Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

EQUIVARIANT NEURAL NETWORKS FOR PROTEIN–LIGAND POSE REFINEMENT USING ATOMIC COORDINATES AND INTERACTION ENERGIES

Ali Hassan[1]*, Noor Siddiqui[1], Bilal Khan[2], Sana Malik[1]

1. ***Department of Pharmaceutical Intelligence Systems, Faculty of Pharmacy, Aga Khan University, Karachi, Pakistan.***
2. ***Department of AI Drug Analytics, Faculty of Engineering, Qatar University, Doha, Qatar.***

ARTICLE INFO

Received:
12 May 2025
Received in revised form:
30 July 2025
Accepted:
05 August 2025
Available online:
28 August 2025

Keywords: Equivariant neural networks, Protein–ligand docking, Pose refinement, Geometric deep learning, Interaction energies, Structure-based drug design

ABSTRACT

Accurate protein–ligand binding poses are crucial for structure-based drug discovery, as downstream interpretation relies on plausible atomic contacts; while docking workflows can generate useful hypotheses, many poses remain geometrically imperfect and require refinement for confident use. Current post-docking refinement tools often rely on either computationally expensive physics-based relaxation or learned scoring models that fail to explicitly preserve three-dimensional molecular symmetry, limiting their reliability when input poses are rotated, translated, or structurally distinct from training examples. To address this, we propose an equivariant neural network that refines docked protein–ligand poses directly from atomic coordinates and interaction energies, ensuring that coordinate corrections respect rotational and translational symmetry while remaining informed by local protein–ligand physics. The protein–ligand complex is represented as a three-dimensional graph, with atoms as nodes and spatial contacts as edges; node features capture atomic identity and chemistry, while edge features encode distances, directions, and interaction energies such as electrostatics, van der Waals forces, and hydrogen bonds. Conceptually, this model is expected to guide docked ligands toward more physically plausible binding geometries without relying on fixed molecular orientations and can provide an interpretable quality score indicating whether the refined pose is suitable for downstream design. By combining symmetry-preserving architecture with energy-aware refinement, equivariant geometric deep learning offers a principled approach bridging rapid docking and computationally intensive physical methods, positioning this model as a valuable post-processing tool in modern drug discovery pipelines.

To Cite This Article: Hassan A, Siddiqui N, Khan B, Malik S. Equivariant Neural Networks for Protein–Ligand Pose Refinement Using Atomic Coordinates and Interaction Energies. Pharmacophore. 2025;16(4):22-31. https://doi.org/10.51847/dVUc4iFnva

Introduction

Accurate protein–ligand binding poses are central to structure-based drug design because predicted contacts guide virtual screening, hit interpretation, and lead optimization. Early convolutional scoring approaches showed that learned models can extract useful patterns from protein–ligand complexes [1], while deep affinity models such as KDEEP demonstrated the value of three-dimensional representations for binding assessment [2]. However, docking remains limited by imperfect sampling and simplified scoring, so generated poses may preserve approximate pocket occupancy while misplacing key hydrogen bonds, hydrophobic contacts, or ring orientations. A pose refinement model should therefore be treated not as a replacement for docking, but as a post-docking geometric corrector that improves structural plausibility.

Classical post-docking minimization can relax steric clashes, yet its behavior is strongly shaped by the chosen force field and the local basin around the input pose. Machine-learning scoring functions such as PotentialNet [3] and OnionNet [4] can learn interaction patterns from complex structures, but many scoring models primarily rank static configurations rather than directly generating coordinate corrections. Convolutional and graph-based tools, including GNINA [5] and graph convolutional drug–target models [6], have made learned scoring more useful for docking workflows, although their geometric treatment can still depend on representation choices. The core limitation is that a model may recognize a poor pose without knowing how ligand atoms should move in a symmetry-consistent way.

Corresponding Author: Ali Hassan; Department of Pharmaceutical Intelligence Systems, Faculty of Pharmacy, Aga Khan University, Karachi, Pakistan. E-mail: ali.hassan@gmail.com

Equivariant neural networks address this gap by ensuring that learned geometric quantities transform predictably under rotations and translations. Tensor field networks introduced a framework for rotation- and translation-equivariant learning on three-dimensional point clouds [7], while SchNet showed that continuous-filter neural networks can learn molecular properties from atomic coordinates [8]. Later atomistic architectures such as PhysNet [9] and GemNet [10] further emphasized the value of distance- and direction-aware message passing for molecular systems. These developments motivate a pose refiner that treats coordinate displacement as a geometric field rather than a scalar score appended to a docked complex.

The proposed MDL model accepts a docked protein–ligand complex and predicts both per-atom ligand displacements and a scalar quality score. Its design is inspired by geometric docking models such as EquiBind [11], TANKBind [12], DiffBindFR [13], and DynamicBind [14], which demonstrate that learned molecular geometry can be coupled to protein–ligand pose prediction. By incorporating interaction energies as edge features, the model would combine symmetry-preserving message passing with physically meaningful cues about clashes, attraction, and directional contacts. The resulting framework is intended to learn a displacement field that could move decoy poses toward crystal-like geometries while preserving the global rotational frame of the input.

Background

Protein–Ligand Pose Prediction and the Refinement Problem

Protein–ligand docking typically generates multiple candidate poses by sampling ligand conformations and placements inside a binding site, then ranking them with a scoring function. Learned docking and scoring systems have improved pose assessment, as illustrated by CNN-based scoring [1], GNINA [5], and virtual screening applications using learned docking workflows [15]. Nevertheless, even a high-ranked pose may contain local defects such as flipped heterocycles, strained torsions, missed hydrogen bonds, or steric clashes that are not fully resolved by the original docking algorithm. Pose refinement is therefore the task of taking a plausible but imperfect docked geometry and predicting targeted coordinate corrections rather than resampling the entire binding event.

