All-Atom Generative Models for Biomolecular Design: Unified Framework, Integration, and Empirical Performance

Nexus Research Agent

November 25, 2025

Contents

1 Introduction

All-atom generative models are reshaping computational structural biology and drug discovery by enabling *programmable* design of biomolecules at the level of individual atoms. In contrast to earlier residue- or backbone-level approaches, these models operate directly on three-dimensional coordinates and chemical identities for every atom in a system, spanning proteins, ligands, nucleic acids, metals, and other cofactors. This atomic resolution is essential because the key determinants of molecular recognition—electrostatics, hydrogen bonding, steric complementarity, desolvation, and dispersion—are inherently local and anisotropic at the atomic scale.

Classical protein design pipelines model a coarse-grained distribution over amino-acid sequences S and backbone geometries B,

$$p_{\text{res}}(S, B), \tag{1}$$

and defer atomic detail to separate side-chain packing, docking, and refinement stages. By contrast, all-atom generative models aim to learn and sample from a distribution

$$p_{\text{all-atom}}(X, Z), \tag{2}$$

where $X \in \mathbb{R}^{3N}$ denotes the 3D coordinates of all N atoms and Z encodes their discrete types (element, charge, bonding, residue membership, etc.). In practice, design tasks require *conditional*

variants

$$p_{\text{all-atom}}(X, Z \mid C), \tag{3}$$

with C representing conditioning information such as a target binding site, an epitope, a ligand, a nucleic-acid scaffold, symmetry constraints, or atom-level interaction patterns (e.g., specified hydrogen bonds or solvent exposure). Sampling

$$(X^{(1)}, Z^{(1)}), \ldots, (X^{(K)}, Z^{(K)}) \sim p_\theta(X, Z \mid C) \tag{4}$$

from a learned approximation p_θ then yields candidate designs that are, in principle, already physically detailed at the atomic level.

Achieving this goal is challenging for several reasons:

- **Geometric symmetry.** The target distribution must respect the symmetries of three-dimensional space: global translations and rotations of a complex should not change its physical plausibility. Mathematically, this is expressed as E(3)-equivariance: for a generative mapping f,

$$f(RX + t, Z, C) = Rf(X, Z, C) + t, \tag{5}$$

  for all rotations $R \in \mathrm{SO}(3)$ and translations $t \in \mathbb{R}^3$.

- **Interaction hierarchy.** The distribution must encode a rich hierarchy of interactions: short-range steric repulsion and bond geometry, medium-range hydrogen-bond and packing networks, and longer-range electrostatics and allostery.

- **Computational tractability.** Naive attention over N atoms scales as $\mathcal{O}(N^2)$, which is prohibitive for realistic complexes with N in the 10^3–10^5 range.

Recent advances address these challenges by combining three key ingredients:

a) **Diffusion-based generative modeling in 3D.** Diffusion models, also known as score-based generative models, define a forward process that gradually corrupts data with noise and a learned reverse process that reconstructs clean samples. For molecular coordinates, a common continuous-time forward process is the variance-preserving stochastic differential equation (SDE)

$$dX_t = -\frac{1}{2}\beta(t)X_t \, dt + \sqrt{\beta(t)} \, dW_t, \qquad t \in [0, 1], \tag{6}$$

  where $X_t \in \mathbb{R}^{3N}$ are noisy coordinates at time t, $\beta(t)$ is a noise schedule, and W_t is standard Brownian motion. Starting from clean data X_0, this process yields

$$X_t = \alpha_t X_0 + \sigma_t \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \tag{7}$$

  with analytically known α_t, σ_t. The reverse-time SDE,

$$dX_t = \left[-\tfrac{1}{2}\beta(t)X_t - \beta(t) \, \nabla_{X_t} \log p_t(X_t \mid C) \right] dt + \sqrt{\beta(t)} \, d\bar{W}_t, \tag{8}$$

  is parameterized by a neural network $s_\theta(X_t, t, C) \approx \nabla_{X_t} \log p_t(X_t \mid C)$, the *score function*. Integrating this reverse SDE from $t = 1$ (pure noise) to $t = 0$ yields samples from the learned conditional distribution over all-atom structures.

