CB7MOL: A PHYSICOCHEMICALLY GROUNDED AI-DRIVEN MODEL FOR CB[7] AFFINITY PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

AI-driven scientific discovery requires more than predictive accuracy: models must remain reliable under scarce supervision while respecting underlying physicochemical principles. We study this challenge through cucurbit[7]uril (CB[7])-based host–guest binding affinity prediction, an important problem in molecular recognition for supramolecular chemistry. A fundamental challenge is to jointly acquire the molecular, geometric, and affinity knowledge required for grounded prediction despite limited CB[7]-specific data. We propose **CB7Mol**, a physicochemically grounded AI-driven model that strategically transfers complementary knowledge from compatible sources to bridge these three gaps: molecular encoder post-training acquires CB[7]-relevant chemical semantics while preserving geometry; host–guest complex pre-training learns binding geometry and conformational flexibility from resolved crystal structures; and cross-host calibration transfers affinity supervision from mechanistically compatible macrocyclic hosts. CB7Mol achieves the best overall average ranking across three CB[7] affinity evaluation sets. More importantly, mechanistic analyses show that its predictions reflect key physicochemical determinants of CB[7] recognition rather than dataset-specific correlations, supporting reliable discovery beyond known guest chemistries. Finally, an AI-to-experiment pipeline screens 304k molecules in GEOM-Drugs and prioritizes candidates for wet-lab validation. CB7Mol recovers established strong-binding scaffolds, uncovers physicochemically consistent candidates with novel scaffolds, and yields high-affinity guests validated by competition NMR with $\log K_a$ up to approximately 12.6, expanding the experimentally validated CB[7] chemical space and molecular toolbox for supramolecular design.

1 INTRODUCTION

AI for science is emerging as a transformative paradigm, with breakthroughs from protein structure prediction (Jumper et al., 2021) to molecular and materials design (Gómez-Bombarelli et al., 2018; Merchant et al., 2023) accelerating the discovery of new structures, properties, and functions. Supramolecular chemistry (Lehn, 1988) is a compelling frontier for AI, harnessing non-covalent interactions to build functional systems for applications from targeted therapeutics and disease diagnosis to sensing and advanced materials (Ma & Zhao, 2015; Williams et al., 2021). Molecular recognition lies at its heart, enabling molecules to selectively bind complementary partners with the specificity needed to construct and control such systems (Mattia & Otto, 2015). Among synthetic hosts, cucurbit[7]uril (CB[7]) stands out for its water solubility, biocompatibility, and exceptionally strong and selective guest binding, making high-affinity guests valuable for targeted drug delivery and displacement-based biosensing (Kraft et al., 2026). For example, CB[7]-based recognition enables broad-spectrum detection of fentanyl and emerging analogs, offering a route to sensing illicit synthetic opioids and addressing the deadly opioid epidemic (Yan et al., 2023; Gao et al., 2025).

However, realizing the potential of CB[7] across these applications depends on high-affinity guests, yet known strong binders are limited to a small family of chemical scaffolds (Barrow et al., 2015). Accurate and scalable affinity prediction is therefore critical for discovering new guest chemistries and expanding CB[7]-enabled applications. Physics-based simulations can estimate affinities but remain computationally expensive for large-scale screening (Gonzalez et al., 2022), while experimental measurements are time-consuming (Thordarson, 2011). Given the vast candidate chemical space, scalable AI offers a promising direction to rapidly identify high-affinity binders and guide

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.

experimental discovery. ***Given its real-world impact, we use CB[7]-based host–guest affinity prediction as a compelling use case to investigate how AI can advance scientific discovery while remaining grounded in physicochemical principles.***

Achieving this goal requires addressing ***three fundamental challenges***. *(1) AI models should be grounded in underlying physical and chemical mechanisms.* CB[7] recognition is governed by well-characterized determinants, including the hydrophobic effect, release of high-energy cavity water, and ion–dipole interactions between guest cations and portal carbonyls (Nau et al., 2011; Biedermann et al., 2012). A model may achieve high accuracy by exploiting correlations in known guests without capturing these determinants, yet fail for new chemistries. More fundamentally, ungrounded models offer limited insight into why a guest binds and may prioritize physicochemically implausible candidates, undermining generalization and reliable discovery. *(2) Capturing these mechanisms requires modeling host–guest binding geometry and conformational flexibility.* Upon complexation, a guest must adopt a binding-compatible conformation within the CB[7] cavity, reducing its conformational freedom. The resulting entropy penalty can offset favorable interactions, making affinity sensitive to geometry and flexibility, particularly upon substantial guest reorganization (Cram, 1986; Taylor et al., 2017). Yet resolving unseen binding geometries remains computationally expensive (Plett & Grimme, 2023). *(3) Learning these coupled determinants is challenging under scarce CB[7] affinity supervision.* Few CB[7]–guest affinities have been measured (Ma et al., 2026), making it difficult to distinguish mechanistic determinants from dataset-specific correlations and balance their contributions. These challenges create three gaps: insufficient task-relevant molecular semantics, limited binding-geometry supervision, and scarce target-host affinity labels. These gaps are coupled: molecular semantics identify recognition determinants, geometry and flexibility govern their realization, and affinity supervision quantifies their competing contributions. Existing AI models can be adapted to CB[7] affinity prediction, but none jointly addresses these gaps: generic molecular encoders lack CB[7]-specific physicochemical semantics (Levine et al., 2025; Rong et al., 2020), molecular relational models capture intermolecular interactions but not binding geometry and flexibility (Du et al., 2024; Lee et al., 2023a), and 3D complex models require structural and affinity supervision scarce for CB[7] (Jiao et al., 2026). Thus, jointly acquiring the molecular, geometric, and affinity knowledge for physicochemically grounded prediction remains an open challenge.

To fill this research gap, we propose **CB7Mol**, a *physicochemically grounded framework* for label-scarce CB[7]-based host–guest binding affinity prediction through three complementary forms of transferable supervision. First, through molecular encoder post-training, we equip the model with CB[7]-relevant knowledge by preserving geometry through self-distillation and aligning representations with chemical semantics. Second, through host–guest complex pre-training, we teach the model 3D host–guest binding from experimentally resolved structures of mechanistically compatible complexes while accounting for guest conformational flexibility. Third, through cross-host calibration, we transfer affinity knowledge from compatible rigid macrocyclic hosts to alleviate scarce CB[7] affinity supervision and learn how recognition determinants jointly govern affinity. These components integrate molecular, geometric, and affinity knowledge for physicochemically grounded and scalable prediction. Beyond retrospective evaluation, we establish an AI-to-experiment discovery pipeline by applying CB7Mol to screen 304k molecules in GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2022) and prioritize candidates for wet-lab validation by competition NMR. This pipeline recovers established strong-binding scaffolds, identifies physicochemically consistent candidates with novel scaffolds, and confirms high-affinity guests. We highlight our contributions:

- We introduce **CB7Mol**, a physicochemically grounded AI-driven model for label-scarce CB[7] affinity prediction, with key technical novelty in jointly bridging three critical knowledge gaps through CB[7]-relevant molecular semantics, experimentally grounded binding geometry and conformational flexibility, and transferable cross-host affinity supervision.
- CB7Mol achieves the **best overall average ranking** across three CB[7] affinity evaluation sets. More importantly, extensive mechanistic analyses demonstrate that its predictions reflect the physicochemical basis of CB[7] recognition rather than merely dataset-specific correlations, providing a stronger foundation for reliable discovery beyond known guest chemistries.
- We establish an **AI-to-experiment discovery pipeline** that screens 304k molecules to recover established strong-binding scaffolds and uncover physicochemically consistent candidates with novel scaffolds for experimental prioritization. Competition NMR validates high-affinity guests with $\log K_a$ up to approximately 12.6, demonstrating that CB7Mol expands the experimentally validated CB[7] guest space and molecular toolbox for supramolecular design.

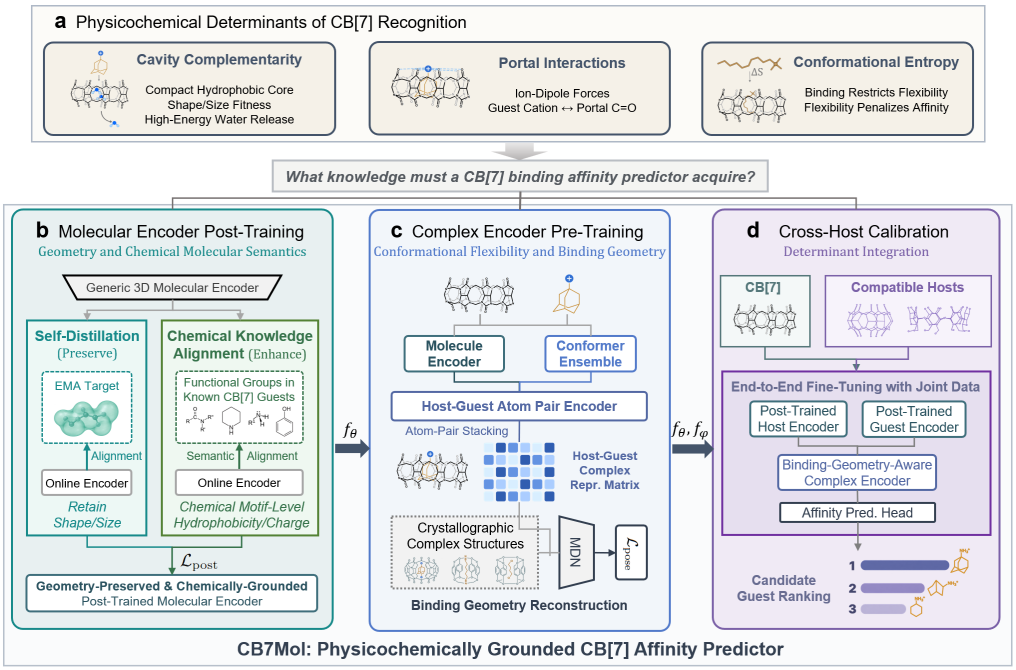


Figure 1: **Overview of CB7Mol.** (a) CB[7] guest recognition is governed by cavity complementarity, portal interactions and the conformational-entropy penalty upon binding. (b) Molecular post-training preserves molecular geometry while transferring CB[7]-relevant chemical semantics. (c) Host–guest complex pre-training incorporates conformational flexibility and transfers binding-geometry supervision from crystallographic complexes. (d) Cross-host calibration further transfers affinity supervision from other macrocyclic hosts with compatible binding mechanisms.