Classical and Physics-Based Refinement

Physics-based refinement uses force-field energies, local minimization, or molecular dynamics to reduce unfavorable contacts and improve local geometry. Neural atomistic potentials such as PhysNet [9] and E(3)-equivariant interatomic potentials such as NequIP [16] show how learned force-like models can represent molecular energy landscapes, but direct physical relaxation can still be computationally demanding for large protein–ligand systems. Molecular dynamics and force-field minimization may also remain trapped near the initial pose when the docked ligand occupies an incorrect local basin. A learned refiner should therefore borrow the local physical intuition of energy minimization while remaining capable of data-driven geometric correction.

Machine-Learning Scoring Functions and Their Geometry Limitations

Machine-learning scoring functions evolved from engineered interaction descriptors toward deep models that process protein–ligand structures more directly. KDEEP [2], OnionNet [4], SE-OnionNet [17], and graph convolutional affinity predictors [18] show that learned features can capture useful patterns in intermolecular contacts. Yet many of these approaches are designed to predict affinity or classify pose quality rather than output equivariant coordinate updates. A pose refinement model must learn not only whether an atom is in an unfavorable local environment, but also the direction in which it should move under the same geometric symmetry constraints as the molecular system. **Figure 1** illustrates how a learned pose-refinement model can transform an initially plausible but locally imperfect docked complex into a geometry-corrected protein–ligand pose while preserving physical plausibility and E(3)-equivariant symmetry constraints.

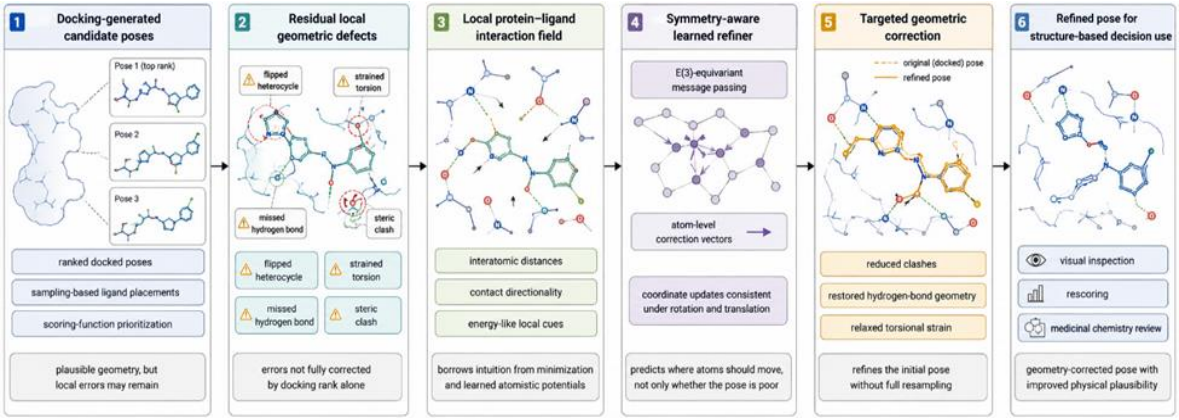

Figure 1. Conceptual workflow for symmetry-aware protein–ligand pose refinement after docking

Equivariant Neural Networks for 3D Geometric Data

Equivariant neural networks represent molecular structures in a way that respects rotations, translations, and sometimes reflections of three-dimensional space. Tensor field networks [7] and general theories of equivariant convolution [19] established core principles for constructing features that transform consistently under symmetry operations. Molecular implementations such as SchNet [8], GemNet [10], and large-scale equivariant atomistic models [20] show how scalar and directional information can be propagated across atomic graphs. For protein–ligand refinement, this means a predicted displacement vector can rotate with the input complex rather than depending on an arbitrary coordinate frame. Recent advances in equivariant and generative molecular modeling further strengthen the rationale for treating protein–ligand pose refinement as a geometry-aware correction task rather than only a scoring problem. DiffDock demonstrates that ligand docking can be formulated as a diffusion process over translational, rotational, and torsional degrees of freedom, supporting the broader idea that learned models can operate directly on pose geometry and provide confidence estimates for predicted binding modes [21]. E(n)-equivariant graph neural networks also show that graph-based models can preserve symmetry under rotations, translations, reflections, and permutations without requiring highly expensive tensor representations, making them relevant to scalable protein–ligand refinement architectures [22]. In addition, higher-order equivariant message passing methods such as MACE illustrate how expressive local geometric messages can improve atomistic force-field learning, which supports the use of richer directional and many-body information when refining molecular coordinates [23].

Incorporating Physics into Deep Learning for Molecular Interactions

Physics-informed molecular learning can use interaction energies, partial charges, distances, and directional contacts as structured inputs rather than forcing a network to infer all relevant forces from atom labels alone. PotentialNet [3] and mathematical deep learning approaches in docking challenges [24] illustrate how learned interaction models can exploit molecular graph structure, while force-aware neural potentials such as PhysNet [9] suggest that energy-related quantities can guide coordinate-sensitive prediction. Extended interaction features also show that chemically meaningful protein–ligand descriptors can improve binding assessment [25]. In an equivariant pose refiner, electrostatic, van der Waals, and hydrogen-bond terms can be embedded as scalar edge features that modulate vector message passing.

