PHARMPD: COUPLING PHARMACOPHORE SEMANTICS AND 3D GEOMETRY FOR MOLECULAR REPRESENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Three-dimensional molecular representations capture spatial relationships, but geometry alone does not specify which of those relationships are chemically relevant to molecular recognition. We introduce PharmPD, an explicit representation that couples interaction semantics and three-dimensional geometry through chemistry-audited pharmacophore centers and their spatial relationships. Across 22 TDC ADMET benchmark tasks, adding PharmPD to a strong 2D baseline improves performance on 18 tasks and on all seven prespecified molecular-recognition endpoints. Controlled ablations show that pharmacophore semantics and geometric information are substantially more predictive when explicitly coupled than when provided separately, while three-dimensional distance provides additional value beyond topological separation. The same explicit structure also makes model attributions chemically inspectable: high-attribution features recover target-specific pharmacophore relationships consistent with established recognition chemistry and can be grounded in molecule-specific structural evidence, improving their verifiability. This complementary value persists on an external OpenADMET PXR task, including against established 3D descriptors and a strong baseline incorporating Uni-Mol v2 representations. Together, these results identify semantic–geometric conjunction as a useful design principle for molecular representations that improves prediction while preserving a direct, verifiable connection to molecular-recognition chemistry.

1 INTRODUCTION

Molecular property prediction increasingly combines complementary sources of chemical information. ADMET-AI augments learned molecular representations with RDKit features (Swanson et al., 2024), MapLight combines multiple fingerprints and physicochemical descriptors (Notwell & Wood, 2023), and recent pretraining methods integrate molecular descriptors, graphs, and three-dimensional structure (Burns et al., 2025; Liu et al., 2022; Ji et al., 2024). These approaches demonstrate the value of richer molecular information. However, combining information sources does not necessarily preserve the relationships between them. For molecular properties that involve recognition by a biological target, one particularly relevant relationship connects the chemical roles of molecular features with their spatial arrangement relative to one another.

Pharmacophore theory describes molecular recognition through interaction features and their geometry (Wermuth et al., 1998; Leach et al., 2010). We refer to the correspondence between pharmacophore roles and their spatial relationships as *semantic–geometric coupling*. Existing approaches already capture aspects of this relation: CATS3D encodes pharmacophore-type distance correlations (Renner et al., 2006), while learned 3D models integrate chemical and geometric information (Liu et al., 2022; Ji et al., 2024). What remains less clear is whether the pair-specific correspondence itself carries predictive information, and whether the same chemically defined relation can remain directly accessible to model attribution. We therefore ask whether explicitly preserving pharmacophore identity together with spatial separation provides information beyond matched pharmacophore and distance summaries while retaining a molecule-grounded basis for interpretation.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We introduce PharmPD, which represents three-dimensional relationships between chemically defined pharmacophore centers as explicit typed distance features. Each relational coordinate retains a specific pharmacophore-pair and distance identity, allowing the same structure to be examined through controlled ablation and model attribution. We audit the underlying pharmacophore assignments against established chemical expectations and evaluate PharmPD across TDC ADMET benchmarks (Huang et al., 2022) and external OpenADMET PXR data (OpenADMET, 2026). Controlled comparisons test the contribution of type-pair-specific distance correspondence against matched marginal and topological controls, while attribution analyses trace influential relations to molecule-specific pharmacophore centers and geometries and compare them with established target-recognition chemistry. Together, these experiments test whether explicit semantic–geometric coupling provides complementary predictive information while retaining a chemically inspectable and directly verifiable basis for model attribution.

2 Related Work

Molecular representations. Molecular property prediction has progressed from fixed fingerprints and physicochemical descriptors to learned and hybrid representations (David et al., 2020). MapLight (Notwell & Wood, 2023) and ADMET-AI (Swanson et al., 2024) combine multiple molecular feature families, while CheMeleon (Burns et al., 2025) learns transferable chemical representations. GraphMVP exploits correspondence between 2D molecular topology and 3D geometry (Liu et al., 2022), whereas Uni-Mol learns directly from large-scale 3D molecular conformations (Zhou et al., 2023).

Pharmacophore and relational representations. Pharmacophore descriptors provide a chemically structured view of molecular interactions. ErG represents pharmacophore roles on reduced molecular graphs (Stiefl et al., 2006), while CATS3D encodes spatial correlations between pharmacophore points (Renner et al., 2006). PharmPD builds on this broader pharmacophore–distance principle while preserving chemically defined center-level relations as explicit predictive and attribution coordinates. Other 3D descriptors, including WHIM (Todeschini & Gramatica, 1997) and USRCAT (Schreyer & Blundell, 2012), capture complementary aspects of molecular geometry.

Molecular explainability. Molecular explanations commonly use feature attribution and atom- or fragment-level importance to relate predictions to chemical structure. SHAP provides Shapley-value-based feature attribution (Lundberg & Lee, 2017), while MolSHAP extends attribution to chemically defined R-group fragments (Tian et al., 2024). Such attribution does not necessarily preserve the semantic–geometric relation underlying the predictive signal.

Our position. PharmPD makes pharmacophore-type–distance relations explicit predictive coordinates that can also serve as units of controlled ablation and model attribution. We evaluate whether this correspondence contributes beyond strong descriptor, pharmacophore, and geometric baselines, and whether attributed relations recover established recognition chemistry and remain traceable to molecule-specific centers, distances, and conformers.

3 Methodology

3.1 PharmPD Representation

PharmPD (Pharmacophore Pair Distance) is a deterministic 200-dimensional molecular representation that encodes the correspondence between pharmacophore interaction type and three-dimensional intercenter distance. Molecular interaction centers are assigned to six pharmacophore classes: positively ionizable (PI), negatively ionizable (NI), aromatic (AR), hydrogen-bond acceptor (HBA), hydrogen-bond donor (HBD), and hydrophobic (HP). Selected ionizable and protonation-dependent centers are weighted according to their expected population near physiological pH. Three-dimensional conformers are then generated, and distances between pharmacophore centers are encoded separately for the 21 unordered pharmacophore-type pairs using nine overlapping Gaussian distance channels. Elementwise maximum pooling across conformers yields 189 pair-distance features, which are supplemented with 11 conformation-independent summaries to form the final 200-dimensional vector.

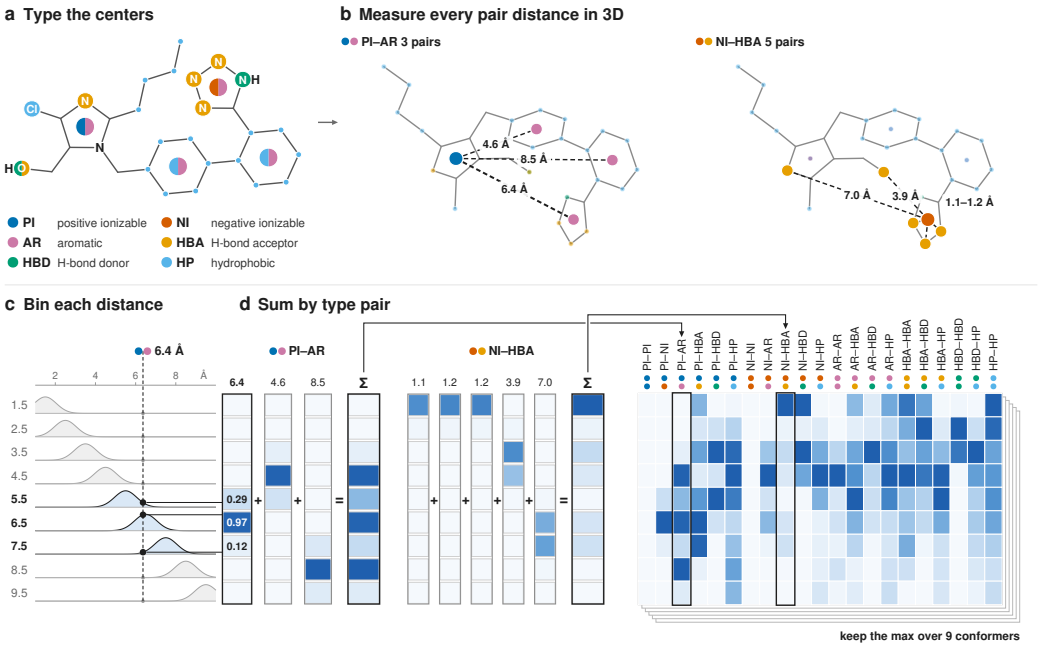

Figure 1: **PharmPD encodes pharmacophore types and 3D distances as a chemically named type-pair × distance map. a.** Atoms are grouped into pharmacophore centers of six types (PI, NI, AR, HBA, HBD, and HP), with chemically defined weights for ionizable and halogen-containing centers. Centers may carry multiple pharmacophore types. **b.** Multiple 3D conformers are generated, and distances between pharmacophore centers are measured within each conformer. **c.** Each distance is expanded over nine Gaussian kernels spanning 1.5–9.5 Å. **d.** Weighted kernel values are summed for each of the 21 pharmacophore type pairs, producing a 21 × 9 map. Max pooling across conformers yields 189 pair–distance features, which are combined with 11 molecular summary features to form the 200-dimensional PharmPD descriptor. Each of the 189 dimensions retains an explicit chemical identity defined by a pharmacophore type pair and distance bin.

PharmPD contains no learned parameters and is generated solely from molecular structure and predefined chemical rules. In the primary predictive experiments, the PharmPD block was concatenated with a 3,092-dimensional adapted MapLight-style two-dimensional representation Notwell & Wood (2023). The principal comparison therefore evaluates whether explicit pharmacophore-resolved three-dimensional information provides predictive value beyond a broad two-dimensional molecular representation while holding the downstream predictive framework fixed.

3.2 Chemistry-Aware Pharmacophore Definition

Pharmacophore centers were initialized from the RDKit `BaseFeatures.fdef` definitions distributed with RDKit version 2025.09.6 RDKit Contributors (2026). The feature families `PosIonizable`, `NegIonizable`, `Aromatic`, `Acceptor`, and `Donor` were mapped to PI, NI, AR, HBA, and HBD, respectively; `Hydrophobe` and `LumpedHydrophobe` were mapped to HP, and `ZnBinder` was excluded. Each center i was characterized by its type t_i, weight w_i, and constituent atom set A_i, with multiatom and overlapping centers permitted.

Because these general-purpose definitions are not necessarily suited to persistent quantitative descriptor channels, the assignments were subjected to a prespecified chemistry audit based on established acid–base behavior, hydrogen-bonding evidence, and medicinal-chemistry knowledge rather than downstream predictive performance Arunan et al. (2011); Laurence et al. (2009); Sipos-Szabó et al. (2026); Petukh et al. (2013). The resulting rules retained the same six pharmacophore classes while correcting selected donor, acceptor, ionizable, and hydrophobic assignments. Full audit procedures and chemical rationales are provided in Appendix C.

Ionizable center weights were made pH dependent using estimated charged-state populations at pH 7.4 from the Henderson–Hasselbalch relationship. These populations were assigned as center weights and propagated multiplicatively into the pair-distance features. Ionizable centers required an estimated charged fraction of at least 0.05, whereas protonation-restored HBD centers required at least 0.20. Other AR, HBA, HBD, and nonhalogen HP centers received unit weight. Full weighting definitions are provided in Appendix D.1.

3.3 Three-Dimensional Pair-Distance Encoding

For multicomponent records, the fragment containing the largest number of heavy atoms was retained, and molecules with fewer than three heavy atoms were not featurized. Up to 10 conformers per molecule were generated using ETKDGv3 Riniker & Landrum (2015); Wang et al. (2020) and subsequently minimized with MMFF94s Halgren (1999). The detailed conformer-generation settings are given in Appendix D.2.