2 Related Work

Existing approaches provide complementary but incomplete knowledge for CB[7] affinity prediction. Physics-based methods such as molecular docking and MM/PBSA can estimate binding poses and affinities, but their computational cost limits large-scale virtual screening (Friesner et al., 2004; Plett & Grimme, 2023; Genheden & Ryde, 2015). Learning-based methods are more scalable, yet generic molecular encoders provide transferable representations without being specifically grounded in CB[7]-relevant physicochemical semantics (Levine et al., 2025; Rong et al., 2020). Molecular relational models capture intermolecular interactions at the atom or substructure level (Pathak et al., 2020; Lee et al., 2023a;b; Du et al., 2024; 2026), but do not explicitly model binding geometry and conformational flexibility. Recent equivariant complex models jointly represent interacting molecules in the same 3D frame (Kong et al., 2024; Fang et al., 2025; Jiao et al., 2026), but rely on structural and affinity supervision that remains scarce for CB[7]. Although existing AI models can be adapted to CB[7] affinity prediction, none is designed to jointly acquire the molecular, geometric, and affinity knowledge required for physicochemically grounded prediction.

3 Methodology

As shown in Figure 1, CB7Mol addresses three knowledge gaps that arise in CB[7] affinity prediction: insufficient CB[7]-relevant molecular semantics, limited supervision for binding geometry and conformational flexibility, and scarce target-host affinity labels. We bridge these gaps through three complementary forms of transferable supervision. Formally, a host $\mathcal{H}$ and a guest $\mathcal{G}$ contain N_h and N_g atoms, respectively. A post-trained molecular encoder f_θ maps them to geometrically and chemically informed per-atom representations $\mathbf{H} \in \mathbb{R}^{N_h \times d}$ and $\mathbf{G} \in \mathbb{R}^{N_g \times d}$. A binding-geometry-pretrained, conformational-flexibility-aware complex encoder f_ϕ then constructs a host–guest interaction matrix $\mathbf{Z} \in \mathbb{R}^{N_h \times N_g \times d_z}$, where $\mathbf{z}_{ij} \in \mathbb{R}^{d_z}$ represents the interaction between host atom i

and guest atom j. After cross-host affinity calibration, an affinity predictor aggregates $\mathbf{Z}$ to estimate CB[7]–guest binding affinity and rank candidate guests for virtual screening.

3.1 Molecular Encoder Post-training

The first knowledge gap is the lack of CB[7]-relevant molecular semantics in generic molecular representations. Although pretrained 3D molecular encoders capture broadly transferable geometric and chemical patterns, they are not specifically optimized for the physicochemical determinants most relevant to CB[7] recognition (Zhou et al., 2023; Levine et al., 2025). Therefore, we post-train the molecular encoder on a large-scale molecular library GEOM (Axelrod & Gomez-Bombarelli, 2022) with two complementary objectives:

(i) Self-distillation. Shape- and size-complementarity governs how completely a guest fills the CB[7] cavity and thereby releases high-energy cavity water (Assaf & Nau, 2014). In order to preserve the molecular geometric knowledge within a molecular encoder, a self-distillation regularizer is introduced to keep the molecular representations consistent with its own geometric-aware outputs.

Realizing such self-distillation usually requires augmented views of the input molecule. Given that common augmentation strategies are destructive for molecular semantics (Lee et al., 2022), we adopt an augmentation-free scheme and use dropout for minimal and semantic-preserving perturbation (Gao et al., 2021). Concretely, we maintain an online encoder f_θ and a target encoder $f_{\theta'}$ whose parameters are an exponential moving average (EMA) version of the online encoder ($\theta' \leftarrow m\,\theta' + (1-m)\,\theta$, where m is the momentum). The same molecule $\mathcal{M}$ is passed through both encoders under dropout. Reading the pooled molecular representation $\mathbf{c} = f_\theta(\mathcal{M})$, an online predictor p_b regresses the target encoder's representation of the same molecule:

$$\mathcal{L}_{\mathrm{SD}} = \ell(p_b(\mathbf{c}),\ \mathrm{sg}[f_{\theta'}(\mathcal{M})])\,, \tag{1}$$

where $\mathrm{sg}[\cdot]$ is the stop-gradient operator. Such a design supplies the asymmetry that keeps the objective non-trivial and prevents representational collapse (Tarvainen & Valpola, 2017).

(ii) Chemical Knowledge Alignment. Generic 3D molecular encoders are not specifically grounded in functional group-level chemical semantics required for affinity prediction, such as hydrophobicity and charge (Levine et al., 2025), which are critical for CB[7] recognition. We therefore align molecular encoder outputs with the representations from another molecular encoder pre-trained with explicit chemical semantic targets. Concretely, a head α_g maps the molecular representation $\mathbf{c}$ to chemically grounded target $\mathbf{v}_{\mathrm{gr}}$, injecting the chemical context that $\mathbf{c}$ underrepresents:

$$\mathcal{L}_{\mathrm{chem}} = \ell(\alpha_g(\mathbf{c}),\ \mathbf{v}_{\mathrm{gr}})\,. \tag{2}$$

where $\mathbf{v}_{\mathrm{gr}}$ is frozen and standardized for stability. Such alignment can inject functional-group knowledge relevant to CB[7] molecular recognition into the molecular representations. The full post-training objective is formulated as:

$$\mathcal{L}_{\mathrm{post}} = \mathcal{L}_{\mathrm{SD}} + \lambda \mathcal{L}_{\mathrm{chem}} \tag{3}$$

Only the online encoder f_θ is retained afterwards for CB[7]–guest complex representation learning and affinity prediction. Together, these objectives equip the molecular encoder with CB[7]-relevant molecular semantics while preserving geometric information. However, molecular representations alone do not specify how a guest adopts a binding-compatible geometry within a host.

3.2 Binding Geometry Learning with Conformational Flexibility

Conformational entropy is a physically significant determinant of binding affinity. Complexation restricts conformational freedom and thus incurs a substantial entropic penalty for flexible guests (Cram, 1986; Taylor et al., 2017). However, the post-trained backbone embeds each molecule from a single conformer and therefore cannot faithfully reflect the ensemble-level flexibility.

Given the post-trained backbone, hosts and guests are embedded separately into $\mathbf{H}$ and $\mathbf{G}$. To model their interactions based on conformational flexibility, we augment molecular embeddings with a scalar per-atom flexibility descriptor, the normalized root-mean-square fluctuation (RMSF) r_i of atom i over a generated conformer ensemble (Arantes et al., 2025), yielding flexibility-aware per-atom representations $\tilde{\mathbf{h}}_i = [\mathbf{h}_i,\, r_i]$ and $\tilde{\mathbf{g}}_j = [\mathbf{g}_j,\, r_j]$. The host–guest atom-pair encoder g_ϕ

combines a host-atom embedding $\tilde{\mathbf{h}}_i$ and a guest-atom embedding $\tilde{\mathbf{g}}_j$ into an atom-pair interaction vector $\mathbf{u}_{ij}$, which is then passed through a stack of residual blocks g_ϕ to yield the atom pair representation $\mathbf{z}_{ij}$:

$$\mathbf{z}_{ij} = g_\phi(\mathbf{u}_{ij}), \qquad \mathbf{u}_{ij} = \big[\, \tilde{\mathbf{h}}_i,\ \tilde{\mathbf{g}}_j,\ \tilde{\mathbf{h}}_i \odot \tilde{\mathbf{g}}_j,\ |\tilde{\mathbf{h}}_i - \tilde{\mathbf{g}}_j| \,\big], \tag{4}$$

where $\mathbf{z}_{ij} \in \mathbb{R}^{d_z}$. Stacking $\mathbf{z}_{ij}$ over all host-atom/guest-atom pairs gives the host–guest complex representation matrix $\mathbf{Z}$. This construction is permutation-equivariant in the atoms and accepts hosts and guests of arbitrary size.

The RMSF-augmented representations provide an atom-wise proxy for the conformational freedom available before complexation. However, molecular flexibility alone does not determine the conformational-entropy penalty. The penalty arises when the conformational ensemble is restricted to conformations compatible with a particular binding geometry, which determines non-covalent contacts realized to stabilize the resulting complex (Grimm et al., 2022). We therefore complement the flexibility-aware atom-pair representations with explicit supervision over binding geometry.

Obtaining accurate CB[7]–guest docking poses is very expensive, while crystallographic geometries can provide informative proxies for solution-state host–guest contacts (Raffaini et al., 2009; Cao et al., 2014). Therefore, we attach a mixture density network (MDN) that maps each interaction vector to a K-component Gaussian mixture over the interatomic distance d_{ij}:

$$p(d_{ij} \mid \mathbf{z}_{ij}) = \sum_{k=1}^{K} \pi_{ij}^k \, \mathcal{N}\big(d_{ij};\, \mu_{ij}^k,\, \sigma_{ij}^k\big)\,, \qquad \sum_k \pi_{ij}^k = 1, \tag{5}$$

where the mixture weights π_{ij}^k are produced by a softmax and the component means μ_{ij}^k and standard deviations σ_{ij}^k are constrained to be positive. Modeling distance distribution allows the network to express the uncertainty of interatomic contacts. The head is supervised with a subset of the Cambridge Structural Database (Wang et al., 2026b; Groom et al., 2016), which contains crystallographically resolved host–guest complexes compatible with CB[7]. For each complex, we read the true interatomic distances $d_{ij}^\star$ and minimize the negative log-likelihood over contacting atom pairs,

$$\mathcal{L}_{\text{pose}} = -\sum_{(i,j)} \log p\left(d_{ij}^\star \mid \mathbf{z}_{ij}\right), \tag{6}$$

Together, the flexibility descriptors and crystallographic structures provide a coarse, transferable prior over conformational constraints and host–guest contacts.

3.3 Cross-Host Calibration

Molecular semantics and binding geometry recognize the presence of determinants, while affinity prediction requires learning how their competing contributions combine to determine realized binding strength. With only a limited number of experimentally measured CB[7] affinities (Ma et al., 2026), directly learning these trade-offs from the target host alone is difficult. Therefore, we incorporate affinity labels from other macrocyclic hosts that share relevant physicochemical recognition determinants with CB[7] (Escobar & Ballester, 2021; Assaf & Nau, 2014; Guo & Liu, 2014). These systems provide diverse and transferable supervision for binding affinity prediction.