Model Development Overview

High-Level Refinement Framework

The proposed framework accepts a protein–ligand complex with docked ligand coordinates and constructs a three-dimensional graph over protein and ligand atoms. This design follows the broader movement from voxelized molecular learning [1, 2] toward graph-based and equivariant molecular modeling [7, 8]. Messages are passed through equivariant layers that update scalar chemical features and vector geometric features, allowing the network to reason about both atom identity and direction. The output consists of ligand-atom displacement vectors and a pose quality score, conceptually linking coordinate refinement with learned scoring in the spirit of docking-aware models such as GNINA [5].

Figure 2 illustrates the proposed equivariant pose-refinement workflow, showing how docked protein–ligand coordinates and interaction-energy features are transformed into symmetry-preserving ligand displacement vectors, a refined pose, and a downstream confidence signal for structure-based drug discovery.

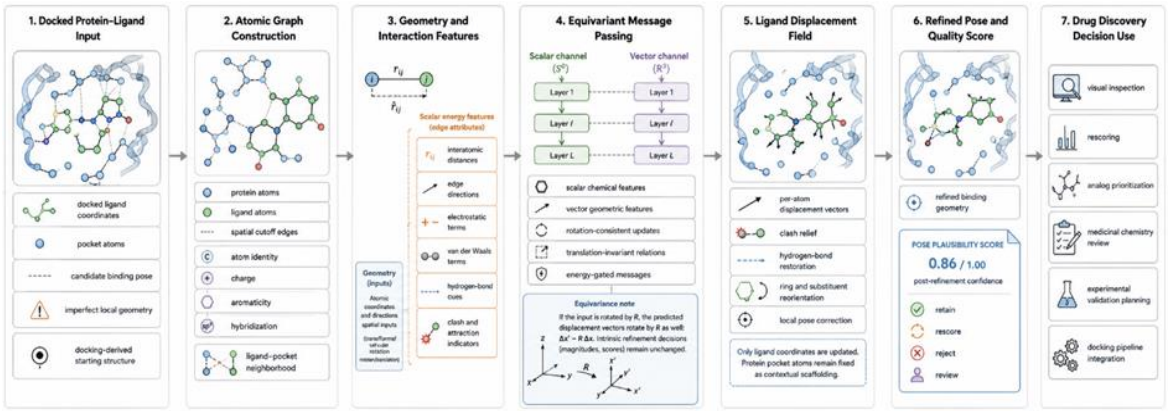

Figure 2. Equivariant Neural Network Workflow for Protein–Ligand Pose Refinement Using Atomic Coordinates and Interaction Energies

Core Input Representations

Each node represents a protein or ligand atom, with scalar attributes such as atom type, aromaticity, hybridization state, formal charge, and estimated partial charge. Edges are defined by spatial proximity, following the contact-based logic used in intermolecular neural models such as OnionNet [4] and SE-OnionNet [17], but extended with directional geometry for equivariant message passing. Edge scalars include interatomic distance, atom-pair category, and pre-computed interaction energies that can indicate attraction, repulsion, or directional complementarity. This representation allows the model to distinguish a close contact that is chemically favorable from one that is geometrically strained. **Table 1** summarizes the core

atom-level and edge-level representations used by the proposed equivariant docking refiner and explains how each input supports chemically meaningful coordinate correction.

Table 1. Compact Representation-to-Function Map for the Equivariant Docking Refiner

Representation component	Example encoded features	Role in the proposed refiner
Protein and ligand atom nodes	Atom type, aromaticity, hybridization, formal charge, estimated partial charge	Defines the chemical identity of each atom so the model can distinguish steric, electronic, and bonding-relevant environments.
Contact-based intermolecular edges	Spatial proximity links between protein and ligand atoms	Captures local binding-pocket contacts in a format consistent with intermolecular neural scoring models such as OnionNet [4] and SE-OnionNet [17].
Edge scalar descriptors	Distance, atom-pair category, pre-computed interaction energy	Helps separate favorable close contacts from repulsive or geometrically strained contacts.
Directional geometric information	Relative orientation and direction-aware contact vectors	Enables equivariant message passing so predicted coordinate corrections remain consistent under rotation or translation.
Ligand displacement output	Atom-wise coordinate correction field	Provides an interpretable refinement signal showing how each ligand atom should move to improve pose consistency.

Design Principles

The central design principle is that global rotation or translation of the entire protein–ligand complex should not change the intrinsic refinement decision. Equivariant architectures such as tensor field networks [7], NequIP [16], and other learned atomistic dynamics models [20] provide the conceptual basis for coordinate updates that transform consistently with the input geometry. The model is also physically grounded because energy-based edge features give the network explicit access to local interaction cues rather than relying only on learned embeddings. Finally, the output is interpretable at the atom level because each ligand displacement can be visualized as part of a learned correction field.

Data Sources and Preprocessing

Training Set Construction

Training would conceptually use curated protein–ligand crystal structures and corresponding docked decoys, with each decoy paired to a native pose to define a target displacement field. PDBbind-style resources are widely used in learned binding and docking models, including KDEEP [2], PotentialNet [3], and several protein–ligand graph models [18]. Docked poses could be generated by established docking tools and then treated as imperfect initial states for refinement, similar to the way GNINA integrates learned scoring into docking workflows [5]. The manuscript does not require experiments, but the data design should support future evaluation of whether the predicted displacement field improves structural plausibility.