Let

$$\mathcal{T} = \mathrm{PI}, \mathrm{NI}, \mathrm{AR}, \mathrm{HBA}, \mathrm{HBD}, \mathrm{HP} \tag{1}$$

denote the six pharmacophore types. Unordered combinations with repetition define 21 type-pair channels,

$$\mathcal{P} = (q,k) \in \mathcal{T}^2 : q \preceq k, \qquad |\mathcal{P}| = 21. \tag{2}$$

For conformer c, the coordinate of center i was defined as the arithmetic mean of the coordinates of its constituent atoms,

$$\mathbf{p}_i^{(c)} = \frac{1}{|A_i|} \sum_{a \in A_i} \mathbf{x} * a^{(c)}, \tag{3}$$

and the corresponding Euclidean intercenter distance was

$$d * ij^{(c)} = \left| \mathbf{p}_i^{(c)} - \mathbf{p}_j^{(c)} \right|_2 . \tag{4}$$

Each intercenter distance was expanded over nine overlapping Gaussian basis functions centered at

$$\mu_m = 0.5 + m, \qquad m = 1, \ldots, 9, \tag{5}$$

corresponding to 1.5–9.5 Å, with fixed width $\sigma = 0.55$ Å. For pharmacophore pair (q,k), distance channel m, and conformer c, the encoded value was

$$H_{(q,k),m}^{(c)} = \sum_{i:t_i=q} \sum_{j:t_j=k} \chi_{ij}^{(c)} w_i w_j \exp\left[-\frac{(d_{ij}^{(c)} - \mu_m)^2}{2\sigma^2} \right], \tag{6}$$

where $\chi_{ij}^{(c)}$ restricts contributions to intercenter distances between 0.3 and 10.5 Å and removes self-pairs defined by identical atom sets. The multiplicative factor $w_i w_j$ couples the chemical state of the two centers directly to their spatial relationship. Evaluation of 21 pharmacophore-type pairs over nine distance channels produces a 189-dimensional representation for each conformer.

Conformer-specific representations were combined by elementwise maximum pooling,

$$H_{(q,k),m} = \max_{c \in \mathcal{C}(M)} H_{(q,k),m}^{(c)}, \tag{7}$$

where $\mathcal{C}(M)$ denotes the generated conformer ensemble for molecule M. Each coordinate therefore records the strongest realization of a given pharmacophore-type–distance pattern among the sampled conformers rather than selecting a single representative conformer.

The resulting 189-dimensional block was supplemented with 11 conformation-independent features: five summaries of positively and negatively ionizable content and six weighted pharmacophore-class abundances. Thus,

$$\phi(M) = \rho(M) \oplus \kappa(M) \oplus \mathrm{vec}(H) \in \mathbb{R}^{5+6+189} = \mathbb{R}^{200}. \tag{8}$$

Definitions of the individual auxiliary coordinates and additional implementation conventions are provided in Appendix D.2. If conformer generation failed, the 189-dimensional three-dimensional block was set to zero while the 11 conformation-independent features were retained.

3.4 PREDICTIVE EVALUATION

The primary two-dimensional baseline was an adapted MapLight-style descriptor block containing 3,092 features Notwell & Wood (2023): a 2,048-bit radius-2 Morgan fingerprint Rogers & Hahn (2010), a 512-bit Avalon fingerprint Gedeck et al. (2006), a 315-dimensional ErG reduced-graph descriptor Stiefl et al. (2006), and 217 RDKit two-dimensional molecular descriptors RDKit Contributors (2026).Undefined, infinite, or missing descriptor values were replaced by zero.

Let

$$\psi(M) \in \mathbb{R}^{3092} \tag{9}$$

denote the baseline representation and $\phi(M) \in \mathbb{R}^{200}$ denote PharmPD. The principal matched comparison was

$$\mathbf{z} * \mathrm{ML}(M) = \psi(M) \tag{10}$$

versus

$$\mathbf{z} * \mathrm{ML} + \mathrm{PharmPD}(M) = \psi(M) \oplus \phi(M) \in \mathbb{R}^{3292}. \tag{11}$$

PharmPD was additionally evaluated alone where representation-level comparisons were required.

CatBoost and XGBoost models were trained independently for each task and descriptor arm Prokhorenkova et al. (2018); Chen & Guestrin (2016), and their predictions were combined by unweighted arithmetic averaging. Predictor families, hyperparameters, validation procedures, and ensemble rules were held constant across descriptor arms, and no task-specific hyperparameter optimization was performed. Ten percent of each training partition was reserved for early stopping, with stratification for binary classification. Consequently, the primary MapLight versus MapLight+PharmPD comparison differed in feature composition rather than in separately optimized downstream model configurations. Full model settings are provided in Appendix D.3.

4 EXPERIMENTS

4.1 DATASETS

We evaluated PharmPD on the 22 ADMET tasks from the Therapeutics Data Commons (TDC) (Huang et al., 2022), using the official scaffold splits and task-specific evaluation metrics. Seven hERG and CYP inhibition/substrate endpoints were additionally grouped by a prespecified target-recognition criterion for controlled representation analyses; this subset was defined by endpoint mechanism rather than predictive performance. External validation used the OpenADMET PXR challenge data, comprising 4,139 retained training molecules and 513 official challenge analogs from two series (OpenADMET, 2026). Because the challenge labels had been released before our analysis, we treat this as protocol-frozen external evaluation rather than prospective blind validation.

4.2 PREDICTION IMPROVEMENT

We evaluated PharmPD both as an additional input to existing predictors and as a standalone representation. These settings address distinct questions: whether PharmPD improves an existing prediction pipeline, and which aspects of its representation are useful when other descriptor blocks are absent.

Table 1: Predictive contribution of PharmPD on target-recognition endpoints.

A. Complementarity to MapLight

Endpoint	Metric	MapLight	+ PharmPD	Δ_{ML}
CYP2C9 sub.	AUPRC	0.3784	**0.4052**	+0.0268
CYP2C9 inh.	AUPRC	0.7874	**0.7994**	+0.0120
CYP2D6 sub.	AUPRC	0.6932	**0.6989**	+0.0057
CYP2D6 inh.	AUPRC	0.7147	**0.7330**	+0.0183
CYP3A4 sub.	AUROC	0.6486	**0.6721**	+0.0235
CYP3A4 inh.	AUPRC	0.8811	**0.8847**	+0.0036
hERG	AUROC	0.8422	**0.8685**	+0.0263

B. Controlled preservation of type–distance correspondence

Endpoint	Metric	Rank-1	PharmPD	Δ_{R1}
hERG	AUROC	0.8492	**0.8753**	+0.0261
CYP2C9 inh.	AUPRC	0.7162	**0.7529**	**+0.0367***
CYP2D6 inh.	AUPRC	0.5886	**0.6564**	**+0.0678***
CYP3A4 inh.	AUPRC	0.7858	**0.8350**	**+0.0492***
CYP2C9 sub.	AUPRC	**0.4221**	0.4135	-0.0086
CYP2D6 sub.	AUPRC	0.6332	**0.7001**	+0.0669
CYP3A4 sub.	AUROC	0.6595	**0.6857**	+0.0262

Panel A reports the original 10-seed TDC benchmark, where PharmPD was added to the 3,092-dimensional MapLight baseline. Δ_{ML} denotes MapLight+PharmPD minus MapLight. Panel B reports the separate 10-seed standalone GPU experiment. Rank-1 preserves the type-pair and distance marginals while removing type-pair-specific distance profiles. Δ_{R1} denotes PharmPD minus Rank-1. *BH-adjusted $q < 0.05$ for the prespecified Full–Rank-1 comparison. Results from the CPU and GPU experiments are not pooled.

4.2.1 Complementary Value across TDC ADMET Tasks

We first evaluated whether PharmPD provides predictive information complementary to a strong descriptor baseline by adding its 200-dimensional representation to a 3,092-dimensional adapted MapLight input (Notwell & Wood, 2023) across the 22 TDC ADMET tasks (Huang et al., 2022). The original benchmark used official scaffold splits, task-specific metrics, and 10 matched CPU training seeds, with predictor configurations held fixed between arms. Mean performance improved on 18 tasks and decreased on four, indicating broad but task-dependent complementarity. Complete taskwise results and seed-level consistency are reported in Appendix A.1.

The effect was particularly consistent across the seven hERG and CYP inhibition/substrate endpoints grouped by their relevance to target recognition. As shown in Table 1A, adding PharmPD to MapLight improved the mean primary metric on all seven endpoints, with five meeting the original significance criterion. The largest improvements were observed for CYP2C9 substrate prediction, hERG, and CYP3A4 substrate prediction. In contrast, the broader 22-task benchmark included decreases on Caco-2 permeability and P-glycoprotein transport. Thus, PharmPD complements existing molecular descriptors across diverse ADMET tasks, while its benefit is neither universal nor restricted to a single endpoint family.

4.2.2 Controlled Analysis of Relational Information

We tested whether PharmPD's predictive value depends specifically on preserving pharmacophore type-pair–distance correspondence. A separate standalone GPU experiment evaluated five representations across seven recognition endpoints and 10 seeds: the 11 auxiliary features (A); A plus separate type-pair and distance totals (*Unjoined*, 41D); a marginal-preserving reconstruction (*Rank-1*, 200D); a matched topological-distance representation (*Topology*, 200D); and full PharmPD (*Full*, 200D). Complete five-arm results are reported in Appendix A.2.

For the conformer-pooled matrix $H \in \mathbb{R}^{21\times 9}$, Rank-1 is defined as

$$\widetilde{H}_{p,k} = \frac{r_p c_k}{Z}, \qquad r_p = \sum_k H_{p,k}, \quad c_k = \sum_p H_{p,k}, \quad Z = \sum_{p,k} H_{p,k}, \tag{12}$$

with $\widetilde{H} = 0$ when $Z = 0$. This preserves type-pair and distance marginals, total mass, and dimensionality while removing type-pair-specific distance profiles, providing a matched control for relational correspondence. Implementation details are given in Appendix D.4. Inference used mean individual-seed metrics with paired Murcko-scaffold bootstrap intervals and Benjamini–Hochberg correction across 21 prespecified contrasts; full procedures are in Appendix D.7.

Table 1B shows the primary Full–Rank-1 comparison. Full had higher mean performance on six of seven endpoints, with the clearest effects on CYP2C9, CYP2D6, and CYP3A4 inhibition (+0.0367, +0.0678, and +0.0492 AUPRC). All three differences were positive in all 10 seeds and remained significant after multiplicity correction ($q = 0.00175$). Thus, type-pair-specific distance structure retained predictive information beyond matched type and distance marginals. The effect was not universal: CYP2C9 substrate favored Rank-1, and Full did not improve over A on CYP2C9 or CYP3A4 substrate prediction.

Topology was also competitive. Although Full had higher mean performance on six endpoints, no Full–Topology contrast remained significant after correction, and CYP2D6 substrate favored Topology, indicating that the incremental value of Euclidean 3D geometry and conformer pooling over a graph-distance relational proxy is task dependent. Detailed contrasts are provided in Appendix Tables 6, 7, and 8. When Full and Rank-1 were instead added above MapLight in a separate five-seed experiment, their differences were small and inconsistent, distinguishing standalone relational value from incremental value above a descriptor-rich baseline.

More broadly, this comparison motivates a relational view of molecular representation learning, in which predictive information may depend on cross-feature structure that must remain identifiable after representation construction; we develop this perspective mathematically in Appendix E.4.

4.2.3 Comparison with Established Molecular Descriptors

We next compared PharmPD with established molecular descriptors to determine whether its predictive contribution was shared by alternative pharmacophore and geometric representations. In the original CPU standalone evaluation against eight published descriptors, CATS3D (Renner et al., 2006) was the strongest or co-strongest published comparator across hERG and the three CYP inhibition tasks. On CYP2C9, for example, CATS3D reached 0.7473 AUPRC compared with 0.7493 for PharmPD. ErG was also competitive, whereas generic geometric descriptors showed more variable taskwise performance. These comparisons establish a strong pharmacophore-distance baseline and are reported separately from the controlled GPU ablations above. Complete standalone descriptor results are reported in Appendix A.3.

On external OpenADMET PXR data, seven 3D descriptors were additionally evaluated as augmentations of MapLight using matched conformer generation and a common evaluation population. PharmPD attained the lowest mean relative absolute error (RAE), 0.6660, compared with 0.6839 for MapLight, 0.6743 with CATS3D, and 0.6783 with WHIM. RDF, MORSE, GETAWAY, AUTOCORR3D, and USRCAT did not improve the MapLight baseline. Five of seven direct PharmPD–comparator contrasts had analog-series bootstrap intervals excluding zero. These results compare the complete descriptor designs rather than isolating any single difference in normalization, pharmacophore assignment, or distance encoding. Complete descriptor results and direct PharmPD–comparator contrasts are reported in Appendix Tables 11 and 12.

4.2.4 External Validation and Complementarity to Richer Baselines

The PXR evaluation used 4,139 retained training molecules and 513 official challenge analogs from two sets (OpenADMET, 2026). Because labels had been released before analysis, this was a protocol-frozen external evaluation rather than prospective blind validation. Adding PharmPD to MapLight reduced pooled RAE by 0.0178 (95% analog-series bootstrap CI: [0.0026, 0.0328]), with improvement in all 10 training seeds and in both analog sets (+0.0245 and +0.0107). Random 200-dimensional features and molecule-permuted PharmPD controls did not reproduce this improvement. Test-set consistency and negative-control results are reported in Appendix Table 15.