To predict affinity, we integrate the post-trained molecular encoders and pretrained complex encoder, reducing $\mathbf{Z}$ to a fixed-size vector $\bar{\mathbf{z}}$ by mean-pooling over all atom pairs. Then, an MLP-based affinity head α_ω regresses the binding affinity $\hat{y} = \alpha_\omega(\bar{\mathbf{z}})$ under a squared-error loss $\mathcal{L}_{\text{aff}} = (\hat{y} - y)^2$. Because the host structure is explicitly encoded, CB7Mol can retain host-specific differences in cavity geometry and portal chemistry while learning from the binding determinants shared across compatible macrocyclic host–guest systems.

4 Experiments

4.1 Experimental Settings

Datasets. We collected host–guest complexes from three sources for binding affinity prediction: CB[7]-based complexes from 14 SAMPL4 entries (Muddana et al., 2014) and 60 expert annotations (Expert), and all the 492 rigid macrocyclic host–guest complexes from SupraBench (Ma et al.,

Table 1: **CB[7]–guest binding affinity prediction results** under 5-fold validation and average ranking (A.R.). The best and second-best results are highlighted in **boldface** and underline, respectively. CB7Mol achieves the strongest A.R..

Method	SupraBench			Expert			SAMPL4			A.R.
	RMSE↓	ρ↑	τ↑	RMSE↓	ρ↑	τ↑	RMSE↓	ρ↑	τ↑	
aISS	–	0.129 ±.31	0.089 ±.21	–	0.063	0.041	–	−0.037	−0.077	–
QSAR-RF	1.640 ±.39	0.718 ±.09	0.547 ±.08	1.963 ±.03	0.323 ±.06	0.243 ±.05	1.831 ±.09	0.599 ±.20	0.400 ±.13	6.00
QSAR-XGB	1.683 ±.46	0.664 ±.11	0.506 ±.11	2.050 ±.08	0.230 ±.05	0.170 ±.04	2.348 ±.29	0.694 ±.14	0.511 ±.12	7.11
GCN	1.990 ±.39	0.573 ±.05	0.421 ±.05	1.781 ±.13	**0.565** ±.03	**0.435** ±.02	1.890 ±.25	0.452 ±.01	0.302 ±.02	7.00
GIN	1.985 ±.27	0.581 ±.06	0.429 ±.06	1.956 ±.14	0.468 ±.01	0.354 ±.02	1.859 ±.61	0.720 ±.07	0.578 ±.08	5.33
MPNN	2.034 ±.32	0.580 ±.07	0.437 ±.05	2.036 ±.24	0.425 ±.03	0.318 ±.03	2.397 ±.56	0.555 ±.19	0.387 ±.20	8.89
CIGIN	1.855 ±.49	0.591 ±.14	0.449 ±.12	2.058 ±.19	0.430 ±.11	0.325 ±.09	2.024 ±.25	0.507 ±.21	0.402 ±.16	7.44
CGIB	2.875 ±.50	0.237 ±.20	0.174 ±.16	2.882 ±1.12	0.115 ±.30	0.087 ±.22	2.674 ±1.03	0.535 ±.16	0.402 ±.11	11.83
CMRL	2.104 ±.30	0.584 ±.11	0.426 ±.09	2.014 ±.12	0.335 ±.11	0.249 ±.09	2.347 ±.87	0.569 ±.14	0.420 ±.11	8.56
MMGNN	1.849 ±.25	0.630 ±.09	0.454 ±.07	1.923 ±.29	0.510 ±.02	0.385 ±.01	3.036 ±.28	0.715 ±.19	0.564 ±.21	5.11
I2Mole	1.946 ±.28	0.607 ±.09	0.449 ±.09	1.785 ±.11	0.493 ±.04	0.373 ±.03	2.038 ±.93	0.717 ±.08	0.562 ±.05	4.83
EPT	2.165 ±.41	0.506 ±.16	0.385 ±.12	1.921 ±.15	0.370 ±.02	0.265 ±.01	**1.244** ±.15	**0.819** ±.03	**0.654** ±.02	6.22
GET	2.738 ±.34	0.021 ±.11	0.001 ±.07	2.340 ±.04	−0.092 ±.04	−0.066 ±.03	2.672 ±.10	0.716 ±.01	0.528 ±.02	10.89
CB7Mol	**1.253** ±.35	**0.834** ±.12	**0.691** ±.12	**1.729** ±.11	0.558 ±.06	0.427 ±.05	1.864 ±.12	0.741 ±.12	0.585 ±.12	**1.78**

2026), spanning CB[7], α-/β-/γ-cyclodextrin, p-SC[4]arene and CB[5]. We construct training and validation set from SupraBench, where only 125 CB[7]-based complexes are used for validation. SAMPL4 and Expert are adopted as independent test sets. See Appendix A.1 for details.

Baselines. We compare CB7Mol against four families of methods: *(i) Physics-driven methods*: in-silico aISS (Plett & Grimme, 2023) and QSAR (RF, XGBoost) (Breiman, 2001; Chen & Guestrin, 2016); *(ii) Graph neural networks*: GCN (Kipf & Welling, 2017), GIN (Xu et al., 2018) and MPNN (Gilmer et al., 2017); *(iii) Molecular relational learning*: CIGIN (Pathak et al., 2020), CGIB (Lee et al., 2023a), CMRL (Lee et al., 2023b), MMGNN (Du et al., 2024) and I2Mole (Du et al., 2026); *(iv) Host–Guest Complex Encoder*: EPT (Jiao et al., 2026) and GET (Kong et al., 2024). Please refer to Appendix A.2 for detailed introduction.

Evaluation Metrics. We report RMSE for affinity regression and Spearman's ρ and Kendall's τ for guest ranking quality (details in Appendix A.3).

Implementation Details. We adopt Uni-Mol (Zhou et al., 2023) and GROVER (Rong et al., 2020) as the molecular encoder backbone and chemical knowledge alignment target, respectively. Please refer to Appendix A.4 for other details.

4.2 Main Results

Table 1 evaluates the affinity-prediction and binder-ranking performance of CB7Mol. We compare all selected methods across three evaluation sets. aISS is not designed for binding affinity prediction, so we only report ranking-based performance for aISS. CB7Mol achieves the best average ranking of 1.78 with best and sub-best performance in ρ and τ across all evaluation sets. CB7Mol achieves the best performance on SupraBench. The performance gain of CB7Mol is limited for guests with unseen scaffolds in Expert and SAMPL4, while CB7Mol still achieves consistently competitive guest ranking quality. These results establish the predictive utility of CB7Mol.

4.3 Design Justification

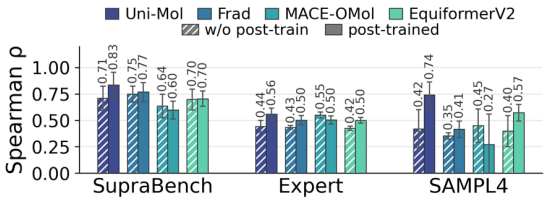
We evaluate whether the three supervision sources provide complementary benefits and whether these gains depend on the chosen molecular backbone and chemical alignment target.

Ablation Study. We ablate five key components of CB7Mol in turn and attempt to replace molecular post-training objectives with per-atom objectives. Experimental setup is summarized in Appendix A.5. As shown in Figure 2a, cross-host calibration provides the largest performance gain, highlighting the importance of transferring affinity supervision from mechanistically compatible hosts. The remaining components yield generally complementary improvements in guest ranking, con-

a Ablation Study for CB7Mol Components

Method	SupraBench RMSE↓	SupraBench ρ↑	SupraBench τ↑	Expert RMSE↓	Expert ρ↑	Expert τ↑	SAMPL4 RMSE↓	SAMPL4 ρ↑	SAMPL4 τ↑
CB7Mol	**1.253** ±.35	**0.834** ±.12	**0.691** ±.12	1.729 ±.11	**0.558** ±.06	**0.427** ±.05	1.864 ±.12	**0.741** ±.12	**0.585** ±.12
w/o chem-align	1.507 ±.33	0.768 ±.13	0.617 ±.14	**1.599** ±.07	0.486 ±.01	0.369 ±.01	1.742 ±.17	0.650 ±.11	0.499 ±.08
w/o SD	1.477 ±.43	0.780 ±.10	0.653 ±.11	1.751 ±.08	0.530 ±.02	0.408 ±.02	2.016 ±.28	0.583 ±.06	0.393 ±.04
w/o RMSF	1.395 ±.39	0.825 ±.07	0.685 ±.09	1.738 ±.11	0.506 ±.03	0.380 ±.03	**1.536** ±.37	0.725 ±.07	0.552 ±.07
w/o pose	1.343 ±.32	0.810 ±.10	0.653 ±.10	1.768 ±.12	0.511 ±.04	0.394 ±.03	1.738 ±.35	0.712 ±.13	0.557 ±.10
w/o cross-host	2.712 ±.34	0.345 ±.40	0.265 ±.31	2.240 ±.02	0.293 ±.17	0.219 ±.13	2.427 ±.14	−0.086 ±.17	−0.069 ±.10
w/ atom-mask	1.390 ±.33	0.815 ±.10	0.680 ±.12	1.688 ±.12	0.501 ±.05	0.380 ±.04	2.149 ±.23	0.550 ±.11	0.411 ±.09
w/ atom-chem-align	1.514 ±.43	0.755 ±.11	0.624 ±.11	1.770 ±.06	0.496 ±.04	0.374 ±.03	1.885 ±.33	0.596 ±.13	0.460 ±.10

b Molecular Encoder Backbone Evaluation



c Chemical Alignment Target Evaluation

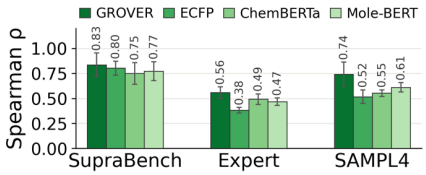


Figure 2: **Design Justification for CB7Mol.** (a) Ablation study over key CB7Mol components. (b) Alternative molecular encoder backbone evaluation with and without post-training. (c) Comparison of alternative chemical knowledge alignment targets.

sistent with our identified knowledge gaps. Moreover, the molecule-level post-training objectives outperform their atom-level alternatives due to high-level semantic modeling.

