HIDDEN IN THE FOLD: CRYPTIC POCKET DISCOVERY WITH LOCAL-GLOBAL GEOMETRIC CONTEXT

Anonymous authors
Paper under double-blind review

ABSTRACT

Cryptic pockets are transient or weakly exposed binding sites that are often invisible in static protein structures, yet they play important roles in ligand recognition, allosteric regulation, and drug discovery. Unlike canonical surface pockets, cryptic pockets may only emerge through subtle conformational rearrangements, making their discovery a challenging problem that requires reasoning beyond apparent local concavity. In this work, we propose **PocketMon**, a context-conditioned equivariant residue field for cryptic pocket discovery. It treats cryptic pocket discovery as a local-global geometric inference problem, where visible residue geometry and latent residue-pair interactions jointly indicate hidden binding opportunities. To this end, **PocketMon** integrates equivariant geometric modeling with general pairwise residue priors, enabling whole-protein residue-level prediction. We further curate and standardize three cryptic-pocket discovery benchmarks to systematically evaluate this problem. Across these benchmarks, **PocketMon** consistently outperforms competitive baselines, with up to 14.83% absolute PR-AUC improvement on PocketMiner, and we further apply it to human proteome-scale cryptic pocket detection to provide a broad map of potential hidden binding sites.

1 INTRODUCTION

Proteins are not rigid molecular objects, but dynamic systems that fluctuate among conformational states. Such conformational plasticity can transiently expose binding sites that are absent from experimentally observed ground-state structures. These hidden sites, known as *cryptic pockets*, have emerged as attractive opportunities for drug discovery because they expand the ligandable proteome beyond canonical active sites and visible surface cavities (Vajda et al., 2018; Amaro, 2019; Beglov et al., 2018; Kuzmanic et al., 2020). They are particularly relevant for proteins traditionally considered difficult to target, where ligand binding may occur only after local rearrangements, side-chain repacking, loop motion, or larger conformational transitions (Bowman & Geissler, 2012; Oleinikovas et al., 2016; Sun et al., 2020).

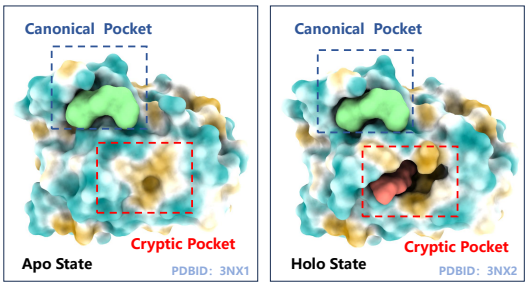

Figure 1: Canonical pocket versus cryptic pocket.

The distinction between canonical and cryptic pockets is illustrated in Figure 1. A conventional pocket is already accessible in the apo structure and can often be detected by geometric pocket finders through surface concavity, alpha spheres, or solvent-accessible probes (Hendlich et al., 1997; Le Guilloux et al., 2009; Krivák & Hoksza, 2018). In contrast, a cryptic pocket is closed, collapsed, or only weakly exposed in the apo state, becoming binding-competent only in an excited or ligand-stabilized conformation. Therefore, cryptic pocket discovery is not merely static cavity detection; it requires predicting where a protein can create a hidden binding site through latent conformational rearrangement. Residue localization alone leaves the ligand-compatible open conformation unresolved. Systematic identification of cryptic pockets remains challenging. Experimental discovery often depends on screening campaigns and ligand-bound structural characterization (Vajda et al., 2018; Meller et al., 2023), while molecular dynamics and enhanced sampling can expose pocket-opening events but are too costly for routine proteome-scale screening (Bowman

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.

& Geissler, 2012; Oleinikovas et al., 2016; Comitani & Gervasio, 2018; Zimmerman et al., 2021; Kuzmanic et al., 2020). Recent learning-based efforts improve scalability by deriving residue-level cryptic-site supervision from apo–holo structural changes or molecular simulations (Cimermancic et al., 2016; Meller et al., 2023; Martinez et al., 2026), and emerging benchmarks further emphasize the need for standardized evaluation of cryptic binding-site prediction (Škrhák et al., 2025).

However, a central difficulty remains: ***cryptic pockets must be inferred from structures in which the binding site may not yet be geometrically visible***. Purely local pocket descriptors can miss collapsed sites, while sequence-only models may overlook the geometric mechanisms that make a residue patch binding-competent. This motivates a different view: cryptic pocket discovery should be treated as a *local-global geometric inference* problem. Locally, pocket formation depends on residue geometry, backbone frames, side-chain orientation, and spatial neighborhood structure. Globally, the opening of a cryptic pocket is constrained by long-range residue couplings, fold-level context, and latent pairwise interactions. This perspective aligns with recent progress in geometric deep learning for proteins, where equivariant neural networks provide coordinate-aware representations of molecular structure (Fuchs et al., 2020; Jing et al., 2021; Satorras et al., 2021), and with pretrained protein models whose internal representations encode transferable structural and functional signals (Jumper et al., 2021; Baek et al., 2021; Lin et al., 2023; Gazizov et al., 2026).

In this work, we propose ***PocketMon***, a pair-conditioned equivariant residue field for cryptic pocket discovery. **PocketMon** predicts cryptic pocket residues over the whole protein by integrating local equivariant geometric modeling with general pairwise residue priors. Importantly, this formulation is not tied to a specific source of pair representations; any informative residue-pair representation can serve as the global prior. We further group predicted residues into spatial patches and use ProbeEBM to rank molecular probes by patch–ligand compatibility. The selected probes and pocket constraints then guide Boltz-2 (Passaro et al., 2025) toward candidate open-pocket complexes.

To systematically evaluate cryptic pocket discovery, we organize three complementary benchmarks covering simulation-derived pocket opening, apo–holo cryptic-site annotations, and pair-representation-based binding-site data. Across these benchmarks, **PocketMon** consistently improves over competitive baselines, including an absolute PR-AUC gain of 14.83% on PocketMiner (Meller et al., 2023). Beyond curated benchmarks, we further apply **PocketMon** to human proteome-scale cryptic pocket detection, producing a broad map of potential hidden binding sites for downstream drug discovery and functional exploration. The main contributions of this work are three-fold:

❶ We formulate cryptic pocket discovery as a local-global geometric inference problem, emphasizing the joint role of visible residue geometry and latent residue-pair interactions.

❷ We introduce **PocketMon**, a context-conditioned equivariant model for whole-protein, residue-level cryptic pocket prediction from a single structure. We further couple a patch-conditioned probe energy with constrained complex prediction to obtain candidate open-state structures.

❸ We curate and standardize three cryptic-pocket discovery benchmarks, demonstrate consistent improvements over strong baselines, and extend prediction to the human proteome scale.

2 Related Work

Cryptic pocket discovery. Cryptic pockets are ligandable sites that are absent or poorly exposed in apo protein structures, but become accessible through conformational rearrangements (Vajda et al., 2018; Beglov et al., 2018; Kuzmanic et al., 2020). Compared with conventional pocket detection, cryptic pocket discovery is more challenging because the target site may not be geometrically visible in the input structure. Classical pocket finders, such as LIGSITE (Hendlich et al., 1997), Fpocket (Le Guilloux et al., 2009), and P2Rank (Krivák & Hoksza, 2018), are effective for identifying exposed cavities, but are not designed to infer collapsed or latent binding sites. Molecular dynamics and enhanced sampling methods can reveal cryptic pocket opening by exploring conformational ensembles (Oleinikovas et al., 2016; Comitani & Gervasio, 2018; Zimmerman et al., 2021; Borsatto et al., 2024), yet their computational cost limits large-scale screening. This motivates learning-based methods that predict cryptic sites directly from static structures or sequences.

Learning-based cryptic pocket prediction. Several data-driven methods have been proposed to make cryptic pocket discovery scalable. CryptoSite learns residue-level cryptic-site predictors

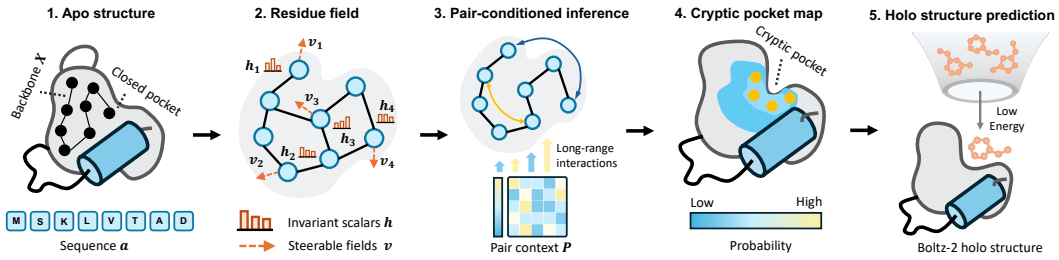

Figure 2: Overview of the workflow of **PocketMon**, which consists of backbone-informed initialization, global representation conditioned inference, and probe-guided structure prediction.

from experimentally observed apo–holo structural changes (Cimermancic et al., 2016). PocketMiner trains a graph neural network to predict residues likely to participate in simulation-derived pocket-opening events from a single protein structure (Meller et al., 2023). CryptoBank constructs a larger apo–holo-derived cryptic-site resource and studies sequence-based prediction with protein language models (Martinez et al., 2026). Recent benchmark efforts such as CryptoBench further emphasize the need for standardized evaluation of cryptic protein–ligand binding-site prediction (Škrhák et al., 2025). Our work follows this line of residue-level prediction, but focuses on modeling cryptic pockets as a local-global geometric inference problem.

Protein representations and geometric learning. Protein representation learning has recently provided strong priors for structure and function prediction. Large pretrained models such as AlphaFold (Jumper et al., 2021), RoseTTAFold (Baek et al., 2021), and ESMFold (Lin et al., 2023) encode rich structural context, while AF2BIND shows that pair representations can be highly informative for small-molecule binding-site prediction (Gazizov et al., 2026). In parallel, equivariant geometric networks, including SE(3)-Transformers (Fuchs et al., 2020), E(n)-equivariant graph neural networks (Satorras et al., 2021), and geometric vector perceptrons (Jing et al., 2021), provide symmetry-aware tools for learning from 3D molecular structures. **PocketMon** connects these two directions by conditioning equivariant residue-level geometry on global pairwise residue priors, enabling whole-protein cryptic pocket prediction from a single structure.

3 Preliminaries

Problem formulation. Given a protein with sequence $\mathbf{a} = (a_1, \ldots, a_L)$ and backbone coordinates $\mathbf{X} = \{(\mathbf{x}_i^N, \mathbf{x}_i^\alpha, \mathbf{x}_i^C)\}_{i=1}^L$, cryptic pocket discovery aims to identify residues forming hidden or transient binding sites, which are often collapsed in the observed conformation. We formulate this as a whole-protein residue-level prediction task:

$$f_\theta : (\mathbf{a}, \mathbf{X}, \mathbf{P}) \mapsto \hat{\mathbf{y}} \in [0,1]^L, \tag{1}$$

where $\hat{y}_i$ is the predicted probability of residue i belonging to a cryptic pocket, and $\mathbf{P} \in \mathbb{R}^{L \times L \times c_p}$ is an optional invariant pair representation. The predictor satisfies $f_\theta(\mathbf{a}, g \cdot \mathbf{X}, \mathbf{P}) = f_\theta(\mathbf{a}, \mathbf{X}, \mathbf{P})$ for any $g \in \mathrm{SE}(3)$.

Protein residue graph. We model the protein as a spatial graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes $\mathcal{V} = \{1, \ldots, L\}$ centered at Cα coordinates $\mathbf{p}_i = \mathbf{x}_i^\alpha$. An edge $(i,j) \in \mathcal{E}$ exists if $r_{ij} = \|\mathbf{p}_j - \mathbf{p}_i\|_2 \le r_{\text{cut}}$. For each residue i, we denote its spatial neighbors by $\mathcal{N}(i)$ and the unit edge direction to neighbor j by $\hat{\mathbf{r}}_{ij} = (\mathbf{p}_j - \mathbf{p}_i)/r_{ij}$.

