sledzieski, S.; Singh, R.; Cowen, L.; Berger, B. D-Script Translates Genome to Phenome with Sequence-Based, Structure-Aware, Genome-Scale Predictions of Protein–Protein Interactions. *Cell Syst.* **2021**, *12*, 969–982.e6. <https://doi.org/10.1016/j.cels.2021.08.010>.
6. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. Erratum in *Nature* **2024**, *636*, E4. <https://doi.org/10.1038/s41586-024-07487-w>.
7. Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Židek, A.; Bates, R.; Blackwell, S.; Yim, J.; et al. Protein Complex Prediction with AlphaFold-Multimer. *bioRxiv* **2021**. <https://doi.org/10.1101/2021.10.04.463034>.
8. Watson, J.L.; Juergens, D.; Bennett, N.R.; Trippe, B.L.; Yim, J.; Eisenach, H.E.; Ahern, W.; Borst, A.J.; Ragotte, R.J.; Milles, L.F.; et al. De Novo Design of Protein Structure and Function with RFDiffusion. *Nature* **2023**, *620*, 1089–1100. <https://doi.org/10.1038/s41586-023-06415-8>.
9. Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C.L.; Ma, J.; et al. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2016239118. <https://doi.org/10.1073/pnas.2016239118>.
10. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **2023**, *379*, 1123–1130. <https://doi.org/10.1126/science.ade2574>.
11. Ingraham, J.B.; Baranov, M.; Costello, Z.; Barber, K.W.; Wang, W.; Ismail, A.; Frappier, V.; Lord, D.M.; Ng-Thow-Hing, C.; Van Vlack, E.R.; et al. Illuminating Protein Space with a Programmable Generative Model. *Nature* **2023**, *623*, 1070–1078. <https://doi.org/10.1038/s41586-023-06728-8>.
12. Jing, B.; Eismann, S.; Suriana, P.; Townshend, R.J.L.; Dror, R. Learning from Protein Structure with Geometric Vector Perceptrons. *arXiv* **2021**, arXiv:2009.01411. <https://doi.org/10.48550/arxiv.2009.01411>.
13. Fuchs, F.B.; Worrall, D.E.; Fischer, V.; Welling, M. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. *arXiv* **2020**, arXiv:2006.10503. <https://doi.org/10.48550/arxiv.2006.10503>.
14. Tsishyn, M.; Pucci, F.; Rومان, M. Quantification of Biases in Predictions of Protein–Protein Binding Affinity Changes upon Mutations. *Brief. Bioinform.* **2024**, *25*, bbad491. <https://doi.org/10.1093/bib/bbad491>.
15. Bock, J.R.; Gough, D.A. Predicting Protein–Protein Interactions from Primary Structure. *Bioinformatics* **2001**, *17*, 455–460. <https://doi.org/10.1093/bioinformatics/17.5.455>.
16. Shen, J.; Zhang, J.; Luo, X.; Zhu, W.; Yu, K.; Chen, K.; Li, Y.; Jiang, H. Predicting Protein–Protein Interactions Based Only on Sequences Information. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4337–4341. <https://doi.org/10.1073/pnas.0607879104>.
17. Ben-Hur, A.; Noble, W.S. Kernel Methods for Predicting Protein–Protein Interactions. *Bioinformatics* **2005**, *21*, i38–i46. <https://doi.org/10.1093/bioinformatics/bti1016>.
18. Guo, Y.; Yu, L.; Wen, Z.; Li, M. Using Support Vector Machine Combined with Auto Covariance to Predict Protein–Protein Interactions from Protein Sequences. *Nucleic Acids Res.* **2008**, *36*, 3025–3030. <https://doi.org/10.1093/nar/gkn159>.
19. Li, Z.W.; You, Z.H.; Chen, X.; Li, L.P.; Huang, D.S.; Yan, G.Y.; Nie, R.; Huang, Y.A. Accurate Prediction of Protein–Protein Interactions by Integrating Potential Evolutionary Information Embedded in PSSM Profile and Discriminative Vector Machine Classifier. *Oncotarget* **2017**, *8*, 23638–23649. <https://doi.org/10.18632/oncotarget.15564>.