Graph Construction and Feature Encoding

Graph construction begins from raw Cartesian coordinates, assigning nodes to all ligand atoms and selected nearby protein atoms in the binding environment. Atom features can follow conventions used in learned scoring and affinity prediction models [6, 18], while edge features can extend contact descriptors used in OnionNet-style architectures [4, 17]. Partial charges and force-field-inspired terms would be computed before message passing so that electrostatic and steric interactions are available as structured scalar inputs. Because the graph preserves spatial coordinates and edge directions, the model can encode both chemical identity and local three-dimensional arrangement.

Data Augmentation with Random Rotations

Random rotations of the entire protein–ligand complex can be used during development to verify that the model behaves consistently under changes of coordinate frame. Equivariant theory predicts that scalar quality scores should remain unchanged while vector displacement outputs should rotate with the input, as established by geometric learning principles [7, 19]. Molecular architectures such as SchNet [8] and GemNet [10] further support the use of coordinate-based representations that do not depend on arbitrary orientation. This augmentation should improve robustness without altering the physical meaning of the protein–ligand interaction.

Equivariant Network Architecture for Pose Refinement

Equivariant Message-Passing Layers

Each equivariant message-passing layer updates scalar and vector features by aggregating information from neighboring atoms in the three-dimensional graph. Tensor field networks provide a natural template in which edge directions and learned feature tensors are combined so that outputs transform correctly under rotation [7]. Directional molecular graph networks such as GemNet [10] similarly show why angular and directional information is important for molecular geometry. In the proposed refiner, messages from protein atoms to ligand atoms would encode how local contacts suggest attraction, repulsion, or reorientation of ligand substructures.

Coordinate Update Head

After the final equivariant layers, a coordinate update head maps learned vector features into a displacement vector for each ligand atom. This head follows the same symmetry logic as equivariant atomistic potentials, where vectorial outputs must rotate consistently with the molecular coordinate frame [16, 20]. Protein–ligand docking models such as EquiBind [11] and DiffBindFR [13] support the idea that learned geometric transformations can be applied directly to binding poses. Because the displacement is assembled from equivariant quantities, rotating the input pose would rotate the predicted correction rather than changing its chemical interpretation.

Quality Score Head

A separate scalar head aggregates ligand and pocket features into a pose quality score that estimates whether the refined geometry is likely to be reliable. This component is conceptually related to learned pose and affinity scoring models such as CNN scoring [1], KDEEP [2], GNINA [5], and mathematical deep learning approaches in docking benchmarks [24]. Unlike a standalone scorer, however, the quality head is coupled to a coordinate refiner, so it can assess the pose after a symmetry-preserving correction has been proposed. The score should be evaluated as a confidence indicator rather than presented as a guaranteed measure of binding affinity or refinement success.

Table 2 defines the architectural logic of the proposed energy-aware equivariant refiner by linking each model component to its symmetry role, molecular-refinement function, interpretive value, and implementation risk.

Table 2. Model Architecture Logic for Energy-Aware Equivariant Protein–Ligand Pose Refinement

Architectural component	Input information used	Symmetry role	Molecular-refinement function	Interpretive value	Main implementation risk
Docked protein–ligand complex	Initial ligand coordinates, protein pocket coordinates, docked pose geometry	Provides the spatial frame that must be handled consistently under rotation and translation	Defines the starting pose that will be corrected rather than fully regenerated	Makes the model explicitly post-docking rather than a replacement for docking	Poor starting poses may fall outside the correctable local refinement basin
Atomic node representation	Atom type, element, formal charge, aromaticity, hybridization, ligand/protein identity, partial charge	Scalar node features remain invariant to global coordinate transformations	Encodes chemical identity needed to distinguish polar atoms, hydrophobic atoms, ring systems, and charged groups	Allows atom-level diagnosis of which ligand regions drive correction	Incorrect protonation, tautomer state, or missing cofactors may distort learned corrections
Spatial edge construction	Protein–ligand and ligand–ligand contacts within a distance cutoff	Edges preserve relational geometry rather than absolute orientation	Identifies local interaction neighborhoods where clashes, missed contacts, or geometric strain may occur	Clarifies which pocket contacts influence each ligand atom	Cutoff selection may omit long-range electrostatics or include irrelevant crowded contacts
Directional geometric features	Interatomic vectors, distances, local directions, angular relations	Directional information supports equivariant vector message passing	Enables the network to infer not only whether a contact is poor, but also where the ligand atom should move	Supports visualization of correction direction as a learned geometric field	Directional features may become noisy for flexible ligands or uncertain pocket conformations
Interaction-energy edge features	Electrostatic, van der Waals, hydrogen-bond, steric clash, and attraction-related terms	Energy scalars modulate equivariant geometric messages without changing transformation rules	Provides physics-informed cues for correcting unfavorable local contacts and preserving favorable anchors	Helps separate chemically favorable proximity from steric or electrostatic conflict	Energy terms may depend strongly on force-field assumptions and preprocessing quality
Equivariant message-passing layers	Scalar chemical features, vector geometric features, energy-weighted edge features	Ensures vector outputs rotate consistently with the input complex and scalar outputs remain orientation-independent	Learns how local protein–ligand environments imply ligand-atom coordinate corrections	Provides a principled symmetry-preserving mechanism for pose refinement	High computational cost and architectural complexity may reduce scalability
Coordinate update head	Final ligand atom vector features	Produces ligand displacement vectors that transform consistently with molecular rotation	Moves ligand atoms toward more plausible binding geometry while respecting the input coordinate frame	Gives directly visualizable atom-level correction arrows	Large displacements may distort ligand geometry if internal constraints are not controlled
Pose quality score head	Aggregated ligand and pocket features after refinement	Produces an invariant scalar confidence signal	Estimates whether the refined pose should be trusted, rescored, rejected, or reviewed	Links geometric correction to practical decision support	The confidence score may be misread as binding affinity unless clearly defined
Human review and downstream use	Refined pose, displacement field, confidence score,	Not a symmetry mechanism, but a	Ensures the model informs, rather than replaces, medicinal	Supports transparent use in docking	Overreliance on refined poses may occur without

	structural plausibility checks	governance boundary for use	chemistry and structural biology judgment	pipelines and lead optimization	experimental or expert validation

