Therapeutics, Pharmacology and Clinical Toxicology

Hypothesis

Ther Pharmacol Clin Toxicol
Volume: 23, Issue: 3, July–September 2025

Received: 11 August 2025
Accepted: 9 September 2025

A hypothesis-driven theoretical framework for a learnable geometric functional in mutation-conditioned ligand–receptor affinity prediction

Ioan-Matei **Rusu**[1], Juliana **Margineanu**[1, 2], Stefan-Rares **Maxim**[1], Radu **Magop**[1], George **Smau**[2], Razvan **Rotaru**[2], Ionel-Bogdan **Tamba**[1]*

[1] Prof. Ostin C. Mungiu Advanced Research and Development Center for Experimental Medicine – CEMEX, Grigore T. Popa University of Medicine and Pharmacy, Iasi, Romania
[2] Alexandru Ioan Cuza University, Iasi, Romania

*** Correspondence to:** Ionel-Bogdan Tamba, Prof. Ostin C. Mungiu Advanced Research and Development Center for Experimental Medicine – CEMEX, Grigore T. Popa University of Medicine and Pharmacy, 16 Universitatii Str., 700115, Iasi, Romania. E-mail: bogdan.tamba@umfiasi.ro

Abstract

Predicting ligand–receptor binding affinity remains a cornerstone of modern drug discovery, enabling the identification of potential therapeutic compounds with high potency and selectivity. We introduce a new, interpretable framework that decomposes the total affinity into a geometric component and an energetic correction. The geometric term is defined as a learnable functional over overlapping atomic density fields and orientation-aware kernels, providing a measure of 3D shape complementarity that is explicitly invariant to global rotations and translations. The energetic component is estimated *via* a lightweight, symmetry-preserving surrogate graph network, capturing effects such as electrostatic interactions and solvation while remaining modular and extensible. A mutation-conditioned message-passing mechanism enables adaptation to receptor variants without recomputing expensive quantum data, addressing challenges in personalized medicine and protein engineering. An active learning loop links model predictions to sparse DFT reference points, balancing physical accuracy and computational efficiency through uncertainty-guided selection. This article is submitted as a hypothesis framework paper: it proposes a conceptual, testable architecture (with a proposed validation roadmap) rather than reporting validated predictive performance.

Keywords: geometric component, ligand–receptor affinity prediction, energetic component

Introduction

In 2025, drug discovery relies increasingly on computational strategies to identify promising ligands prior to synthesis, significantly reducing time and costs associated with experimental screening. Central to this process is predicting the binding affinity (ΔG, K_d) between a small molecule (ligand) and its target receptor (typically a protein) — a quantity that dictates potency, selectivity, and overall therapeutic efficacy [1–8].

Classical approaches, such as molecular docking and molecular mechanics (*e.g.*, MM-PBSA/MM-GBSA), offer rapid approximations but often lack the precision needed for complex biomolecular interactions and can suffer from scoring-function bias and limited generalizability [9–11]. In contrast, quantum mechanical methods like Density

Functional Theory (DFT) provide high accuracy by modeling electron densities and energies, but at a prohibitive computational cost, limiting their scalability to large datasets or high-throughput screening [12, 13].

Recent advances in graph neural networks (GNNs), equivariant 3D neural architectures, and geometric deep learning have enabled data-driven models to approximate these relationships with remarkable efficiency, both from sequence and structure-level representations [1–4]. These models represent molecules as graphs, where atoms or residues are nodes and bonds or contacts are edges, learning to predict affinities from structural features. However, most deep learning architectures remain black boxes, offering limited interpretability into the underlying physical mechanisms. This opacity hinders trust in predictions and complicates integration with domain knowledge, especially in safety-critical settings [5].

Moreover, many existing models are trained on static complexes from curated datasets such as PDBbind and CASF-2016 [6–8], and often fail to generalize to:

- Mutated or engineered receptors (*e.g.*, variants arising in protein design or genomics-driven personalized medicine);
- New chemical series far from the training data;
- Data-poor regimes where high-quality experimental or quantum labels are sparse.

To address these limitations, we hypothesize that the total binding affinity can be represented as a sum of two interpretable contributions:

$$A_{\text{total}} = A_{\text{geom}} + A_{\text{ener}}$$

where: A_{geom} encodes 3D geometric complementarity (shape and orientation matching). A_{ener} accounts for energetic effects (electrostatics, van der Waals forces, solvation, polarization etc.).

This separation draws inspiration from physical chemistry principles, such as the decomposition into force fields and implicit-solvent models, and provides a route to physical interpretability and modular learning.

By making the geometric term a learnable yet analytically grounded functional, and the energetic term a lightweight surrogate, our approach aims to bridge data-driven AI with physics-based modeling. Concretely, we propose a hybrid framework that:

- Defines a learnable geometric functional over overlapping atomic Gaussian densities, allowing for differentiable optimization, explicit control over shape complementarity, and built-in rotational and translational invariance;
- Introduces a mutation-conditioned message-passing network that adapts embeddings locally to handle receptor variants efficiently, rather than treating mutations only as global features;
- Employs an active DFT feedback loop for selective quantum calibration, leveraging uncertainty estimates from ensembles to minimize computational overhead while ensuring accuracy;
- Maintains a physically interpretable decomposition by enforcing consistency between a geometric pseudo-potential and an energetic surrogate, with regularization to avoid degeneracy between the two contributions.