20. Hashemifar, S.; Neyshabur, B.; Khan, A.A.; Xu, J. Predicting Protein–Protein Interactions through Sequence-Based Deep Learning. *Bioinformatics* **2018**, *34*, i802–i810. <https://doi.org/10.1093/bioinformatics/bty573>.
21. Gao, Z.; Jiang, C.; Zhang, J.; Jiang, X.; Li, L.; Zhao, P.; Yang, H.; Huang, Y.; Li, J. Hierarchical Graph Learning for Protein–Protein Interaction. *Nat. Commun.* **2023**, *14*, 1093. <https://doi.org/10.1038/s41467-023-36736-1>.
22. Réau, M.; Renaud, N.; Xue, L.C.; Bonvin, A.M.J.J. DeepRank-GNN: A Graph Neural Network Framework to Learn Patterns in Protein–Protein Interfaces. *Bioinformatics* **2023**, *39*, btac759. <https://doi.org/10.1093/bioinformatics/btac759>.
23. Ko, Y.; Parkinson, J.; Liu, C.; Wang, W. TUNA: An Uncertainty-Aware Transformer Model for Sequence-Based Protein–Protein Interaction Prediction. *Brief. Bioinform.* **2024**, *25*, bbae359. <https://doi.org/10.1093/bib/bbae359>.
24. Song, B.; Luo, X.; Luo, X.; Liu, Y.; Niu, Z.; Zeng, X. Learning Spatial Structures of Proteins Improves Protein–Protein Interaction Prediction. *Brief. Bioinform.* **2022**, *23*, bbab558. <https://doi.org/10.1093/bib/bbab558>.
25. Singh, R.; Devkota, K.; Sledzieski, S.; Berger, B.; Cowen, L. Topsy-Turvy: Integrating a Global View into Sequence-Based PPI Prediction. *Bioinformatics* **2022**, *38*, i264–i272. <https://doi.org/10.1093/bioinformatics/btac258>.
26. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. <https://doi.org/10.1093/nar/28.1.235>.
27. Dominguez, C.; Boelens, R.; Bonvin, A.M.J.J. HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737. <https://doi.org/10.1021/ja026939x>.

28. Chen, R.; Li, L.; Weng, Z. ZDOCK: An Initial-Stage Protein-Docking Algorithm. *Proteins* **2003**, *52*, 80–87. <https://doi.org/10.1002/prot.10389>.
29. Kozakov, D.; Hall, D.R.; Xia, B.; Porter, K.A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein–Protein Docking. *Nat. Protoc.* **2017**, *12*, 255–278. <https://doi.org/10.1038/nprot.2016.169>.
30. Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S.Y. HDock: A Web Server for Protein–Protein and Protein–DNA/RNA Docking Based on a Hybrid Strategy. *Nucleic Acids Res.* **2017**, *45*, W365–W373. <https://doi.org/10.1093/nar/gkx407>.
31. Park, T.; Baek, M.; Lee, H.; Seok, C. GalaxyTongDock: Symmetric and Asymmetric Ab Initio Protein–Protein Docking Web Server with Improved Energy Parameters. *J. Comput. Chem.* **2019**, *40*, 2413–2417. <https://doi.org/10.1002/jcc.25874>.
32. Gainza, P.; Sverrisson, F.; Monti, F.; Rodolà, E.; Boscaini, D.; Bronstein, M.M.; Correia, B.E. Deciphering Interaction Fingerprints from Protein Molecular Surfaces Using Geometric Deep Learning. *Nat. Methods* **2020**, *17*, 184–192. <https://doi.org/10.1038/s41592-019-0666-6>.
33. Yuan, Q.; Chen, J.; Zhao, H.; Zhou, Y.; Yang, Y. Structure-Aware Protein–Protein Interaction Site Prediction Using Deep Graph Convolutional Network. *Bioinformatics* **2022**, *38*, 125–132. <https://doi.org/10.1093/bioinformatics/btab643>.