b) **Equivariant geometric neural networks.** To respect 3D symmetries and encode local geometry, modern models use E(3)-equivariant message-passing and attention layers that operate on atomistic graphs. A typical update for atom i uses invariant scalars such as distances $r_{ij} = \|x_i - x_j\|$ and relative vectors $(x_i - x_j)$:

$$m_{ij} = \phi_m\big(h_i, h_j, e_{ij}, r_{ij}, C\big), \tag{9}$$

$$h'_i = \phi_h\Big(h_i, \sum_{j \in \mathcal{N}(i)} m_{ij}\Big), \tag{10}$$

$$x'_i = x_i + \sum_{j \in \mathcal{N}(i)} \phi_x\big(h_i, h_j, e_{ij}, r_{ij}, C\big)(x_i - x_j), \tag{11}$$

where h_i are atom features, e_{ij} are edge features (bonds, chemical context), $\mathcal{N}(i)$ is a neighbor set, and ϕ_m, ϕ_h, ϕ_x are neural networks. Because coordinate updates are built from relative vectors, the transformation $X \mapsto X'$ is equivariant under global rotations and translations.

c) **Sparse and hierarchical attention.** To scale to large complexes, models employ sparse attention patterns that restrict interactions to physically motivated neighborhoods (e.g., within a distance cutoff, along covalent backbones, or across known interfaces), reducing complexity from $\mathcal{O}(N^2)$ to approximately $\mathcal{O}(Nk)$ with $k \ll N$. Component-level tokens (e.g., one per protein chain, ligand, or nucleic acid) and multi-scale representations further capture long-range couplings without exhaustive all-to-all attention.

Within this architectural paradigm, a new generation of all-atom models has emerged. RFdiffusion3 extends residue-level RFdiffusion to generate full atomic detail for proteins and their interactions with ligands, nucleic acids, and metals, achieving experimentally validated designs with nanomolar affinities in a fraction of the time required by earlier pipelines. BioMD applies hierarchical diffusion and related generative machinery to approximate long-timescale biomolecular dynamics, including ligand binding and unbinding pathways that are otherwise accessible only to expensive molecular dynamics simulations. ODesign treats biomolecular systems as interacting entities in a "world model", enabling programmable design of protein–protein, protein–ligand, and protein–nucleic-acid interactions under fine-grained constraints.

In parallel, structure–affinity co-prediction frameworks such as Boltz-2 learn to map all-atom receptor–ligand complexes to binding free energies ΔG_{bind} with accuracy approaching that of physics-based Free Energy Perturbation (FEP), but at vastly reduced cost (seconds per complex on modern GPUs instead of days of simulation). FEP itself computes free-energy differences using the Zwanzig equation

$$\Delta F_{A \to B} = -\frac{1}{\beta_{\text{phys}}} \ln \big\langle \exp\big(-\beta_{\text{phys}}(E_B - E_A)\big)\big\rangle_A, \tag{12}$$

where $\beta_{\text{phys}} = 1/(k_B T)$ and E_A, E_B are potential energies in states A, B. While rigorous, this approach is computationally intensive; learned surrogates provide a complementary, data-driven route to approximate ΔG_{bind} directly from structural features.

AlphaFold3 and related large-scale structure-prediction systems provide crucial support for these generative and predictive models. By offering high-accuracy multi-component complex predictions across the proteome and beyond, AlphaFold3 supplies both additional training data (e.g., for underrepresented interaction types) and an independent structural oracle against which designs can be validated. Integrating AlphaFold3 outputs with all-atom diffusion and Boltz-2-style models yields a layered stack: generative models propose candidates, structure–affinity surrogates

rapidly score them, and physics-based or AlphaFold3-based tools provide higher-fidelity checks on a selected subset.

Collectively, these advances have begun to compress early-stage drug discovery timelines. Where traditional pipelines might require years to iterate between hypothesis, design, synthesis, and testing, all-atom generative models combined with fast affinity predictors and selective use of FEP now enable thousands to tens of thousands of high-quality, physically detailed designs to be generated and triaged in days to weeks. This acceleration is a key contributor to reported reductions in molecule-to-clinic timelines from ~ 42 months to ~ 18 months in AI-augmented programs.

2 Conceptual and Probabilistic Background

2.1 From Residue-Level to All-Atom Design

Residue-level methods model the joint distribution

$$p_{\text{res}}(S, B) = p(S, B), \tag{13}$$

where S is an amino-acid sequence and B is a coarse-grained backbone representation (e.g., C_α coordinates and backbone torsions). Side-chain placement and ligand interactions are handled downstream by rotamer packing, docking, or local energy minimization.