This work should be read as a *conceptual and theoretical framework* since we do not present numerical experiments. However, we specified the model family, training objective, symmetry considerations, and an evaluation plan on benchmarks such as PDBbind and CASF-2016. The objective is to provide a physically motivated blueprint for future implementations rather than a finalized predictive tool.

Related work

a. Structure-based deep learning for binding affinity

Structure-based deep learning for binding affinity prediction has progressed rapidly over the last decade, driven by improved structural databases and advances in representation learning [1–3, 5]. Many models operate directly on 3D structures from PDBbind and related datasets [6, 7], achieving competitive performance to or surpassing classical scoring functions on CASF-type benchmarks [8].

Graph neural networks (GNNs) encode atom-level or residue-level interactions through message passing and have achieved state-of-the-art performance on PDBbind core and refined sets. For instance, the Geometric Interaction Graph Neural Network (GIGN) incorporates 3D spatial information and physical interaction terms to model covalent and noncovalent interactions in a unified message-passing scheme [14]. SS-GNN adopts a simple-structured GNN with a distance-threshold-based graph, reducing graph size while maintaining high predictive accuracy [15]. GraphScoreDTA combines GNNs with Vina-inspired distance terms and bi-transport mechanisms to improve docking and scoring power simultaneously [16].

Hybrid models that incorporate physics-based constraints or energy terms have been proposed to improve interpretability and generalization. PIGNet, for example, predicts atom–atom

pairwise interaction energies *via* physics-informed equations parameterized by neural networks and sums them to obtain the total binding affinity [17]. Physics-guided neural networks couple implicit-solvent models with GNNs to better match experimental and simulation-based free energies [18]. These approaches show that embedding physical structure into the network can reduce overfitting and improve extrapolation to new chemotypes.

Recent works like the Edge-Enhanced Interaction Graph Network (EIGN) emphasize rich edge features and adaptive encoders for protein–ligand complexes, demonstrating competitive performance on PDBbind and CASF-2016 [19]. PLAIG proposes a generalized GNN-based framework that converts 3D protein–ligand complexes into graph-structured data and uses ensemble regressors to mitigate overfitting [20]. Such models highlight the trend toward more physically grounded and structurally rich neural architectures.

b. Handling mutations and data bias

Handling receptor mutations remains underexplored relative to wild-type modeling. GraphPPI introduces a pre-trained graph neural network augmented with gradient boosting trees to predict changes in protein– protein binding affinity upon mutations [21–23]. While successful in capturing mutation effects, such approaches typically treat mutations as features in a downstream predictor rather than integrating them directly into the local message-passing dynamics for protein–ligand interactions.

Concurrently, several studies have shown that benchmark leakage and dataset bias can inflate apparent performance of deep models and hinder generalization across targets and chemotypes. Recent reviews and benchmark analyses emphasize the need for physically motivated inductive biases, robust evaluation splits, and uncertainty quantification in affinity prediction [3–5]. Our formulation, which separates geometric and energetic contributions and explicitly models mutation-conditioned messages, aims to be compatible with improved, bias-reduced datasets and evaluation protocols.

c. Relation to equivariant GNNs and density-based methods

Equivariant neural networks for molecules and materials—including SchNet [24], DimeNet [25], EGNN [26], and SE(3)- and E(3)-Transformers [27, 28]—have demonstrated the importance of encoding 3D geometry in a way that respects physical symmetries. More recently, architectures such as TorchMD-Net and related models have leveraged equivariant message passing for molecular dynamics and potential energy surfaces [29]. These methods often represent interactions *via* learned continuous filters over interatomic distances and sometimes angular information, enforcing rotational and translational invariance or equivariance by construction.

Our geometric functional shares conceptual similarities with density-overlap and shape-complementarity approaches from classical molecular recognition, where overlapping Gaussian representations or smooth surface approximations are used to quantify 3D complementarity. By formulating A_{geom} as a learnable functional over overlapping atomic Gaussian densities, we bridge classical density-based methods and modern neural architectures. Unlike standard equivariant GNNs, which often rely on message passing over discrete graphs, the present framework starts from continuous density fields whose pairwise overlaps are modulated by a neural kernel. In addition, we explicitly maintain a separate energetic surrogate E_{sur} and enforce a consistency relation between the geometric and energetic contributions, which is less common in existing equivariant GNNs.

Finally, our use of DFT-calibrated active learning connects to Δ-learning and orbital-informed models (*e.g.*, OrbNet and related schemes) that augment lower-level or empirical methods with quantum mechanical corrections [30]. In our case, quantum data is employed selectively *via* uncertainty-driven acquisition, with the goal of aligning a physically interpretable decomposition rather than directly regressing total energies.