Equivariant geometric representation. Our local geometric encoder follows the invariant-coefficient and steerable-basis principle of GotenNet (Aykent & Xia, 2025). It maintains SE(3)-invariant scalar features and steerable geometric features, using invariant features to modulate directional bases. We instantiate this principle at the residue level with backbone-derived fields.

4 METHODOLOGY

To connect residue-level discovery to explicit structural hypotheses, we first predict cryptic-site residues from the apo structure and then rank molecular probes for constrained complex prediction. The geometric details and symmetry proof appear in Sections B and C.

4.1 BACKBONE-INFORMED EQUIVARIANT RESIDUE FIELD

We begin by embedding protein geometry into a residue-level state comprising $\mathrm{SE}(3)$-invariant scalars and steerable vectors. The residue construction is central to our design: in addition to inter-residue distances, it makes local chemical composition, backbone conformation, and side-chain orientation available before equivariant message passing.

Invariant scalar states. For each interacting residue pair (i, j), we define an edge descriptor that combines physical proximity and sequence separation:

$$\boldsymbol{\epsilon}_{ij} = [\boldsymbol{\varphi}(r_{ij}); \mathrm{PE}(j-i)], \qquad \varphi_k(r) = c(r)\exp\{-\beta_k(e^{-\alpha r} - \mu_k)^2\}, \tag{2}$$

where $c(r) = \frac{1}{2}\{\cos(\pi r/r_{\mathrm{cut}}) + 1\}\mathbf{1}\{r < r_{\mathrm{cut}}\}$ is a smooth cutoff and PE is a learnable relative positional embedding. The invariant residue state starts from amino acid identity and backbone dihedral angles:

$$\mathbf{h}^0_{i,\mathrm{raw}} = \mathrm{Emb}(a_i) + W_\tau[\cos\phi_i, \cos\psi_i, \cos\omega_i, \sin\phi_i, \sin\psi_i, \sin\omega_i]. \tag{3}$$

To expose local chemical context prior to deep interaction, we perform a one-hop, distance-weighted aggregation:

$$\mathbf{h}^0_i = \mathrm{LN}\left(W_h\left[\mathbf{h}^0_{i,\mathrm{raw}}; \sum_{j\in\mathcal{N}(i)} \mathrm{Emb}_{\mathrm{nbr}}(a_j) \circ W_r\boldsymbol{\varphi}(r_{ij}) \circ c(r_{ij})\right]\right), \qquad \mathbf{t}^0_{ij} = (\mathbf{h}^0_i + \mathbf{h}^0_j) \circ W_e\boldsymbol{\epsilon}_{ij}. \tag{4}$$

Here $\circ$ denotes channel-wise multiplication. Consequently, an edge carries both geometric separation and the chemical states of its endpoint residues.

Steerable backbone fields. Scalar descriptors alone cannot distinguish orientations of a collapsed residue patch. Let $\hat{\mathbf{c}}_i$ and $\hat{\mathbf{n}}_i$ denote the normalized Cα–C and Cα–N directions, and let $\mathrm{norm}(\mathbf{u}) = \mathbf{u}/\|\mathbf{u}\|_2$. We construct a pseudo-Cβ direction

$$\hat{\mathbf{b}}_i = -\sqrt{\frac{1}{3}}\,\mathrm{norm}(\hat{\mathbf{c}}_i + \hat{\mathbf{n}}_i) - \sqrt{\frac{2}{3}}\,\mathrm{norm}(\hat{\mathbf{c}}_i \times \hat{\mathbf{n}}_i), \qquad \mathbf{V}^{(1),0}_i = [\hat{\mathbf{b}}_i, \hat{\mathbf{u}}^+_i, \hat{\mathbf{u}}^-_i]W_B, \tag{5}$$

alongside the forward and reverse chain directions $\hat{\mathbf{u}}^+_i$ and $\hat{\mathbf{u}}^-_i$. The bias-free W_B mixes multiplicity channels without mixing spatial coordinates. This initialization provides a directional proxy for the side-chain axis even when the pocket is not visible in the apo structure.

Equivariant interaction. The subsequent interaction follows the invariant-coefficient, steerable-basis construction of GotenNet (Aykent & Xia, 2025). Each layer updates $\mathbf{h}^t_i$, $\mathbf{t}^t_{ij}$, and $\mathbf{V}^{(\ell),t}_i$ with geometry-aware tensor attention, invariant endpoint-field contraction, and an equivariant feed-forward block. In compact form,

$$(\mathbf{h}^{t+1}, \mathbf{V}^{t+1}, \mathbf{t}^{t+1}) = \mathcal{F}_{\mathrm{eq}}(\mathbf{h}^t, \mathbf{V}^t, \mathbf{t}^t; \mathcal{G}). \tag{6}$$

Its scalar coefficients are functions of invariant quantities; thus $\mathbf{h}^t_i$ and $\mathbf{t}^t_{ij}$ remain invariant while $\mathbf{V}^{(\ell),t}_i$ transforms under the degree-ℓ representation of $\mathrm{SO}(3)$. The GotenNet-style layer equations and symmetry proof are given in Sections B.1 and C.

4.2 GLOBAL CONTEXT-CONDITIONED POCKET PREDICTION

While the equivariant model captures local structural geometry, cryptic pockets can also depend on long-range residue coupling and fold-level flexibility (Gazizov et al., 2026). After T layers, we reduce the steerable vectors to an invariant local descriptor:

$$\mathbf{z}_i = \left[\mathbf{h}^T_i; \bigoplus_{\ell=1}^{\ell_{\max}} \left\|\mathbf{V}^{(\ell),T}_i\right\|_2\right]. \tag{7}$$

To represent interactions beyond the cutoff graph, we query an optional external pair representation $\mathbf{P}$, e.g., from AlphaFold2 or ESMFold. Each pair vector is projected to $\mathbf{q}_{ij}$, and an attention-weighted summary forms the global context for residue i:

$$\mathbf{q}_{ij} = W_P \mathbf{P}_{ij}, \qquad \pi_{ij} = \frac{\exp\big(\mathbf{u}_P^\top \sigma(W_A \mathbf{q}_{ij})\big)}{\sum_{k=1}^{L} \exp\big(\mathbf{u}_P^\top \sigma(W_A \mathbf{q}_{ik})\big)}, \qquad \mathbf{g}_i = \sum_{j=1}^{L} \pi_{ij} \mathbf{q}_{ij}. \tag{8}$$

The pair representation may be extracted from AlphaFold2 (Jumper et al., 2021) or another pretrained model; the readout does not assume a particular source. We fuse local geometry and global context to predict

$$\hat{y}_i = \mathrm{sigmoid}(\psi_{\mathrm{cls}}([\mathbf{z}_i; \mathbf{g}_i])), \; \mathcal{L}_{\mathrm{pocket}} = -\frac{1}{|\mathcal{V}_{\mathrm{valid}}|} \sum_{i \in \mathcal{V}} w_i \left[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \right], \tag{9}$$

where w_i balances the sparse positive residues and $\mathcal{V}$ excludes ambiguous labels. Without $\mathbf{P}$, the classifier uses $\mathbf{z}_i$ alone. Since $\mathbf{z}_i$ and $\mathbf{g}_i$ are invariant, the pocket probabilities do not depend on the global coordinate frame.

4.3 Probe-Guided Open-Pocket Structure Prediction

Predicting pocket-lining residues identifies where a cryptic site may form, but ligand design also requires a structure in which the site is accessible. We use the protein–ligand complex prediction capability of Boltz-2 (Passaro et al., 2025) to obtain such candidate conformations. Its residue-level pocket constraints provide a natural interface for the sites predicted by **PocketMon**. Complex prediction also requires a ligand. However, an arbitrary molecule may be poorly matched to the proposed site. We therefore select pocket-specific molecular probes before predicting the complex.

Let $\mathcal{S}$ be a connected patch of high-scoring residues and m a probe molecule. We encode the residue patch and molecular graph as $\mathbf{z}_{\mathcal{S}}$ and $\mathbf{z}_m$, respectively. To retain more than a comparison of pooled representations, it also summarizes interactions between residue states $\mathbf{v}_i$ and atom states $\mathbf{u}_a$:

$$c(m, \mathcal{S}) = \frac{1}{|\mathcal{S}||m|} \sum_{i \in \mathcal{S}} \sum_{a \in m} \left[\frac{\langle W_r \mathbf{v}_i, W_a \mathbf{u}_a \rangle}{\sqrt{d}} + b_{\kappa(i), \tau(a)} \right], \tag{10}$$

$$E_\eta(m, \mathcal{S}) = -\psi_E([\mathbf{z}_m; \mathbf{z}_{\mathcal{S}}; \mathbf{z}_m \circ \mathbf{z}_S; c(m, \mathcal{S})]). \tag{11}$$

Here $b_{\kappa(i),\tau(a)}$ is a learned residue-class–atom-type compatibility term. Lower E_η indicates a better learned match between the probe and patch; it is not a physical binding free energy. The residue patch encoder uses residue identity and relative sequence position, so probe ranking does not require an open-pocket structure or a ligand pose.

We train this energy-based model (LeCun et al., 2006) on observed protein–ligand complexes. Each ligand m^+ is paired with its contacting residue patch $\mathcal{S}$, and a property-matched molecule m^- with a different scaffold provides a contrastive decoy. The objective favors the observed pairing:

$$\mathcal{L}_{\mathrm{probe}} = \mathbb{E}_{(\mathcal{S}, m^+, m^-)} \left[\mathrm{softplus}\left(\mathbb{E}_\eta(m^+, \mathcal{S}) - \mathbb{E}_\eta(m^-, \mathcal{S}) \right) \right]. \tag{12}$$

At inference, we rank candidate molecules by $\mathbb{E}_\eta(m, \mathcal{S})$ and retain a small, scaffold-diverse set. For each selected probe, we map the predicted pocket residues to Boltz-2 sequence positions and specify the corresponding pocket constraints $C(\mathcal{S}, m)$. With the protein sequence $\mathbf{a}$ fixed, the frozen complex predictor produces

$$(\mathcal{P}^{(m)}, \mathcal{L}^{(m)}) \sim p_\phi(\cdot \mid \mathbf{a}, m, C(\mathcal{S}, m)), \tag{13}$$

where $\mathcal{P}^{(m)}$ and $\mathcal{L}^{(m)}$ are the predicted protein and probe coordinates. The protein structure is a candidate ligand-compatible open conformation, which can be assessed against the apo input for local pocket opening and preservation of the overall fold.

5 Experimental Settings

Benchmarks and baselines. We evaluate **PocketMon** on PocketMiner (Meller et al., 2023), CryptoBank (Martinez et al., 2026), and AF2BIND (Gazizov et al., 2026), which span simulation-derived pocket opening, apo–holo crypticity, and ligand-contact recovery. We compare against

Table 1: Overall residue-level binding-site prediction performance. Best results are in **boldface**.