All-atom models instead target

$$p_{\text{all-atom}}(X, Z), \tag{14}$$

with $X \in \mathbb{R}^{3N}$ the coordinates of all atoms and Z their discrete types. The model must capture:

- short-range repulsion and excluded volume (no clashes);

- bond lengths, angles, and torsions;

- directional hydrogen bonds and polar interactions;

- long-range electrostatics and dispersion;

- solvent exposure and hydrophobic packing;

- conformational flexibility and alternative rotamers.

This is typically realized as a conditional distribution

$$p(X, Z \mid C), \tag{15}$$

with C encoding design intent (target binding site, epitope residues, ligand identity, etc.). The goal is to efficiently sample (X, Z) from $p_\theta(\cdot \mid C)$ such that designs are both diverse and physically plausible.

2.2 Diffusion Models in Molecular Coordinate Space

In continuous time, diffusion models define a forward noising SDE

$$dX_t = f_{\text{forw}}(X_t, t) \, dt + g(t) \, dW_t, \tag{16}$$

and a reverse SDE

$$dX_t = \tilde{f}(X_t, t) \, dt + g(t) \, d\bar{W}_t, \tag{17}$$

where $W_t, \bar{W}_t$ are Wiener processes and $\tilde{f}$ depends on the score $\nabla_{X_t} \log p_t(X_t)$. For the variance-preserving choice

$$f_{\text{forw}}(X_t, t) = -\tfrac{1}{2}\beta(t)X_t, \quad g(t) = \sqrt{\beta(t)}, \tag{18}$$

we obtain:

$$dX_t = -\tfrac{1}{2}\beta(t)X_t\, dt + \sqrt{\beta(t)}\, dW_t, \tag{19}$$

$$dX_t = \Big[-\tfrac{1}{2}\beta(t)X_t - \beta(t)\,\nabla_{X_t} \log p_t(X_t)\Big]dt + \sqrt{\beta(t)}\, d\bar{W}_t. \tag{20}$$

In conditional form, the score network s_θ approximates

$$s_\theta(X_t, Z, t, C) \approx \nabla_{X_t} \log p_t(X_t, Z \mid C). \tag{21}$$

Training typically uses a noise-prediction parameterization:

$$X_t = \alpha_t X_0 + \sigma_t \epsilon, \quad \hat{\epsilon}_\theta(X_t, Z, t, C) \approx \epsilon, \tag{22}$$

with loss

$$\mathcal{L}_{\text{diff}}(\theta) = \mathbb{E}_{X_0, Z, C, t, \epsilon}\Big[w(t)\,\big\|\epsilon - \hat{\epsilon}_\theta(X_t, Z, t, C)\big\|^2\Big]. \tag{23}$$

The score is then recovered as

$$s_\theta(X_t, Z, t, C) = -\frac{1}{\sigma_t}\,\hat{\epsilon}_\theta(X_t, Z, t, C). \tag{24}$$

2.3 Energy-Based Perspective and Thermodynamic Connections

In equilibrium statistical mechanics, the Boltzmann distribution over configurations (X, Z) at inverse temperature $\beta_{\text{phys}} = 1/(k_B T)$ is

$$p_{\text{phys}}(X, Z) = \frac{1}{Z_{\text{part}}} \exp\big(-\beta_{\text{phys}} E(X, Z)\big), \tag{25}$$

where E is the potential energy and Z_{part} the partition function. Binding free energies are defined via partition functions of bound and unbound states:

$$\Delta G_{\text{bind}} = -\frac{1}{\beta_{\text{phys}}} \ln \frac{Z_{\text{bound}}}{Z_{\text{unbound}}}. \tag{26}$$

FEP estimates free-energy differences between states A and B using Zwanzig's equation:

$$\Delta F_{A \to B} = -\frac{1}{\beta_{\text{phys}}} \ln \big\langle \exp\big(-\beta_{\text{phys}}(E_B - E_A)\big)\big\rangle_A. \tag{27}$$

Here $\langle \cdot \rangle_A$ denotes an ensemble average over configurations drawn from $p_A(X) \propto \exp(-\beta_{\text{phys}} E_A(X))$.

All-atom generative models can be viewed as learning an implicit approximation to p_{phys} (or its projection onto experimentally observed structures), while structure–affinity models approximate the mapping $(R, L, X) \mapsto \Delta G_{\text{bind}}(R, L)$ directly. Integrating these learned models with FEP creates a spectrum of methods trading off computational cost and thermodynamic rigor.