Theoretical framework and conceptual formulation

The following section formalizes the proposed model family at the level of a theoretical and conceptual specification. Any datasets, baselines, metrics, and benchmarking are presented later strictly as a proposed future validation strategy.

a. Overview and problem formulation

We consider a ligand L and a receptor R forming a complex $C = (L,R)$, with a given binding pose obtained from crystallography, docking, or structure prediction. We assume access to:

- Atomic coordinates $\{\mathbf{r}_i\}$ and features $\{t_i\}$ for all atoms in L and R;
- Mutation annotations for residues in R (*e.g.*, wild-type *vs.* mutant amino acids and positions);
- Experimental binding affinity labels A_{exp} for a subset of complexes;
- Optional high-fidelity quantum reference values A_{DFT} for a small set of complexes.

The learning task is to construct a model

$$(L,R,mutations) \mapsto A_{\text{pred}}(L,R) \approx A_{\text{total}}(L,R) \quad (1)$$

where Atotal is the true binding affinity (*e.g.*, $-\Delta G$ in kcal/mol or a transformed pK_d).

We decompose

$$A_{\text{pred}} = \alpha A_{\text{geom}} + \beta E_{\text{sur}} + b \quad (2)$$

where A_{geom} a geometric functional, E_{sur} an energetic surrogate output, α, β, b learned scalar parameters.

b. Geometric functional formulation

To capture shape complementarity in a smooth and differentiable manner, we represent each molecule by atomic Gaussian densities, approximating atomic volumes as continuous fields rather than discrete points. This approach is inspired by electron-density-based methods in quantum chemistry and the Gaussian product theorem, which allows analytic evaluation of overlap integrals in Kohn–Sham DFT with Gaussian basis sets [12, 13].

The density fields are defined as:

$$\rho L(r) = \sum_{i\in L} w_i exp(-||r - r_i||^2/(2\sigma_i{}^2)) \quad (3)$$

$$\rho R(r) = \sum_{j\in R} w_j exp(-||r - r_j||^2/(2\sigma_j{}^2)) \quad (4)$$

where w_i are weights (*e.g.*, functions of atomic type or hybridization); σ_i are widths (*e.g.*, related to van der Waals radii); $\mathbf{r}_i$ are atomic positions.

We define a continuous geometric affinity as:

$$A_{\text{geom}} = \int \rho L(r)\rho R(r) K_\theta(d(r;L,R))dr, \quad (5)$$

where K_θ is a learnable kernel (*e.g.*, a small neural network) depending on local distances d and possibly orientation-derived invariants. Intuitively, $\rho L \rho R$ encodes how much the two density fields overlap at position $\mathbf{r}$, and K_θ reweights this overlap based on local geometry and chemistry.

Using standard identities for products of Gaussians, the integral can be discretized analytically into a sum over atomic pairs (i, j):

$$A_{geom} \approx \sum_{i\epsilon L} \sum_{j\epsilon R} \omega_i \omega_j C(\sigma_i\sigma_j) exp[-\frac{r_{ij}^2}{2(\sigma_i^2+\sigma_j^2)}] \times \phi_\theta(g_{ij}, t_i, t_j) \quad (6)$$

where $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$; g_{ij} denotes a collection of rotationally and translationally invariant geometric quantities (*e.g.*, distances, possibly angle-based invariants constructed over local triplets); t_i, t_j are atomic features (*e.g.*, element type, partial charge), $C(\sigma_i,\sigma_j)$ is a normalization constant; ϕ_θ is a neural kernel (*e.g.*, an MLP) that modulates the Gaussian overlap based on contextual features.

A typical choice for the normalization is:

$$C(\sigma_i,\sigma_j) = (2\pi)^{-3/2}(\sigma_i^2+\sigma_j^2)^{-3/2} \quad (7)$$

which arises from the integral of the product of two 3D Gaussians. Because ϕ_θ depends only on invariant quantities g_{ij} and local features, A_{geom} is invariant to global rigid motions of the complex. The sum can be truncated with a cutoff radius r_c to maintain tractable complexity, using neighbor lists or cell lists to reduce the effective number of interacting pairs.

Detailed mathematical derivations, normalization constants, and gradient expressions underlying the geometric functional are provided in Appendix A (mathematical details and derivations).

c. Energetic component and consistency

The geometric term alone cannot capture long-range electrostatics, solvation, polarization effects, or subtle entropic contributions that are implicitly handled in methods like MM-PBSA/MM-GBSA and implicit solvent DFT calculations [10–12]. To account for these non-geometric effects, we introduce a secondary lightweight, symmetry-preserving GNN that acts as an energetic surrogate E_{sur}.

The complex C is represented as a graph whose nodes are atoms (or residues) and edges encode covalent bonds and noncovalent contacts. Node features may include element type, formal charge, partial charges, hybridization, solvent-accessible surface area, and simple environment descriptors (*e.g.*, degree, local density). Edge features can include interatomic distances, contact types (hydrogen bond, π–π stacking, salt bridge etc.), and bond orders.

A compact message-passing network processes this graph to output a scalar:

$$E_{\text{sur}} = f_{\text{GNN}}(C;\psi), \quad (8)$$

where ψ are parameters of the surrogate network.