Dataset	Model	ROC-AUC	PR-AUC	Acc	Prec	Recall	Recovery
PocketMiner	Fpocket	71.09	65.77	74.21	**100.00**	15.45	22.05
	P2Rank	62.41	56.15	75.30	92.80	20.60	**44.81**
	CryptoSite	76.97±1.23	64.88±1.47	75.51±1.08	66.87±1.79	39.08±2.46	21.50±2.17
	PocketMiner	78.90±1.14	66.39±1.36	77.19±0.93	69.72±1.57	44.58±2.28	24.33±2.04
	PocketMon	**88.31±1.07**	**81.22±1.29**	**83.42±0.84**	75.65±1.38	**67.32±2.13**	39.22±1.92
CryptoBank	Fpocket	56.97	6.41	92.21	9.92	11.05	10.06
	P2Rank	59.08	6.71	**92.83**	10.92	10.43	10.07
	CryptoSite	64.73±0.92	7.02±0.73	80.46±0.82	7.50±1.24	33.00±2.76	10.80±1.87
	PocketMiner	65.93±0.86	8.76±0.77	80.64±0.74	9.04±1.27	37.94±3.08	12.41±2.14
	PLM	70.00±0.53	11.00±0.94	92.70±0.43	**16.00±1.83**	18.16±2.24	15.07±1.68
	PocketMon	**74.39±0.71**	**16.78±1.12**	80.32±0.67	11.78±1.58	**54.07±3.47**	**17.73±2.36**
AF2BIND	Fpocket	76.34±0.24	41.72±0.44	92.51±0.11	59.98±0.71	36.95±1.02	45.44±0.43
	P2Rank	88.15±0.20	60.56±0.56	93.74±0.12	**71.72±0.46**	44.31±0.55	59.82±0.45
	CryptoSite	73.00±0.69	18.90±0.45	91.34±0.20	28.00±4.00	1.20±0.40	26.50±1.20
	PocketMiner	75.00±1.46	20.24±1.38	91.49±0.11	38.35±8.30	0.64±0.51	28.43±1.09
	AF2BIND	93.05±0.11	67.77±0.38	94.00±0.26	69.15±3.45	56.00±4.24	62.55±0.31
	PocketMon	**93.69±0.17**	**68.63±0.21**	**94.10±0.19**	66.90±1.70	**60.30±1.92**	**63.88±3.20**

Fpocket (Le Guilloux et al., 2009), P2Rank (Krivák & Hoksza, 2018), CryptoSite (Cimermancic et al., 2016), PocketMiner, the CryptoBank PLM, and AF2BIND, with structure-only **PocketMon** variants as ablation controls.

Training and evaluation. **PocketMon** variants are optimized with Adam (Kingma & Ba, 2014) using benchmark-specific input features, loss functions, and weighting schemes. We report ROC-AUC, PR-AUC, Accuracy, Precision, Recall, and Recovery@k using each benchmark's aggregation protocol. Accuracy, Precision, and Recall are computed at a fixed threshold of 0.5, and each checkpoint is selected on its validation metric before any test evaluation. Dataset construction, baseline configurations, model hyperparameters, metric definitions, model selection, and reporting protocols are detailed in section D.

Probe ranking and constrained generation. ProbeEBM is trained on 3,671 filtered PDBbind complexes. Each positive patch and ligand pair is contrasted with one property-matched decoy that uses a different scaffold and has ECFP4 similarity below 0.4, and candidates in a fixed pool are ranked by ascending learned energy. For structure prediction, we keep Boltz-2 frozen and compare ligand-only generation with generation conditioned on soft constraints from **PocketMon**. Within each pair, the protein sequence, MSA, ligand identity, and random seed stay fixed, with a maximum constraint distance of 4.5 Å. Full training and evaluation settings are provided in section D.

6 Empirical Studies

6.1 Performance Comparison and Cross-Dataset Transfer

We evaluate **PocketMon** from two complementary perspectives: direct comparison against representative baselines on each benchmark and transfer across datasets with different supervision semantics. Table 1 summarizes the benchmark comparison, while Table 2 evaluates models trained on different sources under a unified PocketMiner test protocol. All table metrics are reported in percent.

Performance comparison. Results in Table 1 suggest the following trends across diverse supervision settings: ***(i)*** ***PocketMon*** *consistently improves ranking performance.* Across all three benchmarks, **PocketMon** achieves the best primary ranking metric and shows the largest gains on recall-oriented measures, a critical criterion for cryptic pocket screening because missed binding sites are difficult to recover in later stages. On PocketMiner, it improves ROC-AUC by nearly 10 points over the strongest prior method and raises recall from 45% to 67% relative to the original PocketMiner baseline. ***(ii)*** *Structural input is especially valuable under sparse cryptic labels.* On CryptoBank, where only 5% of residues are cryptic, recall rises from 18% to 54%, highlighting the value of explicit structural input beyond sequence features alone. Fpocket and P2Rank can achieve high Accuracy through

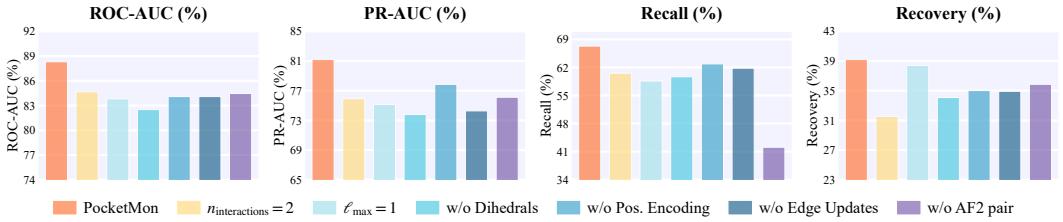

Figure 3: Ablation study on PocketMiner. Each panel reports one evaluation metric.

Table 2: Cross-dataset transfer to the PocketMiner test set across different training sources and label definitions. Best and second-best results are in **boldface** and underlined.

Training Source	ROC-AUC	PR-AUC	Acc	Prec	Recall	Recovery
PocketMiner, 38 proteins with MD labels	88.31±1.07	81.22±1.29	83.42±0.84	75.65±1.38	67.32±2.13	39.22±1.92
CryptoBank, 6,006 proteins with holo/apo labels	89.38±1.53	83.32±1.74	**86.73±1.22**	83.97±2.43	**69.80±4.17**	44.49±3.08
AF2BIND, 672 proteins with ligand-contact labels	**91.73±1.82**	**84.07±2.03**	83.21±1.31	**92.03±2.14**	49.20±4.76	**50.14±3.57**

conservative prediction, but their low PR-AUC and Recall on CryptoBank reveal limited sensitivity to sparse cryptic residues. ***(iii)*** *Local geometry and global pair priors are complementary.* On AF2BIND, **`PocketMon`** slightly improves over the specialized logistic regression model in both ROC-AUC and Recovery, suggesting that equivariant geometric processing provides a complementary signal alongside pretrained pair features. Overall, the results support the design of **`PocketMon`**, where equivariant message passing captures local cavity geometry and AlphaFold2 pair features supply broader residue-interaction priors invisible from local coordinates alone.

Cross-dataset transfer. Table 2 shows that **`PocketMon`** learns structural signals not tied to a single labeling strategy. The CryptoBank-trained model obtains higher recall and PR-AUC than the in-domain PocketMiner model, suggesting that its larger cryptic annotations cover broader latent pocket patterns than the 38-protein MD training set. The AF2BIND-trained model achieves the highest ROC-AUC and Precision but lower recall, consistent with the stricter ligand-contact label definition. This transfer pattern reflects the supervision semantics of each dataset: AF2BIND labels direct ligand contact and therefore yields precise but narrow predictions, while CryptoBank captures broader conformational exposure and better matches the PocketMiner goal of finding latent pocket-lining residues. These results indicate that, with a sufficiently expressive representation, larger and more diverse structural supervision can compensate for differences in label definition. The PocketMiner test set shares no protein with AF2BIND and only 3P53A with the CryptoBank. Removing 3P53A changes every transfer metric by less than two points, ruling out training-set overlap (section D.5).

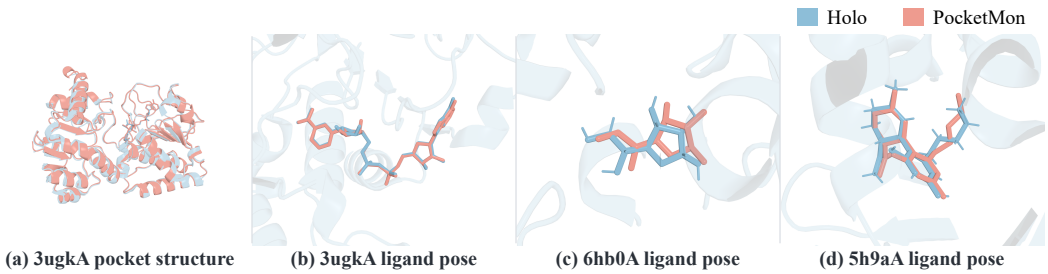

Figure 4: PocketMon-conditioned Boltz-2 case studies. (a) Full-structure overlay for 3ugkA. (b–d) Pocket-aligned ligand poses for 3ugkA, 6hb0A, and 5h9aA. Blue and coral denote holo structures and **`PocketMon`**-conditioned predictions, respectively.

6.2 Ablation Analysis

We conduct single-component ablations on PocketMiner, following a one-change-at-a-time protocol. Figure 3 summarizes the results across four metrics most informative for cryptic pocket screening:

ROC-AUC, PR-AUC, Recall, and Recovery. Every ablated variant degrades at least one primary metric, indicating that **PocketMon**'s gains do not come from a single isolated module. The largest sensitivity loss occurs after removing AlphaFold2 pair fusion, while geometry-related components and message-passing refinements mainly affect ranking quality and site recovery. Overall, the ablation study supports the intended design: global pair priors improve hidden-residue sensitivity, expressive local geometry improves discrimination, and iterative message passing helps recover extended pocket regions. Component definitions and additional interpretation are provided in section D.

6.3 Probe Ranking and Pocket-Constrained Complex Prediction

Patch-conditioned molecular retrieval.

We evaluate ProbeEBM on 323 test patches against a fixed pool of 308 molecules. Each candidate is scored by learned energy and ranked in ascending order, and retrieval is measured by the rank of the co-crystallized ligand. As shown in Table 3, ProbeEBM reaches a mean rank of 48.1 and an MRR of 0.1077, compared with random expectations of 154.5 and 0.0205. It places the co-crystallized ligand in the top 10 and top 20 for 28.2% and 48.3% of queries, compared with 3.25% and 6.49% under random ordering. The pocket-blind control stays near chance, with a mean rank of 153.9 and Hit@10 of 3.10%, showing that the enrichment depends on compatibility between each query patch and ligand rather than a general preference for certain ligands.

Table 3: ProbeEBM retrieval on the test split.

Condition	Mean rank ↓	MRR ↑	Hit@10 ↑	Hit@20 ↑
Random ordering	154.5	0.0205	3.25%	6.49%
Pocket-blind control	153.9	0.0202	3.10%	6.50%
ProbeEBM	**48.1**	**0.1077**	**28.2%**	**48.3%**

PocketMon-conditioned Boltz-2 case studies. We evaluate three representative PocketMiner proteins using ten paired Boltz-2 seeds per target. Conditioning on **PocketMon** predictions improves ligand placement across all three targets. The strongest gain occurs for 3ugkA, where mean ligand RMSD after pocket alignment decreases from 0.764 to 0.349 Å and the generated pocket conformation also moves closer to the holo structure. All generated structures pass the geometry checks. Representative overlays are shown in fig. 4, with complete target-specific results provided in section D.

6.4 Human Proteome-Wide Pocket Discovery

To assess whether residue-level signals generalize to proteome scale, we apply **PocketMon**, AF2BIND, and P2Rank to all 20,550 AlphaFold-predicted single-chain human protein structures.

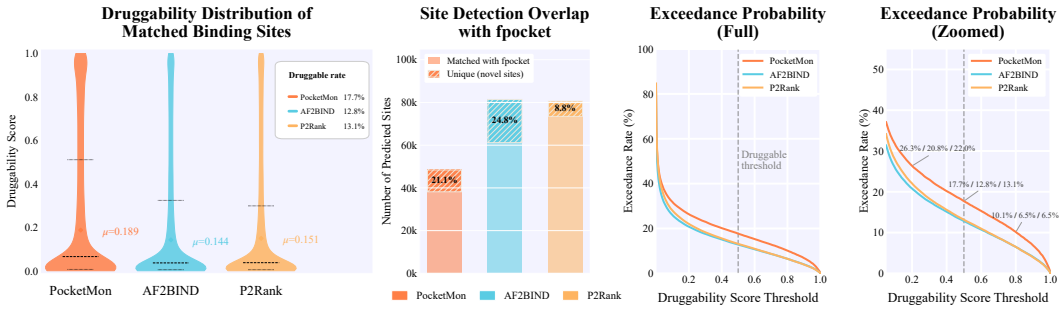

Figure 5: Proteome-wide druggability assessment of predicted binding sites. **PocketMon** achieves consistently higher Fpocket druggability scores than the baselines.