Complementarity persisted when the baseline already contained richer learned representations or assay information. PharmPD reduced RAE by 0.0295 above CheMeleon (Burns et al., 2025), 0.0160 above MapLight+CheMeleon, and 0.0117 above MapLight with auxiliary multi-fidelity PXR in-

formation. Above Uni-Mol v2+MapLight+multi-fidelity (Ji et al., 2024), PharmPD reduced RAE by 0.0067 (95% CI: [0.0029, 0.0107]) and MAE from 0.4450 to 0.4399. Two richer combinations showed smaller positive RAE differences whose intervals included zero. Full RAE and MAE results for the richer baselines are reported in Appendix Tables 13 and 14.

Together, these results show that explicit pharmacophore–distance information can provide complementary predictive value, while controlled ablations further identify type-pair-specific distance patterns as useful beyond matched marginal summaries on several recognition endpoints. The magnitude of this contribution depends on both the endpoint and the information already available to the predictor. We next ask whether the same explicit relational structure also supports chemically inspectable model attribution.

4.3 Explainability Improvement

Beyond predictive performance, we evaluated whether PharmPD supports chemically inspectable model attribution through four complementary questions: whether explanations reflect the model, whether attributed features can be grounded in the molecule, whether their computational claims can be verified, and whether high-attribution relations are consistent with established recognition chemistry. Table 2 summarizes role-level recovery across representations, with detailed analyses reported in Appendix B.

Table 2: Recovery of pharmacophore roles associated with established recognition chemistry.

Target	MapLight	CATS3D-style	PharmPD	Δ PP–ML
hERG	38.19%	43.13%	**49.68%**	**+11.49 pp**
CYP2D6	32.84%	**60.53%**	46.39%	**+13.55 pp**
CYP3A4	37.55%	**68.30%**	62.82%	**+25.27 pp**
CYP2C9[†]	**15.76%**	8.84%	14.67%	-1.09 pp

[†]The CYP2C9 NI–HBD signature was originally motivated by substrate chemistry rather than established inhibition chemistry and is therefore not interpreted as mechanistic support. MapLight does not explicitly encode the corresponding 3D type–distance relations; relation-level recovery is evaluated separately between PharmPD and CATS3D-style.

Values report the fraction of attribution mass assigned to the two pharmacophore roles defining each prespecified recognition relation. MapLight is evaluated using its ErG block, whereas the CATS3D-style and PharmPD models include the common MapLight backbone together with their respective 3D representations. Δ denotes PharmPD minus MapLight in percentage points. This role-level comparison does not measure recovery of the spatial relation itself. Additional role-level ranking and attribution-mass statistics are reported in Appendix Table 25.

4.3.1 Faithfulness to model attribution.

We first tested whether generated explanations were driven by model attribution rather than chemical plausibility alone. In a controlled SHAP input experiment, explanations followed correctly assigned SHAP values, abstained from assigning contributions when SHAP was withheld, and followed deliberately permuted SHAP assignments when these were provided (Appendix B.1; Table 16). Independent replay of the frozen CatBoost models further reproduced feature-level SHAP values and model outputs exactly (Appendix B.3). In a separate cross-representation analysis, fixed-complexity global surrogate explanations using PharmPD features achieved higher held-out fidelity than MapLight-only explanations in 6 of 12 prespecified comparisons, although the advantage was not universal and was more limited relative to a CATS3D-style representation (Appendix B.2; Table 18).

4.3.2 Grounding attribution in molecular structure.

We next asked whether an attributed PharmPD feature could be traced beyond its feature label to the molecular computation that produced it. For held-out molecules, we reconstructed the full provenance chain from atoms to pharmacophore centers, center–center distances, Gaussian pair contributions, conformer-specific channel values, and conformer-max aggregation. The independently

recomputed values reproduced the stored PharmPD features for all evaluated molecules. Thus, an attributed pharmacophore-type–distance channel can be traced to its constituent molecular centers and distances and to the conformer attaining the channel maximum, rather than remaining an abstract descriptor index. Full reconstruction checks and a representative molecule-specific trace are reported in Appendix B.3 and Table 19.

4.3.3 Computational verifiability.

We then tested whether claims derived from this provenance could be checked directly. Six classes of structured claims—center membership, center distance, winning conformer, pair aggregation, joint conformer realization, and feature attribution—were evaluated by recomputation from molecular records. All evaluable claims were correctly verified in the held-out verification set. Claim-family counts and verification results are reported in Appendix B.4 and Table 20. Providing precomputed feature-to-geometry links did not further improve verification accuracy under the tested exhaustive solver, indicating that the result establishes *checkability* of the attributed molecular claims rather than an efficiency advantage of explicit links.

4.3.4 Consistency with established recognition chemistry.

Finally, we asked whether high-attribution pharmacophore relations recover independently established target-recognition signatures (Appendix Table 21). Using the prespecified relations unchanged, we evaluated hERG PI–AR, CYP2D6 PI–HP, and CYP3A4 AR–HP against matched pharmacophore-pair decoys across ten frozen model seeds. PharmPD showed positive matched-decoy enrichment for all three relations, which ranked within the top three pair types in 80%, 100%, and 100% of seeds, respectively (Appendix Tables 22 and 24). Recovery was target dependent: PharmPD was stronger than CATS3D-style for hERG, whereas CATS3D-style was stronger for CYP2D6 and CYP3A4 (Appendix Table 23). Recovery was target dependent, with PharmPD stronger for hERG and CATS3D-style stronger for CYP2D6 and CYP3A4. PharmPD nevertheless provides an explicit path from recognition-aligned attribution to molecule-specific structural evidence that can be directly verified.

Together, these analyses show that PharmPD attribution can be tied to the model, resolved to molecule-specific structural evidence, independently recomputed, and, for multiple recognition endpoints, aligned with established pharmacophore relationships. These results support chemically inspectable and molecule-grounded attribution, while not implying universal superiority over other pharmacophore representations or establishing causal binding mechanisms.

5 Conclusion and Discussion

We introduced PharmPD, an explicit pharmacophore–distance representation that provides complementary predictive value while retaining chemically inspectable model attribution. Controlled experiments showed predictive value beyond matched marginal summaries on several recognition endpoints, while attributed relations could be traced to molecule-specific structural evidence and recovered established recognition signatures for hERG, CYP2D6, and CYP3A4. These effects were task- and comparator-dependent rather than universal, suggesting that the value of semantic–geometric coupling depends on the extent to which a given endpoint is governed by spatially organized chemical roles.

Notably, these results were obtained with a deliberately coarse representation that preserves pharmacophore identity and spatial relationships while omitting much atomistic detail. This suggests that, for some recognition-related tasks, retaining chemically meaningful relations can preserve useful signal even without representing every molecular degree of freedom. We therefore view PharmPD as a simple prototype of a broader class of relational molecular representations, in which richer physicochemical information could be explicitly coupled to geometry while remaining accessible to chemically grounded analysis. A broader discussion of endpoint dependence, representational abstraction, limitations, and future extensions is provided in Appendix E.

Reproducibility statement

Implementation code for PharmPD and the experiments reported in this paper is provided as anonymized supplementary material. Additional details on descriptor construction, model training, evaluation procedures, and hyperparameter settings are provided in the Methods and Appendix.

References

Keir Adams, Kento Abeywardane, Jenna Fromer, and Connor W. Coley. ShEPhERD: Diffusing shape, electrostatics, and pharmacophores for bioisosteric drug design. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=KSLkFYHlYg.

Alex M. Aronov. Common pharmacophores for uncharged human ether-a-go-go-related gene (hERG) blockers. *Journal of Medicinal Chemistry*, 49(23):6917–6921, 2006. doi: 10.1021/jm060500o.

Nongnuch Artrith, Keith T. Butler, François-Xavier Coudert, Seungwu Han, Olexandr Isayev, Anubhav Jain, and Aron Walsh. Best practices in machine learning for chemistry. *Nature Chemistry*, 13:505–508, 2021. doi: 10.1038/s41557-021-00716-z.

Elangannan Arunan, Gautam R. Desiraju, Roger A. Klein, Joanna Sadlej, Steve Scheiner, Ibon Alkorta, David C. Clary, Robert H. Crabtree, Joseph J. Dannenberg, Pavel Hobza, Henrik G. Kjaergaard, Anthony C. Legon, Benedetta Mennucci, and David J. Nesbitt. Definition of the hydrogen bond (IUPAC recommendations 2011). *Pure and Applied Chemistry*, 83(8):1637–1641, 2011. doi: 10.1351/PAC-REC-10-01-02.

Simon Axelrod and Rafael Gómez-Bombarelli. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9:185, 2022. doi: 10.1038/s41597-022-01288-4.

Riccardo Baron and J. Andrew McCammon. Molecular recognition and ligand association. *Annual Review of Physical Chemistry*, 64:151–175, 2013. doi: 10.1146/annurev-physchem-040412-110047.

Jackson Burns, Akshat Zalte, and William Green. Descriptor-based foundation models for molecular property prediction, 2025. URL https://arxiv.org/abs/2506.15792v1. arXiv:2506.15792v1.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016. doi: 10.1145/2939672.2939785.

Laurianne David, Amol Thakkar, Rocío Mercado, and Ola Engkvist. Molecular representations in ai-driven drug discovery: a review and practical guide. *Journal of Cheminformatics*, 12:56, 2020. doi: 10.1186/s13321-020-00460-5.

Marika Ekroos and Tove Sjögren. Structural basis for ligand promiscuity in cytochrome p450 3a4. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (37):13682–13687, 2006. doi: 10.1073/pnas.0603236103.

Peter Gedeck, Bernhard Rohde, and Christian Bartels. QSAR—how good is it in practice? comparison of descriptor sets on an unbiased cross section of corporate data sets. *Journal of Chemical Information and Modeling*, 46(5):1924–1936, 2006. doi: 10.1021/ci050413p.

Michael K. Gilson and Huan-Xiang Zhou. Calculation of protein–ligand binding affinities. *Annual Review of Biophysics and Biomolecular Structure*, 36:21–42, 2007. doi: 10.1146/annurev.biophys.36.040306.132550.

Thomas A. Halgren. MMFF VI. MMFF94s option for energy minimization studies. *Journal of Computational Chemistry*, 20(7):720–729, 1999. doi: 10.1002/(SICI)1096-987X(199905)20:7⟨720::AID-JCC7⟩3.0.CO;2-X.

Kexin Huang, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, and Marinka Zitnik. Artificial intelligence foundation for therapeutic science. *Nature Chemical Biology*, 18(10):1033–1036, 2022. doi: 10.1038/s41589-022-01131-2. URL https://doi.org/10.1038/s41589-022-01131-2.

Xiaohong Ji, Zhen Wang, Zhifeng Gao, Hang Zheng, Linfeng Zhang, Guolin Ke, and Weinan E. Exploring molecular pretraining model at scale. In *Advances in Neural Information Processing Systems*, volume 37, pp. 46956–46978, 2024. doi: 10.52202/079017-1489. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/53923bb44655a7defb31c7744c01b62b-Abstract-Conference.html.

José Jiménez-Luna, Francesca Grisoni, and Gisbert Schneider. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2:573–584, 2020. doi: 10.1038/s42256-020-00236-4.

Fleming Kretschmer, Jan Seipp, Marcus Ludwig, Gunnar W. Klau, and Sebastian Böcker. Coverage bias in small molecule machine learning. *Nature Communications*, 16:554, 2025. doi: 10.1038/s41467-024-55462-w.

Christian Laurence, Ken A. Brameld, Jérôme Graton, Jean-Yves Le Questel, and Eric Renault. The pK(BHX) database: Toward a better understanding of hydrogen-bond basicity for medicinal chemists. *Journal of Medicinal Chemistry*, 52(14):4073–4086, 2009. doi: 10.1021/jm801331y.

Andrew R. Leach, Valerie J. Gillet, Richard A. Lewis, and Robin Taylor. Three-dimensional pharmacophore methods in drug discovery. *Journal of Medicinal Chemistry*, 53(2):539–558, 2010. doi: 10.1021/jm900817u. URL https://doi.org/10.1021/jm900817u.

Shengchao Liu, Hanchen Wang, Weiyang Liu, Joan Lasenby, Hongyu Guo, and Jian Tang. Pre-training molecular graph representation with 3D geometry. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=xQUe1pOKPam.

Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

John S. Mitcheson, Jun Chen, Monica Lin, Chris Culberson, and Michael C. Sanguinetti. A structural basis for drug-induced long qt syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22):12329–12333, 2000. doi: 10.1073/pnas.210244497.

James H. Notwell and Michael W. Wood. ADMET property prediction through combinations of molecular fingerprints, 2023. URL https://arxiv.org/abs/2310.00174. arXiv:2310.00174.

OpenADMET. PXR challenge train/test dataset. Hugging Face dataset, 2026. URL https://huggingface.co/datasets/openadmet/pxr-challenge-train-test. Accessed 2026-09-22.

Marharyta Petukh, Shannon Stefl, and Emil Alexov. The role of protonation states in ligand-receptor recognition and binding. *Current Pharmaceutical Design*, 19(23):4182–4190, 2013. doi: 10.2174/1381612811319230004.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, volume 31, pp. 6638–6648, 2018.

RDKit Contributors. RDKit: Open-source cheminformatics, 2026. URL https://www.rdkit.org/. Version 2025.09.6.

S. Renner, C. H. Schwab, J. Gasteiger, and G. Schneider. Impact of conformational flexibility on three-dimensional similarity searching using correlation vectors. *Journal of Chemical Information and Modeling*, 46:2324–2332, 2006. doi: 10.1021/ci050075s. URL https://doi.org/10.1021/ci050075s.

Sereina Riniker and Gregory A. Landrum. Better informed distance geometry: Using what we know to improve conformation generation. *Journal of Chemical Information and Modeling*, 55 (12):2562–2574, 2015. doi: 10.1021/acs.jcim.5b00654.

Raquel Rodríguez-Pérez and Jürgen Bajorath. Explainable machine learning for property predictions in compound optimization. *Journal of Medicinal Chemistry*, 64(24):17744–17752, 2021. doi: 10.1021/acs.jmedchem.1c01789.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. doi: 10.1021/ci100050t.

Paul Rowland, Frank E. Blaney, Martin G. Smyth, Jo J. Jones, Vaughan R. Leydon, Amanda K. Oxbrow, Ceri J. Lewis, Mike G. Tennant, Sandeep Modi, Drake S. Eggleston, Richard J. Chenery, and Angela M. Bridges. Crystal structure of human cytochrome p450 2d6. *Journal of Biological Chemistry*, 281(11):7614–7622, 2006. doi: 10.1074/jbc.M511232200.

Adrian M. Schreyer and Tom Blundell. USRCAT: real-time ultrafast shape recognition with pharmacophoric constraints. *Journal of Cheminformatics*, 4:27, 2012. doi: 10.1186/1758-2946-4-27.

Irina F. Sevrioukova and Thomas L. Poulos. Dissecting cytochrome p450 3a4–ligand interactions using ritonavir analogues. *Biochemistry*, 52(26):4474–4481, 2013. doi: 10.1021/bi4005396.

Levente Sipos-Szabó, Dávid Bajusz, György T. Balogh, and György M. Keserű. Benchmarking pK_a prediction algorithms against an extensive, public data set. *Journal of Chemical Information and Modeling*, 66(8):4607–4619, 2026. doi: 10.1021/acs.jcim.6c00107.

Nikolaus Stiefl, Ian A. Watson, Knut Baumann, and Andrea Zaliani. ErG: 2d pharmacophore descriptions for scaffold hopping. *Journal of Chemical Information and Modeling*, 46(1):208–220, 2006. doi: 10.1021/ci050457y.

Gunter R. Strobl, Stephanie von Kruedener, Joachim Stöckigt, F. Peter Guengerich, and Thomas Wolff. Development of a pharmacophore for inhibition of human liver cytochrome p-450 2d6: Molecular modeling and inhibition studies. *Journal of Medicinal Chemistry*, 36(9):1136–1145, 1993. doi: 10.1021/jm00061a004.

Kyle Swanson, Parker Walther, Jeremy Leitz, Souhrid Mukherjee, Joseph C. Wu, Rabindra V. Shivnaraine, and James Zou. ADMET-AI: A machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics*, 40(7):btae416, 2024. doi: 10.1093/bioinformatics/btae416. URL https://doi.org/10.1093/bioinformatics/btae416.

Tingzhong Tian, Shuya Li, Meng Fang, Dan Zhao, and Jianyang Zeng. MolSHAP: Interpreting quantitative structure–activity relationships using shapley values of R-groups. *Journal of Chemical Information and Modeling*, 64(7):2236–2249, 2024. doi: 10.1021/acs.jcim.3c00465.

Roberto Todeschini and Paola Gramatica. The WHIM theory: New 3d molecular descriptors for QSAR in environmental modelling. *SAR and QSAR in Environmental Research*, 7(1-4):89–115, 1997. doi: 10.1080/10629369708039126.

Anna Vuorinen and Daniela Schuster. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. *Methods*, 71:113–134, 2015. doi: 10.1016/j.ymeth.2014.10.013.

Shuzhe Wang, Jagna Witek, Gregory A. Landrum, and Sereina Riniker. Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. *Journal of Chemical Information and Modeling*, 60(4):2044–2058, 2020. doi: 10.1021/acs.jcim.0c00025.

C. G. Wermuth, C. R. Ganellin, P. Lindberg, and L. A. Mitscher. Glossary of terms used in medicinal chemistry (IUPAC recommendations 1998). *Pure and Applied Chemistry*, 70(5): 1129–1143, 1998. doi: 10.1351/pac199870051129. URL https://doi.org/10.1351/pac199870051129.

Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, and Guolin Ke. Uni-Mol: A universal 3d molecular representation learning framework. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=6K2RM6wVqKu.

A Additional Prediction Results

The following tables expand the headline prediction results reported in the main text. Historical CPU experiments, newer GPU ablations, and external PXR analyses are kept separate because they use different estimands and inferential procedures.

A.1 Complete TDC ADMET Benchmark

PharmPD was added to the 3,092-dimensional MapLight baseline across all 22 TDC ADMET tasks. The table uses the seed-deduplicated historical export: six source tasks contained exact duplicate seed rows, but deduplication changed no mean or delta and restored ten unique paired seeds per task. No scaffold-bootstrap confidence interval or BH-adjusted q value was stored for this historical experiment; the reported p values are nominal two-sided paired-seed t tests and are not converted into adjusted values.

Table 3: Complete 22-task TDC benchmark: predictive performance.

Task	Metric	MapLight	+ PharmPD	Δ
Ames	AUROC	0.8659	**0.8668**	+0.0010
BBB	AUROC	0.9113	**0.9132**	+0.0018
Bioavailability	AUROC	0.7070	**0.7154**	+0.0085
Caco-2	MAE	**0.2854**	0.2954	-0.0100*
Hepatocyte clearance	Spearman	0.4595	**0.4683**	+0.0088
Microsome clearance	Spearman	0.6001	**0.6195**	+0.0194*
CYP2C9 substrate	AUPRC	0.3784	**0.4052**	+0.0268*
CYP2C9 inhibition	AUPRC	0.7874	**0.7994**	+0.0120*
CYP2D6 substrate	AUPRC	0.6932	**0.6989**	+0.0057
CYP2D6 inhibition	AUPRC	0.7146	**0.7330**	+0.0183*
CYP3A4 substrate	AUROC	0.6486	**0.6721**	+0.0235
CYP3A4 inhibition	AUPRC	0.8811	**0.8847**	+0.0036*
DILI	AUROC	0.8783	**0.9070**	+0.0287*
Half-life	Spearman	**0.5623**	0.5408	-0.0216*
hERG	AUROC	0.8422	**0.8685**	+0.0263*
HIA	AUROC	0.9666	**0.9909**	+0.0243*
LD50	MAE	0.5932	**0.5894**	+0.0038*
Lipophilicity	MAE	0.5080	**0.4892**	+0.0189*
P-gp	AUROC	**0.9196**	0.9142	-0.0054*
PPBR	MAE	8.0167	**7.5950**	+0.4217*
Solubility	MAE	0.7425	**0.7355**	+0.0071*
VDss	Spearman	**0.6967**	0.6898	-0.0069*

Δ is defined so that positive values favor MapLight+PharmPD; for lower-is-better MAE tasks it is MapLight minus MapLight+PharmPD, and otherwise it is MapLight+PharmPD minus MapLight. *Nominal paired-seed $p < 0.05$; no multiple-testing correction or scaffold-bootstrap interval was stored for this historical experiment. Means are over 10 unique paired seeds. Seed consistency and nominal p values are listed in Table 4.

Table 4: Complete 22-task TDC benchmark: seed consistency and nominal tests.

Task	**Positive seeds**	**Nominal p**
Ames	5/10	0.308
BBB	7/10	0.190
Bioavailability	6/10	0.369
Caco-2	2/10	0.00605
Hepatocyte clearance	6/10	0.202
Microsome clearance	8/10	0.0448
CYP2C9 substrate	7/10	0.0392
CYP2C9 inhibition	10/10	4.47e-05
CYP2D6 substrate	7/10	0.477
CYP2D6 inhibition	10/10	2.51e-08
CYP3A4 substrate	7/10	0.0528
CYP3A4 inhibition	10/10	0.000451
DILI	10/10	0.00512
Half-life	1/10	0.0173
hERG	10/10	0.000157
HIA	10/10	0.000487
LD50	9/10	0.00634
Lipophilicity	10/10	3.53e-08
P-gp	2/10	0.0244
PPBR	10/10	3.50e-08
Solubility	9/10	0.0188
VDss	2/10	0.0363

A.2 Controlled Representation Ablations

A separate standalone GPU experiment evaluated the 11 auxiliary features (A), the auxiliary features plus unjoined type-pair and distance summaries (Unjoined), the matched-marginal Rank-1 control, the graph-distance Topology control, and full PharmPD across seven recognition endpoints and ten seeds. The primary estimator was the mean individual-seed metric.

Table 5: **Mean performance of all five standalone representation arms.**

Endpoint	**Metric**	$\boldsymbol{A}$	**Unjoined**	**Rank-1**	**Topology**	**Full**
hERG	AUROC	0.8377	0.8506	0.8492	0.8545	**0.8753**
CYP2C9 inh.	AUPRC	0.6677	0.7110	0.7162	0.7523	**0.7529**
CYP2D6 inh.	AUPRC	0.5368	0.5837	0.5886	0.6465	**0.6564**
CYP3A4 inh.	AUPRC	0.7248	0.7830	0.7858	0.8317	**0.8350**
CYP2C9 sub.	AUPRC	**0.4227**	0.4219	0.4221	0.4057	0.4135
CYP2D6 sub.	AUPRC	0.5518	0.6365	0.6332	**0.7321**	0.7001
CYP3A4 sub.	AUROC	**0.6972**	0.6475	0.6595	0.6656	0.6857

Values are means over ten independently fitted seeds. A contains 11 auxiliary features; Unjoined adds separate marginal summaries; Rank-1 preserves pair-type and distance marginals but removes type-pair-specific distance profiles; Topology uses the matched graph-distance proxy; Full is the complete PharmPD representation.

Table 6: Primary Full–Rank-1 contrasts.

Endpoint	**Metric**	**Full−Rank-1 [95% CI];** q
hERG	AUROC	+0.0261 [0.0036, 0.0651]; 0.2214
CYP2C9 inh.	AUPRC	**+0.0367 [0.0219, 0.0517]; 0.0017**
CYP2D6 inh.	AUPRC	**+0.0678 [0.0392, 0.1019]; 0.0017**
CYP3A4 inh.	AUPRC	**+0.0492 [0.0346, 0.0649]; 0.0017**
CYP2C9 sub.	AUPRC	-0.0086 [-0.0733, 0.0433]; 0.8578
CYP2D6 sub.	AUPRC	+0.0669 [-0.0194, 0.1322]; 0.2214
CYP3A4 sub.	AUROC	+0.0262 [-0.0086, 0.0644]; 0.2912

Entries are mean individual-seed metric differences with paired 95% scaffold-bootstrap intervals. Positive values favor Full. BH correction covered 21 prespecified contrasts across Full–Rank-1, Full–Topology, and Full–A.

Table 7: Primary Full–Topology contrasts.