To maintain invariance to global rigid motions, messages are restricted to depend on invariant quantities such as interatomic distances and internally constructed angle features, following the design principles of SchNet- or DimeNet-type architectures [24, 25].

To maintain interpretability of the decomposition $A_{\text{geom}} + A_{\text{ener}}$, we introduce a consistency loss that regularizes the relationship between A_{pred}, A_{geom}, and E_{sur}:

$$A_{pred} = \alpha A_{geom} + \beta E_{sur} + b, \quad (9)$$

$$\underline{A_{total}} = A_{geom} + f_\psi(E_{sur}), \quad (10)$$

$$\mathcal{L}_{consist} = (A_{pred} - \underline{A_{total}})^2 \quad (11)$$

where f_ψ is a possibly non-linear mapping (*e.g.*, a 1–2 layer MLP).

To limit degeneracy between A_{geom} and E_{sur} — where one term could, in principle, mimic the other — we additionally regularize the complexity of the energetic surrogate *via* weight decay or direct penalties on its variance:

$$\mathcal{L}_{reg}=\lambda_{sur} E_x (E_{sur}(x)^2) \quad (12)$$

encouraging E_{sur} to act primarily as a correction to a dominant geometric contribution when the data supports such a decomposition.

d. Mutation-conditioned message passing

Receptor mutations alter local structures and energies, requiring adaptive modeling that can generalize across variant receptors. We introduce mutation embeddings m_r for each affected residue r, which encode aminoacid changes (*e.g.*, wild-type *vs.* mutant identity, biochemical properties, and position). These embeddings can be initialized from amino-acid descriptors or pre-trained protein language models and are then fine-tuned jointly with the rest of the model.

Let R_{mut} denote the set of mutated residues in the receptor. For each receptor node i, we define:

$$m_i = \{m_r \quad if\ i \epsilon r \epsilon R_{mut}, 0 \quad otherwise. \quad (13)$$

Node features are initialized as:

$$h_i^{(0)} = [f_i, m_i], \quad (14)$$

where $\mathbf{f}_i$ includes standard structural and chemical descriptors (element, charge, SASA, secondary structure etc.).

Message passing in the receptor (and receptor–ligand interface) uses attention that is conditioned on mutation embeddings:

$$h_i^{(l+1)} = \sigma(\sum_{j\epsilon N(i)} \alpha_{ij}^{(l)} W_v h_j^{(l)}) \quad (15)$$

$$\alpha_{ij}^{(l)} \propto exp(a_\theta(h_i^{(l)}, h_j^{(l)}, m_i, m_j, e_{ij})) \quad (16)$$

where e_{ij} are edge features (*e.g.*, distance, contact type); a_θ is a small neural network.

In practice, the attention scores are constructed from invariant functions of geometric quantities (*e.g.*, distances and angles) to preserve symmetry. This allows the local propagation of mutation effects through the interaction graph, modifying effective interaction strengths in a learned, context-dependent way without recomputing expensive quantum properties for each variant. For mutation-focused tasks (*e.g.*, ΔG prediction), one can compute differences in predicted affinities between wild-type and mutant receptors using a shared parameterization of the underlying model.

e. Active learning with DFT feedback

To integrate high-fidelity quantum data sparingly, we consider an active learning strategy where the model queries DFT calculations only for the most informative complexes. Let D_{exp} denote experimental affinity data (*e.g.*, from PDBbind) and D_{DFT} denote a small set of complexes with reliable quantum reference values (*e.g.*, relative free energies from DFT-based protocols).

We maintain an ensemble of K models $\{f^{(k)}\}_{k=1}^{K}$ trained with different random seeds or bootstrap resampling. For a given unlabeled complex x, the predictive mean and variance are:

$$\mu(x) = \frac{1}{K}\sum_{k=1}^{K} f^{(k)}(x), \quad (17)$$

$$u(x) = \frac{1}{K-1}\sum_{k=1}^{K} (f^{(k)}(x) - \mu(x))^2 \quad (18)$$

where $u(x)$ serves as an uncertainty estimate in the spirit of deep ensembles [23].

High-uncertainty examples are prioritized for DFT evaluation.

The active learning loop proceeds as follows: this loop prioritizes complexes where the model is most uncertain, aiming to maximize information gain per DFT evaluation and gradually align the surrogate decomposition with quantum reference data.

f. Training objective

Let $A_{exp}(x)$ denote experiental affinities and $A_{DFT}(x)$ quantum-based references. The total loss combines contributions from experimental regression, quantum calibration, consistency, ranking, and regularization:

$$\mathcal{L} = \lambda_{exp} E_{x\epsilon D_{exp}} (A_{pred}(x) - A_{exp}(x))^2 + \\ + \lambda_{DFT} E_{x\epsilon D_{exp}} (A_{pred}(x) - A_{DFT}(x))^2 + \lambda_{cons} E_x \mathcal{L}_{consist}(x) + \\ + \lambda_{rank}\mathcal{L}_{rank} + \lambda_{sur}\mathcal{L}_{reg}, \quad (19)$$

where L_{rank} can be a margin-based loss enforcing correct ordering of affinities for ligand series against a given receptor, as in CASF-style scoring and ranking power evaluations [8].