Molecular Encoder Evaluation. We replace Uni-Mol (Zhou et al., 2023) with three advanced molecular encoders: Frad (Feng et al., 2023), MACE-OMol (Levine et al., 2025) and EquiformerV2 (Liao et al., 2024) (See Appendix A.6). We also evaluate each backbone with and without post-training to verify the general applicability of our post-training framework. As shown in Figure 2b, post-trained Uni-Mol performs best across all evaluation sets, while post-training improves most alternative backbones. This indicates that the gains primarily arise from adapting generic representations toward CB[7]-relevant knowledge rather than from a particular architecture.

Chemical Knowledge Target Evaluation. We replace GROVER (Rong et al., 2020) target with three representative molecular representations spanning distinct chemical knowledge modalities: ECFP (Rogers & Hahn, 2010), ChemBERTa (Ahmad et al., 2022) and Mole-BERT (Xia et al., 2023). Their details are provided in Appendix A.7. Figure 2c shows that none consistently surpasses GROVER, whose advantage is strongest on unseen SAMPL4 scaffolds. Its contextualized motif-level knowledge appears better aligned with CB[7]-relevant chemical semantics, suggesting that transfer quality depends more on task-relevant determinants than representation complexity.

4.4 Physicochemical Grounding Analysis

Beyond predictive performance, Figure 3 answers whether CB7Mol reflects key physicochemical determinants of CB[7] recognition across three dimensions: molecular semantics, binding geometry and conformational flexibility, and transferable affinity signals.

Functional-Group Knowledge Decoding. We select a small set of chemical motifs prevalent in known CB[7] guests, and measure how well each functional group can be recovered from molecular representations. Experimental setup can be found in Appendix A.8. Figure 3a reports post-training gains in decoding accuracy together with motif prevalence among known CB[7] guests. We can find that the original molecular encoder systematically under-represents the nitrogen-rich and protonatable groups that are common in CB[7] guests (e.g., primary amines), while post-training makes these functional-group knowledge more accessible to CB7Mol.

Docking Pose Reconstruction. We recover the predicted interatomic distances from the pre-trained MDN and compare them with reference structures. Figure 3b shows that the pretrained encoder reconstructs the binding geometry of in-domain CSD complexes well. For unseen CB[7]-guest com-

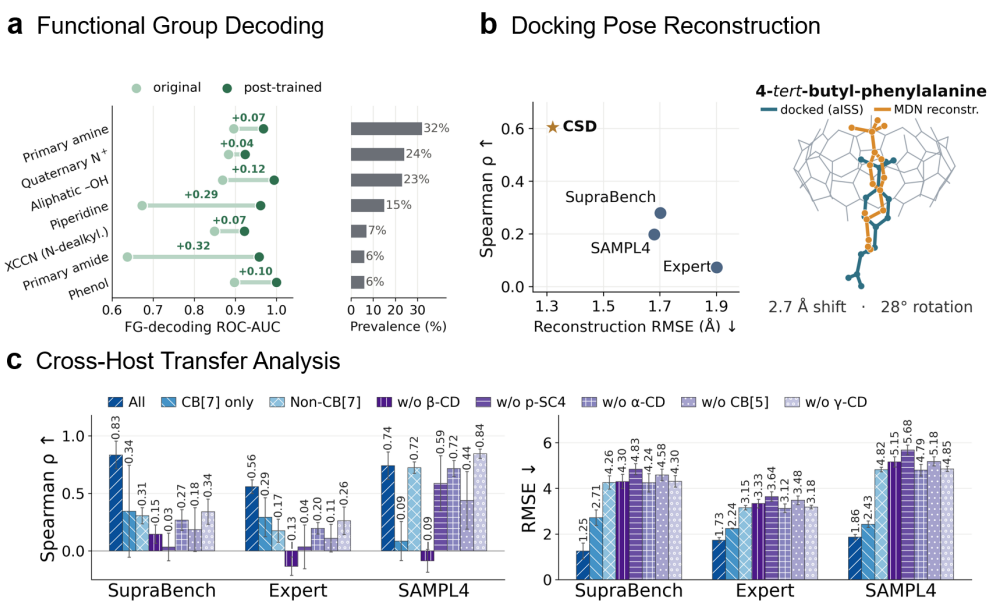


Figure 3: **Physicochemical Grounding Analysis of CB7Mol.** (a) CB[7]-relevant functional-group decoding from the original and post-trained molecular representations. (b) Binding-geometry reconstruction on crystallographic and CB[7]-guest complexes, along with case study. (c) Cross-host transfer evaluation under different calibration settings.

plexes, binding-geometry-aware pre-training transfers a coarse but non-trivial prior over plausible host-guest contacts. In the visualized case, the pre-trained MDN predicts a chemically reasonable pose despite deviations from the aISS-docked pose. This does not imply exact structure recovery, but supports the intended role of crystallographic supervision as transferable geometric guidance. Experimental setup and detailed analysis can be found in Appendix B.1.

Cross-Host Transfer Analysis. To validate the contribution of non-CB[7] affinity supervision, we calibrate CB7Mol with different data settings and report the results in Figure 3c. *Non-CB[7]* calibrates CB7Mol with all the non-CB[7] complexes, while each *w/o* * removes the designated host from training data to demonstrate the contribution of a single host. Even without any CB[7] supervision, calibrating on non-CB[7] hosts alone transfers substantially to CB[7] evaluation sets. This indicates that affinity information can transfer across hosts when the source systems share compatible binding mechanisms with CB[7], supporting cross-host calibration for integration of determinant influence over affinity. Detailed analysis can be found in Appendix B.2.

4.5 AI-to-Experiment CB[7] Guest Discovery

Retrospective benchmarking is necessary but not sufficient to establish whether physicochemical grounding can translate affinity prediction into experimentally actionable discovery. We therefore build an AI-to-experiment pipeline that applies CB7Mol to screen 304k molecules in GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2022), ranks and prioritizes high-affinity candidates, analyzes the resulting scaffold space, and advances selected candidates to competition NMR validation.

Scaffold Analysis for Top-Ranked Candidates. CB7Mol narrows the given search space to 349 strong-binder candidates with predicted $\log K_a > 8$. We summarize the dominant generic scaffolds of these candidates based on scaffold definition in Appendix A.9. Figure 4a presents a subset of extracted scaffolds by its included candidate molecules and mean predicted affinity, distinguishing scaffolds already observed among known CB[7] guests from the novel ones. CB7Mol recovers established structural patterns of strong CB[7] binders, mostly adamantane and its derivatives. There are also several novel scaffolds prioritized with compatible physicochemical features. As illustrated in the scaffold examples in Figure 4a, the known adamantane and novel cyclohexane-cycloheptane scaffolds can both provide a hydrophobic core and an protonatable group can be placed near the

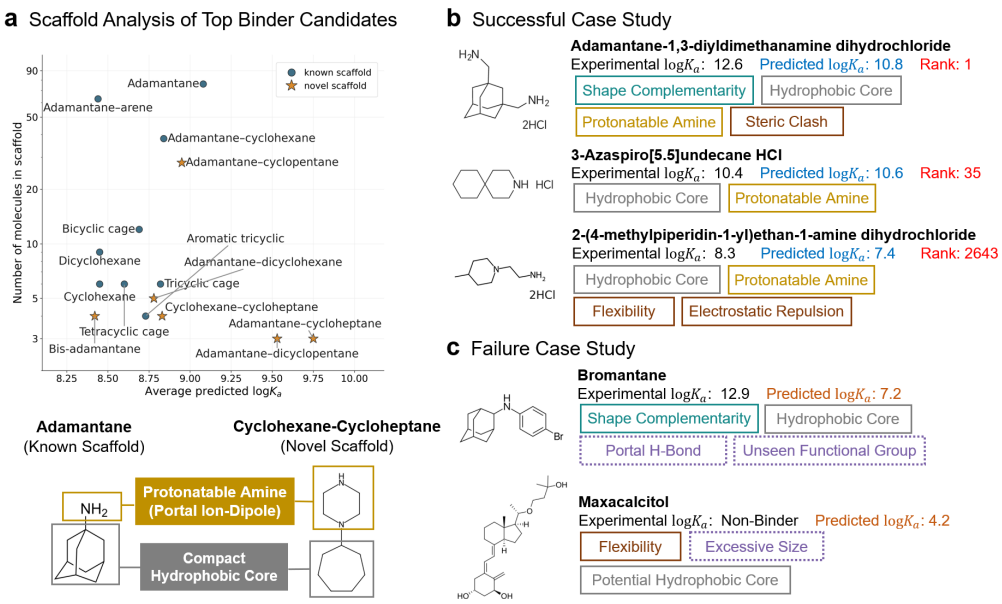


Figure 4: **AI-to-Experiment CB[7] Guest Discovery in GEOM-Drugs.** (a) Prioritized scaffolds among 349 GEOM-Drugs top candidates with predicted $\log K_a > 8$. (b) Three successful case studies demonstrate that CB7Mol can integrate the effect of favorable and harmful determinants. (c) Two failure cases indicate that CB7Mol learns central binding determinants, but not the complete causal chain of CB[7] recognition. Red and dashed boxes indicate the negative determinants and the determinants CB7Mol fails to identify, respectively.

CB[7] portal. Such reduction makes the candidate space substantially more tractable for expert inspection. See Appendix C.1 and C.3 for complete scaffold analysis and comparison with heuristic virtual screening respectively.

Experimental Validation and Case Study. We next validate whether model-prioritized candidates translate into experimentally confirmed CB[7] binders, completing the AI-to-experiment path from large-scale virtual screening to wet-lab affinity measurement. Selected candidates are evaluated by competition NMR (details in Appendix C.5). Figure 4b presents three cases from both known and novel scaffolds with predicted affinities reasonably close to the experimentally measured values. These cases suggest that CB7Mol can identify both favorable binding determinants (e.g., protonatable groups and cavity complementarity) and unfavorable determinants (e.g., geometric and conformational penalties), balancing their competing contributions instead of simply measuring scaffold similarity. Figure 4c presents two failure cases, where CB7Mol underestimates a guest involving an uncommon binding mechanism, while overestimating a non-binder for which favorable features cannot overcome the prerequisite of cavity compatibility. The findings suggest that CB7Mol learns *central determinants* of CB[7] recognition but not the complete *causal chain* from binding feasibility to realized affinity. Detailed analysis is provided in Appendix C.4.