Druggability assessment. Per-residue binding probabilities are clustered into multi-residue sites via DBSCAN; P2Rank directly outputs pocket predictions. **PocketMon** identifies 48,396 candidate sites across 16,601 proteins, substantially fewer than AF2BIND with 81,401 or P2Rank with 80,596, reflecting a stricter, geometry-aware selection. Using Fpocket (Le Guilloux et al., 2009) as an independent pocket-quality oracle, we match predicted sites to Fpocket pockets and inherit their druggability scores. As shown in Figure 5, **PocketMon** achieves the highest mean druggability among matched sites and dominates both baselines across all thresholds. Among **PocketMon**'s matched sites, 17.7% exceed the druggable threshold, compared to 13.1% for P2Rank and 12.8% for AF2BIND, consistent with fewer but higher-quality pocket predictions. Full scan protocol details are provided in section D.

Representation-level analysis. UMAP projections of 500,000 residue embeddings in Figure 6 reveal a clear divergence in how the two models encode pocket-specific information. AF2BIND disperses high-probability residues across diffuse clusters, whereas **PocketMon** organizes them into *sharply localized* manifolds with smooth probability gradients. This suggests that **PocketMon** disentangles pocket-lining signals from proteomic background in latent space, enabling reliable prioritization at proteome scale. Together with the druggability results, this underscores a precision-oriented screening logic: **PocketMon** is more selective, yet its retained candidates are substantially enriched for geometrically plausible and actionable pockets.

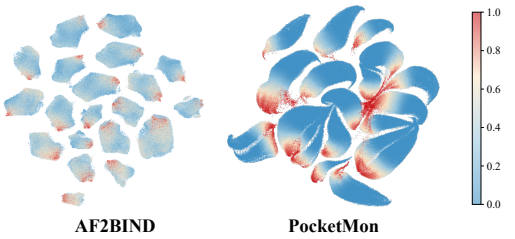

Figure 6: UMAP projections of human proteome residue embeddings.

6.5 Case Studies

We further examine two representative dynamic systems, Pasteurella dagmatis sialyltransferase (PdST) (Schmölzer et al., 2015) and the AfuABC transporter (Sit et al., 2015), to assess whether **PocketMon** can recover such latent binding sites from apo conformations.

In both systems, baseline approaches exhibit complementary failure modes: PocketMiner tends to overpredict shallow surface depressions, leading to elevated false-positive rates, whereas P2Rank frequently misses functionally critical residues, resulting in severe false negatives. Structural comparison between apo and holo states further reveals that many of these overlooked residues participate in substantial domain rearrangements and flexible loop remodeling during pocket formation, highlighting the limitations of static local geometric descriptors for detecting conformationally masked sites. Through the integration of equivariant local residue representations with global pairwise interaction modeling, **PocketMon** captures long-range residue coupling and latent structural plasticity associated with pocket opening. As a result, we successfully identify adaptive binding regions without requiring computationally expensive conformational ensemble sampling, demonstrating strong potential for structure-based targeting of dynamic proteins.

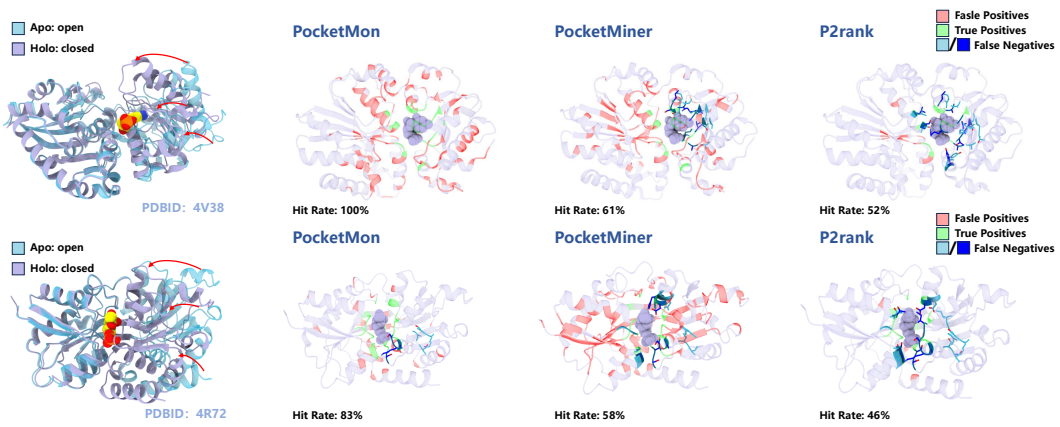

Figure 7: Visualization of representative cryptic pockets on two proteins: *4V38* and *4R72*.

7 Conclusion

In this work, we formulate cryptic pocket discovery as a local-global geometric inference problem and propose **PocketMon**, a pair-conditioned equivariant residue field that integrates local backbone geometry with residue-pair priors for hidden-site prediction. Across three diverse benchmarks, it consistently outperforms bioinformatics, sequence, and geometry-based baselines. We hope this work facilitates future research on cryptic pocket prediction and structure-based drug discovery.

AI USE STATEMENT

We used OpenAI Codex to assist with grammar checking and LaTeX formatting. It was not used to generate research data, design or run experiments, implement the reported method, formulate mathematical claims, assist with proofs, or compute the reported results. All AI-assisted edits were checked against the source manuscript and reported tables. The authors take full responsibility for the final text, claims, and artifacts.

ETHICS STATEMENT

This work studies cryptic pocket discovery using existing public protein-structure and benchmark datasets. We do not collect new human-subject data, and the analyses do not infer personal attributes or behavior. Predicted pockets and generated complexes are computational hypotheses; predictive accuracy alone does not establish biological activity, therapeutic utility, or safety. Any downstream experimental or therapeutic use requires independent validation of pocket opening, ligand binding, and failure modes, together with compliance with applicable biosafety and biomedical research guidelines.

REPRODUCIBILITY STATEMENT

Section 4 describes the proposed framework, including the backbone-informed equivariant residue field, global context-conditioned pocket prediction, and probe-guided open-pocket structure prediction. Appendix B details the equivariant blocks, radial basis, whole-protein processing, and pair-edge conditioning, and Appendix C provides the symmetry proof. Appendix D specifies dataset construction, baseline reproduction, metric definitions, model selection, implementation settings, ablation variants, the ProbeEBM and pocket-constrained Boltz-2 protocols, and the human proteome scan. Threshold sensitivity and calibration are reported in Appendix D.8.

REFERENCES

Rommie E Amaro. Will the real cryptic pocket please stand out? *Biophysical Journal*, 116(5): 753–754, 2019.

Sarp Aykent and Tian Xia. Gotennet: Rethinking efficient 3d equivariant graph neural networks. In *The Thirteenth International Conference on Learning Representations*, 2025.

Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.

Dmitri Beglov, David R Hall, Amanda E Wakefield, Lingqi Luo, Karen N Allen, Dima Kozakov, Adrian Whitty, and Sandor Vajda. Exploring the structural origins of cryptic sites on proteins. *Proceedings of the National Academy of Sciences*, 115(15):E3416–E3425, 2018.

Alberto Borsatto, Eleonora Gianquinto, Valerio Rizzi, and Francesco Luigi Gervasio. Swish-x, an expanded approach to detect cryptic pockets in proteins and at protein–protein interfaces. *Journal of Chemical Theory and Computation*, 20(8):3335–3348, 2024.

Gregory R Bowman and Phillip L Geissler. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proceedings of the National Academy of Sciences*, 109(29):11681–11686, 2012.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, et al. Jax: composable transformations of python+ numpy programs, 2018.

Peter Cimermancic, Patrick Weinkam, T Justin Rettenmaier, Leon Bichmann, Daniel A Keedy, Rahel A Woldeyes, Dina Schneidman-Duhovny, Omar N Demerdash, Julie C Mitchell, James A Wells, et al. Cryptosite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. *Journal of molecular biology*, 428(4):709–719, 2016.

Federico Comitani and Francesco Luigi Gervasio. Exploring cryptic pockets formation in targets of pharmaceutical interest with swish. *Journal of chemical theory and computation*, 14(6):3321–3331, 2018.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

Fabian B Fuchs, Daniel E Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 1970–1981, 2020.

Artem Gazizov, Anna Lian, Casper Goverde, Jody Mou, Sergey Ovchinnikov, and Nicholas F Polizzi. Af2bind: predicting small-molecule binding sites using the pair representation of alphafold2. *Nature Methods*, pp. 1–10, 2026.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Manfred Hendlich, Friedrich Rippmann, and Gerhard Barnickel. Ligsite: automatic and efficient detection of potential small molecule-binding sites in proteins. *Journal of Molecular Graphics and Modelling*, 15(6):359–363, 1997.

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael JL Townshend, and Ron O Dror. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations*, 2021.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Radoslav Krivák and David Hoksza. P2rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of cheminformatics*, 10(1):39, 2018.

Antonija Kuzmanic, Gregory R Bowman, Jordi Juarez-Jimenez, Julien Michel, and Francesco L Gervasio. Investigating cryptic binding sites by molecular dynamics simulations. *Accounts of chemical research*, 53(3):654–661, 2020.

Vincent Le Guilloux, Peter Schmidtke, and Pierre Tuffery. Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics*, 10(1):168, 2009.

Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, Fujie Huang, et al. A tutorial on energy-based learning. *Predicting structured data*, 1(0), 2006.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

Pedro Febrer Martinez, Thorben Fröhlking, Alberto Borsatto, and Francesco L Gervasio. Cryptobank: A resource for the identification and prediction of cryptic sites in proteins. *Science Advances*, 12 (17):eady6364, 2026.

Robert T McGibbon, Kyle A Beauchamp, Matthew P Harrigan, Christoph Klein, Jason M Swails, Carlos X Hernández, Christian R Schwantes, Lee-Ping Wang, Thomas J Lane, and Vijay S Pande. Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical journal*, 109(8):1528–1532, 2015.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Artur Meller, Michael Ward, Jonathan Borowsky, Meghana Kshirsagar, Jeffrey M Lotthammer, Felipe Oviedo, Juan Lavista Ferres, and Gregory R Bowman. Predicting locations of cryptic pockets from single protein structures using the pocketminer graph neural network. *Nature communications*, 14 (1):1177, 2023.

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold: making protein folding accessible to all. *Nature methods*, 19(6):679–682, 2022.

Vladimiras Oleinikovas, Giorgio Saladino, Benjamin P Cossins, and Francesco L Gervasio. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *Journal of the American Chemical Society*, 138(43):14257–14263, 2016.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, et al. Boltz-2: Towards accurate and efficient binding affinity prediction. *BioRxiv*, 2025.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, pp. 9323–9332. PMLR, 2021.

Katharina Schmölzer, Tibor Czabany, Christiane Luley-Goedl, Tea Pavkov-Keller, Doris Ribitsch, Helmut Schwab, Karl Gruber, Hansjörg Weber, and Bernd Nidetzky. Complete switch from α-2, 3-to α-2, 6-regioselectivity in pasteurella dagmatis β-d-galactoside sialyltransferase by active-site redesign. *Chemical communications*, 51(15):3083–3086, 2015.

Brandon Sit, Shauna M Crowley, Kirandeep Bhullar, Christine Chieh-Lin Lai, Calvin Tang, Yogesh Hooda, Charles Calmettes, Husain Khambati, Caixia Ma, John H Brumell, et al. Active transport of phosphorylated carbohydrates promotes intestinal colonization and transmission of a bacterial pathogen. *PLoS pathogens*, 11(8):e1005107, 2015.

Vít Škrhák, Marian Novotnỳ, Christos P Feidakis, Radoslav Krivák, and David Hoksza. Cryptobench: cryptic protein–ligand binding sites dataset and benchmark. *Bioinformatics*, 41(1):btae745, 2025.

Zhuyezi Sun, Amanda Elizabeth Wakefield, Istvan Kolossvary, Dmitri Beglov, and Sandor Vajda. Structure-based analysis of cryptic-site opening. *Structure*, 28(2):223–235, 2020.

Sandor Vajda, Dmitri Beglov, Amanda E Wakefield, Matthew Egbert, and Adrian Whitty. Cryptic binding sites on proteins: definition, detection, and druggability. *Current Opinion in Chemical Biology*, 44:1–8, 2018.