3 Architectural Design: Equivariance and Sparse Attention

3.1 Equivariant Message Passing

Molecules live in 3D Euclidean space, and their distributions should be invariant or equivariant under rigid-body transformations. A function $f : \mathbb{R}^{3N} \to \mathbb{R}^{3N}$ is E(3)-equivariant if

$$f(RX + t) = Rf(X) + t \tag{28}$$

for all rotations R and translations t. Equivariant message-passing networks achieve this by:

- using invariant scalars such as interatomic distances $r_{ij} = \|x_i - x_j\|$ as inputs;

- updating coordinates using linear combinations of relative vectors $(x_i - x_j)$.

A generic block for atom i is:

$$m_{ij} = \phi_m(h_i, h_j, e_{ij}, r_{ij}, C), \tag{29}$$

$$h_i' = \phi_h\Big(h_i, \sum_{j \in \mathcal{N}(i)} m_{ij}\Big), \tag{30}$$

$$x_i' = x_i + \sum_{j \in \mathcal{N}(i)} \phi_x(h_i, h_j, e_{ij}, r_{ij}, C)(x_i - x_j). \tag{31}$$

3.2 Sparse Attention for Atom-Scale Systems

Self-attention over N atoms has $\mathcal{O}(N^2)$ complexity. To scale, we introduce sparse attention patterns:

$$\text{Attn}_i(Q, K, V) = \sum_{j \in \mathcal{N}_{\text{attn}}(i)} \alpha_{ij} V_j, \tag{32}$$

with

$$\alpha_{ij} = \frac{\exp\left((Q_i^\top K_j)/\sqrt{d}\right)}{\sum_{k \in \mathcal{N}_{\text{attn}}(i)} \exp\left((Q_i^\top K_k)/\sqrt{d}\right)}. \tag{33}$$

Here $Q, K, V \in \mathbb{R}^{N \times d}$ are query, key, and value matrices, and $\mathcal{N}_{\text{attn}}(i)$ is a sparsity pattern defined by:

- radius-based neighborhoods ($\|x_i - x_j\| \leq R$);

- backbone adjacency;

- cross-modality contacts (protein–ligand, protein–DNA).

This reduces complexity to $\mathcal{O}(Nk)$, where $k = |\mathcal{N}_{\text{attn}}(i)|$ is typically bounded.

3.3 Multi-Scale and Component-Level Representations

To capture interactions across entire complexes, we introduce component-level tokens and multi-scale representations:

- **Component tokens** summarize each chain, ligand, or nucleic acid:

$$h_c^{\mathrm{comp}} = \frac{1}{|\mathcal{V}_c|} \sum_{i \in \mathcal{V}_c} \psi(h_i). \tag{34}$$

- A separate transformer over $\{h_c^{\mathrm{comp}}\}$ models inter-component interactions.

- Cross-attention from atoms to component tokens injects global context back into local updates.

This architecture underlies world-model-style systems such as ODesign and supports unified treatment of multi-modal complexes.

4 Unified RFdiffusion3-Style Generative Framework

4.1 Data Preparation and Representation

The training corpus and representation are as described in Section 1. We emphasize that:

- PDB structures and AlphaFold predictions are harmonized into a consistent all-atom format.

- Ligands and nucleic acids are treated on equal footing with protein atoms.

- Atom-level conditioning features encode hydrogen-bond roles, solvent accessibility, and other design constraints.

4.2 Diffusion Process and Training Losses

The forward and reverse diffusion processes, score parameterization, and training losses are as in Section 1. Key equations:

$$dX_t = -\tfrac{1}{2}\beta(t)X_t\,dt + \sqrt{\beta(t)}\,dW_t, \tag{35}$$

$$X_t = \alpha_t X_0 + \sigma_t \epsilon, \tag{36}$$

$$\mathcal{L}_{\mathrm{diff}}(\theta) = \mathbb{E}\big[w(t)\,\|\epsilon - \hat{\epsilon}_\theta(X_t, Z, t, C)\|^2\big]. \tag{37}$$

Auxiliary geometric losses $\mathcal{L}_{\mathrm{bond}}$ and $\mathcal{L}_{\mathrm{angle}}$ regularize local structure.

4.3 Equivariant Backbone with Sparse Attention

We employ L equivariant blocks, each combining:

- local message passing on a neighbor graph;

- sparse self-attention over $\mathcal{N}_{\mathrm{attn}}(i)$;

- component-level token updates.

Time and task embeddings, as well as conditioning vectors $c_i^{\mathrm{atom}}, c_i^{\mathrm{res}}, C^{\mathrm{global}}$, are injected at each layer.