5 Conclusion

We introduce CB7Mol, a physicochemically grounded framework for label-scarce CB[7]-based binding affinity prediction. CB7Mol jointly bridges three knowledge gaps through CB[7]-relevant molecular semantics, binding geometry and conformational flexibility, and affinity supervision transferred from mechanistically compatible macrocyclic hosts. Beyond strong predictive performance, our analyses show that its predictions reflect key physicochemical determinants while not implying a complete causal model of recognition. Finally, an AI-to-experiment pipeline screens 304k molecules and validates high-affinity guests by competition NMR, expanding the experimentally supported CB[7] guest space. More broadly, our results suggest that scientific principles can be operationalized as complementary transferable supervision when direct task labels are scarce.

AI USE STATEMENT

In this work, we used generative AI tools for assisting with translation and implementing methods. We have not used generative AI tools for generating synthetic datasets, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, cleaning and reformatting dataset, supporting qualitative and thematic data analysis, or interpreting results. Developing theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, and writing of proofs are not applicable to this work. Additionally, we used generative AI tools for paper polishing and literature summarization. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

All authors have read and adhered to the ICLR Code of Ethics. This work uses scientific data obtained from existing scientific datasets and our own real-world experiments. It does not involve human participants, personal data, or animal subjects. We respect the licenses and access conditions of all data sources, including restrictions on the redistribution of records obtained from licensed databases. The wet-lab validation involves only small-molecule host-guest supramolecular systems and was conducted following applicable laboratory safety procedures.

CB7Mol is intended as a virtual screening tool that prioritizes candidate guests for subsequent expert-supervised experimental evaluation. The predicted binding affinity should not be interpreted as evidence of biological efficacy, toxicity, pharmacokinetic properties, or clinical safety. CB7Mol can not replace established chemical risk assessment.

REPRODUCIBILITY STATEMENT

We have provided the information needed to reproduce the experimental results throughout the main content and appendices, including data settings, baseline settings, CB7Mol hyperparameter settings, chemical analytical frameworks, and competition-NMR validation protocol. We will provide the complete source code and our synthesized Expert dataset upon publication.

REFERENCES

Walid Ahmad, Elana Simon, Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. Chemberta-2: Towards chemical foundation models. *arXiv preprint arXiv:2209.01712*, 2022.

Pablo Ricardo Arantes, Rodrigo Ligabue-Braun, and Conrado Pedebos. ermsf: A python package for ensemble-based rmsf analysis of biomolecular systems. *Journal of chemical information and modeling*, 65(23):12648, 2025.

Khaleel I Assaf and Werner M Nau. Cucurbiturils: from synthesis to high-affinity binding and catalysis. *Chemical Society Reviews*, 44(2):394–418, 2014.

Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific data*, 9(1):185, 2022.

Steven J Barrow, Setu Kasera, Matthew J Rowland, Jesús Del Barrio, and Oren A Scherman. Cucurbituril-based molecular recognition. 2015.

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in neural information processing systems*, 35:11423–11436, 2022.

Guy W Bemis and Mark A Murcko. The properties of known drugs. 1. molecular frameworks. *Journal of medicinal chemistry*, 39(15):2887–2893, 1996.

Frank Biedermann, Vanya D Uzunova, Oren A Scherman, Werner M Nau, and Alfonso De Simone. Release of high-energy water as an essential driving force for the high-affinity binding of cucurbit [n] urils. *Journal of the American Chemical Society*, 134(37):15318–15323, 2012.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Michael Brocidiacono, Konstantin I Popov, and Alexander Tropsha. An improved metric and benchmark for assessing the performance of virtual screening models. *ArXiv*, pp. arXiv–2403, 2024.

Liping Cao, Marina Šekutor, Peter Y Zavalij, Kata Mlinarić-Majerski, Robert Glaser, and Lyle Isaacs. Cucurbit[7]uril · guest pair with an attomolar dissociation constant. *Angewandte Chemie*, 126(4):1006–1011, 2014.

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.

Donald J Cram. Preorganization—from solvents to spherands. *Angewandte Chemie International Edition in English*, 25(12):1039–1057, 1986.

Wenjie Du, Shuai Zhang, Di Wu 0057, Jun Xia, Ziyuan Zhao, Junfeng Fang, and Yang Wang. Mmgnn: A molecular merged graph neural network for explainable solvation free energy prediction. In *IJCAI*, pp. 5808–5816, 2024.

Wenjie Du, Jiahui Zhang, Xuqiang Li, Sihan Wang, Zhengyang Zhou, Hongxin Xiang, Jun Xia, Ye Wei, and Yang Wang. I2mole: Interaction-aware invariant molecular learning for generalizable property prediction. In *International Conference on Learning Representations*, volume 2026, pp. 1710–1739, 2026.

Luis Escobar and Pablo Ballester. Molecular recognition in water using macrocyclic synthetic receptors. *Chemical Reviews*, 121(4):2445–2514, 2021.

Ada Fang, ZAIXI ZHANG, and Marinka Zitnik. Atomica: Learning universal representations of molecular interactions. In *ICML 2025 Generative AI and Biology (GenBio) Workshop*, 2025.

Shikun Feng, Yuyan Ni, Yanyan Lan, Zhi-Ming Ma, and Wei-Ying Ma. Fractional denoising for 3d molecular pre-training. In *International Conference on Machine Learning*, pp. 9938–9961. PMLR, 2023.

Richard A Friesner, Jay L Banks, Robert B Murphy, Thomas A Halgren, Jasna J Klicic, Daniel T Mainz, Matthew P Repasky, Eric H Knoll, Mee Shelley, Jason K Perry, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7):1739–1749, 2004.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. Simcse: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pp. 6894–6910, 2021.

Yanjing Gao, Farbod Shirinichi, Audrey Hansrisuk, Runyao Zhu, Sijie Xian, Marya Lieberman, Matthew J Webber, and Yichun Wang. A supramolecular–quantum dot system for broad-spectrum detection of fentanyl analogs. *Small*, 21(16):2407702, 2025.

Samuel Genheden and Ulf Ryde. The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, 10(5):449–461, 2015.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. Pmlr, 2017.

Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268, 2018.

Duvan Gonzalez, Luis Macaya, Carlos Castillo-Orellana, Toon Verstraelen, Stefan Vogt-Geisse, and Esteban Vohringer-Martinez. Nonbonded force field parameters from minimal basis iterative stockholder partitioning of the molecular electron density improve cb7 host–guest affinity predictions. *Journal of Chemical Information and Modeling*, 62(17):4162–4174, 2022.

Laura M Grimm, Sebastian Spicher, Boryslav Tkachenko, Peter R Schreiner, Stefan Grimme, and Frank Biedermann. The role of packing, dispersion, electrostatics, and solvation in high-affinity complexes of cucurbit [n] urils with uncharged polar guests. *Chemistry–A European Journal*, 28 (38):e202200529, 2022.

Colin R Groom, Ian J Bruno, Matthew P Lightfoot, and Suzanna C Ward. The cambridge structural database. *Structural Science*, 72(2):171–179, 2016.

Dong-Sheng Guo and Yu Liu. Supramolecular chemistry of p-sulfonatocalix [n] arenes and its biological applications. *Accounts of chemical research*, 47(7):1925–1934, 2014.

Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

Rui Jiao, Xiangzhe Kong, Li Zhang, Ziyang Yu, Fangyuan Ren, Wenjuan Tan, Wenbing Huang, and Yang Liu. An equivariant pretrained transformer for unified 3d molecular representation learning. *Nature Communications*, 17(1):2606, 2026.

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

Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. Pubchem 2025 update. *Nucleic acids research*, 53(D1):D1516–D1525, 2025.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Craig Knox, Mike Wilson, Christen M Klinger, Mark Franklin, Eponine Oler, Alex Wilson, Allison Pon, Jordan Cox, Na Eun Chin, Seth A Strawbridge, et al. Drugbank 6.0: the drugbank knowledgebase for 2024. *Nucleic acids research*, 52(D1):D1265–D1275, 2024.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. Generalist equivariant transformer towards 3D molecular interaction learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 25149–25175. PMLR, 21–27 Jul 2024.

Albert Kraft, Connor R Schmidt, Christian Akakpo, Frank Biedermann, Matthew J Webber, and Lyle Isaacs. Cucurbit[n]uril-type molecular containers. *Nature Reviews Methods Primers*, 6(1): 40, 2026.

Namkyeong Lee, Junseok Lee, and Chanyoung Park. Augmentation-free self-supervised learning on graphs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 7372–7380, 2022.

Namkyeong Lee, Dongmin Hyun, Gyoung S Na, Sungwon Kim, Junseok Lee, and Chanyoung Park. Conditional graph information bottleneck for molecular relational learning. In *International Conference on Machine Learning*, pp. 18852–18871. PMLR, 2023a.

Namkyeong Lee, Kanghoon Yoon, Gyoung S Na, Sein Kim, and Chanyoung Park. Shift-robust molecular relational learning with causal substructure. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1200–1212, 2023b.

Jean-Marie Lehn. Supramolecular chemistry—scope and perspectives molecules, supermolecules, and molecular devices (nobel lecture). *Angewandte Chemie International Edition in English*, 27 (1):89–112, 1988.

Daniel S Levine, Muhammed Shuaibi, Evan Walter Clark Spotte-Smith, Michael G Taylor, Muhammad R Hasyim, Kyle Michel, Ilyes Batatia, Gábor Csányi, Misko Dzamba, Peter Eastman, et al. The open molecules 2025 (omol25) dataset, evaluations, and models. *arXiv preprint arXiv:2505.08762*, 2025.

Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. In *International Conference on Learning Representations*, volume 2024, pp. 39282–39309, 2024.

Simin Liu, Christian Ruspic, Pritam Mukhopadhyay, Sriparna Chakrabarti, Peter Y Zavalij, and Lyle Isaacs. The cucurbit [n] uril family: prime components for self-sorting systems. *Journal of the American Chemical Society*, 127(45):15959–15967, 2005.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Tianyi Ma, Yijun Ma, Zehong Wang, Weixiang Sun, Ziming Li, Connor R Schmidt, Chuxu Zhang, Matthew J Webber, and Yanfang Ye. Suprabench: A benchmark for supramolecular chemistry. *arXiv preprint arXiv:2606.13477*, 2026.

Xing Ma and Yanli Zhao. Biomedical applications of supramolecular systems based on host–guest interactions. *Chemical reviews*, 115(15):7794–7839, 2015.

Elio Mattia and Sijbren Otto. Supramolecular systems chemistry. *Nature nanotechnology*, 10(2): 111–119, 2015.

Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023.

Hari S Muddana, Andrew T Fenley, David L Mobley, and Michael K Gilson. The sampl4 host–guest blind prediction challenge: an overview. *Journal of computer-aided molecular design*, 28 (4):305–317, 2014.