34. Sverrisson, F.; Feydy, J.; Correia, B.E.; Bronstein, M. Fast End-to-End Learning on Protein Surfaces. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual Event, 20–25 June 2021.
35. Tubiana, J.; Schneidman-Duhovny, D.; Wolfson, H.J. ScanNet: An Interpretable Geometric Deep Learning Model for Structure-Based Protein Binding Site Prediction. *Nat. Methods* **2022**, *19*, 730–739. <https://doi.org/10.1038/s41592-022-01490-7>.
36. Bryant, P.; Pozzati, G.; Elofsson, A. Improved Prediction of Protein–Protein Interactions Using AlphaFold2. *Nat. Commun.* **2022**, *13*, 1265. Erratum in *Nat. Commun.* **2022**, *13*, 1694. <https://doi.org/10.1038/s41467-022-28865-w>.
37. Guerois, R.; Nielsen, J.E.; Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More than 1000 Mutations. *J. Mol. Biol.* **2002**, *320*, 369–387. [https://doi.org/10.1016/S0022-2836\(02\)00442-4](https://doi.org/10.1016/S0022-2836(02)00442-4).
38. Jankauskaitė, J.; Jiménez-García, B.; Dapkūnas, J.; Fernández-Recio, J.; Moal, I.H. SKEMPI 2.0: An Updated Benchmark of Changes in Protein–Protein Binding Energy, Kinetics and Thermodynamics upon Mutation. *Bioinformatics* **2019**, *35*, 462–469. <https://doi.org/10.1093/bioinformatics/bty635>.
39. Ketata, M.A.; Laue, C.; Mammadov, R.; Stärk, H.; Wu, M.; Corso, G.; Marquet, C.L.; Barzilay, R.; Jaakkola, T.S. DiffDock-PP: Rigid Protein–Protein Docking with Diffusion Models. *arXiv* **2023**, arXiv:2304.03889. <https://doi.org/10.48550/arxiv.2304.03889>.
40. Kewalramani, N.; Emili, A.; Crovella, M. State-of-the-Art Computational Methods to Predict Protein–Protein Interactions with High Accuracy and Coverage. *Proteomics* **2023**, *23*, e2200292. <https://doi.org/10.1002/pmic.202200292>.
41. Smialowski, P.; Pagel, P.; Wong, P.; Brauner, B.; Dunger, I.; Fobo, G.; Frishman, G.; Montrone, C.; Rattei, T.; Frishman, D.; et al. The Negatome Database: A Reference Set of Non-Interacting Protein Pairs. *Nucleic Acids Res.* **2010**, *38*, D540–D544. <https://doi.org/10.1093/nar/gkp1026>.
42. Blohm, P.; Frishman, G.; Smialowski, P.; Goebels, F.; Wachinger, B.; Ruepp, A.; Frishman, D. Negatome 2.0: A Database of Non-Interacting Proteins Derived by Literature Mining, Manual Annotation and Protein Structure Analysis. *Nucleic Acids Res.* **2014**, *42*, D396–D400. <https://doi.org/10.1093/nar/gkt1079>.
43. Huttlin, E.L.; Bruckner, R.J.; Navarrete-Perea, J.; Cannon, J.R.; Baltier, K.; Gebreab, F.; Gygi, M.P.; Thornock, A.; Zarraga, G.; Tam, S.; et al. Dual Proteome-Scale Networks Reveal Cell-Specific Remodeling of the Human Interactome. *Cell* **2021**, *184*, 3022–3040.e28. <https://doi.org/10.1016/j.cell.2021.04.011>.
44. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; del-Toro, N.; et al. The MIntAct Project—IntAct as a Common Curation Platform for 11 Molecular Interaction Databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. <https://doi.org/10.1093/nar/gkt1115>.
45. Hu, X.; Feng, C.; Ling, T.; Chen, M. Deep Learning Frameworks for Protein–Protein Interaction Prediction. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 3223–3233. <https://doi.org/10.1016/j.csbj.2022.06.025>.