Integrating Interaction Energies Into the Message-Passing Framework

Pre-Computed Interaction Energies as Edge Features

Before equivariant message passing, each protein–ligand atom pair within the interaction cutoff can be assigned scalar energy descriptors representing electrostatic, van der Waals, and hydrogen-bond contributions. Energy-aware molecular learning is motivated by models such as PhysNet, which learns energies and related molecular quantities from atomic structure [9], and by PotentialNet, which uses staged molecular graph propagation to encode interactions relevant to binding [3]. In the proposed refiner, these energy terms would not replace learned geometry but would act as physically interpretable edge attributes. The network could therefore distinguish a short contact caused by favorable complementarity from a short contact caused by steric conflict.

Learning to Weight Energy Contributions

The model can learn to modulate messages according to the physical character of each contact, giving stronger influence to edges that suggest clashes, unsatisfied polar contacts, or favorable anchoring interactions. Classical interaction-aware descriptors and learned binding models, including OnionNet [4], SE-OnionNet [17], and extended connectivity interaction features [25], show that protein–ligand contact patterns can be informative when encoded systematically. In an equivariant framework, these scalar energy cues can gate vector messages that propose local ligand displacements. This creates a conceptual bridge between force-field-like reasoning and data-driven coordinate correction.

Ablation of Physics vs. Learned Features

The architecture should be designed so that interaction-energy features can be removed during evaluation, allowing conceptual comparison between purely learned geometric refinement and physics-augmented refinement. Learned scoring models such as Pafnucy-style three-dimensional affinity prediction [26] and CNN-guided docking approaches [27] illustrate how deep models can exploit structural input, but they do not by themselves resolve whether explicit energy features improve geometric correction. An ablation framework would ask whether distance, atom identity, and direction alone are sufficient, or whether electrostatic and van der Waals terms provide a more stable refinement signal. Such analysis would help clarify whether the model learns chemically meaningful corrections rather than only memorizing local coordinate patterns.

Model Interpretability and Assessment of Refined Geometries

Visualising the Displacement Field

The predicted ligand-atom displacement vectors can be visualized directly on the input pose to show how the model proposes to resolve local geometric conflicts. Prior visualization work on convolutional protein–ligand scoring [28] demonstrated the value of inspecting learned spatial responses rather than treating docking networks as opaque black boxes. In the proposed model, arrows overlaid on ligand atoms could indicate whether the network is separating clashing groups, restoring directional hydrogen bonds, or rotating aromatic fragments into more plausible contact geometry. This atom-level interpretability is a natural advantage of predicting an equivariant vector field rather than only a scalar score.

Confidence Score as a Predictor of Refinement Success

The scalar quality score can serve as a confidence estimate for whether the refined pose should be retained, rescored, or rejected. Learned docking systems such as GNINA [5] and virtual screening workflows built around neural scoring [15] already show how model-derived scores can be used to prioritize poses, while PoseBusters emphasizes the need to assess whether predicted complexes are physically valid rather than merely plausible to a scoring function [29]. In this framework, a low score after refinement would indicate uncertainty about the corrected geometry or possible failure of the learned displacement field. The score should therefore be interpreted as a decision-support signal for downstream review, not as a standalone claim of binding affinity.

Integration Into Structure-Based Drug Design Pipelines

Post-Processing Tool for Docking Workflows

The proposed refiner can be positioned after conventional docking, where it would consume raw docked poses and return symmetry-consistent coordinate corrections together with confidence scores. CNN-guided docking [27] and GNINA-based docking [5] provide precedents for integrating learned models into practical docking pipelines, while newer geometric docking systems such as EquiBind [11], TANKBind [12], and FlowDock [30] show that learned protein–ligand geometry can be used more directly. A post-processing role is especially attractive because it does not require replacing existing docking engines. Instead, the model could act as an energy-aware geometric correction layer before visual inspection, rescoring, or medicinal chemistry review. **Figure 3** illustrates how predicted ligand-atom displacement vectors, refinement-confidence scoring, and downstream docking-pipeline decisions can be integrated to assess whether a corrected protein–ligand pose should be retained, rescored, visually reviewed, or rejected.