Werner M Nau, Mara Florea, and Khaleel I Assaf. Deep inside cucurbiturils: physical properties and volumes of their inner cavity determine the hydrophobic driving force for host–guest complexation. *Israel Journal of Chemistry*, 51(5-6):559–577, 2011.

Yashaswi Pathak, Siddhartha Laghuvarapu, Sarvesh Mehta, and U Deva Priyakumar. Chemically interpretable graph interaction network for prediction of pharmacokinetic properties of drug-like molecules. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 873–880, 2020.

Christoph Plett and Stefan Grimme. Automated and efficient generation of general molecular aggregate structures. *Angewandte Chemie*, 135(4):e202214477, 2023.

Paul Raccuglia, Katherine C Elbert, Philip DF Adler, Casey Falk, Malia B Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A Friedler, Joshua Schrier, and Alexander J Norquist. Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533(7601):73–76, 2016.

Giuseppina Raffaini, Fabio Ganazzoli, Luciana Malpezzi, Claudio Fuganti, Giovanni Fronza, Walter Panzeri, and Andrea Mele. Validating a strategy for molecular dynamics simulations of cyclodextrin inclusion complexes through single-crystal x-ray and nmr experimental data: a case study. *The Journal of Physical Chemistry B*, 113(27):9110–9122, 2009.

Mikhail V Rekharsky and Yoshihisa Inoue. Complexation thermodynamics of cyclodextrins. *Chemical reviews*, 98(5):1875–1918, 1998.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.

Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. *Advances in neural information processing systems*, 33:12559–12571, 2020.

József Szejtli. Introduction and general overview of cyclodextrin chemistry. *Chemical reviews*, 98 (5):1743–1754, 1998.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017.

Christopher GP Taylor, William Cullen, Olivia M Collier, and Michael D Ward. A quantitative study of the effects of guest flexibility on binding inside a coordination cage host. *Chemistry–A European Journal*, 23(1):206–213, 2017.

Pall Thordarson. Determining association constants from titration experiments in supramolecular chemistry. *Chemical Society Reviews*, 40(3):1305–1323, 2011.

Zehong Wang, Yijun Ma, Connor R Schmidt, Tianyi Ma, Weixiang Sun, Ziming Li, Xiaoguang Guo, Chuxu Zhang, Matthew J Webber, and Yanfang Ye. Mdforge: Agentic molecular dynamics pipeline design under sparse simulator feedback. *arXiv preprint arXiv:2606.12916*, 2026a.

Zidi Wang, Tao Zhang, Muyao Yu, Chuyi Zhou, Zezhao Xu, Huiyu Liu, Yuzhen Wen, Linjiang Chen, Jie Zheng, and Shan Jiang. Learning to dock: Geometric deep learning for predicting supramolecular host-guest complexes. *arXiv preprint arXiv:2601.12268*, 2026b.

George T Williams, Cally JE Haynes, Mohamed Fares, Claudia Caltagirone, Jennifer R Hiscock, and Philip A Gale. Advances in applied supramolecular technologies. *Chemical Society Reviews*, 50(4):2737–2763, 2021.

Jun Xia, Chengshuai Zhao, Bozhen Hu, Zhangyang Gao, Cheng Tan, Yue Liu, Siyuan Li, and Stan Z. Li. Mole-BERT: Rethinking pre-training graph neural networks for molecules. In *The Eleventh International Conference on Learning Representations*, 2023.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Kun Yan, Lancheng Wang, Zhihang Zhu, Shiqi Duan, Zhendong Hua, Peng Xu, Hui Xu, Chi Hu, Youmei Wang, and Bin Di. Cucurbituril-protected dual-readout gold nanoclusters for sensitive fentanyl detection. *Analyst*, 148(6):1253–1258, 2023.

Andrew F Zahrt, Jeremy J Henle, Brennan T Rose, Yang Wang, William T Darrow, and Scott E Denmark. Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science*, 363(6424):eaau5631, 2019.

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 *The eleventh international conference on learning representations*, 2023.

Table 2: Ligand-only descriptors composing the two heuristic CB[7] binding scores used to select the post-training molecules; each score is the weighted sum of its components. Symbols: $\mathrm{PC} = V/V_c$ (packing coefficient; V = vdW volume, $V_c = 279\,\text{Å}^3$ = cavity volume); $V_h = V f_h$ (hydrophobic volume); f_h = hydrophobic-atom fraction; f_R = ring-atom fraction; n_r = rotatable bonds; n_+ = cationic centers; $q^+ = \max(0, q_{\max})$ ($q_{\max}$ = max Gasteiger charge); n_a, n_ℓ = aromatic/aliphatic rings; Q = net formal charge; n_{a^-} = anionic groups; MW = molecular weight; σ = logistic; $\mathbb{1}[\cdot]$ = indicator.

Score	Component	Weight	Definition
Physical	Packing	0.30	$\exp[-8(\mathrm{PC}-0.6)^2]$
	Hydrophobicity	0.25	$0.5\min(1, V_h/180) + 0.3 f_h + 0.2\,\sigma(\mathrm{logP}-2)$
	Cationic charge	0.30	$\min(1, n_+/2) \times \min(1, q^+/0.3)$
	Rigidity	0.15	$0.6/(1+n_r) + 0.4\min(1, f_R/0.5)$
Chemical	Positive charge	0.25	$\min(1, n_+/2) \times \min(1, q^+/0.4)$
	Hydrophobicity	0.20	$0.4\min(1, V_h/180) + 0.3 f_h + 0.3\min(1, (n_a+n_\ell)/2)$
	LogP	0.15	$\exp[-0.1(\mathrm{logP}-2.5)^2]$
	Rigidity	0.15	$\exp(-0.3\,n_r)$
	Ring content	0.10	$\min(1, f_R/0.5)$
	Formal charge	0.10	$\mathbb{1}[Q>0] + 0.5\,\mathbb{1}[Q=0]$
	Polarity penalty	−0.15	$\min\big(1, \max(0, (\mathrm{TPSA}-60)/60)\big)$
	Liability penalty	−0.20	$0.4\min(1, n_{a^-}) + 0.2\min(1, \frac{\mathrm{TPSA}}{150}) + 0.2\,\mathbb{1}[n_r>6] + 0.2\,\mathbb{1}[V>V_c \vee \mathrm{MW}>400]$

A Experimental Setup Details

A.1 Dataset Curation

Molecular Encoder Post-Training Corpus. Post-training corpus is drawn from GEOM, which is a subset assembled by a sampling strategy focusing on likely CB[7] binders while preserving chemical diversity. Based on binder-intrinsic features and CB[7] binding mechanisms (Nau et al., 2011; Biedermann et al., 2012), we define two heuristic affinity scores: a physical score and a chemical score, where each is a weighted combination of descriptors that encode the geometric and physicochemical hallmarks of CB[7] recognition. Descriptors are summarized in Table 2. Then, we rank all GEOM molecules by the two scores respectively and retain the top 5% molecules from each ranking list. To retain diversity, we randomly sample additional 5% molecules from GEOM. The three 5% subsets are deduplicated against one another, so we have the final 15% molecule subset from GEOM. For each selected molecule, we treat it as candidate guests of CB[7] and compute their 3D structures by aISS docking, integrating the top-scored guest poses into post-training corpus.

Pose-Aware Complex Encoder Pre-Training Corpus. We adopt the crystallographic complex corpus from DeepHostGuest (Wang et al., 2026b) for complex encoder pre-training. This corpus is a set of 1,499 host-guest inclusion complexes curated from the Cambridge Structural Database (CSD), each taken in its experimental crystal frame (Groom et al., 2016). The raw crystallographic entries in CSD were parsed to identify their molecular components, which are extracted and classified as host or guest. Geometric and intermolecular criteria are then applied to retain genuine inclusion complexes that are compatible with CB[7] complexes, followed by manual curation to remove artifacts and misassignments.

Expert Dataset Curation. The molecules annotated by experts are filtered out by MDForge (Wang et al., 2026a) over DrugBank (Knox et al., 2024). We first rank candidate molecules based on their physically simulated CB[7] binding affinity. Then, we retain 60 molecules from the top of this ranking that domain experts judged different from known CB[7] guests and valuable for new guest discovery. These affinity labels are annotated by the experts and used as an independent test set.

Molecule Distributions of Evaluation Sets. The three evaluation sets probe significantly different regions of chemical space. The 125 SupraBench CB[7] guests are dominated by small, common CB[7] guest chemotypes such as cyclohexane and cyclopentane. The 60 Expert guests, by contrast, are entirely cyclic and skew toward bulkier fused or linked polycyclic frameworks (e.g., cyclohexyl-cyclohexane, decalin- and adamantane-linked cages, and steroids), which hardly exist in SupraBench. Moreover, SAMPL4 is featured by graded cycloalkane/bicycloalkane series. It is notable that cyclohexane, norbornane, and adamantane are shared across all three sets and represent privileged rigid hydrophobic cores for CB[7] guests.

Data Preprocess. For every host-guest pair used in calibration and evaluation, the 3D guest structure is generated by aISS docking, and the top-scored pose is retained. Macrocyclic host structures are all retrieved from PubChem (Kim et al., 2025). To obtain the per-atom conformational-flexibility channel supplied to the complex encoder, we rebuild the covalent bond graph from the stored 3D molecular geometry, embed a 15-conformer ETKDGv3 ensemble, relax it with MMFF, superpose the conformers on their heavy atoms, and take each heavy atom's RMSF about its mean position. Deduplication is applied to all the datasets involved in CB7Mol to avoid data leakage.

Data Split. 5-fold cross-validation is adopted for affinity prediction evaluation. Non-CB[7] complexes are fixed in training set, while the 125 CB[7] complexes are partitioned into five random folds. In each round, one CB[7] fold is held out for validation, and the remaining four are combined with the non-CB[7] complexes for training, such that every CB[7] fold serves as the validation set for once. We do not adopt scaffold split given limited affinity labels and the existence of two out-of-distribution evaluation sets (Expert and SAMPL4).