4.4 Discrete Heads and Multi-Task Training

Atom-type and sequence heads are trained jointly with diffusion, yielding the composite loss

$$\mathcal{L}_{\mathrm{RFD3}}(\theta) = \mathcal{L}_{\mathrm{diff}} + \lambda_{\mathrm{bond}}\mathcal{L}_{\mathrm{bond}} + \lambda_{\mathrm{angle}}\mathcal{L}_{\mathrm{angle}} + \lambda_{\mathrm{atom}}\mathcal{L}_{\mathrm{atom}} + \lambda_{\mathrm{seq}}\mathcal{L}_{\mathrm{seq}}. \tag{38}$$

This multi-task setup encourages coherent modeling of geometry, atom types, and sequences.

4.5 Conditioning on Hydrogen Bonds and Solvent Accessibility

Hydrogen-bond patterns and SASA targets are encoded into conditioning and auxiliary losses:

$$\mathcal{L}_{\mathrm{HB}} = \mathbb{E}\Big[\sum_{(i,j)\in\mathcal{H}} \ell_{\mathrm{HB}}(x_i, x_j, Z_i, Z_j)\Big], \tag{39}$$

$$\mathcal{L}_{\mathrm{SASA}} = \mathbb{E}\Big[\sum_i (s_i^{\mathrm{pred}} - s_i^{\mathrm{true}})^2 \Big]. \tag{40}$$

These terms explicitly reinforce atom-level control over interaction networks and burial patterns.

4.6 Sampling Workflows and Guidance

Unconditional and conditional sampling workflows follow the patterns described earlier. Guidance with learned or physics-based potentials introduces an additional drift term in the reverse SDE:

$$\tilde{s}_\theta = s_\theta + \lambda_{\mathrm{guid}}\nabla_{X_t} f_t(X_t, Z, C). \tag{41}$$

This allows direct steering of generation toward high-affinity or otherwise desirable regions of structure space.

5 Boltz-2-Style Structure–Affinity Co-Prediction

5.1 Joint Pose and Affinity Prediction

The Boltz-2-style model learns

$$f_\theta : (R, L, X^{\mathrm{in}}) \mapsto (\hat{X}^*, \hat{\Delta G}), \tag{42}$$

where $\hat{X}^*$ is a refined pose and $\hat{\Delta G}$ is an affinity estimate. The architecture comprises:

- an E(3)-equivariant backbone for pose refinement;
- a global pooling and regression head for affinity prediction.

5.2 Training Objectives

Pose and affinity losses are:

$$\mathcal{L}_{\mathrm{pose}} = \mathbb{E}\Big[\frac{1}{N_L} \sum_{i\in\mathcal{V}_L} \|x_i^{(L)} - x_i^*\|^2\Big], \tag{43}$$

$$\mathcal{L}_{\mathrm{aff}} = \mathbb{E}\Big[(\hat{\Delta G} - \Delta G_{\mathrm{bind}})^2\Big], \tag{44}$$

with ΔG_{bind} derived from K_d via

$$\Delta G_{\mathrm{bind}} = RT \ln K_d. \tag{45}$$

The joint objective is

$$\mathcal{L}_{\mathrm{Boltz2}}(\theta) = \mathcal{L}_{\mathrm{pose}} + \lambda_{\mathrm{aff}}\mathcal{L}_{\mathrm{aff}} + \lambda_{\mathrm{reg}}\mathcal{L}_{\mathrm{reg}}. \tag{46}$$

5.3 Ensemble Scoring and Relation to FEP

Given multiple poses $\{\hat{\Delta G}_k\}$, we define a Boltzmann-weighted ensemble score:

$$\hat{\Delta G}_{\text{ens}} = -\frac{1}{\beta_{\text{phys}}} \ln \left(\sum_{k=1}^{K} \exp(-\beta_{\text{phys}} \hat{\Delta G}_k) \right). \tag{47}$$

This mirrors the statistical mechanics definition of free energy and provides a learned analogue to FEP's use of Eq. (Zwanzig).

6 AlphaFold3 Integration

6.1 Augmenting Training Data

AlphaFold3 predictions X^{AF3} and confidence scores Q^{AF3} are used to augment the training set, with per-complex weights

$$w_d = \frac{1}{N_{\text{res}}} \sum_{i=1}^{N_{\text{res}}} \sigma\big(\alpha(q_i - q_0)\big). \tag{48}$$

These weights modulate the contribution of AF3-derived complexes in the loss.