Maxwell I Zimmerman, Jason R Porter, Michael D Ward, Shachi Singh, Neha Vithani, Artur Meller, Usha L Mallimadugula, Cassandra E Kuhn, Jonathan H Borowsky, Rafal P Wiewiora, et al. Sars-cov-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nature Chemistry*, 13(7):651–659, 2021.

A Predicted Pocket Visualization

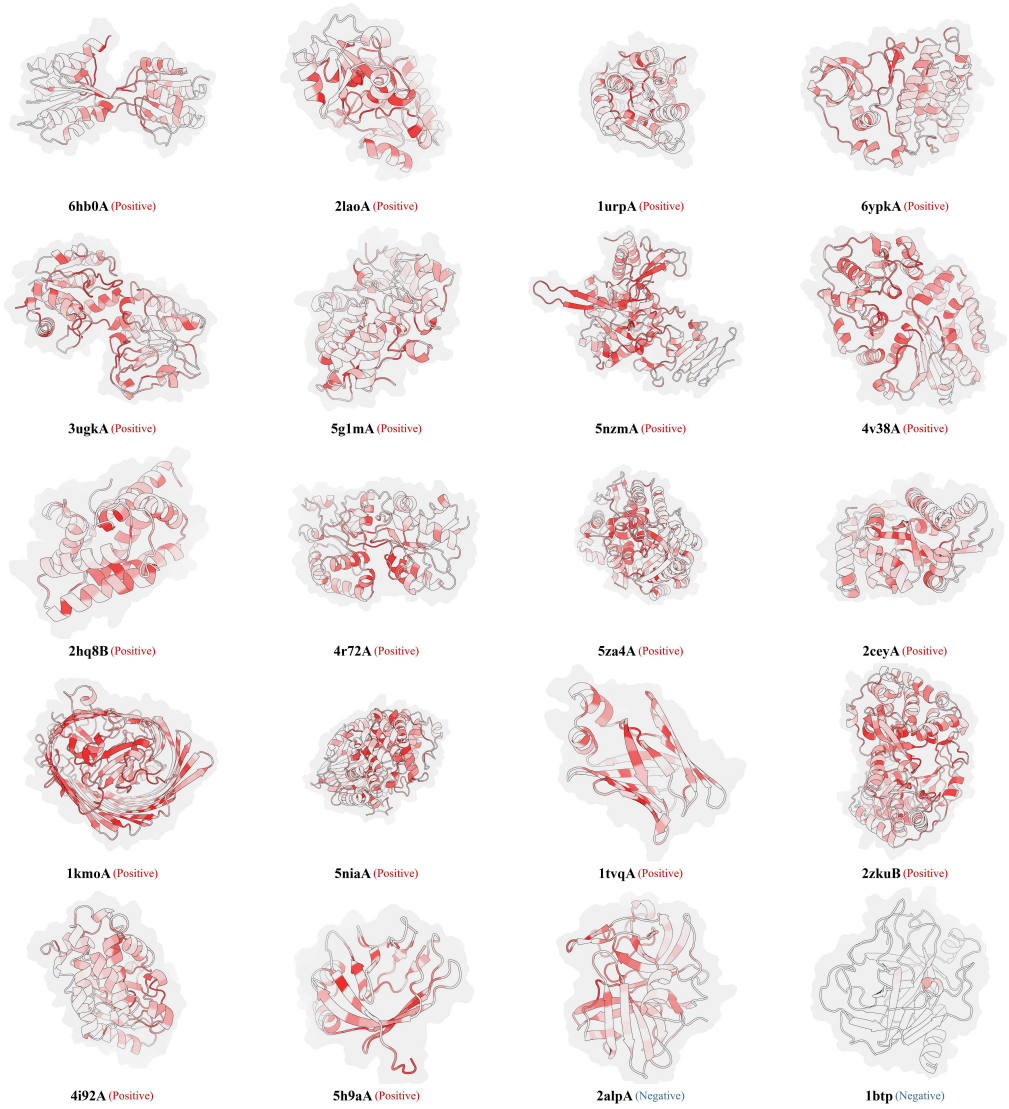

Figure 8: 3D visualization of predicted cryptic pockets of 20 targets sampled from PocketMiner.

B Additional Architectural Details

B.1 GotenNet Equivariant Blocks

PocketMon adopts three functional blocks from the GotenNet-style geometric design (Aykent & Xia, 2025). We briefly summarize them here to make the notation self-contained. In our residue-level setting, each layer maintains invariant residue features $\mathbf{h}_i^t$, invariant edge features $\mathbf{t}_{ij}^t$, and steerable residue fields $\mathbf{V}_i^{(\ell),t}$ of degree ℓ.

Geometry-aware tensor attention (GATA). GATA performs message passing over the residue graph. It computes attention weights from invariant node and edge features, and then produces invariant coefficients that control three types of updates: scalar residue updates, injection of steerable edge-direction bases, and transport of neighboring steerable fields. In our notation, the resulting coefficients are $\boldsymbol{\omega}_{ij}^{0,t}$, $\boldsymbol{\omega}_{ij}^{r,\ell,t}$, and $\boldsymbol{\omega}_{ij}^{v,\ell,t}$, respectively. The crucial property is that these coefficients

are invariant scalars; therefore, multiplying them with steerable quantities such as $\mathbf{R}_{ij}^{(\ell)} = Y_\ell(\hat{\mathbf{r}}_{ij})$ or $\mathbf{V}_j^{(\ell),t}$ preserves equivariance.

For completeness, the residue-level attention weights and routed coefficients are

$$\alpha_{ij}^{m,t} = \mathrm{softmax}_{j\in\mathcal{N}(i)}\left[\frac{\left\langle \mathbf{h}_i^t W_q^m, (\mathbf{h}_j^t W_k^m)\circ\sigma(\mathbf{t}_{ij}^t W_{re}^m)\right\rangle}{\sqrt{d_m}}\right],$$

$$\begin{aligned}\boldsymbol{\Omega}_{ij}^t &= \mathrm{split}\left(\bigoplus_{m=1}^{M}\alpha_{ij}^{m,t}\gamma_v^m(\mathbf{h}_j^t) + (\mathbf{t}_{ij}^t W_{rs})\circ\gamma_s(\mathbf{h}_j^t)\circ W_\varphi\boldsymbol{\varphi}(r_{ij})\right)\\ &= \left(\boldsymbol{\omega}_{ij}^{0,t}, \{\boldsymbol{\omega}_{ij}^{r,\ell,t}\}_{\ell=1}^{\ell_{\max}}, \{\boldsymbol{\omega}_{ij}^{v,\ell,t}\}_{\ell=1}^{\ell_{\max}}\right).\end{aligned} \tag{14}$$

The three coefficient groups update scalar states, edge-direction bases, and transported fields, respectively:

$$\begin{aligned}\mathbf{h}_i^{t+\frac{1}{2}} &= \mathbf{h}_i^t + \sum_{j\in\mathcal{N}(i)}\boldsymbol{\omega}_{ij}^{0,t},\\ \mathbf{V}_i^{(\ell),t+\frac{1}{2}} &= \mathbf{V}_i^{(\ell),t} + \sum_{j\in\mathcal{N}(i)}\left[\mathbf{R}_{ij}^{(\ell)}(\boldsymbol{\omega}_{ij}^{r,\ell,t})^\top + \mathbf{V}_j^{(\ell),t}\,\mathrm{Diag}(\boldsymbol{\omega}_{ij}^{v,\ell,t})\right].\end{aligned} \tag{15}$$

Hierarchical tensor refinement (HTR). HTR refines edge states using the current steerable fields at the two endpoint residues. In our residue interaction, the edge refinement step constructs invariant geometric evidence

$$\mathbf{s}_{ij}^t = \bigoplus_{\ell=1}^{\ell_{\max}}\sum_{m=1}^{2\ell+1}\left(\mathbf{V}_i^{(\ell),t+\frac{1}{2}}W_Q^\ell\right)_{m,:}\circ\left(\mathbf{V}_j^{(\ell),t+\frac{1}{2}}W_K^\ell\right)_{m,:}. \tag{16}$$

The summation over representation coordinates contracts two degree-ℓ steerable fields into an invariant edge descriptor. This descriptor updates the edge state through $\mathbf{t}_{ij}^{t+1} = \mathbf{t}_{ij}^t + \gamma_w(\mathbf{s}_{ij}^t)\circ\gamma_t(\mathbf{t}_{ij}^t)$. Thus, HTR allows edge features to evolve with the learned residue geometry while keeping edge channels invariant.

Equivariant feed-forward network (EQFF). EQFF mixes scalar and steerable channels within each residue after message passing. It first extracts invariant summaries from the steerable fields by taking norms over the representation dimension:

$$[\boldsymbol{\rho}_i^t; \{\boldsymbol{\lambda}_{i,\ell}^t\}_{\ell=1}^{\ell_{\max}}] = \gamma_m\left(\mathbf{h}_i^{t+\frac{1}{2}}, \bigoplus_{\ell=1}^{\ell_{\max}}\left\|\mathbf{V}_i^{(\ell),t+\frac{1}{2}}W_U^\ell\right\|_2\right). \tag{17}$$

The scalar gate $\boldsymbol{\rho}_i^t$ updates the invariant stream, while $\boldsymbol{\lambda}_{i,\ell}^t$ channel-wise modulates the projected steerable field:

$$\mathbf{h}_i^{t+1} = \mathbf{h}_i^{t+\frac{1}{2}} + \boldsymbol{\rho}_i^t, \qquad \mathbf{V}_i^{(\ell),t+1} = \mathbf{V}_i^{(\ell),t+\frac{1}{2}} + \left(\mathbf{V}_i^{(\ell),t+\frac{1}{2}}W_U^\ell\right)\mathrm{Diag}(\boldsymbol{\lambda}_{i,\ell}^t). \tag{18}$$

Because the gates are invariant and the linear maps act only on multiplicity channels, EQFF preserves the transformation law of every steerable field.

B.2 Radial Basis and Cutoff

For edge distance r, the radial basis in eq. (2) is

$$\varphi_k(r) = c(r)\exp\left[-\beta_k(e^{-\alpha r} - \mu_k)^2\right], \qquad k = 1,\dots,K, \tag{19}$$

where $\alpha = 5/r_{\mathrm{cut}}$, the centers μ_k are fixed and uniformly spaced in $[e^{-r_{\mathrm{cut}}}, 1]$, every scale is the constant $\beta_k = \left(2/K\,(1-e^{-r_{\mathrm{cut}}})\right)^{-2}$, and

$$c(r) = \frac{1}{2}\left[\cos\left(\frac{\pi r}{r_{\mathrm{cut}}}\right) + 1\right]\mathbf{1}\{r < r_{\mathrm{cut}}\}. \tag{20}$$

Neither the centers nor the scales are trained. On $[0, r_{\mathrm{cut}}]$, the coordinate $e^{-\alpha r}$ ranges over $[e^{-5}, 1]$; the implementation places μ_k on $[e^{-r_{\mathrm{cut}}}, 1]$ rather than on that image.

B.3 Dense-to-Sparse Whole-Protein Processing

For a padded batch with mask $\mathbf{M} \in \{0,1\}^{B\times L}$, valid residues are collected as

$$\{(b,i) : M_{bi} = 1\} \longrightarrow n = 1, \ldots, N_{\text{tot}}. \tag{21}$$

All graph operations are performed on the sparse node set, while predictions are scattered back to $[B, L]$ using the same mask. This preserves the whole-protein input-output interface while avoiding padded-node computation.

B.4 Optional Pair Edge Conditioning

When full pair tensors are available, pair priors can also condition cutoff edges:

$$\tilde{\mathbf{p}}_{ij} = W_{pe}\mathbf{P}_{ij}, \qquad \boldsymbol{\lambda}_{ij} = \text{sigmoid}\left(\gamma_\lambda([\mathbf{t}_{ij}; \boldsymbol{\varphi}(r_{ij})])\right), \qquad \mathbf{t}_{ij} \leftarrow \mathbf{t}_{ij} + \boldsymbol{\lambda}_{ij} \circ \tilde{\mathbf{p}}_{ij}. \tag{22}$$

This operation modifies only invariant edge channels and therefore does not affect equivariance.