A.2 BASELINES

(i) Physics-Driven Methods

- **aISS** (Plett & Grimme, 2023): The automated Interaction Site Screening docking method, a fast semi-empirical (GFN-FF/`xtb`) protocol that screens candidate binding sites and scores poses by their computed interaction energy; we directly rank guests by the predicted CB[7] binding energy.
- **QSAR (RF, XGB)**: A random forest (Breiman, 2001) and an XGBoost (Chen & Guestrin, 2016) model trained to predict binding affinity from guest-side CB[7]-specific physicochemical descriptors listed in Table 2. These descriptors are concatenated with Morgan fingerprints of the guest to form model inputs. QSAR is guest-only and trained on CB[7] data alone.

(ii) Graph Neural Networks. GNNs are designed to encode host and guest molecules separately, followed by a global readout of the representations from both sides for affinity regression.

- **GCN** (Kipf & Welling, 2017): A graph convolutional network that aggregates information over first-order spectral neighborhoods.
- **GIN** (Xu et al., 2018): Graph isomorphism network, a maximally expressive (Weisfeiler-Lehman-equivalent) message-passing GNN.
- **MPNN** (Gilmer et al., 2017): A general message-passing framework with edge-conditioned messages, originally introduced for molecular property prediction.

(iii) Molecular relational learning. These methods model the interaction between input molecule pairs explicitly. Most of them are originally designed for solute-solvent solvation free-energy prediction and drug-drug interaction prediction, and we adapt each to host-guest affinity regression.

- **CIGIN** (Pathak et al., 2020): A chemically interpretable graph interaction network that learns an atom-level cross-molecular interaction map between the two molecules.
- **CGIB** (Lee et al., 2023a): A conditional graph information-bottleneck model designed to extract the molecular substructures most relevant to interaction of the input molecule pair.
- **CMRL** (Lee et al., 2023b): A shift-robust molecular relational learner that isolates a causal interacting substructure to remain robust under distribution shift.
- **MMGNN** (Du et al., 2024): A molecular GNN that forms connections between the two molecules' atoms on a merged graph and refines them with attention-based aggregation, using node/edge interpreters to surface the interaction-relevant subgraph.
- **I2Mole** (Du et al., 2026): An interaction-aware invariant molecular learner that models intermolecular interactions on a merged graph and learns interaction representations via an improved graph-information-bottleneck with an environment codebook. The interaction representations are invariant across environments and generalizable.

Table 3: Hyperparameter settings for three training stages of CB7Mol.

Stage	Hyperparameter	Value
Molecular Encoder Post-Training	Learning rate	1×10^{-4}
	Weight decay	1×10^{-4}
	Batch size	16
	EMA decay	$0.996 \rightarrow 0.9999$ (ramped)
Pose-Aware Pre-Training	Pair encoder (dim/hidden/blocks/dropout)	64 / 128 / 3 / 0.1
	MDN components K	10
	Learning rate	1×10^{-3}
	Weight decay	1×10^{-4}
	Gradient accumulation	16
Affinity Calibration	Affinity head	LN $\rightarrow 64\times128 \rightarrow$ ReLU $\rightarrow$ dropout $0.1 \rightarrow 1$
	Learning rate	1×10^{-4}
	Weight decay	1×10^{-4}
	Gradient accumulation	8
	Gradient-norm clip	10

(iv) Host-guest complex encoders. These foundation models encode the full host-guest complex directly and are fine-tuned for affinity regression.

- **EPT** (Jiao et al., 2026): An all-atom SE(3)-equivariant pretrained transformer that jointly encodes every atom of the bound complex.
- **GET** (Kong et al., 2024): A generalist equivariant transformer for 3D molecular interaction that represents the complex as sets of atom blocks and models their geometric interactions.

Calibration Data Setting. For guest-only baselines, only CB[7] complexes are applied for training and validation. Otherwise, the cross-host calibration is also applied to the eligible baselines with the same data split as CB7Mol.

A.3 Evaluation Metrics

In virtual screening, accurate absolute affinities allow a physically meaningful cutoff, while a faithful ranking determines how efficiently true binders can be advanced to costly experimental validation (Brocidiacono et al., 2024; Wang et al., 2026a). Accurate affinity prediction and faithful candidate ranking are both important. We therefore adopt two rank-correlation metrics as our primary measures and report an absolute-error metric for calibration. Let $\{y_i\}_{i=1}^N$ and $\{\hat{y}_i\}_{i=1}^N$ denote the reference and predicted $\log K_a$ over a set of N guests.

- **Spearman's** ρ: the Pearson correlation between the rank variables of $\hat{y}$ and y, measuring how well the predicted *ordering* of guests matches the experimental ordering.
- **Kendall's** τ: $\tau = (C - D)/\binom{N}{2}$, where C and D are the numbers of concordant and discordant guest pairs. This provides a pairwise-agreement view of ranking quality.
- **RMSE**: $\sqrt{\frac{1}{N}\sum_i(\hat{y}_i - y_i)^2}$, the absolute prediction error in $\log K_a$ units.

A.4 Implementation Details

All models are implemented in PyTorch 2.8, trained on a single NVIDIA A40 GPU and optimized with AdamW (Loshchilov & Hutter, 2019). $\lambda = 0.3$ is set for $\mathcal{L}_{\text{post}}$. Other hyperparameter settings are summarized in Table 3.

A.5 Ablation Study Setup

CB7Mol Component ablation: *w/o chem-align* and *w/o SD* remove the chemical knowledge alignment and the self-distillation task from molecular encoder post-training respectively; *w/o RMSF*

removes the per-atom RMSF channel from the complex encoder input; *w/o pose* skips the binding geometry pre-training of the complex encoder; *w/o cross-host* performs binding affinity calibration on CB[7] data alone.

Post-training objective ablation: We replace molecule-level self-distillation and chemical representation alignment with atom masking (*w/ atom-mask*) and per-atom alignment (*w/ atom-chem-align*) respectively. More specifically, in *w/ atom-mask*, 15% of the atom tokens are masked on the online branch, and the online per-atom representations at the masked positions are regressed onto the EMA teacher's per-atom representations at the same positions. In *w/ atom-chem-align*, each heavy-atom representation is regressed onto chemical target's corresponding per-atom contextual embedding. Both thus change only the granularity of a single post-training objective.

A.6 Alternative Molecular Encoders

- **Uni-Mol** (Zhou et al., 2023): a 3D molecular representation Transformer pretrained on $\sim$209M conformers with masked-atom prediction and 3D coordinate denoising. We take its per-heavy-atom atomic representations for downstream complex modeling.
- **Frad** (Feng et al., 2023): A 3D pretraining method whose fractional denoising objective decomposes atomic noise into a dihedral-angle and a coordinate component, yielding a closer approximation to the molecular force field than plain coordinate denoising. We use the released QM9-pretrained model and take its per-heavy-atom representation.
- **MACE-OMol** (Batatia et al., 2022; Levine et al., 2025): A higher-order equivariant message-passing network used as a machine-learning interatomic potential. We use the MACE-OMol foundation model (the extra-large checkpoint trained on Open Molecules 2025) and extract the invariant part as per-atom descriptor.
- **EquiformerV2** (Liao et al., 2024): An equivariant Transformer that uses $\mathrm{SO}(2)$-based attention to scale to high-degree equivariant features. We use the released 31M-parameter checkpoint trained on OC20 S2EF and take the invariant part of its per-atom node embeddings.

A.7 Alternative Chemical Alignment Targets

- **GROVER** (Rong et al., 2020): A self-supervised graph Transformer pretrained on $\sim$10M molecules with contextual-property and motif-prediction tasks; we use its molecule embedding as chemistry-aware alignment target.
- **ECFP** (Rogers & Hahn, 2010): Rule-based circular-substructure fingerprints; we use a 2048-bit radius-2 (ECFP4) fingerprint directly as the target.
- **ChemBERTa** (Ahmad et al., 2022): a RoBERTa-style masked language model over SMILES strings. We use the released `DeepChem/ChemBERTa-77M-MLM` checkpoint and mean-pool its final-layer token embeddings as alignment target.
- **Mole-BERT** (Xia et al., 2023): a 2D graph model that pretrains a GIN with a context-aware and tokenizer-based masked-atom objective. We use the released GIN checkpoint and global-mean-pool its node embeddings as alignment target.

A.8 Functional-Group Decoding Setup

We use RDKit SMARTS fragments to represent functional groups, from which we retain only those distinctive to CB[7] guests. More specifically, we keep the functional groups enriched among the CB[7] guests relative to the general molecular background, and exclude ubiquitous drug-like motifs. For each retained group, we cast decoding as binary classification (presence vs. absence) and fit a simple ridge-regression linear probe (Hoerl & Kennard, 1970) on top of molecular encoder's mean-pooled frozen atom embeddings. Decodability reflects information already present in the molecular representation. The ridge-regression is trained on GEOM molecules and evaluated on all evaluation CB[7] guests without fold split. We report ROC-AUC for the encoder before and after post-training, whose gap measures the functional-group knowledge that post-training injects.

A.9 Scaffold Definition

Considering that CB[7] recognition is governed by shape and size complementarity, we group the guest candidates by a shape-oriented scaffold representation. For each molecule we first extract its Bemis-Murcko scaffold (Bemis & Murcko, 1996), retaining only the ring systems and the linkers that connect them. We then decompose the scaffold into ring systems (maximal sets of fused rings) and represent the molecule by the multiset of these ring systems, discarding the linkers. This merges guests that share the same rings but differ only in how those rings are linked. Each ring system is abstracted to its shape by ignoring heteroatoms but preserving whether it is aromatic or saturated, and is then labeled as the adamantane cage, by its ring size (for a single ring), or by its number of fused rings (for a polycyclic cage).Two molecules fall in the same scaffold class when their ring-system multisets are identical. Finally, to assess scaffold novelty, we compute the same representation for all experimentally known CB[7] guests from SupraBench and SAMPL4 to identify absent compositions.

B Additional Experimental Results

B.1 Docking Pose Assessment

We reconstruct complex structures with MDN prediction in CSD validation set and all three affinity evaluation sets. For each host-atom/guest-atom pair (i, j), the pre-trained MDN head outputs a K-component Gaussian mixture over their interatomic distance, which we summarize by its expected value $\hat{d}_{ij} = \sum_{k=1}^{K} \pi_{ijk}\, \mu_{ijk}$. Given that the bond structure of guest and the CB[7] structure are both fixed, we can recover the guest structure from the expected interatomic distances.