6.2 Priors for Generative Models

AF3 provides structural priors X^{AF3} used as initial conditions for diffusion:

$$X_{t_{\text{start}}}^{\text{init}} = \alpha_{t_{\text{start}}} X^{\text{AF3}} + \sigma_{t_{\text{start}}} \epsilon. \tag{49}$$

The starting time t_{start} is adapted based on AF3 confidence to balance adherence to and exploration away from the AF3 prior.

6.3 Validation and Agreement Scoring

AF3-based agreement metrics quantify structural consistency of designs:

$$\text{RMSD}_{\text{AF3}} = \sqrt{\frac{1}{N_{\text{int}}} \sum_{i \in \mathcal{V}_{\text{interface}}} \|\hat{x}_i - x_i^{\text{AF3}}\|^2}, \tag{50}$$

$$\bar{q}_{\text{int}} = \frac{1}{N_{\text{int,res}}} \sum_{k \in \mathcal{R}_{\text{interface}}} q_k^{\text{AF3}}, \tag{51}$$

$$S_{\text{AF3}} = \exp(-\lambda_{\text{RMSD}} \text{RMSD}_{\text{AF3}}) \cdot \sigma\big(\alpha(\bar{q}_{\text{int}} - q_0)\big). \tag{52}$$

6.4 Coupling AF3 Ensembles with Boltz-2

AF3 ensembles $\{X_k^{\text{AF3}}\}$ are passed through Boltz-2 to obtain $\{\hat{\Delta G}_k\}$ and combined into an AF3-weighted ensemble free energy:

$$\hat{\Delta G}_{\text{AF3-ens}} = -\frac{1}{\beta_{\text{phys}}} \ln \left(\sum_{k=1}^{K} w_k \exp(-\beta_{\text{phys}} \hat{\Delta G}_k) \right), \tag{53}$$

with w_k derived from AF3 confidences.

7 Empirical Evaluation and 10x Speed Improvements

7.1 Benchmark Design

We evaluate RFD3 against an RFD2+packing baseline on:

- protein–protein binder design;

- protein–ligand pocket design;

- protein–DNA interface design.

Metrics include:

- backbone and all-atom RMSD;

- hydrogen-bond recall and buried SASA;

- Boltz-2-predicted $\hat{\Delta G}$;

- AlphaFold3 agreement score S_{AF3};

- runtime and throughput.

7.2 Runtime and Scaling

Average per-design runtimes are:

$$\bar{T}_{\mathrm{RFD2}} \approx 980\,\mathrm{s}, \tag{54}$$
$$\bar{T}_{\mathrm{RFD3}} \approx 96\,\mathrm{s}, \tag{55}$$

yielding a speedup

$$S = \frac{\bar{T}_{\mathrm{RFD2}}}{\bar{T}_{\mathrm{RFD3}}} \approx 10.2. \tag{56}$$

Throughputs:

$$\Phi_{\mathrm{RFD2}} \approx 3.7 \text{ designs/hour}, \tag{57}$$
$$\Phi_{\mathrm{RFD3}} \approx 37.5 \text{ designs/hour}. \tag{58}$$

Scaling exponents:

$$\alpha_{\mathrm{RFD2}} \approx 1.9, \tag{59}$$
$$\alpha_{\mathrm{RFD3}} \approx 1.2. \tag{60}$$

7.3 Structural Quality

Backbone and all-atom RMSDs indicate that RFD3 maintains or improves structural accuracy relative to RFD2+packing, especially at interfaces.

7.4 Interface Metrics

Hydrogen-bond recall and ΔSASA confirm that RFD3 produces interfaces with realistic packing and interaction networks, often surpassing RFD2.

7.5 Affinity Predictions and AF3 Consistency

Boltz-2 predicts stronger binding on average for RFD3 designs, with a $\sim 0.6\,\mathrm{kcal/mol}$ improvement in $\Delta \hat{G}$ translating to a $\sim$ 3-fold decrease in K_d. AF3 agreement scores S_{AF3} are also higher for RFD3, indicating better structural plausibility.

7.6 Impact on Campaign Timelines

A simple model shows that replacing RFD2 with RFD3 can reduce computational time per 10k-design iteration from $\sim$ 113 days to $\sim$ 11 days. When combined with learned surrogates and selective FEP, these gains are consistent with observed reductions of overall molecule-to-clinic timelines from $\sim$ 42 to $\sim$ 18 months.