C Proof of Symmetry

To avoid notational conflict with the invariant edge state $\mathbf{t}_{ij}$, we write a rigid motion as $g = (R, \boldsymbol{\tau}) \in \text{SE}(3)$, where $R \in \text{SO}(3)$ and $g \cdot \mathbf{x} = R\mathbf{x} + \boldsymbol{\tau}$.

Proof. We prove the statement by induction over the interaction layers. Throughout the proof, scalar quantities transform under the trivial representation, while degree-ℓ steerable fields transform by the Wigner matrix $D^{(\ell)}(R)$.

Input invariants and steerable bases. For any coordinate difference $\mathbf{d}_{ij} = \mathbf{x}_j - \mathbf{x}_i$,

$$\mathbf{d}_{ij}(g \cdot \mathbf{X}) = R\mathbf{d}_{ij}(\mathbf{X}), \qquad \|\mathbf{d}_{ij}(g \cdot \mathbf{X})\|_2 = \|\mathbf{d}_{ij}(\mathbf{X})\|_2. \tag{23}$$

Therefore r_{ij}, $\boldsymbol{\varphi}(r_{ij})$, $\text{PE}(j-i)$, and $\boldsymbol{\epsilon}_{ij}$ are invariant. The unit edge direction transforms as $\hat{\mathbf{r}}_{ij}(g \cdot \mathbf{X}) = R\hat{\mathbf{r}}_{ij}(\mathbf{X})$, and the degree-$\ell$ spherical harmonic basis satisfies

$$\mathbf{R}_{ij}^{(\ell)}(g \cdot \mathbf{X}) = Y_\ell(R\hat{\mathbf{r}}_{ij}) = D^{(\ell)}(R)Y_\ell(\hat{\mathbf{r}}_{ij}) = D^{(\ell)}(R)\mathbf{R}_{ij}^{(\ell)}(\mathbf{X}). \tag{24}$$

Backbone-informed initialization. The scalar initialization uses residue identity, backbone torsions, distances, and sequence offsets. Residue identity and sequence offsets are independent of coordinates, and torsion encodings are invariant under $\text{SE}(3)$ because they can be written using inner products and oriented triple products of normalized backbone differences. Thus,

$$\mathbf{h}_i^0(g \cdot \mathbf{X}) = \mathbf{h}_i^0(\mathbf{X}), \qquad \mathbf{t}_{ij}^0(g \cdot \mathbf{X}) = \mathbf{t}_{ij}^0(\mathbf{X}). \tag{25}$$

For the steerable initialization, define $\hat{\mathbf{n}}_i = \text{norm}(\mathbf{x}_i^N - \mathbf{x}_i^\alpha)$ and $\hat{\mathbf{c}}_i = \text{norm}(\mathbf{x}_i^C - \mathbf{x}_i^\alpha)$. Both transform as vectors:

$$\hat{\mathbf{n}}_i(g \cdot \mathbf{X}) = R\hat{\mathbf{n}}_i(\mathbf{X}), \qquad \hat{\mathbf{c}}_i(g \cdot \mathbf{X}) = R\hat{\mathbf{c}}_i(\mathbf{X}). \tag{26}$$

Since $R \in \text{SO}(3)$, we have $R(\mathbf{u} \times \mathbf{v}) = (R\mathbf{u}) \times (R\mathbf{v})$. Hence the pseudo-C$\beta$ direction and backbone chain directions satisfy

$$\hat{\mathbf{b}}_i(g \cdot \mathbf{X}) = R\hat{\mathbf{b}}_i(\mathbf{X}), \qquad \hat{\mathbf{u}}_i^+(g \cdot \mathbf{X}) = R\hat{\mathbf{u}}_i^+(\mathbf{X}), \qquad \hat{\mathbf{u}}_i^-(g \cdot \mathbf{X}) = R\hat{\mathbf{u}}_i^-(\mathbf{X}). \tag{27}$$

Because W_B acts only on multiplicity channels,

$$\mathbf{V}_i^{(1),0}(g \cdot \mathbf{X}) = [R\hat{\mathbf{b}}_i, R\hat{\mathbf{u}}_i^+, R\hat{\mathbf{u}}_i^-]W_B = R[\hat{\mathbf{b}}_i, \hat{\mathbf{u}}_i^+, \hat{\mathbf{u}}_i^-]W_B = D^{(1)}(R)\mathbf{V}_i^{(1),0}(\mathbf{X}). \tag{28}$$

For $\ell \geq 2$, $\mathbf{V}_i^{(\ell),0} = \mathbf{0}$ is trivially equivariant. Thus, the theorem holds at initialization.

Geometry-aware tensor attention. Assume the theorem holds at layer t. Since $\mathbf{h}_i^t$, $\mathbf{h}_j^t$, and $\mathbf{t}_{ij}^t$ are invariant by the induction hypothesis, all query-key logits and edge-conditioned attention weights are invariant:

$$\alpha_{ij}^{m,t}(g \cdot \mathbf{X}) = \alpha_{ij}^{m,t}(\mathbf{X}). \tag{29}$$

The coefficient tensors generated by the split operation are also invariant:

$$\boldsymbol{\omega}_{ij}^{0,t}(g \cdot \mathbf{X}) = \boldsymbol{\omega}_{ij}^{0,t}(\mathbf{X}), \qquad \boldsymbol{\omega}_{ij}^{r,\ell,t}(g \cdot \mathbf{X}) = \boldsymbol{\omega}_{ij}^{r,\ell,t}(\mathbf{X}), \qquad \boldsymbol{\omega}_{ij}^{v,\ell,t}(g \cdot \mathbf{X}) = \boldsymbol{\omega}_{ij}^{v,\ell,t}(\mathbf{X}). \tag{30}$$

The scalar update is therefore invariant. For the steerable update, each summand transforms equivariantly:

$$\mathbf{R}_{ij}^{(\ell)}(g \cdot \mathbf{X})(\boldsymbol{\omega}_{ij}^{r,\ell,t})^\top = D^{(\ell)}(R)\mathbf{R}_{ij}^{(\ell)}(\mathbf{X})(\boldsymbol{\omega}_{ij}^{r,\ell,t})^\top, \tag{31}$$

$$\mathbf{V}_j^{(\ell),t}(g \cdot \mathbf{X})\,\mathrm{Diag}(\boldsymbol{\omega}_{ij}^{v,\ell,t}) = D^{(\ell)}(R)\mathbf{V}_j^{(\ell),t}(\mathbf{X})\,\mathrm{Diag}(\boldsymbol{\omega}_{ij}^{v,\ell,t}). \tag{32}$$

Since summation over neighbors commutes with the linear action of $D^{(\ell)}(R)$,

$$\mathbf{h}_i^{t+\frac{1}{2}}(g \cdot \mathbf{X}) = \mathbf{h}_i^{t+\frac{1}{2}}(\mathbf{X}), \qquad \mathbf{V}_i^{(\ell),t+\frac{1}{2}}(g \cdot \mathbf{X}) = D^{(\ell)}(R)\mathbf{V}_i^{(\ell),t+\frac{1}{2}}(\mathbf{X}). \tag{33}$$

Hierarchical tensor refinement. The edge refinement step contracts steerable fields at the two endpoint residues. Let $\mathbf{A}_i^{(\ell)} = \mathbf{V}_i^{(\ell),t+\frac{1}{2}}W_Q^\ell$ and $\mathbf{B}_j^{(\ell)} = \mathbf{V}_j^{(\ell),t+\frac{1}{2}}W_K^\ell$. The channel-only linear maps commute with the group action, so both projected fields transform by $D^{(\ell)}(R)$. Using the orthogonality of Wigner matrices,

$$\sum_{m=1}^{2\ell+1}\left(D^{(\ell)}(R)\mathbf{A}_i^{(\ell)}\right)_{m,:} \circ \left(D^{(\ell)}(R)\mathbf{B}_j^{(\ell)}\right)_{m,:} = \sum_{m=1}^{2\ell+1}\mathbf{A}_{i,m,:}^{(\ell)} \circ \mathbf{B}_{j,m,:}^{(\ell)}. \tag{34}$$

Thus $\mathbf{s}_{ij}^t$ is invariant. Since γ_w and γ_t are applied to invariant inputs,

$$\mathbf{t}_{ij}^{t+1}(g \cdot \mathbf{X}) = \mathbf{t}_{ij}^{t+1}(\mathbf{X}). \tag{35}$$

Equivariant feed-forward network. For each degree ℓ, channel-only projection preserves equivariance:

$$\mathbf{V}_i^{(\ell),t+\frac{1}{2}}(g \cdot \mathbf{X})W_U^\ell = D^{(\ell)}(R)\mathbf{V}_i^{(\ell),t+\frac{1}{2}}(\mathbf{X})W_U^\ell. \tag{36}$$

Taking the norm over the representation dimension removes the orientation:

$$\left\|D^{(\ell)}(R)\mathbf{V}_i^{(\ell),t+\frac{1}{2}}W_U^\ell\right\|_2 = \left\|\mathbf{V}_i^{(\ell),t+\frac{1}{2}}W_U^\ell\right\|_2. \tag{37}$$

Therefore the gates $\boldsymbol{\rho}_i^t$ and $\boldsymbol{\lambda}_{i,\ell}^t$ are invariant. The scalar stream remains invariant, and the steerable stream satisfies

$$\mathbf{V}_i^{(\ell),t+1}(g \cdot \mathbf{X}) = D^{(\ell)}(R)\mathbf{V}_i^{(\ell),t+1}(\mathbf{X}). \tag{38}$$

This completes the induction.

Invariant readout. After T layers, the readout descriptor uses invariant scalar states and norms of steerable fields:

$$\left\|\mathbf{V}_i^{(\ell),T}(g \cdot \mathbf{X})\right\|_2 = \left\|D^{(\ell)}(R)\mathbf{V}_i^{(\ell),T}(\mathbf{X})\right\|_2 = \left\|\mathbf{V}_i^{(\ell),T}(\mathbf{X})\right\|_2, \qquad \mathbf{z}_i(g \cdot \mathbf{X}) = \mathbf{z}_i(\mathbf{X}). \tag{39}$$

By assumption, the pair representation $\mathbf{P}$ is coordinate-frame invariant, so the pair summary also satisfies $\mathbf{g}_i(g \cdot \mathbf{X}, \mathbf{P}) = \mathbf{g}_i(\mathbf{X}, \mathbf{P})$. The classifier ψ_{cls} and sigmoid act only on invariant quantities, giving

$$\hat{y}_i(g \cdot \mathbf{X}, \mathbf{P}) = \hat{y}_i(\mathbf{X}, \mathbf{P}). \tag{40}$$

□

Remark. The theorem formalizes the role separation used throughout the model: neural networks learn only invariant routing and gating coefficients, while all orientation-dependent information is carried by steerable bases or steerable residue fields. This is why adding backbone-derived vector initialization, invariant edge descriptors, edge refinement, and coordinate-frame-invariant pair representations does not break the symmetry required by structure-based residue prediction.

D EXPERIMENTAL SETTINGS

D.1 ENERGY-BASED MODEL TRAINING AND EVALUATION SCOPE

The probe-ranking prototype uses the PDBbind v2020 refined source. After structure, contact, and decoy filtering, 3,671 complexes remain (3,027/321/323 train/validation/test). A complex is retained only if its ligand and a single-chain contact patch are parseable and a valid decoy exists within its split. Each positive patch–ligand pair is contrasted with one property-matched decoy. Property matching uses molecular weight, cLogP, hydrogen-bond donors and acceptors, and rotatable-bond count; a different scaffold and ECFP4 similarity below 0.4 are required. The model is optimized with the softplus ranking loss in eq. (12). For one positive ligand and one decoy this is the binary NCE objective, and the checkpoint is selected by that validation loss. The ligand encoder is a three-layer bond-aware GIN, the patch encoder is a single-layer two-head attention module, and both use 128-dimensional hidden states. Validation and test ranking use fixed within-split pools of 305 and 308 unique molecules, respectively. For each query patch, all candidates are scored by learned energy and ranked in ascending order; retrieval is measured by the rank of the co-crystallized ligand. These results use observed contact patches and precede scaffold diversification; they do not establish that a selected probe induces pocket opening.