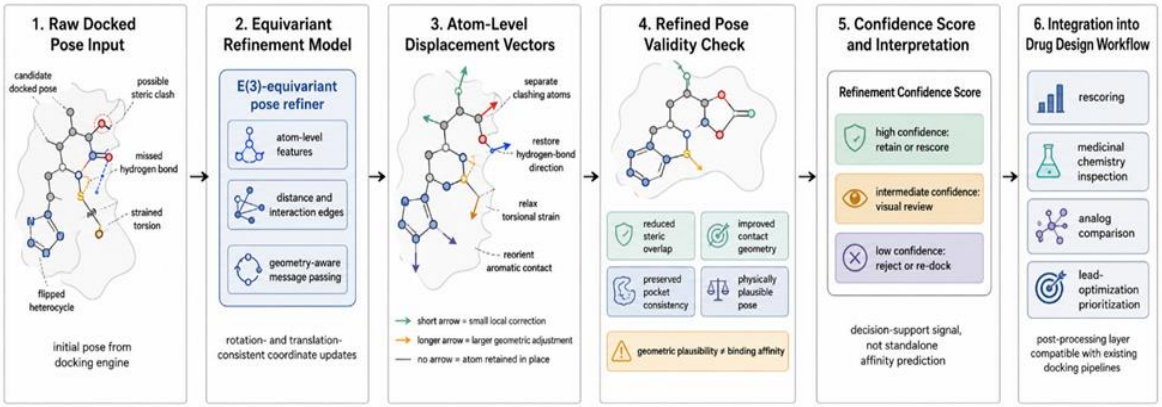

Figure 3. Interpretable Ligand-Pose Refinement as a Displacement-Field and Confidence-Scored Post-Processing Layer

Application in Lead Optimisation

During lead optimization, medicinal chemists often compare related analogs whose predicted binding modes differ by subtle local rearrangements. Deep learning approaches for flexible protein–ligand interactions [31] and equivariant flexible docking models such as DiffBindFR [13] motivate the use of geometry-aware refinement when ligand modifications alter torsions, substituent placement, or pocket contacts. The proposed model could refine a series of analog poses in a consistent coordinate framework and report whether each corrected pose remains chemically plausible. This would support prioritization based on binding-mode confidence while preserving the need for expert inspection and experimental validation. **Table 3** summarizes how the proposed post-processing refiner would add value at key decision points in docking-based lead optimization workflows.

Table 3. Practical Use Points for a Geometry-Aware Docking Pose Refiner

Workflow point	What the refiner adds	Practical value for lead optimization
After conventional docking	Corrects raw docked poses using symmetry-consistent geometric refinement	Improves pose consistency without replacing existing docking engines
Analog-series comparison	Aligns related ligands within a shared coordinate framework	Helps chemists distinguish meaningful binding-mode shifts from docking noise
Confidence scoring	Reports plausibility of corrected poses and local contact stability	Supports prioritization of compounds with more reliable predicted binding modes
Medicinal chemistry review	Flags poses requiring visual inspection or experimental follow-up	Preserves expert oversight while reducing low-confidence pose interpretation

Evaluation Strategy

Pose Refinement Accuracy

Pose refinement accuracy should be evaluated by comparing input docked poses with refined poses against experimental reference geometries, using RMSD-based improvement and near-native recovery as conceptual metrics rather than reporting invented outcomes. Atom3D provides a useful example of task-oriented molecular benchmarks for three-dimensional learning [32], while PoseBusters highlights the importance of checking physical validity in addition to geometric proximity [29]. Recent biomolecular prediction systems such as AlphaFold 3 [33] and RoseTTAFold All-Atom [34] further underscore that structural predictions should be assessed at the level of chemically meaningful complexes. For the proposed model, evaluation should focus on whether the refinement step improves geometry without introducing invalid bond lengths, clashes, or implausible contacts.

Pose Quality Scoring Performance

The quality score should be assessed by how well it distinguishes poses that remain plausible after refinement from poses that should be rejected or reprocessed. Earlier deep scoring systems, including CNN protein–ligand scoring [1], KDEEP [2], and D3R-oriented mathematical deep learning models [24], motivate comparisons between learned confidence and conventional scoring behavior. The evaluation could include rank-based and classification-style measures, but the manuscript should treat them as planned criteria rather than completed results. The key question is whether the scalar head provides useful confidence information beyond the coordinate update itself.

Generalization to Novel Proteins and Chemotypes

Generalization should be tested using temporally separated complexes, unseen protein families, and chemically distinct ligand scaffolds to determine whether the model learns transferable geometry rather than dataset-specific pose patterns. State-specific and generative protein–ligand structure models such as DynamicBind [14], a multiscale deep generative model for complexes [35], and FlowDock [30] show the increasing importance of evaluating learned docking systems beyond familiar training

distributions. Generalized biomolecular modeling frameworks such as AlphaFold 3 [33] and RoseTTAFold All-Atom [34] also suggest that broad structural coverage is essential for practical adoption. The proposed refiner should therefore be judged not only by average behavior, but also by whether it remains physically sensible for novel pockets, ligand chemotypes, and difficult starting poses.

Table 4 provides an evaluation and failure-mode framework for determining whether the proposed equivariant refiner produces geometrically improved, physically valid, symmetry-consistent, and practically useful protein–ligand poses.