The hyperparameters $\lambda_\cdot$ are tuned based on validation performance and desired trade-offs between geometric dominance, energetic corrections, and quantum calibration.

Proposed validation strategy

Because this manuscript is a theoretical contribution, we outline below a concrete and testable

validation roadmap for a future implementation, including dataset selection, baselines, evaluation protocols, and quantum-calibrated active-learning experiments (Figures 1–3).

a. Proposed datasets

PDBbind and CASF. The PDBbind database provides curated protein–ligand complexes with experimental binding affinities. Standard splits (general, refined, and core sets) are commonly used for training and evaluation [6, 7]. CASF-2016 offers a curated subset of 285 complexes with standardized scoring, ranking, docking, and screening power metrics [8].

Mutation datasets

To evaluate mutation-conditioned behavior, one could:

- Mine protein–ligand complexes with known point mutations from the PDB and literature;
- Construct synthetic receptor variants by introducing point mutations in binding pockets using modeling tools (*e.g.*, PyMOL, Rosetta, or AlphaFold-based protocols) followed by structural relaxation;
- Where available, leverage experimental datasets of binding affinity changes upon mutation.

Quantum reference data. For a small subset of complexes (*e.g.*, tens to hundreds), DFT calculations could be performed to estimate binding free energies or relative energy differences between poses and mutants. These would populate D_{DFT} and be used for calibration *via* L_{DFT} and active learning.

b. Proposed baselines

Baseline models should include:

- GIGN for geometric interaction modeling [14];
- SS-GNN as a simple yet strong GNN baseline on PDBbind [15];
- GraphScoreDTA with Vina-inspired distance terms [16];
- PIGNet and physics-guided neural networks for hybrid physics–ML comparison [17, 18];
- PLAIG and EIGN for recent structure-based GNNs with strong CASF performance [19, 20];
- GraphPPI and related mutation-focused models as references for mutation conditioning [21];
- Equivariant 3D GNNs such as SchNet, DimeNet, EGNN, and SE(3)-Transformers to benchmark the impact of the explicit geometric functional relative to end-to-end equivariant architectures [24–28].

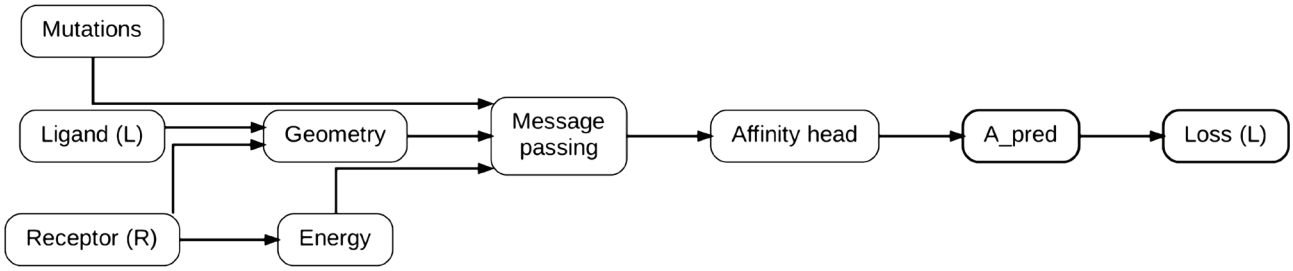

Figure 1. Model pipeline for affinity prediction.

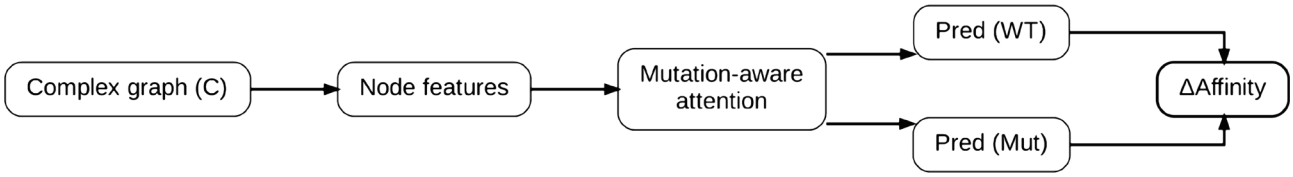

Figure 2. Mutation-aware attention and affinity difference estimation.

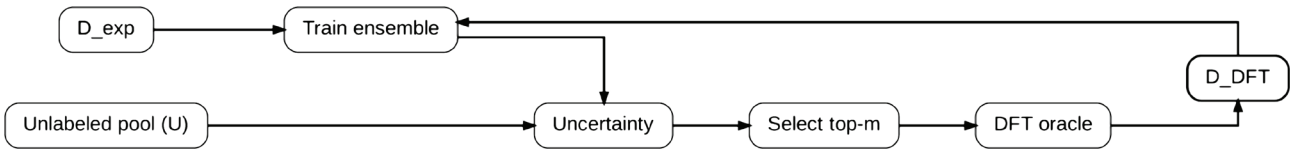

Figure 3. Iterative uncertainty-driven learning with DFT oracle.

c. Proposed metrics and evaluation protocols

Standard quantitative measures include RMSE and MAE for affinity prediction, Pearson and Spearman correlation coefficients for trend agreement, and CASF-style metrics for scoring, ranking, and docking power [8]. For mutation-focused experiments, one can additionally measure ΔG prediction error between wildtype and mutant complexes.