D.2 POCKET-CONSTRAINED BOLTZ-2 PROTOCOL

We keep Boltz-2 frozen and compare two paired conditions: ligand-only generation and generation conditioned on soft pocket constraints derived from **`PocketMon`** predictions. Within each pair, the protein sequence, MSA, ligand identity, and random seed are identical. We use ten paired seeds (501–510) per target and set the maximum constraint distance to 4.5 Å.

Target-specific results. For 3ugkA, conditioning reduces the mean ligand RMSD after pocket alignment from 0.764 to 0.349 Å, with improvements in 9 of 10 seeds; the full ground-truth pocket Cα RMSD concurrently decreases from 2.139 to 0.581 Å. For 6hb0A and 5h9aA, ligand RMSD after pocket alignment decreases from 0.947 to 0.770 Å and from 1.015 to 0.807 Å, respectively, with improvements in all 10 seeds for both targets. Pocket-backbone accuracy is preserved for these targets (0.343 to 0.343 Å for 6hb0A and 0.401 to 0.382 Å for 5h9aA), and all generated structures pass the geometry checks.

D.3 COMPUTING INFRASTRUCTURES

Software infrastructures. All experiments are implemented in Python 3.10. The main training and evaluation environment uses the following libraries: PyTorch 2.6.0 (Paszke et al., 2019), PyG 2.7.0 (Fey & Lenssen, 2019), e3nn 0.6.0 (Geiger & Smidt, 2022), and MDTraj 1.10.3 (McGibbon et al., 2015). For offline extraction of AlphaFold2 pair representations, we use a separate environment based on JAX 0.4.28 (Bradbury et al., 2018) and ColabDesign 1.1.1 (Mirdita et al., 2022).

Hardware infrastructures. We conduct all experiments on a computing cluster equipped with NVIDIA A100-SXM4-40GB GPUs (with 40GB memory each) and 2 Intel Xeon Platinum 8358 CPUs (with 64 cores and 128 threads per node). Each training run uses a single GPU.

D.4 DATASET CONSTRUCTION AND STATISTICS

PocketMiner (Meller et al., 2023) evaluates whether a model can infer cryptic pockets from apo structures by learning from simulated pocket-opening events. Its training set is constructed from molecular dynamics simulations of 38 proteins and contains 3,290 trajectory-derived samples. Each sample is defined by a trajectory file, a corresponding structure file, and a frame index, so each protein can contribute multiple conformational snapshots or trajectory segments. Residue-level labels are aligned to the sampled conformation: each residue is scored by the maximum increase in nearby LIGSITE (Hendlich et al., 1997) probe-grid points within a 40 ns simulation window, and the continuous pocket-expansion score is binarized at the original threshold of 20. Validation and test evaluation use experimentally curated apo structures with 26 and 35 proteins, respectively, where residues are labeled as cryptic-pocket (1), non-pocket (0), or ambiguous (2; excluded from metrics).

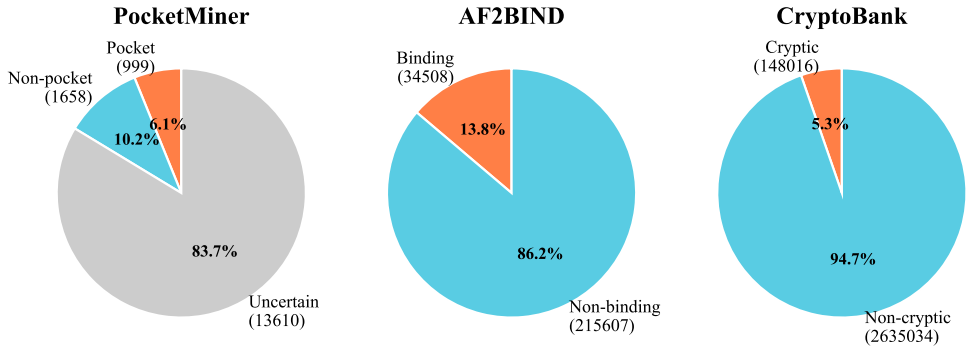

Figure 9: Residue-level label distribution across the three benchmarks. All datasets are highly imbalanced, and PocketMiner additionally contains a large fraction of uncertain labels.

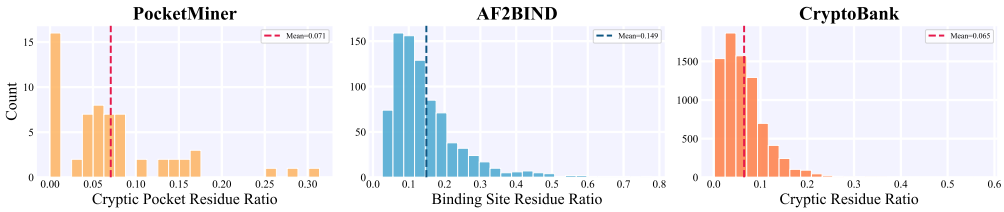

Figure 10: Per-protein positive residue ratio distributions. Most proteins contain only a small fraction of positive residues, especially in PocketMiner and CryptoBank.

This benchmark directly tests whether conformationally sampled training signals transfer to latent pocket detection in static apo conformations.

CryptoBank (Martinez et al., 2026) tests large-scale generalization of cryptic-site prediction across diverse protein families. The benchmark is constructed from holo–apo structural comparisons and provides residue-level binary crypticity annotations. We adopt the original PLM splits and map sequence-level annotations to parsed PDB residues; after filtering structures with incomplete backbone representations, the processed set contains ∼6,000 training, 750 validation, and 770 test proteins. The task is highly imbalanced, with positive cryptic-site residues accounting for only about 5% of valid residues. Since the test split contains protein families with low sequence similarity to the training and validation sets, this benchmark emphasizes out-of-family generalization rather than interpolation within homologous clusters.

AF2BIND (Gazizov et al., 2026) evaluates whether pretrained AlphaFold2 pair representations contain transferable signals for ligand-contact recovery. The benchmark is built from experimentally resolved protein–ligand co-crystal structures, where a residue is labeled as binding if its nearest ligand heavy-atom distance is below 5 Å; ambiguous low-confidence residues are masked following the original protocol. We use 832 proteins across ten predefined cross-validation splits, with about 13.8% positive residues, corresponding to a positive-to-negative ratio of approximately 1:6. Although AF2BIND is not restricted to cryptic pockets, it is directly relevant to our setting because it probes de novo binding-residue prediction from apo-like AlphaFold2 representations and 5,120-dimensional binder-pair features, which constitute the same class of global interaction priors exploited by PocketMon.

Label distribution and class imbalance. Figure 9 summarizes the residue-level label composition across the three benchmarks. PocketMiner annotates only 6.1% of residues as cryptic-pocket positives and 10.2% as confirmed non-pocket negatives, with the remaining 83.7% labeled as uncertain—a direct consequence of the difficulty of obtaining confident residue-level labels from molecular dynamics trajectories. AF2BIND has the highest positive ratio, with binding-site residues accounting for 13.8% of all valid residues—a positive-to-negative ratio of approximately 1:6—because its annotations are derived by binarizing residue–ligand distances under a fixed threshold. CryptoBank

exhibits the most extreme imbalance, with cryptic residues accounting for only 5.3% of all residues, making it a particularly challenging benchmark for residue-level prediction.

Figure 10 further reports the per-protein positive residue ratio. PocketMiner has a mean ratio of 0.071, with most proteins containing fewer than 15% positive residues; AF2BIND shows a broader distribution with a mean of 0.149, indicating moderate binding-site coverage across its 832 proteins; CryptoBank displays a comparably sparse signal, with a mean ratio of 0.065 and the majority of proteins below 10% cryptic residues. Together, the two views confirm that cryptic pocket prediction is an inherently imbalanced task at the residue level, where positive signals are sparse and locally concentrated. This motivates the use of weighted loss functions and class-aware sampling in our training pipeline.

D.5 Cross-Dataset Protein Overlap

For the transfer evaluation in Table 2, we compared the 35 PocketMiner test proteins with the proteins used to train the transferred models. The AF2BIND set shares none of these test proteins. The CryptoBank training set shares only 3P53A. That protein is not shared supervision: CryptoBank labels it by holo–apo structural change, whereas PocketMiner labels it by MD pocket opening. Re-evaluating the CryptoBank-trained model after removing 3P53A changes ROC-AUC from 89.38 to 89.87, PR-AUC from 83.32 to 83.70, Accuracy from 86.73 to 87.27, Precision from 83.97 to 83.84, Recall from 69.80 to 71.12, and Recovery from 44.49 to 46.42. Every change is below two points, and excluding the shared protein does not reduce transfer performance. This comparison matches PDB chain identifiers and does not include a sequence-homology clustering of differently numbered chains.

D.6 Baseline Methods and Reproduction

We compare **`PocketMon`** with three complementary baseline families. **Bioinformatics tools** include Fpocket (Le Guilloux et al., 2009) and P2Rank (Krivák & Hoksza, 2018), which detect exposed pockets from surface geometry. **Sequence-based prediction** is represented by the CryptoBank PLM built on ProtT5-XL-UniRef50 (Martinez et al., 2026), which operates without three-dimensional structural input. **Geometry-based prediction** includes CryptoSite (Cimermancic et al., 2016) and PocketMiner (Meller et al., 2023), both trained with structure-derived cryptic-pocket labels. AF2BIND (Gazizov et al., 2026) provides a direct pair-representation baseline by applying logistic regression to 5,120-dimensional binder-pair features. We additionally report structure-only **`PocketMon`** variants as ablation controls on AF2BIND.

Each baseline follows the configuration validated in its original publication and official implementation; method-specific settings are retained rather than artificially sharing hyperparameters across different architectures. Fpocket and P2Rank are evaluated as deterministic tools, while learned baselines use their prescribed training and split protocols.

D.7 Evaluation Protocol

ROC-AUC and PR-AUC. For PocketMiner and CryptoBank, we pool all valid residues in the test split before computing ROC-AUC and PR-AUC. PocketMiner contains an ambiguous label (2), which is removed before evaluation; only confirmed pocket residues (1) and non-pocket residues (0) are used. CryptoBank already provides binary labels, so we only remove padding positions labeled as -1. For AF2BIND, we follow the original protocol and compute ROC-AUC for each protein before averaging with TM-score-based inverse homology weights. We also report global PR-AUC over pooled residues so that the same precision–recall view is available across datasets.

Accuracy, Precision, and Recall. We compute Accuracy, Precision, and Recall from the model's sigmoid output using a fixed threshold of 0.5. After pooling all valid test residues, we use: $\mathrm{Acc} = (TP+TN)/(TP+FP+TN+FN)$, $\mathrm{Prec} = TP/(TP+FP)$, and $\mathrm{Recall} = TP/(TP+FN)$. The same threshold and pooling rule are used for all three datasets.

Recovery@k. Recovery@k measures whether the highest-ranked residues recover the true binding or cryptic site. For each protein, residues are sorted by predicted probability and the top k residues are selected, where k is the number of positive residues in that protein. Recovery is the fraction of

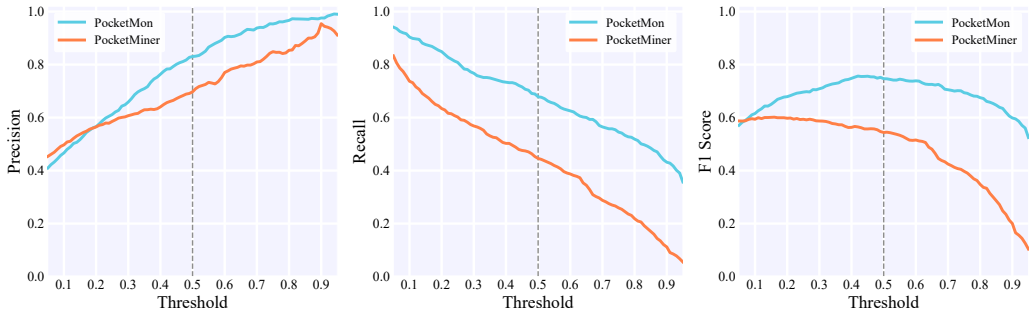

Figure 11: Threshold sensitivity on the PocketMiner test set. Precision, Recall, and F1 are evaluated from thresholds 0.05 to 0.95; the dashed line marks the default threshold of 0.5.

positives recovered in this selected set, and proteins without any positive residues are excluded. On PocketMiner, the ranking pool includes all residues, including ambiguous label-2 residues. We do not restrict the pool to labels 0 and 1 because that would artificially inflate recovery for proteins dominated by label-1 residues. On CryptoBank, the ranking pool contains all valid residues after removing padding. On AF2BIND, Recovery@k is averaged across proteins using TM-score weights.