Endpoint	**Metric**	**Full−Topology [95% CI];** q
hERG	AUROC	+0.0208 [-0.0181, 0.0645]; 0.4949
CYP2C9 inh.	AUPRC	+0.0006 [-0.0164, 0.0186]; 0.9480
CYP2D6 inh.	AUPRC	+0.0099 [-0.0149, 0.0310]; 0.5604
CYP3A4 inh.	AUPRC	+0.0033 [-0.0082, 0.0152]; 0.7609
CYP2C9 sub.	AUPRC	+0.0078 [-0.0598, 0.0711]; 0.8784
CYP2D6 sub.	AUPRC	-0.0320 [-0.0826, 0.0081]; 0.2767
CYP3A4 sub.	AUROC	+0.0201 [-0.0232, 0.0662]; 0.5562

Table 8: Primary Full–auxiliary contrasts.

Endpoint	**Metric**	**Full−A [95% CI];** q
hERG	AUROC	+0.0376 [-0.0031, 0.0867]; 0.2214
CYP2C9 inh.	AUPRC	**+0.0852 [0.0590, 0.1098]; 0.0017**
CYP2D6 inh.	AUPRC	**+0.1196 [0.0794, 0.1601]; 0.0017**
CYP3A4 inh.	AUPRC	**+0.1102 [0.0849, 0.1367]; 0.0017**
CYP2C9 sub.	AUPRC	-0.0092 [-0.0662, 0.0390]; 0.8384
CYP2D6 sub.	AUPRC	**+0.1483 [0.0358, 0.2356]; 0.0165**
CYP3A4 sub.	AUROC	-0.0115 [-0.0714, 0.0493]; 0.8384

A.3 Established Molecular Descriptor Comparisons

Two descriptor comparisons are kept distinct. The standalone table is descriptive because the PharmPD and comparator runs used the same official test sets but different inner-validation ordering. A separate MapLight-augmentation experiment used matched arms and paired scaffold inference.

Table 9: **Standalone descriptor performance on hERG and CYP inhibition tasks.**

Descriptor	hERG AUROC	CYP2C9 AUPRC	CYP2D6 AUPRC	CYP3A4 AUPRC
PharmPD	**0.8718**	**0.7493**	**0.6646**	**0.8361**
CATS3D-style	0.8576	0.7473	0.6378	0.8236
ErG	0.8317	0.7118	0.6142	0.8069
WHIM	0.7959	0.6425	0.4180	0.7398
RDF	0.7623	0.6941	0.5349	0.8025
MoRSE	0.7848	0.7200	0.5720	0.8006
GETAWAY	0.8203	0.7069	0.5679	0.8236
AUTOCORR3D	0.7877	0.6623	0.5258	0.7531
USRCAT	0.7976	0.6800	0.4453	0.7833

Values are ten-seed means. These standalone means are descriptive only: comparator and PharmPD runs shared the official test sets but used different inner-validation ordering, so no paired confidence interval or q value is claimed here. The CATS3D entry denotes the local CATS3D-style implementation defined in Appendix D.6.

Table 10: Descriptor augmentations above the common MapLight backbone.

Input	hERG AUROC	CYP2C9 AUPRC	CYP2D6 AUPRC	CYP3A4 AUPRC
MapLight	0.8312	0.7862	0.7163	0.8805
MapLight+RDF	0.8297	0.7875	0.7119	0.8826
MapLight+MORSE	0.8268	0.7797	0.7068	0.8781
MapLight+AUTOCORR3D	0.8349	0.7860	0.7150	0.8789
MapLight+WHIM	0.8414	0.7945	0.7206	0.8783
MapLight+GETAWAY	0.8367	0.7834	0.7103	0.8803
MapLight+CATS3D-style	0.8517	0.7922	0.7246	0.8815
MapLight+USRCAT	0.8427	0.7908	0.7223	0.8795
MapLight+PharmPD	**0.8718**	**0.7948**	**0.7320**	**0.8852**

Values are ten-seed means from the matched MapLight-augmentation experiment. Paired inference for the 32 prespecified PharmPD-versus-comparator contrasts used 2,000 paired scaffold-bootstrap draws with BH correction across all 32 comparisons. The standalone table above is a separate descriptive experiment and should not be pooled with these results.

A.4 OpenADMET PXR External Evaluation

The PXR evaluation used 4,139 retained training molecules and 513 challenge analogs (253 and 260 in the two test sets), grouped into 96 analog series for clustered bootstrap. Because challenge labels were available before this analysis, the study is described as protocol-frozen external evaluation rather than prospective blind validation.

Table 11: PXR 3D descriptor augmentations above MapLight.

Augmentation	Dim.	Pooled RAE	Δ vs ML
PharmPD	200	**0.6660**	+0.0179
CATS3D-style	210	0.6743	+0.0095
WHIM	114	0.6783	+0.0056
RDF	210	0.6862	-0.0024
MORSE	224	0.6892	-0.0053
GETAWAY	273	0.6898	-0.0059
AUTOCORR3D	80	0.6916	-0.0078
USRCAT	60	0.6951	-0.0112

Lower RAE is better. Δ vs MapLight is MapLight RAE minus augmented-model RAE, so positive values indicate improvement. Direct PharmPD-versus-comparator intervals are reported in Table 12.

Table 12: PXR PharmPD-versus-descriptor RAE contrasts.

Comparator	PharmPD advantage in RAE [95% CI]
CATS3D-style	+0.0084 [-0.0040, 0.0200]
WHIM	+0.0123 [-0.0087, 0.0320]
RDF	+0.0203 [0.0015, 0.0383]
MORSE	+0.0232 [0.0059, 0.0396]
GETAWAY	+0.0238 [0.0030, 0.0452]
AUTOCORR3D	+0.0257 [0.0089, 0.0426]
USRCAT	+0.0291 [0.0098, 0.0478]

Table 13: Complementarity of PharmPD to richer PXR baselines: RAE.

Base representation	Base RAE	+ PP RAE	ΔRAE	95% CI
CheMeleon alone	0.7104	0.6809	+0.0295	[0.0110, 0.0483]
MapLight	0.6822	0.6644	+0.0178	[0.0026, 0.0328]
MapLight + CheMeleon	0.6813	0.6653	+0.0160	[0.0043, 0.0276]
MapLight + multi-fidelity	0.6371	0.6254	+0.0117	[0.0033, 0.0197]
ML + CheMeleon + mf	0.6214	0.6185	+0.0029	[-0.0035, 0.0091]
Uni-Mol v2 84M + ML + mf	0.5848	0.5781	+0.0067	[0.0029, 0.0107]
Uni-Mol v2 + ML + CheMeleon + mf	0.5779	0.5757	+0.0022	[-0.0009, 0.0057]

ΔRAE is base RAE minus base+PharmPD RAE; positive values favor adding PharmPD. Confidence intervals are analog-series bootstrap intervals.

Table 14: Complementarity of PharmPD to richer PXR baselines: MAE.

Base representation	Base MAE	+ PharmPD MAE
CheMeleon alone	0.5405	0.5181
MapLight	0.5190	0.5055
MapLight + CheMeleon	0.5183	0.5062
MapLight + multi-fidelity	0.4847	0.4758
ML + CheMeleon + mf	0.4728	0.4706
Uni-Mol v2 84M + ML + mf	0.4450	0.4399
Uni-Mol v2 + ML + CheMeleon + mf	0.4397	0.4380

Table 15: PXR test-set consistency and negative controls.

Condition	Δ RAE	95% CI	Seed wins
Analog set 1	+0.0245	[0.0011, 0.0485]	10/10
Analog set 2	+0.0107	[-0.0125, 0.0315]	8/10
Pooled	+0.0178	[0.0026, 0.0328]	10/10
PharmPD values permuted across molecules	-0.0216	[-0.0301, -0.0144]	1/10
200 random columns	-0.0531	[-0.0701, -0.0387]	0/10

For the test-set rows, positive values indicate improvement from adding PharmPD. The negative controls used molecule-permuted PharmPD values or 200 random columns and did not reproduce the observed benefit.

B Additional Explainability Results

These analyses correspond to the four questions used in the main text: model fidelity, molecule-specific grounding, computational verification, and consistency with established recognition chemistry. Exploratory studies that do not directly support the final argument are not included here.

B.1 Controlled SHAP Attribution Checks

We performed a controlled attribution experiment to test whether structured molecule-level explanations tracked the SHAP values actually supplied to the explanation procedure. The evaluation covered four endpoints, with three molecules per endpoint and three attribution conditions per molecule, for 36 total responses. Each response received the same 200 PharmPD feature values, feature ordering, and atom–distance grounding information. The conditions differed only in the supplied attribution information: the original SHAP values (*True SHAP*), no SHAP values (*No SHAP*), or SHAP values permuted across features (*Permuted SHAP*).

The attribution target was the raw logit of the seed-1 CatBoost component. Accordingly, this experiment does not evaluate explanations of the complete CatBoost+XGBoost ensemble probability. Reports were evaluated for output-schema validity, agreement with the original-model SHAP attribution, agreement with the SHAP values actually supplied to the explanation procedure, and verification of molecule-specific grounding claims.

With the true attribution values available, the generated reports recovered 48/72 scored SHAP assignments. When SHAP values were withheld, the procedure abstained rather than inventing original-model attribution assignments. When the attribution values were permuted across features, no scored assignment matched the original-model SHAP attribution, whereas 42/72 matched the permuted values that had actually been supplied. Grounding accuracy remained measurable without SHAP but decreased under the permuted condition. These results support sensitivity of the generated attribution statements to the provided feature-attribution evidence, while also showing that correct SHAP input alone does not guarantee a fully valid or fully grounded report.

This controlled experiment is an attribution-consistency check rather than a test of human interpretability, explanation usefulness, or correctness of the complete ensemble-level prediction.

B.2 Cross-Representation Explanation Fidelity

Fixed CatBoost teachers were summarized by small decision-tree surrogates using three explanation feature spaces: MapLight, MapLight+CATS3D-style, or MapLight+PharmPD. The primary complexity was eight leaves (seven internal conditions), and the target was the frozen teacher class-1 probability. Test teacher predictions and SHAP values were not provided as explanation inputs.

Table 16: Controlled SHAP attribution checks.

Condition	Valid	Original SHAP	Supplied SHAP	Grounding
True SHAP	8/12	48/72 (66.7%)	48/72 (66.7%)	42/72 (58.3%)
No SHAP	9/12	0/72†	0/72†	44/72 (61.1%)
Permuted SHAP	7/12	0/72	42/72 (58.3%)	28/72 (38.9%)

Each condition contains 12 reports and 72 scored attribution or grounding slots. "Original SHAP" measures agreement with the attribution of the unmodified model, whereas "Supplied SHAP" measures agreement with the attribution values provided to the explanation procedure. †In the No-SHAP condition, the system abstained from making SHAP assignments rather than reconstructing them independently.

Table 17: Eight-leaf global-surrogate fidelity: test MSE.

Task	Fixed teacher	ML	CATS3D-style	PharmPD
hERG	MapLight	0.013090	0.011306	0.011148
hERG	ML+CATS3D-style	0.009214	0.008400	0.011244
hERG	ML+PharmPD	0.020773	0.016920	0.010382
CYP2C9 inh.	MapLight	0.034009	0.034394	0.035071
CYP2C9 inh.	ML+CATS3D-style	0.035737	0.035912	0.035525
CYP2C9 inh.	ML+PharmPD	0.035869	0.036220	0.034212
CYP2D6 inh.	MapLight	0.026081	0.024444	0.024336
CYP2D6 inh.	ML+CATS3D-style	0.028515	0.022717	0.023698
CYP2D6 inh.	ML+PharmPD	0.026842	0.023115	0.021924
CYP3A4 inh.	MapLight	0.043922	0.044971	0.043958
CYP3A4 inh.	ML+CATS3D-style	0.048059	0.049269	0.043859
CYP3A4 inh.	ML+PharmPD	0.050523	0.053539	0.042692

Lower MSE indicates a more faithful fixed-complexity global surrogate. Each row fixes the same teacher; only the explanation feature space changes. Paired contrasts are listed in Table 18.

Table 18: Eight-leaf global-surrogate fidelity: PharmPD-space contrasts.