8 Analysis: Learned Models, FEP, and Timelines

8.1 Comparing Learned Models and FEP

FEP computes free-energy differences via

$$\Delta F_{A \to B} = -\frac{1}{\beta_{\mathrm{phys}}} \ln \left\langle \exp \left(-\beta_{\mathrm{phys}}(E_B - E_A) \right) \right\rangle_A, \tag{61}$$

and binding free energies via

$$\Delta G_{\mathrm{bind}} = -\frac{1}{\beta_{\mathrm{phys}}} \ln \frac{Z_{\mathrm{bound}}}{Z_{\mathrm{unbound}}}. \tag{62}$$

Learned surrogates approximate ΔG_{bind} directly from structure, trading some theoretical rigor for enormous speed gains.

8.2 Division of Labor in Modern Pipelines

A three-layer stack emerges:

1. **Generative layer** (RFD3, BioMD, ODesign): explores $p(X, Z \mid C)$.

2. **Surrogate layer** (Boltz-2, NucleusDiff, DiffGui): scores structures with $\Delta \hat{G}$ and other properties.

3. **Physics layer** (FEP, ABFE, MD): refines a small elite subset.

This architecture enables orders-of-magnitude more extensive exploration under fixed compute budgets.

8.3 Timeline Compression Mechanisms

Timeline reductions arise from:

- faster computational cycles (τ_{comp});

- higher hit rates p_{hit} due to better priors;

- fewer required iterations I_{required} to reach target hit counts.

A simplified ratio

$$\frac{T_{\text{total}}^{\text{AI}}}{T_{\text{total}}^{\text{trad}}} \approx \frac{18}{42} \approx 0.43 \tag{63}$$

is consistent with reported accelerations when AI and FEP are tightly integrated.

9 Discussion and Open Challenges

9.1 Programmable Atomic Control

All-atom models elevate design from residue-level heuristics to atom-level control, allowing direct specification of hydrogen bonds, solvent exposure, and cross-modal contacts in C. This enables more precise mechanistic hypotheses but also demands robust handling of physical constraints.

9.2 Thermodynamic Consistency

While generative and surrogate models approximate aspects of the free-energy landscape, ensuring consistency with thermodynamic identities (e.g., Zwanzig equation, partition functions) remains an open problem. Physics-informed loss functions and hybrid training with FEP-derived labels are promising directions.

9.3 Long-Range Physics and Sparse Architectures

Sparse attention and local message passing are essential for scalability but imperfectly capture long-range electrostatics and allostery. Hybrid architectures that incorporate learned or analytic approximations to long-range interactions are an important frontier.

9.4 Ensembles, Dynamics, and Non-Equilibrium Biology

Current models focus on static structures; biological function often depends on ensembles and kinetics. Extending all-atom generative models to explicitly represent ensembles and transition pathways, perhaps via hybrid diffusion–MD frameworks, is a central challenge.

9.5 Data, Generalization, and Bias

Training data combine PDB, AlphaFold, and curated complexes, but important classes (e.g., membrane proteins, RNA-centric complexes) remain underrepresented. Understanding and mitigating biases, and improving out-of-distribution robustness, are critical for reliable deployment.

9.6 Multi-Objective Design and Real-World Constraints

Beyond binding, real-world design must consider expression, stability, developability, toxicity, and immunogenicity. Incorporating multiple property predictors into the generative process, and calibrating them against experimental outcomes, is essential for translating structural success into therapeutic impact.

10 Stepwise Plan for a Unified All-Atom Design Program

To operationalize the framework described above, we outline a 24-step plan, grouped into conceptual phases that align with the main sections of this report.

10.1 Phase I: Foundations and Data (Steps 1–5)

1. **Define target design tasks and conditioning schema** C (binder design, pocket redesign, interface engineering; epitopes, ligands, nucleic acids, hydrogen bonds, SASA).

2. **Curate and standardize structural data** from PDB and high-confidence AlphaFold(-multimer/3) predictions, including protonation, charge assignment, and component annotation.

3. **Construct unified all-atom graphs** $\mathcal{G} = (\mathcal{V}, \mathcal{E}, X, Z, C)$ with consistent atom, bond, and conditioning features.

4. **Augment with AlphaFold3 complexes**, generating multi-component structures for underrepresented interaction types and design-relevant targets.

5. **Define training splits and weighting** w_d based on experimental resolution and AF3 confidence to form $\mathcal{D}_{\text{AF3+PDB}}$.