Evaluation protocols should follow community standards:

- Random splits and scaffold-based splits to assess generalization to novel chemotypes;
- Target-level splits to assess performance on unseen proteins;
- Mutation-based splits to test generalization from wild-type to mutants.

d. Ablation studies

Ablations would isolate the contribution of:

1. The geometric functional A_{geom} *vs.* standard GNN encodings;
2. Mutation conditioning *vs.* non-conditioned message passing;
3. Active learning with DFT *vs.* training only on experimental data;
4. Consistency and regularization losses *vs.* naive linear combination of A_{geom} and E_{sur}.

Discussion and outlook

The proposed decomposition into geometric and energetic terms offers several advantages:

- *Interpretability*: Researchers can inspect A_{geom} as a pseudo-potential of shape complementarity and E_{sur} as an energetic correction, visualizing where shape or energy dominate binding. This can be used to diagnose failure modes and generate mechanistic hypotheses;
- *Modularity*: New physical effects (*e.g.*, more elaborate implicit solvation models or entropic estimators) can be incorporated into E_{sur} without redesigning the geometric functional, or vice versa. The functional and surrogate can also be replaced or extended independently;
- *Extensibility*: The same framework can be adapted to different modalities (*e.g.*, RNA–ligand, protein– protein) by modifying density parameters and kernels, as well as by changing node and edge features in the surrogate network.

Our mutation-aware message passing provides a structured way to propagate the effect of amino-acid changes through local interaction neighborhoods. This is crucial for precision medicine applications where polymorphisms or engineered mutations significantly alter drug binding but remain underrepresented in training data. Importantly, the design is compatible with bias-reduced datasets and robust evaluation splits [3, 5].

The active learning loop offers a practical route to combining DFT-level accuracy with data-driven scalability. By querying quantum calculations only where the ensemble is uncertain, the framework can gradually align its learned decomposition with physically grounded references, echoing recent work on physics guided and uncertainty-aware models in computational chemistry [17, 18, 23].

Limitations

The current framework assumes effectively rigid molecular structures, potentially overlooking conformational flexibility that is important for induced fit and allostery. Future iterations could incorporate conformational ensembles (*e.g.*, *via* MD snapshots or multiple docking poses) and average A_{geom} and E_{sur} over ensembles. For very large proteins, computational complexity may require hierarchical approximations, cutoff schemes, or GPU-accelerated implementations of the pairwise Gaussian sums.

Moreover, the active learning loop assumes access to reliable and consistent DFT protocols. In practice, the choice of functional, basis set, and solvation model introduces its own uncertainties, and reconciling experimental and quantum scales is nontrivial. Finally, as with any data-driven method, training data bias and coverage remain critical: the framework should be combined with careful dataset curation and evaluation practices.

Ethical considerations

This work advances computational drug discovery, potentially accelerating the identification of therapeutic candidates, including for underserved diseases. However, predictions must be validated experimentally to mitigate the risk of false positives or misleading structure–activity trends. Wider ethical concerns include:

- Ensuring that training datasets reflect diverse targets and chemotypes to avoid reinforcing existing biases in therapeutic focus;
- Considering the downstream impact of accelerated design on accessibility and affordability of resulting therapies;
- Maintaining transparency regarding the theoretical nature of the method and its current limitations.

Conclusion

This study proposes a hypothesis-driven, physically motivated framework for ligand–receptor affinity prediction that decomposes binding into an interpretable geometric component and a complementary energetic correction. By formulating the geometric term as a learnable, symmetry-consistent functional over atomic density overlaps and integrating mutation-conditioned message passing with uncertainty-guided quantum calibration, the framework offers a modular and transparent alternative to end-to-end black-box models.

The work is intentionally conceptual and does not report validated predictive performance. Instead, it defines a coherent model family, learning objectives, and a concrete validation roadmap designed to support future empirical implementation and benchmarking. By bridging geometric deep learning, physics-informed modeling, and active learning, this framework provides a foundation for developing interpretable, mutation-aware affinity predictors applicable to drug discovery, protein engineering, and precision medicine.

Acknowledgments

Conflict of interest

The authors declare no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors used ChatGPT (GPT-5.1, OpenAI, San Francisco, CA; accessed August 2025) to support language refinement and improve the clarity and coherence of the manuscript. All content was subsequently reviewed and revised by the authors, who take full responsibility for the accuracy and integrity of the final work.