Task	Teacher	Δ PP−ML (q)	Δ PP−CATS (q)
hERG	MapLight	-0.0019 (0.243)	-0.0002 (0.977)
hERG	ML+CATS3D-style	+0.0020 (0.212)	+0.0028 (0.180)
hERG	ML+PharmPD	**-0.0104 (0.012)**	-0.0065 (0.074)
CYP2C9 inh.	MapLight	+0.0011 (0.185)	+0.0007 (0.373)
CYP2C9 inh.	ML+CATS3D-style	-0.0002 (0.881)	-0.0004 (0.701)
CYP2C9 inh.	ML+PharmPD	-0.0017 (0.136)	**-0.0020 (0.048)**
CYP2D6 inh.	MapLight	**-0.0017 (0.048)**	-0.0001 (0.980)
CYP2D6 inh.	ML+CATS3D-style	**-0.0048 (0.012)**	+0.0010 (0.787)
CYP2D6 inh.	ML+PharmPD	**-0.0049 (0.012)**	-0.0012 (0.551)
CYP3A4 inh.	MapLight	+0.0000 (0.980)	-0.0010 (0.701)
CYP3A4 inh.	ML+CATS3D-style	**-0.0042 (0.014)**	-0.0054 (0.063)
CYP3A4 inh.	ML+PharmPD	**-0.0078 (0.006)**	**-0.0108 (0.006)**

Δ is PharmPD-explanation MSE minus comparator-explanation MSE, so negative values favor the PharmPD explanation space. Bold contrasts satisfy both the paired 95% scaffold-bootstrap criterion and BH-adjusted $q < 0.05$ over 24 prespecified comparisons. This experiment evaluates global surrogate fidelity, not human understanding or local SHAP correctness.

Across the 120 primary PharmPD-space surrogate trees, 118/120 (98.33%) used at least one PharmPD relation channel. Of 840 total internal conditions, 218 used relation channels, 5 used

PharmPD auxiliary features, and 617 used MapLight features. The complexity sweep did not establish a general advantage in the minimum number of leaves required to reach a common fidelity target.

B.3 Molecule-Specific Attribution Provenance

A separate 64-molecule cohort was used to reconstruct the complete computational provenance of the PharmPD relation block. Native float32 reconstruction was bit-exact for 64/64 molecules, and an independent float64 implementation passed the prespecified numerical tolerance for 64/64. The maximum absolute center-distance discrepancy was 1.751e-06 Å, the maximum aggregate feature discrepancy was 1.025e-03, and the maximum pooled-feature discrepancy was 8.454e-04. Replayed feature SHAP values, base values, raw logits, and probabilities matched the frozen model records exactly, with maximum SHAP-additivity residual 2.665e-15. These checks establish frozen computational provenance, not an experimental bound pose or causal binding mechanism.

Table 19: Representative molecule-specific PharmPD provenance trace.

Field	**Value**
Task	herg
SMILES	Nc1nc(N)nc(CCCc2nc(N)nc(N)n2)n1
Attributed feature	PharmPD:AR-HBD @6.5A
Feature value	1.978833
Feature SHAP	+0.090285
Winning conformer	2
Center A	HBD atoms [0]
Center B	AR atoms [1, 2, 3, 5, 6, 18]
Example constituent-pair distance	2.6556 Angstrom
Model output	raw logit -0.8264; probability 0.3044

The PharmPD channel aggregates all eligible ordered center pairs of the given types; the listed center pair is one constituent contribution, not the complete channel. Feature attribution is therefore not interpreted as a causal SHAP value for this individual pair.

B.4 Computational Verification of Structured Attribution Claims

Six structured claim families were checked by recomputation from molecular records: center membership, center distance, winning conformer, pair aggregation, joint winner realization, and feature attribution. Each arm had 384 planned claim slots and 382 evaluable claims. Both raw-evidence and explicitly linked conditions verified all 382/382 evaluable claims; on the fixed denominator this corresponds to 99.479%. The two unavailable slots arose because one single-conformer molecule could not support a nontrivial incorrect winner or joint-winner set and was not replaced.

Table 20: Structured claim verification by claim family.

Claim family	B evaluable	B correct	C evaluable	C correct
Center	64	64	64	64
Distance	64	64	64	64
Winner	63	63	63	63
Aggregate	64	64	64	64
Joint	63	63	63	63
Attribution	64	64	64	64

Strict correctness required the correct verdict, error location, supporting evidence identifiers, and recomputed value or set. The linked condition did not improve accuracy in this exhaustive verification setting; this experiment therefore establishes checkability of the computational claims rather than a comparative efficiency advantage for explicit links.

B.4.1 Literature Basis of the Prespecified Recognition Signatures

The recognition relations evaluated in the attribution-recovery analysis were defined from target-specific pharmacophore and structural literature rather than selected from the attribution results. They represent coarse pharmacophore-pair signatures rather than universal binding requirements or molecule-specific distance constraints. Their literature basis and interpretation are summarized in Table 21.

Table 21: **Literature basis of the prespecified recognition signatures.**

Target	Relation	Literature basis
hERG	PI–AR	hERG blocker pharmacophore studies commonly identify basic/positive and aromatic or hydrophobic features, while mutagenesis implicates the aromatic Tyr652/Phe656-lined inner cavity in drug recognition (Aronov, 2006; Mitcheson et al., 2000).
CYP2D6	PI–HP	CYP2D6 recognition has been associated with a basic nitrogen together with a hydrophobic or aromatic region, with acidic residues such as Asp301 and Glu216 contributing to recognition of basic ligands (Strobl et al., 1993; Rowland et al., 2006).
CYP3A4	AR–HP	Structural and functional studies describe CYP3A4 as a large and adaptable predominantly hydrophobic cavity in which aromatic and nonpolar ligand interactions contribute to recognition (Ekroos & Sjögren, 2006; Sevrioukova & Poulos, 2013).
CYP2C9[†]	NI–HBD	This legacy signature was motivated by acidic *substrate* chemistry rather than an established CYP2C9 inhibition mechanism and was retained unchanged as a negative or uncertain reference.

The relations specify pharmacophore-pair identity only. They do not assert a universal causal binding mechanism, a fixed centroid-distance window, or a required molecule-specific atom pair. [†]The CYP2C9 NI–HBD relation is therefore not interpreted as established inhibition chemistry.

B.5 Recovery of Established Recognition Chemistry

Four previously defined pharmacophore-pair signatures were retained without selecting new relations after inspecting the final comparison: hERG PI–AR, CYP2D6 PI–HP, CYP3A4 AR–HP, and a legacy CYP2C9 NI–HBD signature. The CYP2C9 signature was originally motivated by substrate chemistry and is therefore retained as an uncertain reference rather than an established inhibition mechanism. For PharmPD and CATS3D-style, absolute SHAP values were summed over distance channels to the common 21 pair types. The primary statistic was E_{excess}, the attribution-mass fraction on the prespecified reference pair minus the mean fraction on matched heterotypic decoy pairs sharing exactly one role and the training pair-presence category.

Table 22: Matched-decoy recovery of prespecified target-recognition relations.

Target	Reference	PharmPD E_{excess} [95% CI]	CATS3D-style E_{excess} [95% CI]
hERG	PI–AR	0.1977 [0.1826, 0.2090]	-0.0507 [-0.0576, -0.0433]
CYP2D6	PI–HP	0.1045 [0.0976, 0.1134]	0.3050 [0.2902, 0.3225]
CYP3A4	AR–HP	0.1278 [0.1256, 0.1299]	0.2099 [0.2071, 0.2128]
CYP2C9	NI–HBD†	-0.0038 [-0.0049, -0.0028]	-0.0027 [-0.0033, -0.0021]

Positive E_{excess} indicates that the prespecified relation carries more relation-block attribution mass than its matched decoys. PharmPD itself showed positive recovery for hERG, CYP2D6, and CYP3A4. †The CYP2C9 NI–HBD relation was retained from substrate-oriented chemistry and is not treated as established inhibition-mechanism support.

Table 23: PharmPD–CATS3D-style matched-decoy contrasts.

Target	Δ PP−CATS [95% CI]	BH q
hERG	+0.2484 [0.2287, 0.2644]	0.00050
CYP2D6	-0.2006 [-0.2104, -0.1919]	0.00050
CYP3A4	-0.0821 [-0.0841, -0.0801]	0.00050
CYP2C9	-0.0012 [-0.0017, -0.0007]	0.00050

Comparative recovery was target dependent: PharmPD exceeded CATS3D-style for hERG, whereas CATS3D-style had greater matched-decoy enrichment for CYP2D6 and CYP3A4. The reported q values are limited by the 2,000-draw Monte Carlo resolution.

Table 24: Rank and top-k recovery of the prespecified recognition relations.

Target	Model	Rank /21	Top-1	Top-3	Top-5	Ref. mass
hERG	CATS3D-style	7.65	0.0%	30.0%	40.6%	6.58%
hERG	PharmPD	2.10	40.0%	80.0%	100.0%	26.42%
CYP2D6	CATS3D-style	1.00	100.0%	100.0%	100.0%	32.39%
CYP2D6	PharmPD	1.20	80.0%	100.0%	100.0%	15.85%
CYP3A4	CATS3D-style	1.00	100.0%	100.0%	100.0%	27.37%
CYP3A4	PharmPD	1.20	80.0%	100.0%	100.0%	18.33%
CYP2C9	CATS3D-style	15.80	0.0%	0.0%	0.0%	0.62%
CYP2C9	PharmPD	17.30	0.0%	0.0%	0.0%	0.95%

Top-k recovery is computed from the global pair-type attribution ranking in each seed. These are pair-type-level results and do not establish recovery of a specific atom pair, protein contact, or causal binding mechanism.

B.6 Scope of the MapLight Chemical-Attribution Comparison

MapLight contains 2,048 Morgan/ECFP4 bits, 512 Avalon bits, 315 ErG values, and 217 RDKit2D descriptors. Only the ErG block was used for the common native pharmacophore-role comparison. Morgan features require post-hoc allocation to contributing atom environments and may collide in the hashed bit vector; the stored Avalon bits do not have an established unique pharmacophore mapping; and many RDKit2D columns are global molecular properties. ErG natively exposes six pharmacophore-role classes and pair structure, making it the defensible common role block. The remaining MapLight attribution is therefore treated as outside this common role-level comparison rather than as zero or failed recovery.

Table 25: Common pharmacophore-role recovery across MapLight, CATS3D-style, and PharmPD models.

Target	Block	Role rank /6	Top-2	Role mass	Block/total
hERG	MapLight (ErG)	2.900	45.0%	38.19%	17.35%
hERG	ML+CATS3D-style	2.625	50.0%	43.13%	22.00%
hERG	ML+PharmPD	2.100	70.0%	49.68%	30.34%
CYP2D6	MapLight (ErG)	3.700	30.0%	32.84%	18.75%
CYP2D6	ML+CATS3D-style	1.500	100.0%	60.53%	25.23%
CYP2D6	ML+PharmPD	2.000	50.0%	46.39%	32.43%
CYP3A4	MapLight (ErG)	3.450	50.0%	37.55%	12.89%
CYP3A4	ML+CATS3D-style	1.500	100.0%	68.30%	23.16%
CYP3A4	ML+PharmPD	1.500	100.0%	62.82%	27.48%
CYP2C9	MapLight (ErG)	5.500	0.0%	15.76%	16.54%
CYP2C9	ML+CATS3D-style	5.500	0.0%	8.84%	17.71%
CYP2C9	ML+PharmPD	5.000	0.0%	14.67%	23.52%

The supported roles are the two endpoints of the corresponding prespecified pair. This is a role-level comparison, not relation-level recovery. In the MapLight-only model, the interpretable ErG block accounts for only part of total model attribution; Morgan, Avalon, and RDKit2D attribution are not assigned post-hoc pharmacophore labels.

C Chemical Audit

The chemical audit contains several different denominators that answer different questions and are therefore reported separately. The site audit contains 108,701 evaluated sites from 38,044 molecules with at least one evaluable site. A separate positive-role coverage analysis contains 40,891 sites across 46,430 unique TDC molecules, and a separate assigned-center coverage sample contains 8,000 molecules. These denominators are not combined.

Table 26: Authoritative chemical-audit strata.

Stratum	Sites	Mols.	Stock err.	Stock %	Depl. err.	Depl. %
Independent controls	66,068	31,297	0	0.00%	0	0.00%
Independent nonarguable traps	6,474	4,741	0	0.00%	0	0.00%
Repair self-check	17,679	11,602	16,868	95.41%	2,688	15.20%
Donor-fix amide-N self-check	6,534	5,704	5,040	77.13%	0	0.00%
Arguable / quarantined	11,946	9,300	10,664	89.27%	3,655	30.60%

The final strata separate independent controls/traps from rule-development self-checks and arguable chemistry. The stale historical trap total of 16,539 sites is not used. Original 25-category family totals were 66,068 controls, 24,812 repairs, and 17,821 traps; final reporting separates donor-fix self-checks and quarantines arguable sites.

Table 27: Positive-role coverage audit for 12 unambiguous chemical groups.