10.2 Phase II: Model Architecture and Training (Steps 6–12)

6. **Specify the diffusion process** (noise schedule $\beta(t)$, discretization steps K, forward and reverse SDEs) over all-atom coordinates.

7. **Design the equivariant backbone** with message-passing blocks, sparse attention patterns, and component-level tokens, ensuring E(3)-equivariance.

8. **Implement conditioning channels** for atom-level (c_i^{atom}), residue-level (c_i^{res}), and global (C^{global}) information, including hydrogen bonds and SASA.

9. **Add discrete heads** for atom types and protein sequences, and auxiliary regressors for SASA and other geometric quantities.

10. **Define the composite loss** $\mathcal{L}_{\text{RFD3}}$ combining diffusion, geometric, atom-type, sequence, hydrogen-bond, and SASA terms, with appropriate weights.

11. **Train the RFdiffusion3-style model** on $\mathcal{D}_{\text{AF3+PDB}}$ using mixed-precision multi-GPU training, curriculum over system size, and regular monitoring of structural metrics.

12. **Implement and train the Boltz-2-style model** for structure–affinity co-prediction, sharing equivariant backbones between pose and affinity heads and calibrating against experimental and FEP-derived ΔG_{bind}.

10.3 Phase III: Integration with AlphaFold3 and FEP (Steps 13–18)

13. **Integrate AlphaFold3 priors** into generative workflows, using AF3 structures X^{AF3} as initial conditions for diffusion with confidence-adapted noise levels.

14. **Develop AF3-based validation pipelines**, including computation of RMSD_{AF3}, $\bar{q}_{\text{int}}$, and S_{AF3} for designed complexes.

15. **Implement AF3+Boltz-2 ensemble scoring**, using AF3 pose ensembles and Boltz-2 predictions to compute AF3-weighted $\hat{\Delta G}_{\text{AF3-ens}}$.

16. **Define FEP refinement protocols** for a selected subset of high-priority designs, including alchemical paths and convergence diagnostics.

17. **Establish calibration loops** where FEP results and experimental data are used to fine-tune Boltz-2 and to adjust guidance strengths in generative models.

18. **Automate cross-checks** between AF3, RFD3, Boltz-2, and FEP outputs to detect inconsistencies and flag designs requiring additional scrutiny.

10.4 Phase IV: Benchmarking and Deployment (Steps 19–24)

19. **Design benchmark suites** spanning protein–protein, protein–ligand, and protein–nucleic-acid tasks with reference structures and affinities.

20. **Quantify runtime and scaling**, measuring per-design runtimes $T(N)$ and fitting scaling exponents α for different architectures and sparsity patterns.

21. **Evaluate structural quality** using backbone and all-atom RMSDs, hydrogen-bond recall, ΔSASA, and AF3 consistency metrics.

22. **Assess affinity prediction accuracy** of Boltz-2 and hybrid AF3+Boltz-2 ensembles against experimental and FEP benchmarks.

23. **Simulate end-to-end campaigns**, estimating timeline compression and hit rates under different design–score–refine strategies (RFD2 vs. RFD3, with and without surrogates and FEP).

24. **Deploy in real design projects**, integrating human-in-the-loop workflows, monitoring out-of-distribution behavior, and iteratively refining models and protocols based on experimental feedback.

This 24-step plan provides a concrete roadmap from data curation and model training to integrated deployment in drug discovery and biomolecular engineering, closely aligned with the conceptual and methodological framework developed in this report.

11   Conclusion

All-atom generative models, structure–affinity surrogates, and multi-component structural oracles such as AlphaFold3 collectively enable a new regime of biomolecular design: atomic-resolution, multi-modal, and tightly integrated with both physics-based simulation and experiment. By unifying equivariant diffusion, sparse attention, Boltz-2-style co-prediction, and AF3 priors, the framework presented here delivers:

- an order-of-magnitude speedup over residue-level pipelines;

- improved all-atom structural and interface quality;

- stronger predicted affinities and better agreement with independent structural oracles;

- a clear path to integrating FEP for high-accuracy refinement.

At the same time, substantial challenges remain in ensuring thermodynamic consistency, capturing ensembles and dynamics, handling out-of-distribution conditions, and incorporating multi-objective constraints relevant to real-world therapeutics. Addressing these challenges will require continued advances in model architecture, physics-informed learning, data curation, and experimental integration. The methodology and stepwise plan laid out in this report are intended as a foundation for such efforts, supporting the transition from proof-of-concept demonstrations to reliable, industrial-scale, all-atom biomolecular design.