References

1. H. Wang, X. Zeng, et al. Prediction of protein–ligand binding affinity via deep learning models. Briefings in Bioinformatics, 25(2):bbae081, 2024.
2. D. D. Wang, Y. Wu, et al. Structure-based, deep-learning models for protein–ligand binding affinity prediction. Journal of Cheminformatics, 16:2, 2024.
3. Y. Wang, J. Cui, et al. Prediction of protein–ligand binding affinity with deep learning. Computational and Structural Biotechnology Journal, 21:1152–1173, 2023.
4. X. Zeng, C. Liu, et al. A comprehensive review of the recent advances on deep learning for drug–target affinity prediction. Briefings in Bioinformatics, 25(2):bbae017, 2024.
5. R. Meli, G. M. Morris, and P. C. Biggin. Scoring functions for protein–ligand binding affinity prediction using structure-based deep learning: a review. Frontiers in Bioinformatics, 2:885983, 2022.
6. Z. Liu, Y. Li, et al. PDB-wide collection of binding data: current status of the PDBbind database. Bioinformatics, 31(3):405–412, 2015.
7. Z. Liu, C. Wang, et al. Forging the basis for developing protein–ligand interaction scoring functions. Accounts of Chemical Research, 50(2):302–309, 2017.
8. M. Su, Q. Yang, et al. Comparative assessment of scoring functions: the CASF-2016 update. Journal of Chemical Information and Modeling, 59(2):895–913, 2019.
9. O. Trott and A. J. Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31(2):455–461, 2010.
10. S. Genheden and U. Ryde. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opinion on Drug Discovery, 10(5):449–461, 2015.
11. C. Wang, D. Greene, et al. Recent developments and applications of the MMPBSA method. Frontiers in Molecular Biosciences, 4:87, 2018.
12. W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. Physical Review, 140(4A):A1133–A1138, 1965.
13. J. A. Pople. Kohn–Sham density-functional theory within a finite basis set. Chemical Physics Letters, 1992.
14. Z. Yang, W. Li, et al. Geometric interaction graph neural network for predicting protein–ligand binding affinities from 3D structures. The Journal of Physical Chemistry Letters, 14(8):2020–2029, 2023.
15. S. Zhang, Y. Jin, et al. SS-GNN: a simple-structured graph neural network for affinity prediction. ACS Omega, 8:13045–13056, 2023.
16. K. Wang, H. Wu, et al. GraphScoreDTA: optimized graph neural network for protein–ligand binding affinity prediction. Bioinformatics, 39(6):btad340, 2023.
17. S. Moon, S. Ryu, et al. PIGNet: a physics-informed deep learning model toward generalized drug–target interaction predictions. Chemical Science, 13:8160–8171, 2022.
18. S. Cain, A. Risheh, and N. Forouzesh. A physics-guided neural network for predicting protein–ligand binding free energy: from host–guest systems to the PDBbind database. Biomolecules, 12(7):919, 2022.
19. D. Yang, J. Zhang, et al. Edge-enhanced interaction graph network for protein–ligand binding affinity prediction. PLOS ONE, 20(4):e0320465, 2025.
20. M. V. Samudrala, et al. PLAIG: Protein–ligand binding affinity prediction using a generalized graph neural network framework. ACS Biomedicine and Chemical Au, 2025.
21. X. Liu, Y. Luo, S. Song, and J. Peng. Pre-training of graph neural network for modeling effects of mutations on protein–protein binding affinity. arXiv preprint arXiv:2008.12473, 2020.
22. T. Voitsitskyi, R. Stratiichuk, et al. 3DProtDTA: a deep learning model for drug–target affinity prediction based on residue-level protein graphs. RSC Advances, 13:10261–10272, 2023.
23. B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In Advances in Neural Information Processing Systems, 30:6402–6413, 2017.
24. K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In Advances in Neural Information Processing Systems, 30:991–1001, 2017.
25. J. Klicpera, J. Groß, and S. Günnemann. Directional message passing for molecular graphs. In International Conference on Learning Representations, 2020.
26. V. Satorras, E. Hoogeboom, and M. Welling. E(n) Equivariant Graph Neural Networks. In International Conference on Machine Learning, 2021.
27. F. Fuchs, D. Worrall, V. Fischer, and M. Welling. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. In Advances in Neural Information Processing Systems, 33:1970–1981, 2020.

28. M. Thölke and G. De Fabritiis. TorchMD-Net: Equivariant Transformers for Neural Network Based Molecular Potentials. Journal of Chemical Theory and Computation, 18(5):3113–3123, 2022.
29. M. Thölke and G. De Fabritiis. TorchMD-Net: Equivariant molecular neural networks for molecular dynamics. arXiv preprint arXiv:2102.03150, 2021.
30. Z. Qiao, Y. Welborn, A. Anandkumar, F. Noé, and T. F. Miller III. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. Journal of Chemical Physics, 153(12):124111, 2020.