Group	Role	Sites	Stock fail	Stock %	Deployed %
Amidine	PI	1,479	633	42.80%	0.41%
Guanidine	PI	432	0	0.00%	0.00%
Tetrazole	NI	29	29	100.00%	0.00%
Phosphate/phosphonate	NI	411	411	100.00%	0.00%
Acylsulfonamide	NI	163	163	100.00%	0.00%
Hydroxamate	NI	53	53	100.00%	0.00%
Nitrile	HBA	2,085	2,085	100.00%	0.00%
N,N-dialkylaniline	HBA	4,237	4,237	100.00%	1.82%
Aryl chloride	ANY	9,574	9,574	100.00%	0.00%
Aryl bromide	ANY	1,268	1,268	100.00%	0.00%
Aryl iodide	ANY	339	339	100.00%	0.00%
Donor assignment on N without H	HBD	20,821	13,173	63.27%	0.00%
Total	mixed	40,891	31,965	78.17%	0.20%

This coverage analysis uses a separate denominator from the site-audit table above. Across 40,891 sites in the 12 unambiguous groups, the stock detector failed on 31,965 sites (78.17%), whereas the deployed PharmPD typing failed on 83 sites (0.203%). This is a targeted positive-role coverage audit rather than a universal chemical-accuracy estimate.

D Representation Controls and Statistical Details

D.1 pH-Aware Pharmacophore Weights

After pharmacophore identities were assigned, PharmPD associated each center with a scalar weight w_i. Ionizable and selected protonation-dependent pharmacophore features were treated as pH dependent because protonation state influences molecular charge, solvation, hydrogen-bonding behavior, membrane permeability, and protein–ligand recognition Sipos-Szabó et al. (2026); Petukh et al. (2013).

$$\frac{1}{1 + 10^{\mathrm{p}K_a - 7.4}}. \tag{13}$$

The resulting population was assigned as the scalar weight of the corresponding pharmacophore center and subsequently propagated into every three-dimensional pair feature involving that center. Thus, for center i, w_i modulates its contribution throughout the pair-distance representation rather than defining a separate atom-level descriptor.

Three numerical thresholds served distinct purposes. First, an ionizable PI or NI feature was emitted by the featurizer only when its estimated charged fraction was at least 0.05. Groups below this threshold contributed no corresponding ionizable center. Second, a protonation-restored HBD center required an estimated protonated fraction of at least 0.20, because the donor role was intended to represent a sufficiently populated state carrying a transferable proton. Third, the independent categorical chemistry audit considered a pharmacophore assignment present only when its center weight was at least 0.20. The audit criterion was therefore deliberately stricter than the feature-emission criterion. Consequently, an ionizable group satisfying

$$0.05 \leq f < 0.20 \tag{14}$$

could in principle contribute a weak PI or NI center to PharmPD while still being scored as absent in the categorical audit. No evaluated audit category occupied this interval in the reported analysis.

Weights were assigned according to the physical interpretation available for each pharmacophore class. For newly defined acidic NI environments with literature-supported group-specific $\mathrm{p}K_a$ values, the corresponding estimated deprotonated fractions were used directly. Protonation-derived HBD centers were weighted by the estimated fraction of molecules carrying the corresponding donor proton. Selected PI environments were represented using structure-dependent or curated ionization fractions. Pre-existing aromatic PI centers were assigned ring-context-dependent fractions, whereas

common aliphatic amine PI centers retained a fixed weight of 0.97 in the reported featurizer. The latter is an approximation rather than a compound-specific estimate, because aliphatic amine basicity can vary substantially with local chemical environment.

Conventional AR, HBA, HBD, and nonhalogen HP centers that were not governed by a protonation-dependent rule received unit weight. Halogen-derived HP centers used the fixed weights described in Appendix D.1. The numerical interpretation of w_i therefore differs across pharmacophore classes. PI and NI weights can approximate ionized-state populations, protonation-restored HBD weights approximate the population carrying a transferable donor proton, halogen HP weights provide a relative nonpolar-contact scale, and most HBA, AR, conventional HBD, and nonhalogen HP centers retain operational unit weights. Equal numerical values across these channels should therefore not be interpreted as physically commensurate interaction strengths.

D.1.1 Quantitative Audit of Pharmacophore Weights

Chemical validity of the pharmacophore identity and validity of the numerical center weight were evaluated separately. The site-level categorical chemistry audit described in Appendix C tested whether a pharmacophore label was appropriately present or absent. A second quantitative audit was restricted to channels for which the assigned weight had a defined physical referent, specifically PI, NI, and protonation-derived HBD centers.

Literature-supported group-specific $\mathrm{p}K_a$ or $\mathrm{p}K_{aH}$ values were converted to expected charged-state populations using the same Henderson–Hasselbalch relationships used by the featurizer. A center weight was considered discrepant when its absolute deviation from the literature-derived population exceeded 0.10 or when the ratio between the assigned and reference values differed by more than a factor of 1.5. When the expected population was below the 0.05 feature-emission threshold, absence of the corresponding center was treated as agreement.

HBA and AR weights were not included in this quantitative weight audit because their unit weights do not correspond to calibrated hydrogen-bond-basicity, aromatic-interaction, or interaction-energy scales. PharmPD therefore represents the presence and spatial arrangement of these interaction classes but does not attempt to encode experimentally calibrated differences in interaction strength among individual HBA or AR environments. This distinction was retained explicitly when interpreting the resulting representation.

D.2 Implementation and Conformer Generation

For multicomponent records, only the fragment containing the largest number of heavy atoms was retained, and molecules with fewer than three heavy atoms were excluded from featurization. Explicit hydrogens were added with `Chem.AddHs` before three-dimensional embedding. For the conformer cache used in the reported experiments, up to 10 conformers per molecule were generated using RDKit ETKDGv3 Riniker & Landrum (2015); Wang et al. (2020), with `randomSeed=42` and an RMSD pruning threshold of 0.5 Å. The ETKDGv3 options `useMacrocycleTorsions`, `useMacrocycle14config`, `useSmallRingTorsions`, and `useBasicKnowledge` were enabled. All other ETKDGv3 parameters were left at their RDKit defaults. If the initial embedding attempt produced no conformers, embedding was retried once with `useRandomCoords=True`. Each successfully embedded conformer was subsequently minimized using MMFF94s Halgren (1999), with at most 500 minimization iterations per conformer. MMFF convergence tolerances were left at their RDKit defaults, and the returned convergence status was not used as an additional filtering criterion. Conformers for which MMFF processing and energy evaluation completed without exception were retained; no energy-window filtering, post-MMFF RMSD pruning, or additional top-k selection was applied. After RMSD pruning and MMFF processing, the cached molecules contained a median of 7 retained conformers and at most 10. PharmPD relation features were evaluated independently on every retained conformer. For each of the 189 pair-distance channels, the final molecular representation was obtained by taking the element-wise maximum across conformers. Thus, PharmPD does not select a single common representative conformer: different feature channels may attain their maxima in different conformers. Conformer generation was treated as failed if both embedding attempts failed, if MMFF94s parameterization was unavailable, or if no conformer yielded a valid MMFF energy.

Table 28: **Downstream model configurations used across descriptor arms.**

Setting	CatBoost	XGBoost
Classification objective	`Logloss`	`binary:logistic`
Regression objective	`RMSE`	`reg:squarederror`
Maximum rounds	`iterations=1000`	`n_estimators=500`
Learning rate	0.05	0.05
Tree depth	`depth=6`	`max_depth=6`
Early stopping	50 rounds	30 rounds
Classification metric	`PRAUC`	`auc`
Regression metric	`RMSE`	`rmse`
Sampling	Library defaults	`subsample=0.8`; `colsample_bytree=0.8`
Explicit L1/L2	None	None
Execution	CPU	`hist`; CPU
Threads	Scheduler: 8 CPUs	`n_jobs=8`
Class weighting	None	None

D.3 Model Training and Hyperparameters

CatBoost and XGBoost models were trained using fixed model-family-specific configurations rather than separately tuning the downstream learner for each descriptor arm. For each training partition, 10% of the data were reserved as an early-stopping set; binary classification splits were stratified. No class weighting was applied. The same downstream configuration was used when comparing descriptor variants within a given model family and task type, so differences between the primary MapLight and MapLight+PharmPD arms arose from their input representations rather than arm-specific model hyperparameter optimization.

For CatBoost, classification used `CatBoostClassifier` with `Logloss`, while regression used `CatBoostRegressor` with `RMSE`. Models were allowed up to 1,000 boosting iterations with a learning rate of 0.05 and tree depth 6. Early stopping used a patience of 50 rounds, with `PRAUC` as the classification evaluation metric and `RMSE` for regression. Sampling and regularization parameters not explicitly specified were left at the CatBoost library defaults. Training was performed on CPU; eight CPUs were allocated by the job scheduler, while `thread_count` was not set explicitly.

For XGBoost, classification used `XGBClassifier` with the `binary:logistic` objective, and regression used `XGBRegressor` with `reg:squarederror`. Models used `n_estimators=500`, `learning_rate=0.05`, and `max_depth=6`, with `subsample=0.8` and `colsample_bytree=0.8`. Early stopping used a patience of 30 rounds, with `auc` for classification and `rmse` for regression. No explicit L1 or L2 regularization parameter was set beyond the library defaults. XGBoost was run on CPU using `tree_method="hist"`, `device="cpu"`, and `n_jobs=8`.

D.4 Rank-1 Matched-Marginal Control

Let $H \in \mathbb{R}^{21\times 9}$ denote the final nonnegative pair-by-distance matrix after conformer-wise max pooling. Define $r_i = \sum_j H_{ij}$, $c_j = \sum_i H_{ij}$, and $Z = \sum_{ij} H_{ij}$. The Rank-1 control is

$$\widetilde{H}_{ij} = \frac{r_i c_j}{Z}, \tag{15}$$

when $Z > 0$, and the all-zero matrix when $Z = 0$. It is computed after conformer max pooling and preserves the row marginals, column marginals, total mass, and all 189 flattened pair-block columns, together with the same 11 auxiliary values for 200 total dimensions. It is not an SVD or a fitted rank-one approximation.

D.5 Matched Topological-Distance Control

The topological control uses the largest heavy-atom fragment, adds explicit hydrogens, performs MMFF94s property setup, and applies the same six pharmacophore-center types and weights as Full. For centers with constituent atom sets A and B, the distance is the arithmetic mean of all shortest bond-path lengths between $a \in A$ and $b \in B$, multiplied by 1.4 Å/bond. This graph distance replaces Euclidean distance in the same 21 pair types and nine Gaussian channels ($\mu = 1.5, \ldots, 9.5$ Å; $\sigma = 0.55$ Å), with the same ordered same-type counting, identical-atom-set self exclusion, weights, and $0.3 \leq d \leq 10.5$ Åcutoff. There is one graph per molecule, so no coordinate-based conformer pooling is performed. The 11 auxiliary values are held identical to Full.

D.6 Local CATS3D-Style Comparator

The comparator is a local CATS3D-style implementation rather than a demonstrated exact reproduction of literature CATS3D. Stock RDKit feature definitions are grouped into PI, NI, AR, HBA, HBD, and HP centers. Multi-atom centers use the arithmetic mean of their member-atom coordinates. The representation contains 21 unordered type pairs and ten hard 1-Åbins per pair, $[0.5 + b, 1.5 + b)$ for $b = 0, \ldots, 9$; exactly 10.5 Åis excluded by the implemented index check. Values are raw counts with no normalization and are max-pooled independently by column across conformers. Same-type identical-atom-set self pairs are excluded, while ordered same-type pairs are both counted. The final dimensionality is $21 \times 10 = 210$.

D.7 Statistical Protocols

Table 29: Primary estimands and inferential procedures across final experiments.

Experiment	Primary estimand and inference
TDC-22 historical prediction	Mean of 10 individual-seed metrics. No stored scaffold interval or multiple-testing correction; historical paired-seed t tests are retained as nominal.
7-task standalone 5-arm ablation	Mean of 10 individual-seed metrics. Paired largest-fragment Murcko-scaffold bootstrap, 2,000 draws; percentile 95% CI. BH over 21 primary contrasts.
4-task descriptor augmentation	Mean of 10 individual-seed metrics. Paired largest-fragment Murcko-scaffold bootstrap, 2,000 draws; percentile 95% CI. BH over 32 prespecified contrasts.
PXR challenge	Metric of seed-mean predictions. Cluster bootstrap over 96 analog series in the pooled $n = 513$ challenge set, 10,000 draws. No BH family.
Cross-descriptor explanation	Molecule-weighted mean-seed test MSE at 8 leaves. Task-specific largest-fragment Murcko-scaffold bootstrap, 2,000 draws. BH over 24 comparisons.
Final recognition recovery	Mean of 10 per-seed E_{excess} ratios. Paired largest-fragment Murcko-scaffold bootstrap, 2,000 draws. BH over 4 frozen target contrasts.