46. Davis, M.J.; Shin, C.J.; Jing, N.; Ragan, M.A. Rewiring the Dynamic Interactome. *Mol. Biosyst.* **2012**, *8*, 2054–2066. <https://doi.org/10.1039/c2mb25050k>.
47. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
48. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871–876. <https://doi.org/10.1126/science.abj8754>.
49. Krishna, R.; Wang, J.; Ahern, W.; Sturmels, P.; Venkatesh, P.; Kalvet, I.; Lee, G.R.; Morey-Burrows, F.S.; Anishchenko, I.; Humphreys, I.R.; et al. Generalized Biomolecular Modeling and Design with RoseTTAFold All-Atom. *Science* **2024**, *384*, ead12528. <https://doi.org/10.1126/science.adl2528>.

50. Bose, A.J.; Akhound-Sadegh, T.; Huguet, G.; Fatras, K.; Rector-Brooks, J.; Liu, C.H.; Nica, A.C.; Korablyov, M.; Bronstein, M.M.; Tong, A. SE(3) Stochastic Flow Matching for Protein Backbone Generation. In Proceedings of the 12th International Conference on Learning Representations, ICLR 2024, Vienna, Austria, 7–11 May 2024.
51. Ullanat, V.; Jing, B.; Sledzieski, S.; Berger, B. Learning the Language of Protein–Protein Interactions. *Nat. Commun.* **2026**, *17*, 1199. <https://doi.org/10.1038/s41467-025-67971-3>.
52. Liu, D.; Young, F.; Lamb, K.D.; Claudio Quiros, A.; Pancheva, A.; Miller, C.J.; Macdonald, C.; Robertson, D.L.; Yuan, K. PLM-Interact: Extending Protein Language Models to Predict Protein–Protein Interactions. *Nat. Commun.* **2025**, *16*, 9012. <https://doi.org/10.1038/s41467-025-64512-w>.
53. Zambaldi, V.; La, D.; Chu, A.E.; Patani, H.; Danson, A.E.; Kwan, T.O.C.; Frerix, T.; Schneider, R.G.; Saxton, D.; Thillaisundaram, A.; et al. De Novo Design of High-Affinity Protein Binders with AlphaProteo. *arXiv* **2024**, arXiv:2409.08022. <https://doi.org/10.48550/arxiv.2409.08022>.
54. Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R.J.; Milles, L.F.; Wicky, B.I.M.; Courbet, A.; de Haas, R.J.; Bethel, N.; et al. Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN. *Science* **2022**, *378*, 49–56. <https://doi.org/10.1126/science.add2187>.
55. Hsu, C.; Verkuil, R.; Liu, J.; Lin, Z.; Hie, B.; Sercu, T.; Lerer, A.; Rives, A. Learning Inverse Folding from Millions of Predicted Structures. In Proceedings of the 39th International Conference on Machine Learning, ICML 2022, Baltimore, MD, USA, 17–23 July 2022.
56. Kortemme, T. De Novo Protein Design—from New Structures to Programmable Functions. *Cell* **2024**, *187*, 526–544. <https://doi.org/10.1016/j.cell.2023.12.028>.
57. Garcia, M.; Dixit, S.M.; Rocklin, G.J. Evaluating Zero-Shot Prediction of Monomeric Protein Design Success by AlphaFold, ESMFold, and ProteinMPNN. *Protein Sci.* **2026**, *35*, e70453. <https://doi.org/10.1002/pro.70453>.
58. Sircar, A.; Kim, E.T.; Gray, J.J. RosettaAntibody: Antibody Variable Region Homology Modeling Server. *Nucleic Acids Res.* **2009**, *37*, W474–W479. <https://doi.org/10.1093/nar/gkp387>.
59. Ruffolo, J.A.; Sulam, J.; Gray, J.J. Antibody Structure Prediction Using Interpretable Deep Learning. *Patterns* **2022**, *3*, 100406. <https://doi.org/10.1016/j.patter.2021.100406>.
60. Ruffolo, J.A.; Chu, L.; Mahajan, S.; Gray, J.J. Fast, Accurate Antibody Structure Prediction from Deep Learning on Massive Set of Natural Antibodies. *Nat. Commun.* **2023**, *14*, 2389. <https://doi.org/10.1038/s41467-023-38063-x>.