Table 4. Evaluation and Failure-Mode Framework for Equivariant Pose Refinement

Evaluation dimension	Core question	Recommended assessment strategy	Evidence generated	Failure mode detected	Practical decision implication
Coordinate refinement accuracy	Does the model move imperfect docked poses closer to experimentally observed binding geometries?	Compare input and refined poses against reference structures using RMSD improvement, near-native recovery, and atom-level displacement error	Demonstrates whether refinement improves geometry rather than merely changing coordinates	False correction, excessive displacement, or failure to improve moderate decoys	Retain model only if corrected poses improve structural plausibility across relevant targets
Physical validity of refined poses	Are corrected structures chemically and sterically plausible?	Check bond lengths, ligand strain, steric clashes, pocket penetration, unsatisfied polar contacts, and PoseBusters-style validity criteria	Separates visually plausible poses from physically invalid geometries	Clash introduction, broken ligand geometry, unrealistic torsions, invalid contacts	Reject or rescore refined poses that pass scoring but fail chemical validity checks
Equivariance verification	Do predicted displacements behave consistently under rotation and translation?	Apply random rotations and translations to the same complex and compare transformed displacement outputs and invariant quality scores	Confirms whether the model respects the intended geometric symmetry	Orientation leakage, coordinate-frame dependence, unstable vector outputs	Prevent use of models whose predictions change because of arbitrary molecular orientation
Contribution of interaction energies	Do explicit energy features improve refinement beyond geometry and atom identity alone?	Conduct ablation studies removing electrostatic, van der Waals, hydrogen-bond, or all energy features	Clarifies whether physics-informed inputs add meaningful correction signal	Model memorizes local geometry without learning chemically grounded correction	Preserve energy features only if they improve accuracy, validity, or generalization
Confidence score calibration	Does the quality score identify poses likely to be reliable after refinement?	Compare score against refinement success, physical validity, and near-native recovery using calibration and ranking analyses	Determines whether the score is useful for triage rather than decorative reporting	Overconfident failed refinements or underconfident successful refinements	Use score to route poses to retain, rescore, reject, or expert review categories
Generalization to unseen proteins	Does performance hold for protein families not represented during training?	Use family-level or target-level splits and evaluate on novel binding pockets	Tests whether learned refinement is transferable across structural contexts	Target memorization or failure on unfamiliar pocket geometry	Require external target-family testing before practical deployment
Generalization to novel chemotypes	Does the model refine ligands with scaffolds unlike those in training?	Use scaffold splits and evaluate chemically distinct ligands, flexible analogs, and unusual functional groups	Tests whether the model learns general molecular geometry rather than training-set pose patterns	Scaffold-specific correction bias or unreliable treatment of new chemotypes	Flag low-confidence chemotypes for manual inspection or physics-based refinement
Sensitivity to starting-pose quality	What range of docking error can the model realistically correct?	Stratify decoys by initial RMSD, clash burden, and pocket-placement error before refinement	Defines the local correction boundary of the model	Attempted repair of completely wrong binding modes	Use pose ensembles and reject poses outside the model's correction range
Special chemical environments	Can the model handle metals, ordered waters, cofactors, covalent ligands, and protonation-dependent contacts?	Evaluate predefined challenge subsets requiring specialized chemistry	Identifies whether default graph features are sufficient for complex binding sites	Mis-modeled metal coordination, water-mediated contacts, covalent geometry, or protonation ambiguity	Require domain-specific extensions before applying the model to chemically complex systems
Pipeline utility	Does refinement improve downstream structure-based design decisions?	Compare docking-only, docking-plus-refinement, and docking-plus-refinement-plus-human-review workflows	Shows whether the method adds practical value beyond benchmark metrics	Metric improvement without actionable medicinal chemistry benefit	Adopt only if refinement improves prioritization, review efficiency, or experimental hypothesis quality

Limitations

Dependence on the Initial Docking Quality

The proposed model would likely correct moderate local pose errors more reliably than completely incorrect binding modes. Equivariant docking systems such as EquiBind [11], TANKBind [12], DiffBindFR [13], and DynamicBind [14] demonstrate the promise of learned geometric placement, but refinement remains bounded by the information present in the starting pose and pocket representation. If a ligand is docked into the wrong subpocket or adopts an incompatible scaffold orientation, a local displacement field may not recover the intended binding mode. The model should therefore be used with pose ensembles and confidence filtering rather than as a guaranteed repair mechanism for arbitrary docking failures.

Computational Cost and Requirements for Metal ions/Water

Equivariant tensor operations and pre-computed interaction energies are more demanding than simple descriptor-based scoring, especially when many protein atoms are included in the graph. Learned atomistic potentials such as NequIP [16] and large-scale equivariant representations for atomistic dynamics [20] illustrate both the expressive power and computational complexity of symmetry-preserving molecular models. Additional challenges arise for metalloproteins, ordered waters, cofactors, covalent ligands, and protonation-dependent interactions, which may require specialized graph features and energy terms. A practical implementation should therefore define clear default assumptions and allow domain-specific extensions for chemically complex binding sites.

Conclusion

The proposed MDL framework describes an equivariant neural network for protein–ligand pose refinement that operates directly on atomic coordinates and interaction energies. Its central output is a ligand displacement field that can adjust docked poses while preserving the rotational and translational structure of molecular space. A coupled scalar confidence score provides a complementary estimate of whether the refined pose is suitable for downstream use.

The main strength of this design is that it treats pose refinement as a geometric learning problem rather than only a scoring problem. By combining equivariant message passing with physics-informed edge features, the model can reason about both molecular symmetry and local interaction energetics. The resulting displacement vectors are also directly interpretable because they show how each ligand atom is proposed to move.

Important challenges remain before such a model could be considered a mature docking-refinement tool. Poor initial poses, unusual cofactors, explicit water networks, metal coordination, protonation ambiguity, and large protein systems may limit the reliability or scalability of the method. These cases should be treated as core development targets rather than minor implementation details.