Seeds are repeated fits rather than independent test cohorts. Multiple-testing families are experiment-specific and are not pooled across unrelated analyses.

E Extended Discussion

E.1 Semantic–Geometric Coupling Is Endpoint Dependent

Our results indicate that preserving semantic–geometric coupling can provide predictive information beyond the pharmacophore content or distance distribution alone, but that this benefit is not uniform across endpoints. The clearest gains from preserving type-pair-specific 3D structure were observed

for hERG and several of the Veith CYP endpoints, whereas other endpoints showed weaker or inconsistent gains. In particular, the comparison with the rank-1 control is informative because this representation preserves the marginal pharmacophore and distance structure while disrupting their pair-specific correspondence. The resulting differences therefore suggest that, for at least a subset of the tasks considered here, predictive signal resides not only in which interaction features are present or which distances occur, but in *which chemical roles occur at those distances*.

This endpoint dependence is also chemically plausible. Different measured properties need not depend on molecular recognition to the same degree, nor must they depend on the same aspects of recognition. The contribution of a particular relational representation may therefore vary with the underlying mechanism, assay definition, composition of the measured chemical space, and other sources of experimental variation. Dataset coverage itself can substantially affect molecular machine-learning behavior and its domain of applicability (Kretschmer et al., 2025). The present experiments do not isolate which of these factors explains the observed endpoint-level differences, and the seven endpoints studied here were intentionally selected because target recognition provides a plausible mechanistic connection to their labels. Consequently, these results should not be interpreted as evidence that semantic–geometric coupling will improve every molecular property.

Instead, they suggest a more conditional principle: explicit coupling is most likely to be useful when relevant predictive signal can be expressed through the spatial organization of chemically meaningful roles. Determining when this condition holds is itself an important direction for future work. Evaluation across broader endpoint classes, independent datasets, and prospective chemical series could help distinguish properties that benefit from relational 3D representations from those dominated by other molecular or experimental factors.

E.2 Why Can a Coarse Representation Remain Informative?

PharmPD is deliberately coarse relative to an atomistic description of a molecule. It does not encode a complete electronic structure, molecular surface, or continuous interaction field. Instead, it compresses a conformer into chemically defined interaction roles and their pairwise spatial relationships. Such abstraction is closely related to the motivation of pharmacophore modeling, in which molecular recognition is represented through the three-dimensional arrangement of chemical functionalities rather than through every atomic degree of freedom (Vuorinen & Schuster, 2015). More generally, molecular representation is itself a consequential modeling choice: hand-crafted chemical descriptors can introduce domain knowledge and useful inductive biases that differ from those available to fully learned representations (Artrith et al., 2021).

One possible interpretation of the present results is therefore that relevant recognition information can exist at an intermediate level of description. Global molecular descriptors may be too coarse to retain the arrangement of local interaction roles, whereas a complete atomistic representation contains substantially more detail than is required to express those relations. PharmPD occupies a level between these extremes: it retains *what type of interaction a region may support* and *where that role lies relative to other roles*, while discarding much of the remaining atomic detail. The performance of this representation suggests that such a chemically structured, mesoscopic description can preserve useful signal for some recognition-related endpoints.

This compression may also provide a statistical advantage, although the current experiments were not designed to establish this directly. A sufficiently expressive neural model could in principle learn interactions between chemical identity and geometry from more primitive inputs. With finite labeled data, however, explicitly providing a chemically motivated relation reduces the number of intermediate relationships that must be discovered solely from supervision. In this sense, semantic–geometric coupling can be viewed as an inductive bias: it does not increase the amount of raw molecular information, but reorganizes selected information into a form that reflects a plausible structure of molecular recognition.

Importantly, the same abstraction also contributes to inspectability. A highly detailed or latent representation may support accurate prediction while making it difficult to translate an attribution into a recognizable SAR statement. In PharmPD, an attributed feature already refers to a chemically named pair of interaction roles and a spatial scale, which can be traced back to concrete centers in a molecular conformer. Thus, the coarseness of the representation is both a limitation and a design

choice. It inevitably omits potentially relevant molecular detail, but the retained quantities remain sufficiently structured to support both prediction and chemically grounded attribution.

E.3 Beyond Pharmacophore Roles: Richer Semantic–Geometric Coupling

The semantic component used here is intentionally narrow. PharmPD couples pharmacophore identity with inter-feature distance, but molecular recognition is governed by a richer combination of physical and chemical factors. Binding depends on interacting energetic, solvation, and conformational contributions (Gilson & Zhou, 2007; Baron & McCammon, 2013), none of which is completely represented by a discrete pharmacophore label and a pairwise distance. Two substructures assigned the same interaction role may differ substantially in size, charge distribution, local electrostatic environment, flexibility, or accessibility, and these differences may alter how the corresponding spatial arrangement contributes to activity.

This limitation suggests a direct generalization of the framework. Future representations could couple geometry not only to pharmacophore identity, but also to steric volume, partial charge or electrostatic potential, hydrophobic character, polarizability, electron-density-derived descriptors, conformational flexibility, or other local physicochemical quantities. The key extension would not simply be to concatenate these properties as additional molecular descriptors. Rather, it would be to retain their correspondence with molecular position and with the other regions to which they are spatially related. A statement such as "this molecule contains a hydrophobic region" is fundamentally different from representing *which* hydrophobic region occurs near *which* charged or hydrogen-bonding environment.

A related direction has already emerged in molecular generative modeling. ShEPhERD jointly represents three-dimensional molecular shape, electrostatic potential, and directional pharmacophores, and uses these explicit interaction profiles to condition the generation of molecules with desired three-dimensional characteristics (Adams et al., 2025). Although its objective is molecular generation rather than property prediction, this work reflects a closely related representation principle: chemically meaningful information can become more useful when its spatial organization is retained rather than treated as an independent molecular summary. The present results suggest that the same principle may also be useful on the predictive side, where explicit semantic–geometric relations can contribute to both property prediction and chemically grounded attribution.

From this perspective, the specific 200-dimensional PharmPD descriptor is only one simple realization of a broader representation principle. The central idea is to preserve relationships between heterogeneous sources of molecular information rather than allowing those relationships to disappear when the information is summarized independently. This principle could support continuous rather than discretized chemical semantics, learned distance kernels, higher-order relations among three or more molecular regions, or representations in which ligand-side relations are coupled directly to complementary properties of a target pocket. Such extensions would move from a pharmacophore–distance descriptor toward a more integrated description of the physicochemical organization that produces molecular recognition.

E.4 From Explicit Coupling to Learned Relational Representations

The principle illustrated by PharmPD can be expressed more generally in terms of interactions among molecular information sources. Let $X = (X_1, \ldots, X_m)$ denote heterogeneous molecular inputs, such as geometric, electronic, and pharmacophoric information, and let $f(X)$ denote the prediction function for an endpoint. An additive representation would be sufficient if

$$f(X) \approx \beta_0 + \sum_i f_i(X_i),$$

but a general molecular prediction function may instead contain cross-source terms,

$$f(X) = \beta_0 + \sum_i f_i(X_i) + \sum_{i<j} f_{ij}(X_i, X_j) + \cdots,$$

such that predictive signal depends on relationships between information sources rather than on either source independently. A sufficiently expressive downstream model can in principle learn such terms, but only if the representation supplied to it preserves the information required to distinguish them. If a representation $R = T(X)$ maps two molecular states X and X' to the same representation despite $P(Y \mid X) \neq P(Y \mid X')$, no downstream predictor operating only on R can recover the discarded distinction. In information-theoretic terms, complete preservation of endpoint-relevant information would require $I(Y; X \mid R) = 0$; whenever this quantity is nonzero, predictive information has been lost before learning begins.

The Full–Rank-1 comparison provides a concrete instance of this distinction. For the PharmPD matrix M, let

$$r_p = \sum_k M_{p,k}, \qquad c_k = \sum_p M_{p,k}, \qquad Z = \sum_{p,k} M_{p,k}.$$

The Rank-1 control is

$$\widetilde{M}_{p,k} = \frac{r_p c_k}{Z},$$

so that the full representation can be written as

$$M_{p,k} = \frac{r_p c_k}{Z} + \Delta_{p,k}, \qquad \Delta_{p,k} = M_{p,k} - \frac{r_p c_k}{Z}.$$

Because $\sum_k \Delta_{p,k} = 0$ and $\sum_p \Delta_{p,k} = 0$, the residual Δ contains neither additional pair-type abundance nor additional distance abundance; it contains their correspondence. Equivalently, after normalizing M to $q_{p,k} = M_{p,k}/Z$, the Rank-1 construction becomes $q_p q_k$, the product of the two marginals. Thus, conditional on the preserved marginals, any additional predictive information available to the full representation must reside in its joint semantic–geometric structure. Formally,

$$I(Y; M) = I(Y; r, c) + I(Y; M \mid r, c),$$

where the final term represents endpoint-relevant information carried by the joint structure beyond the two marginals. The empirical Full–Rank-1 comparison should not be interpreted as a direct estimate of this information quantity, but it provides a controlled test of whether such additional relational information is useful for prediction.

This principle need not remain tied to a fixed descriptor. PharmPD explicitly specifies one relation,

$$\psi(X_{\text{pharmacophore}}, X_{\text{geometry}}),$$

whereas future representation-learning approaches could learn relational functions

$$R_\theta(X) = \left[\{\phi_i(X_i)\}, \{\psi_{ij}^\theta(X_i, X_j)\}_{i<j}\right]$$

from data. Such models could replace fixed pharmacophore definitions, distance channels, or pairwise kernels with learned components while retaining the central inductive bias that chemically meaningful information remains associated with the geometric, electronic, or molecular context in which it occurs. From this perspective, the broader design problem is not simply how many molecular information sources should be added to a representation, but which relationships among them must remain identifiable after representation learning.

E.5 PharmPD as a Prototype for Predictive and Explainable Relational Representations

PharmPD should therefore be viewed as a deliberately simple prototype rather than a complete representation of molecular recognition. Several approximations in the current implementation make this distinction important. First, the 3D representation depends on computationally generated conformers and their subsequent aggregation. Molecules generally occupy conformational ensembles,

and a single geometry or simplified aggregation cannot fully represent the distribution of accessible states (Axelrod & Gómez-Bombarelli, 2022). The current max-pooling strategy provides a practical way to retain strong geometric signals across conformers, but it does not model conformer populations, energetic preferences, or uncertainty. Future implementations could incorporate energy-aware weighting, learned conformer aggregation, or explicit estimates of conformational uncertainty.

Second, PharmPD is ligand based. It represents internal organization of the ligand but does not explicitly model the receptor environment. Consequently, an attributed pharmacophore relation should not be interpreted as evidence of a particular binding pose or as direct proof of a specific protein–ligand interaction. Similarly, pharmacophore assignments, ionization estimates, and distance discretization introduce modeling assumptions of their own. Integrating target-pocket information and evaluating attributed relations against experimentally resolved complexes would provide a stronger test of whether the relations highlighted by the model correspond to the interactions realized in binding.

The scope of the explainability claims is likewise intentionally limited. Explainable machine learning is particularly relevant in medicinal chemistry because useful predictions often need to be rationalized in terms that can support compound optimization and chemical decision making (Jiménez-Luna et al., 2020; Rodríguez-Pérez & Bajorath, 2021). Our analyses show that PharmPD attribution can be traced to concrete molecular features, computationally audited, compared with established recognition chemistry, and used as an explanatory vocabulary for approximating learned predictions. They do not establish that these explanations are easier for medicinal chemists to understand, nor that they improve human SAR reasoning or compound-design decisions. Expert-centered studies measuring hypothesis generation, error detection, or prospective design utility would therefore be a necessary next step.

Taken together, these limitations also define the opportunity suggested by the present results. The goal is not necessarily to construct an increasingly detailed descriptor that attempts to encode every molecular property. Instead, the results motivate representations that identify chemically relevant information and preserve the relationships among those quantities explicitly. In this view, PharmPD is an initial demonstration of a broader direction: semantic–geometric coupling can provide a bridge between predictive molecular modeling and chemically grounded SAR reasoning. Richer implementations that couple additional physicochemical semantics with geometry, conformational context, and eventually target information may provide a path toward models whose predictive signals remain directly connected to molecular relationships that can be inspected, challenged, and used to formulate chemical hypotheses.