61. Kenlay, H.; Dreyer, F.A.; Cutting, D.; Nissley, D.; Deane, C.M. ABodyBuilder3: Improved and Scalable Antibody Structure Predictions. *Bioinformatics* **2024**, *40*, btac576. <https://doi.org/10.1093/bioinformatics/btac576>.
62. Abanades, B.; Wong, W.K.; Boyles, F.; Georges, G.; Bujotzek, A.; Deane, C.M. ImmuneBuilder: Deep-Learning Models for Predicting the Structures of Immune Proteins. *Commun. Biol.* **2023**, *6*, 575. <https://doi.org/10.1038/s42003-023-04927-7>.
63. Gao, M.; Skolnick, J. Improved Deep Learning Prediction of Antigen–Antibody Interactions. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2410529121. <https://doi.org/10.1073/pnas.2410529121>.
64. Raybould, M.I.J.; Marks, C.; Krawczyk, K.; Taddese, B.; Nowak, J.; Lewis, A.P.; Bujotzek, A.; Shi, J.; Deane, C.M. Five Computational Developability Guidelines for Therapeutic Antibody Profiling. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4025–4030. <https://doi.org/10.1073/pnas.1810576116>.
65. Deng, J.; Gu, M.; Zhang, P.; Dong, M.; Liu, T.; Zhang, Y.; Liu, M. Nanobody–Antigen Interaction Prediction with Ensemble Deep Learning and Prompt-Based Protein Language Models. *Nat. Mach. Intell.* **2024**, *6*, 1594–1604. <https://doi.org/10.1038/s42256-024-00940-5>.
66. Barlow, K.A.; Ó Conchúir, S.; Thompson, S.; Suresh, P.; Lucas, J.E.; Heinonen, M.; Kortemme, T. Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein–Protein Binding Affinity upon Mutation. *J. Phys. Chem. B* **2018**, *122*, 5389–5399. <https://doi.org/10.1021/acs.jpcc.7b11367>.
67. Rodrigues, C.H.; Myung, Y.; Pires, D.E.; Ascher, D.B. mCSM-PPI2: Predicting the Effects of Mutations on Protein–Protein Interactions. *Nucleic Acids Res.* **2019**, *47*, W338–W344. <https://doi.org/10.1093/nar/gkz383>.
68. 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.
69. Dehouck, Y.; Kwasigroch, J.M.; Rومان, M.; Gilis, D. BeAtMuSiC: Prediction of Changes in Protein–Protein Binding Affinity on Mutations. *Nucleic Acids Res.* **2013**, *41*, W333–W339. <https://doi.org/10.1093/nar/gkt450>.
70. Huang, X.; Zheng, W.; Pearce, R.; Zhang, Y. SSIPe: Accurately Estimating Protein–Protein Binding Affinity Change upon Mutations Using Evolutionary Profiles in Combination with an Optimized Physical Energy Function. *Bioinformatics* **2020**, *36*, 2429–2437. <https://doi.org/10.1093/bioinformatics/btz926>.
71. Yu, G.; Zhao, Q.; Bi, X.; Wang, J. DDAffinity: Predicting the Changes in Binding Affinity of Multiple Point Mutations Using Protein 3D Structure. *Bioinformatics* **2024**, *40*, i418–i427. <https://doi.org/10.1093/bioinformatics/btae232>.

72. Xu, X.; Zhang, H.; Liu, Q.; Cheng, Q.; Guo, Y.; Yang, W.; Chen, X. SSIF-Affinity: Multimodal Deep Learning of Sequence-Structure Features for Precise Protein–Protein Binding Affinity Prediction. *J. Chem. Inf. Model.* **2026**, *66*, 74–87. <https://doi.org/10.1021/acs.jcim.5c01734>.
73. Cho, Y.; Dauparas, J.; Tsuboyama, K.; Rocklin, G.; Ovchinnikov, S. Implicit Modeling of the Conformational Landscape and Sequence Allows Scoring and Generation of Stable Proteins. *bioRxiv* **2024**. <https://doi.org/10.1101/2024.12.20.629706>.
74. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In Proceedings of the 34th Conference on Neural Information Processing Systems, NeurIPS 2020, Virtual Event, 6–12 December 2020.
75. Yim, J.; Trippe, B.L.; De Bortoli, V.; Mathieu, E.; Doucet, A.; Barzilay, R.; Jaakkola, T. SE(3) Diffusion Model with Application to Protein Backbone Generation. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023.
76. Krapp, L.F.; Abriata, L.A.; Rodriguez, F.C.; Dal Peraro, M. PeSTo: Parameter-Free Geometric Deep Learning for Accurate Prediction of Protein Binding Interfaces. *Nat. Commun.* **2023**, *14*, 2175. <https://doi.org/10.1038/s41467-023-37701-8>.
77. Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; et al. ProtTrans: Toward Understanding the Language of Life through Self-Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 7112–7127. <https://doi.org/10.1109/tpami.2021.3095381>.
78. Chen, B.; Cheng, X.; Li, P.; Geng, Y.A.; Gong, J.; Li, S.; Bei, Z.; Tan, X.; Wang, B.; Zeng, X.; et al. xTrimoPGLM: Unified 100B-Scale Pre-Trained Transformer for Deciphering the Language of Protein. *arXiv* **2024**, arXiv:2401.06199. <https://doi.org/10.48550/arxiv.2401.06199>.
79. Kalifa, D.; Singer, U.; Radinsky, K. GoProteinGNN: Leveraging Protein Knowledge Graphs for Protein Representation Learning. *arXiv* **2024**, arXiv:2408.00057. <https://doi.org/10.48550/arxiv.2408.00057>.
80. Su, J.; Han, C.; Zhou, Y.; Shan, J.; Zhou, X.; Yuan, F. SaProt: Protein Language Modeling with Structure-Aware Vocabulary. In Proceedings of the 12th International Conference on Learning Representations, ICLR 2024, Vienna, Austria, 7–11 May 2024.
81. Schmirler, R.; Heinzinger, M.; Rost, B. Fine-Tuning Protein Language Models Boosts Predictions Across Diverse Tasks. *Nat. Commun.* **2024**, *15*, 7407. <https://doi.org/10.1038/s41467-024-51844-2>.
82. Salwinski, L.; Miller, C.S.; Smith, A.J.; Pettit, F.K.; Bowie, J.U.; Eisenberg, D. The Database of Interacting Proteins: 2004 Update. *Nucleic Acids Res.* **2004**, *32*, D449–D451. <https://doi.org/10.1093/nar/gkh086>.
83. Chatr-aryamontri, A.; Ceol, A.; Palazzi, L.M.; Nardelli, G.; Schneider, M.V.; Castagnoli, L.; Cesareni, G. MINT: The Molecular Interaction Database. *Nucleic Acids Res.* **2007**, *35*, D572–D574. <https://doi.org/10.1093/nar/gkl950>.
84. Trepte, P.; Secker, C.; Olivet, J.; Blavier, J.; Kostova, S.; Maseko, S.B.; Minia, I.; Silva Ramos, E.; Cassonnet, P.; Golusik, S.; et al. AI-guided Pipeline for Protein–Protein Interaction Drug Discovery Identifies a SARS-CoV-2 Inhibitor. *Mol. Syst. Biol.* **2024**, *20*, 428–457. <https://doi.org/10.1038/s44320-024-00019-8>.
85. Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603. <https://doi.org/10.1103/physrevlett.100.020603>.
86. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151. [https://doi.org/10.1016/s0009-2614\(99\)01123-9](https://doi.org/10.1016/s0009-2614(99)01123-9).
87. Jussupow, A.; Kaila, V.R.I. Effective Molecular Dynamics from Neural Network-Based Structure Prediction Models. *J. Chem. Theory Comput.* **2023**, *19*, 1965–1975. <https://doi.org/10.1021/acs.jctc.2c01027>.