Appendix A. Mathematical details and derivations

A.1 Atomic densities and overlap

We start from Gaussian atomic densities, which provide a continuous representation suitable for overlap integrals:

$$\rho L(r) = \sum_{i \in L} w_i exp(-||r - r_i||^2/(2\sigma_i^2)) \quad (20)$$

$$\rho R(r) = \sum_{j \in R} w_j exp(-||r - r_j||^2/(2\sigma_j^2)) \quad (21)$$

The overlap integral without kernel simplifies to:

$$A_{geom} = \int \rho L(r)\, \rho R(r)\, dr \quad (22)$$

$$= \sum_{i,j} w_i w_j \int exp(-||r - r_i||^2/(2\sigma_i^2) - ||r - r_j||^2/(2\sigma_j^2))dr \quad (23)$$

Using Gaussian identities, the product of two Gaussians is another Gaussian centered at

$$\mu_{ij} = \frac{\sigma_i^2 r_j + \sigma_j^2 r_i}{\sigma_i^2 + \sigma_j^2}$$

with combined variance $\sigma_i^2 + \sigma_j^2$. Integration yields:

$$A_{geom} = \sum_{i,j} w_i w_j (2\pi)^{3/2} (\sigma_i^2 + \sigma_j^2)^{3/2} exp(-r_i j^2/(2(\sigma_i^2 + \sigma_j^2))) \quad (24)$$

where: $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$.

A.2 Normalization constant

To ensure proper scaling for 3D integrals:

$$C(\sigma_i,\sigma_j)=(2\pi)^{-3/2} (\sigma_i^2+\sigma_j^2)^{-3/2},$$

so that $C(\sigma_i,\sigma_j)^{-1}$ reproduces the Gaussian volume factor $(2\pi)^{3/2} (\sigma_i^2+\sigma_j^2)^{3/2}$.

A.3 Learnable kernel extension

Extending the basic overlap with a neural kernel:

$$A_{\text{geom}} = \sum_{i,j} w_i w_j C(\sigma_i,\sigma_j) exp(-r_{ij}^2/(2(\sigma_i^2+\sigma_j^2))) \phi_\theta(g_{ij},t_i,t_j).$$

where g_{ij} collects invariant geometric descriptors such as distances and angle-based features.

Here, ϕ_θ can be parameterized as a multi-layer perceptron taking inputs like g_{ij} and atomic features, allowing the model to learn complex interaction potentials while preserving invariance to global rigid motions.

A.4 Gradients

For optimization, positional gradients are:

$$\partial Ageom/\partial r_i = -\sum_{j \in R} w_i w_j S_{ij} (r_i - r_j)/(\sigma_i^2 + \sigma_j^2) \phi\theta(g_{ij}) +$$

$$+ \sum_{j \in R} w_i w_j S_{ij} (\partial \phi\theta/\partial r_{ij})(r_i - r_j)/r_{ij} \quad (25)$$

where $S_{ij}=C(\sigma_i,\sigma_j)exp(-r_{ij}^2/2(\sigma_i^2+\sigma_j^2))$. This enables gradient-based docking or optimization.

A.5 Energy surrogate and consistency

The surrogate prediction and loss:

$$A_{\text{pred}} = \alpha\, A_{\text{geom}} + \beta E_{\text{sur}} + b \quad (26)$$

$$\mathcal{L}_{\text{consist}} = (A_{\text{pred}} - (A_{\text{geom}} + f_\psi (E_{\text{sur}})))^2 \quad (27)$$

with f_ψ potentially a non-linear mapping to align scales.

The additional regularization L_{reg} constrains E_{sur} to act as a correction term.

A.6 Mutation conditioning

Initialization and update:

$$h_i^{(0)} = [f_i; m_i] \quad (28)$$

$$h_i^{(l+1)} = \sum_{j \in N(i)} Attn(h_i^{(l)}, h_j^{(l)}, m_i, m_j, e_{ij}) \quad (29)$$

where attention weights depend on mutation embeddings and invariant geometric features.

A.7 Active learning DFT loop

Selection and update:

$$x^* = arg\ max_{x \in U} u(x), \quad (30)$$

$$D_{\text{DFT}} \leftarrow D_{\text{DFT}} \cup (x^*, A_{\text{DFT}}(x^*)) \quad (31)$$

$$L_{\text{DFT}} = (A_{\text{pred}}(x^*) - A_{\text{DFT}}(x^*))^2 \quad (32)$$

Uncertainty $u(x)$ is based on ensemble predictive variance.

A.8 Total loss function

The combined objective:

$$\mathcal{L} = \lambda_{\text{exp}} \mathcal{L}_{\text{exp}} + \lambda_{\text{DFT}} \mathcal{L}_{\text{DFT}} + \lambda_{\text{cons}} \mathcal{L}_{\text{consist}} + \\ + \lambda_{\text{rank}} \mathcal{L}_{\text{rank}} + \lambda_{\text{sur}} \mathcal{L}_{\text{reg}} \quad (33)$$

where L_{exp} is derived from experimental data and L_{rank} enforces ranking of ligands per target.

A.9 Computational complexity and interpretation

The pairwise sum scales as $O(N_L N_R)$ but can be reduced using cutoff radii or approximate methods like neighbor lists or fast multipole expansions. A_{geom} acts as a pseudo-potential measuring 3D complementarity between ligand and receptor, interpretable *via* visualization of overlap fields; E_{sur} refines it energetically, allowing decomposition analysis and targeted model debugging.