Future work should emphasize transparent benchmarking, open-source implementation, and integration with widely used docking pipelines. Community-wide pose refinement benchmarks would help separate genuine geometric improvement from superficial rescoring. A shared evaluation culture would also make it easier to compare physics-based refinement, learned scoring, and equivariant coordinate correction under consistent assumptions.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR. Protein-ligand scoring with convolutional neural networks. J Chem Inf Model. 2017;57(4):942-57.
2. Jiménez J, Skalic M, Martinez-Rosell G, De Fabritiis G. KDEEP: protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Inf Model. 2018;58(2):287-96.
3. Feinberg EN, Sur D, Wu Z, Husic BE, Mai H, Li Y, et al. PotentialNet for molecular property prediction. ACS Cent Sci. 2018;4(11):1520-30.
4. Zheng L, Fan J, Mu Y. OnionNet: a multiple-layer intermolecular-contact-based convolutional neural network for protein-ligand binding affinity prediction. ACS Omega. 2019;4(14):15956-65.
5. McNutt AT, Francoeur P, Aggarwal R, Masuda T, Meli R, Ragoza M, et al. GNINA 1.0: molecular docking with deep learning. J Cheminform. 2021;13(1):43.
6. Torng W, Altman RB. Graph convolutional neural networks for predicting drug-target interactions. J Chem Inf Model. 2019;59(10):4131-49.
7. Thomas N, Smidt T, Kearnes S, Yang L, Li L, Kohlhoff K, et al. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. arXiv. 2018;1802.08219.

8. Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR. SchNet-a deep learning architecture for molecules and materials. J Chem Phys. 2018;148(24).
9. Unke OT, Meuwly M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. J Chem Theory Comput. 2019;15(6):3678-93.
10. Gasteiger J, Becker F, Günnemann S. GemNet: universal directional graph neural networks for molecules. Adv Neural Inf Process Syst. 2021;34:6790-802.
11. Stärk H, Ganea O, Pattanaik L, Barzilay R, Jaakkola T. EquiBind: geometric deep learning for drug binding structure prediction. In: Proc Int Conf Mach Learn. 2022. p. 20503-21.
12. Lu W, Wu Q, Zhang J, Rao J, Li C, Zheng S. TankBind: trigonometry-aware neural networks for drug-protein binding structure prediction. Adv Neural Inf Process Syst. 2022;35:7236-49.
13. Zhu J, Gu Z, Pei J, Lai L. DiffBindFR: an SE(3) equivariant network for flexible protein-ligand docking. Chem Sci. 2024;15(21):7926-42.
14. Lu W, Zhang J, Huang W, Zhang Z, Jia X, Wang Z, et al. DynamicBind: predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. Nat Commun. 2024;15(1):1071.
15. Sunseri J, Koes DR. Virtual screening with Gnina 1.0. Molecules. 2021;26(23):7369.
16. Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nat Commun. 2022;13(1):2453.
17. Wang S, Liu D, Ding M, Du Z, Zhong Y, Song T, et al. SE-OnionNet: a convolution neural network for protein-ligand binding affinity prediction. Front Genet. 2021;11:607824.
18. Son J, Kim D. Development of a graph convolutional neural network model for efficient prediction of protein-ligand binding affinities. PLoS One. 2021;16(4):e0249404.
19. Kondor R, Trivedi S. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In: Proc Int Conf Mach Learn. 2018. p. 2747-55.
20. Musaelian A, Batzner S, Johansson A, Sun L, Owen CJ, Kornbluth M, et al. Learning local equivariant representations for large-scale atomistic dynamics. Nat Commun. 2023;14(1):579.
21. Corso G, Stärk H, Jing B, Barzilay R, Jaakkola T. DiffDock: diffusion steps, twists, and turns for molecular docking. Int Conf Learn Represent. 2023.
22. Satorras VG, Hoogeboom E, Welling M. E(n) equivariant graph neural networks. In: Proc Int Conf Mach Learn. 2021;139:9323-32.
23. Batatia I, Kovács DP, Simm GNC, Ortner C, Csányi G. MACE: higher order equivariant message passing neural networks for fast and accurate force fields. Adv Neural Inf Process Syst. 2022.
24. Nguyen DD, Cang Z, Wu K, Wang M, Cao Y, Wei GW. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. J Comput Aided Mol Des. 2019;33(1):71-82.
25. Sánchez-Cruz N, Medina-Franco JL, Mestres J, Barril X. Extended connectivity interaction features: improving binding affinity prediction through chemical description. Bioinformatics. 2021;37(10):1376-82.
26. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. Bioinformatics. 2018;34(21):3666-74.
27. Jiang H, Fan M, Wang J, Sarma A, Mohanty S, Dokholyan NV, et al. Guiding conventional protein-ligand docking software with convolutional neural networks. J Chem Inf Model. 2020;60(10):4594-602.
28. Hochuli J, Helbling A, Skaist T, Ragoza M, Koes DR. Visualizing convolutional neural network protein-ligand scoring. J Mol Graph Model. 2018;84:96-108.
29. Buttenschoen M, Morris GM, Deane CM. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. Chem Sci. 2024;15(9):3130-9.
30. Morehead A, Cheng J. FlowDock: geometric flow matching for generative protein-ligand docking and affinity prediction. Bioinformatics. 2025;41(Suppl_1):i198-206.
31. Lee J, Hao Nguyen C, Mamitsuka H. Beyond rigid docking: deep learning approaches for fully flexible protein-ligand interactions. Brief Bioinform. 2025;26(5):bbaf454.
32. Townshend RJ, Vögele M, Suriana P, Derry A, Powers A, Laloudakis Y, et al. Atom3D: tasks on molecules in three dimensions. arXiv. 2020;2012.04035.
33. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630(8016):493-500.
34. Krishna R, Wang J, Ahern W, Sturmfels P, Venkatesh P, Kalvet I, et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science. 2024;384(6693):eadl2528.
35. Qiao Z, Nie W, Vahdat A, Miller TF III, Anandkumar A. State-specific protein-ligand complex structure prediction with a multiscale deep generative model. Nat Mach Intell. 2024;6(2):195-208.