Reporting details. Fpocket and P2Rank are deterministic. We therefore report single-run values without standard deviations on the fixed PocketMiner and CryptoBank test sets. On AF2BIND, their reported standard deviations come from evaluation across different predefined test splits.

Model selection. Checkpoint selection follows the validation metric used for each benchmark. For PocketMiner, we select the best checkpoint by validation PR-AUC, reflecting the sparse positive labels in cryptic pocket detection. For CryptoBank, we select checkpoints by validation ROC-AUC. For AF2BIND, checkpoints are selected by validation TM-score-weighted ROC-AUC. After checkpoint selection, all reported test metrics on the three datasets are summarized by the mean and standard deviation across several random seeds. Test labels are used only for the final evaluation.

D.8 Threshold Sensitivity, Calibration, and High-Confidence Screening

To examine sensitivity to the fixed threshold used for auxiliary classification metrics, we perform an operating-point analysis on one later combined **`PocketMon`** checkpoint and the PocketMiner baseline. The analysis pools 1,846 confirmed residues from the PocketMiner test set, including 563 positives and 1,283 negatives; residues with the uncertain label 2 are excluded. This analysis complements the threshold-independent ROC-AUC, PR-AUC, and Recovery@k metrics used for the primary comparison.

The PocketMiner numbers in the main text, including Table 1 and the PocketMiner-trained transfer row, use fixed-backbone AlphaFold2 pair features and are means over random seeds. AF2BIND instead uses binder-protocol features, the same 20-decoy representation as its own baseline. A later PocketMiner run with binder features alone stayed close to the fixed-backbone model, with complementary strengths across metrics. Fusing both streams from one binder-protocol pass improved this follow-up checkpoint. The threshold, calibration, and high-confidence results below are that single combined checkpoint.

Threshold sensitivity. Figure 11 shows the expected precision–recall tradeoff as the decision threshold increases. **`PocketMon`** maintains substantially higher recall than PocketMiner throughout the scan and provides a stronger overall tradeoff across practical operating points. At the default threshold of 0.5, **`PocketMon`** obtains 83.0% Precision, 67.9% Recall, and 74.7% F1, compared with 69.7%, 44.6%, and 54.4% for PocketMiner. The ranking-based conclusions therefore do not depend on a favorable choice of classification threshold.

High-confidence screening. The left panels of fig. 12 quantify the tradeoff most relevant to experimental prioritization. At a threshold of 0.7, **`PocketMon`** reaches 93.8% Precision while retaining 56.5% of true pocket residues, whereas PocketMiner reaches 81.0% Precision with 28.8% coverage. At the stricter threshold of 0.9, **`PocketMon`** reaches 97.6% Precision with 43.2% coverage, com-

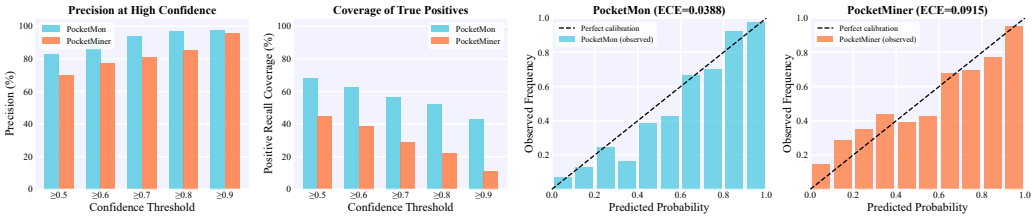

Figure 12: High-confidence screening and calibration on the PocketMiner test set. The left panels report Precision and positive Recall coverage at increasingly strict confidence thresholds. The right panels show reliability diagrams for **PocketMon** and PocketMiner; the dashed diagonal denotes perfect calibration.

pared with 95.5% Precision and 11.2% coverage for PocketMiner. Thus, **PocketMon** preserves substantially more true sites when predictions are restricted to high-confidence residues.

Probability calibration. The reliability diagrams in the right panels of fig. 12 compare raw predicted probabilities with observed positive frequencies across ten bins. **PocketMon** achieves an expected calibration error of 0.039, compared with 0.092 for PocketMiner, a 57.6% reduction. This closer agreement between confidence and empirical frequency supports the use of **PocketMon** scores for threshold-based downstream prioritization.

D.9 Implementation Details

Architecture. PocketMon employs an SE(3)-equivariant graph neural network augmented with AlphaFold2-derived pair or binder-pair fusion. Proteins are represented as residue graphs built from backbone atoms (N, Cα, C, O), with edges constructed by a Cα distance cutoff of 15 Å and at most 128 neighbors per residue. The network uses 128 hidden channels, 32 radial basis functions, 16 positional encoding dimensions, 4 message-passing layers, 4 attention heads, and spherical harmonics up to $l_{\max} = 2$. The relative positional term $\mathrm{PE}(j - i)$ is a fixed sinusoidal encoding of the residue-index offset; it has no embedding parameters, and any learned transformation is applied later by the edge network. Local equivariant embeddings and global pair summaries are combined via concatenation-based late fusion; we found simple concatenation to outperform element-wise multiplicative interaction, likely because the limited training data favors lower-capacity fusion heads. A two-layer MLP with hidden dimension 256 serves as the per-residue classifier. All models are optimized with Adam and trained on a single GPU.

Per-dataset configuration. On PocketMiner, the main-text results fuse fixed-backbone AlphaFold2 pair representations via late fusion and train with weighted BCE loss (threshold 20), dropout 0.1, learning rate 2×10^{-5}, and weight decay 10^{-5}. The combined checkpoint used for threshold and calibration analysis is described in Appendix D.8. On AF2BIND, the 5,120-dimensional binder-pair representation is projected and fused with local graph neural network embeddings; we train for up to 200 epochs with learning rate 5×10^{-5}, dropout 0.5, and TM-score sample weighting. On CryptoBank, we align sequence-level crypticity labels to parsed structural residues, use AF2 binder-pair features when available, and train with batch size 1, learning rate 2×10^{-5}, dropout 0.1, and LR reduction on validation PR-AUC plateaus.

D.10 Ablation Component Descriptions

We organize the ablation variants into three functional groups.

Geometric expressiveness. Removing steerable initialization removes backbone-derived initialization of steerable features. In the full model, pseudo-C_β vectors, forward and reverse backbone orientations are projected to initialize the $l = 1$ channels of the equivariant representation. Without this initialization, all steerable channels start from zero and the network must build directional information purely from message passing. Restricting to $\ell_{\max} = 1$ removes the $l = 2$ traceless symmetric tensor channels, testing whether higher-order angular features, which can represent planar and ellipsoidal geometric motifs, contribute to pocket recognition beyond first-order directional information. Removing dihedral features removes the projection of backbone dihedral angles (ϕ, ψ, ω and their

sine/cosine encodings) from the initial node embedding. Dihedrals encode local secondary-structure context; this variant tests whether the model benefits from explicit torsion information or can infer equivalent signals from spatial coordinates alone.

Information flow. Removing positional encoding sets the 16 fixed sinusoidal sequence-offset dimensions to zero, so edge features rely on radial basis functions alone. This probes whether sequence proximity, a strong inductive bias for local secondary structure, is necessary when spatial distance is already available. Reducing to $n_{\text{interactions}} = 2$ halves the number of equivariant message-passing layers from 4 to 2, halving the effective receptive field. Cryptic pockets often span 10–20 residues; this variant tests whether the reduced receptive field limits the model's ability to capture extended pocket geometry. Removing edge updates freezes edge features after initialization, preventing them from evolving through message-passing iterations. In the full model, edge features are updated at each layer to reflect changing pairwise relationships; removing this tests whether static distance-based edges suffice.

Global–local coupling. Removing AlphaFold2 pair fusion removes the pair representation from the fusion head, so the per-residue classifier relies solely on local equivariant graph embeddings. This directly tests whether global residue–residue interaction priors are necessary or whether local geometric features alone suffice for cryptic pocket detection.

D.11 Human Proteome Scan Protocol

For each of the 20,550 AlphaFold-predicted single-chain human protein structures (v6), we first mask low-confidence residues (pLDDT < 50) by setting their predicted probabilities to zero. PocketMon and AF2BIND each produce per-residue binding probabilities from a single forward pass using the precomputed AlphaFold2 binder-pair representations; residues above a probability threshold of 0.28 are selected and clustered into multi-residue binding sites via DBSCAN (Cα distance $\varepsilon = 6$ Å, min_samples = 3), discarding clusters with fewer than five residues. P2Rank is run in its AlphaFold configuration and directly outputs pocket-level predictions with per-residue membership. PocketMon identifies 48,396 candidate binding sites across 16,601 proteins, AF2BIND identifies 81,401 sites across 18,863 proteins, and P2Rank identifies 80,596 sites across 20,550 proteins. All three methods use the same set of input structures, so differences in the number of predicted sites reflect intrinsic model behavior.

For druggability assessment, Fpocket (Le Guilloux et al., 2009) is run separately on all 20,550 structures; for each protein it returns pockets with associated druggability scores in $[0, 1]$ (based on alpha-sphere decomposition, hydrophobic density, pocket volume, and polarity balance). A predicted binding site from PocketMon, AF2BIND, or P2Rank is matched to an Fpocket pocket if they share at least three residues. Matched sites inherit the Fpocket druggability score; unmatched sites are flagged as novel predictions that Fpocket did not detect. Of PocketMon's 48,396 sites, 38,184 (78.9%) overlap with an Fpocket pocket and 10,212 (21.1%) are unique. AF2BIND has 61,216 matched and 20,185 unique (75.2%/24.8%), while P2Rank has 73,485 matched and 7,111 unique (91.2%/8.8%).

For the UMAP representation-level analysis, we extract pre-classification-head embeddings (384-dimensional fused local-global representations for PocketMon; 5,120-dimensional standardized binder-pair features for AF2BIND, reduced to 1,608 dimensions via PCA retaining 95% variance) for 500,000 stratified-sampled residues spanning 20,494 proteins and project them to two dimensions using UMAP (McInnes et al., 2018).

E Limitations and Broader Impact

E.1 Limitations

Energy-based ranking model and constrained Boltz-2 prediction provide candidate ligand-compatible conformations for predicted sites. Their ability to recover experimentally observed open states from PocketMon-predicted patches has not yet been systematically established. For structure-based design, generated complexes should therefore be treated as hypotheses until pocket opening, pose quality, and ligand binding are assessed independently. In addition, the quality of the prediction can depend on the accuracy of the input structure and the informativeness of the available pair representation; proteins with low-confidence regions, missing loops, large disorder, or poorly represented conformational

states may remain challenging. Finally, our benchmarks are still limited by the available cryptic-pocket annotations, which are sparse and heterogeneous across simulation-derived, apo–holo, and ligand-contact definitions.

E.2 Broader Impact

PocketMon may facilitate the discovery of hidden binding opportunities in proteins that are difficult to target with conventional pocket detection, potentially accelerating early-stage drug discovery, target prioritization, and functional annotation. At the same time, computational pocket predictions should be treated as hypotheses rather than validated binding sites, especially when used to guide costly experimental campaigns or therapeutic development. Because improved binding-site discovery can be dual-use in principle, for example by enabling modulation of biologically important proteins, responsible use should include expert review, experimental validation, and compliance with relevant biosafety and biomedical research guidelines.