88. Huang, S.Y.; Zou, X. Efficient Molecular Docking of NMR Structures: Application to HIV-1 Protease. *Protein Sci.* **2007**, *16*, 43–51. <https://doi.org/10.1110/ps.062501507>.
89. Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Židek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; et al. Highly Accurate Protein Structure Prediction for the Human Proteome. *Nature* **2021**, *596*, 590–596. <https://doi.org/10.1038/s41586-021-03828-1>.
90. National Research Council. *Biotechnology Research in an Age of Terrorism*; National Academies Press: Washington, DC, USA, 2004. <https://doi.org/10.17226/10827>.
91. Sockolosky, J.T.; Trotta, E.; Parisi, G.; Picton, L.; Su, L.L.; Le, A.C.; Chhabra, A.; Silveria, S.L.; George, B.M.; King, I.C.; et al. Selective Targeting of Engineered T Cells Using Orthogonal IL-2 Cytokine-Receptor Complexes. *Science* **2018**, *359*, 1037–1042. <https://doi.org/10.1126/science.aar3246>.
92. Overwijk, W.; Tagliaferri, M.; Zalevsky, J. Engineering IL-2 to Give New Life to T Cell Immunotherapy. *Annu. Rev. Med.* **2021**, *72*, 281–311. <https://doi.org/10.1146/annurev-med-073118-011031>.
93. Saini, J.P.S.; Thakur, A.; Yadav, D. AI-driven Innovations in Pharmaceuticals: Optimizing Drug Discovery and Industry Operations. *RSC Pharm.* **2025**, *2*, 437–454. <https://doi.org/10.1039/d4pm00323c>.
94. Wang, Y.; Wei, T.; Zhao, M.; Huang, A.; Sun, F.; Chen, L.; Lin, R.; Xie, Y.; Zhang, M.; Xu, S.; et al. Alkenyl Oxindole Is a Novel PROTAC Moiety that Recruits the CRL4DCAF11 E3 Ubiquitin Ligase Complex for Targeted Protein Degradation. *PLoS Biol.* **2024**, *22*, e3002550. <https://doi.org/10.1371/journal.pbio.3002550>.

95. Shan, K.S.; Musleh Ud Din, S.; Dalal, S.; Gonzalez, T.; Dalal, M.; Ferraro, P.; Hussein, A.; Vulfovich, M. Bispecific Antibodies in Solid Tumors: Advances and Challenges. *Int. J. Mol. Sci.* **2025**, *26*, 5838. <https://doi.org/10.3390/ijms26125838>.
96. Wouters, O.J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *JAMA* **2020**, *323*, 844–853. Erratum in *JAMA* **2022**, *328*, 1111. <https://doi.org/10.1001/jama.2020.1166>.
97. Helma, J.; Cardoso, M.C.; Muyldermans, S.; Leonhardt, H. Nanobodies and Recombinant Binders in Cell Biology. *J. Cell Biol.* **2015**, *209*, 633–644. <https://doi.org/10.1083/jcb.201409074>.
98. Morsut, L.; Roybal, K.T.; Xiong, X.; Gordley, R.M.; Coyle, S.M.; Thomson, M.; Lim, W.A. Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* **2016**, *164*, 780–791. <https://doi.org/10.1016/j.cell.2016.01.012>.
99. Goldenzweig, A.; Fleishman, S.J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018**, *87*, 105–129. <https://doi.org/10.1146/annurev-biochem-062917-012102>.
100. Kim, C.; Chandrasekaran, A.; Jha, A.; Ramprasad, R. Active-Learning and Materials Design: The Example of High Glass Transition Temperature Polymers. *MRS Commun.* **2019**, *9*, 860–866. <https://doi.org/10.1557/mrc.2019.78>.
101. Hicks, D.R.; Kennedy, M.A.; Thompson, K.A.; DeWitt, M.; Coventry, B.; Kang, A.; Bera, A.K.; Brunette, T.J.; Sankaran, B.; Stoddard, B.; et al. *De Novo* Design of Protein Homodimers Containing Tunable Symmetric Protein Pockets. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2113400119. <https://doi.org/10.1073/pnas.2113400119>.