We adopt experimentally validated structures in CSD (Wang et al., 2026b) and aISS-docked CB[7]-guest complex structures as reference labels. The evaluation is two-fold: For quantitative evaluation, we report distance RMSE and the Spearman correlation ρ between predicted and reference contact distances to reflect geometric fidelity and the consistency of identified contact interface. However, since the aISS-docked structure is itself a computational approximation rather than an experimental ground truth, we complement our analysis with qualitative evaluation in visualized case study. Given that CB[7] is a rigid and approximately symmetric host, a reconstructed pose may undergo a moderate translation or rotation while preserving an equivalent binding mode. We therefore assess pose quality primarily through its chemically relevant contact topology. First, the guest's hydrophobic core should be inserted into the CB[7] cavity at a physically plausible depth, providing effective cavity filling without penetrating the host wall. Second, polar or protonated groups should remain near the carbonyl-lined portals and accessible to solvent, where they can form ion-dipole forces. Third, the pose should be sterically and conformationally feasible, without severe host-guest atomic overlap, unrealistically short non-bonded contacts, or strongly distorted guest geometry. In Figure 3b, the reconstructed pose is displaced and rotated relative to the aISS pose, but it retains the central cavity inclusion and portal-directed orientation of the guest. 4-*tert*-butyl-phenylalanine contains a 4-*tert*-butylphenyl rigid hydrophobic core and a protonated α-ammonium, where the hydrophobic core is inserted into the cavity while the cationic anchor sits at the carbonyl portal. It can therefore represent a chemically plausible binding geometry, demonstrating that pre-trained MDN has learned non-trivial binding patterns.

B.2 Additional Cross-Host Transfer Analysis

The non-CB[7] hosts used for calibration share CB[7]-relevant recognition determinants to differing degrees. The cyclodextrins (α-, β-, γ-CD) are rigid macrocycles that bind through hydrophobic cavity inclusion and differ mainly in cavity size (Rekharsky & Inoue, 1998; Szejtli, 1998): β-CD's cavity is closest to that of CB[7] and accommodates the same adamantane-like fillers, whereas α-CD is too small and γ-CD too large for CB[7]-type guests. p-Sulfonatocalix[4]arene (p-SC4) pairs a hydrophobic aromatic cavity with an anionic sulfonate rim that attracts cations (Guo & Liu, 2014), mirroring CB[7] recognition mechanisms with its anionic rim acting as the electrostatic analogue of the carbonyl portal. CB[5] shares CB[7]'s exact cucurbituril chemistry with carbonyl portals and a hydrophobic cavity, but its much smaller cavity admits only very small guests (Barrow et al., 2015).

Based on our empirical verification, the transferable binding signal is strongest from hosts that share multiple CB[7]-relevant recognition determinants and a compatible cavity size. Removing β-CD or p-SC4 causes the largest performance degradation, consistent with their greater similarity to CB[7]. In contrast, the removal of other hosts has little impact on the guest ranking quality. Cross-host supervision is most useful when the source host lies within the physicochemical validity range of the transferred affinity signal.

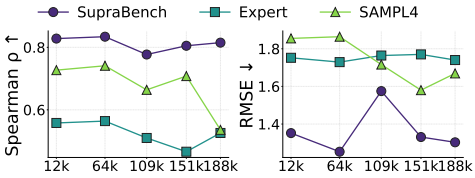


Figure 5: Data scaling analysis. We visualize Spearman ρ and RMSE CB[7]-guest binding affinity prediction results on all three evaluation sets under different post-training corpus scale. Enlarging the corpus yields no consistent gain.

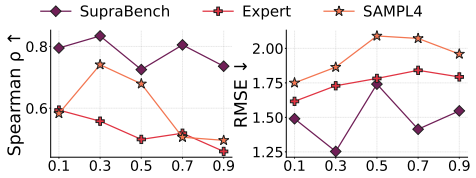


Figure 6: Chemical knowledge alignment coefficient λ analysis. We visualize Spearman ρ and RMSE CB[7]-guest binding affinity prediction results on different λ values.

B.3 Data Scaling.

To verify whether the generalizability of CB7Mol is bottlenecked by the scale of injected knowledge through molecular encoder post-training, we progressively enlarge the post-training corpus with additional unlabeled molecules from GEOM. We follow the same data sampling strategy as post-training corpus construction and adjust the sampling ratio for different corpus scale. Performance evolving trend is visualized in Figure 5, where we can find that enlarging the post-training corpus yields no consistent gain. All three ρ curves peak at 64k corpus size and then flatten or decline, where SAMPL4 suffers significant degradation after corpus size reaches 188k; RMSE curves show similar patterns, but larger corpus can help with the predictive accuracy in SAMPL4. To this end, CB7Mol is not limited by post-training data volume but by the scarcity of downstream binding-affinity labels. Learning grounded molecular representations is not as data-consuming as general foundation models.

B.4 Hyperparameter Analysis

To further evaluate the trade-off between self-distillation and chemical knowledge alignment during molecular encoder post-training, we sweep $\lambda \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and visualize Spearman's ρ and RMSE changes on all three evaluation sets in Figure 6. Performance degrades markedly for $\lambda \geq 0.5$ on both metrics, indicating that over-weighting the chemical target is not helpful, and geometrical information is more critical for CB[7]-based binding affinity prediction. $\lambda = 0.1$ yields the best absolute accuracy on Expert and SAMPL4, whereas $\lambda = 0.3$ yields the best ranking quality on SupraBench and SAMPL4, which is more important for faithful guest ranking.

C Additional Virtual Screening and Guest Discovery Analysis

C.1 Complete Scaffold Analysis over Strong Candidates

Table 4 lists all 21 generic-scaffold classes shared by at least three $\log K_a > 8$ candidates. The recovered scaffolds are overwhelmingly rigid, size-complementary saturated hydrocarbons, consistent with the shape-driven recognition of CB[7]. The distribution is dominated by the adamantane family: the Adamantane class alone accounts for 76 candidates, and together with Adamantane-arene, Adamantane-cyclohexane and Adamantane-cyclopentane. The next tier comprises the simple saturated cages, including bicyclic, tricyclic and tetracyclic cages. Aromatic or heteroatom-rich ring systems are essentially absent from the top of the list.

Table 4: Generic scaffolds of the predicted strong binder candidates ($\log K_a > 8$, 349 molecules) from GEOM-Drugs, listing all scaffolds shared by $n \geq 3$ candidates. Generic scaffolds with the same ring-system composition but different linkers are merged. Avg. $\log K_a$ is the mean predicted affinity over molecules sharing the scaffold, and AffRank is the scaffold ranking by this average among scaffolds shared by at least three candidates.

#	Generic Scaffold	SMILES	n	Avg. $\log K_a$	AffRank
1	Adamantane	C1C2CC3CC1CC(C2)C3	76	9.08	3
2	Adamantane–arene	c1ccc(C23CC4CC(CC(C4)C2)C3)cc1	63	8.44	14
3	Adamantane–cyclohexane	CC(C1CCCCC1)C12CC3CC(CC(C3)C1)C2	38	8.84	5
4	Adamantane–cyclopentane	CC(C1CCCC1)C12CC3CC(CC(C3)C1)C2	28	8.95	4
5	Bicyclic cage	C1CC2CCC1C2	12	8.69	10
6	Dicyclohexane	C1CCC(CC2CCCCC2)CC1	9	8.45	13
7	Dicyclohexane–arene	CC(Cc1cccc(CC(C)C2CCCCC2)c1)C1CCCCC1	7	8.18	21
8	Tricyclic cage	C1CCC2C(C1)CCC1CCCC12	6	8.82	7
9	Tetracyclic cage	C1CC[C@@H]2C(C1)C[C@H]1C[C@@H]2CC2CCCC[C@@H]21	6	8.60	11
10	Cyclohexane	C1CCCCC1	6	8.45	12
11	Tricyclohexane	CC(C1CCCCC1)C1CCC(C(C)C2CCCCC2)CC1	6	8.37	16
12	Adamantane–dicyclohexane	CC(CC12CC3CC(CC(C3)C1)C2)C1CCC(C2CCCCC2)CC1	5	8.78	8
13	Cyclohexane–cyclopentane	CC(CC1CCCCC1)CC1CCCC1	5	8.19	18
14	Cyclohexane–cycloheptane	C1CCCC(C2CCCCC2)CC1	4	8.83	6
15	Aromatic tricyclic	c1ccc2c(c1)CCC1(CCCCC1)C2	4	8.73	9
16	Bis-adamantane	CC(CC12CC3CC(CC(C3)C1)C2)CC12CC3CC(CC(C3)C1)C2	4	8.42	15
17	Adamantane–cycloheptane	C1CCCC(CC23CC4CC(CC(C4)C2)C3)CC1	3	9.75	1
18	Adamantane–dicyclopentane	CC(CCC12CC3CC(C1)CC(CCC(C)C1CCCC1)(C3)C2)C1CCCC1	3	9.53	2
19	Adamantane–biaryl	CC(C1CCc2ccccc21)C12CC3CC(CC(C3)C1)C2	3	8.24	17
20	Cyclohexane–bicyclic cage	C1CCC(C2CCC3CCC2C3)CC1	3	8.19	19
21	Adamantane–cyclohexane–arene	CC(c1cccs1)C1CCC(C23CC4CC(CC(C4)C2)C3)CC1	3	8.18	20

Most recurrent scaffolds are the expected CB[7] chemotypes among experimentally known guests, including adamantane, the bi-/tri-/tetra-cyclic cages, and the di-/tri-cyclohexanes. This confirms that the CB7Mol ranking recovers established binders. Nine scaffolds are absent from these known guests and are flagged as novel. Seven are new recombinations built on the privileged adamantane core (adamantane linked to a cyclopentane, a cycloheptane, a second adamantane, a biaryl, or additional cyclohexane/cyclopentane rings); the remaining two are non-adamantane scaffolds, including a dicyclohexyl-arene and a cyclohexane-cycloheptane bicyclic, where the rigid hydrophobic core is assembled from fused or linked saturated rings rather than an adamantane cage. Both non-adamantane novel scaffolds are predicted to have lower affinity ranking due to shape non-complementarity.

C.2 Physicochemical and Functional-Group Profile of Strong Candidates

To characterize the candidates beyond their ring scaffolds, we profile the functional groups and physicochemical properties of the 349 strong candidates in Table 5. Consistent with the CB[7] cavity, the strong candidates are mostly small, rigid and highly aliphatic with median molecular weight 265, median fraction of sp^3 carbons 0.90 and a median of only two rotatable bonds. They can fall into two established CB[7] binding modes: about half carry a protonatable or cationic nitrogen (a primary/secondary/tertiary amine or ammonium) that can anchor at the carbonyl portals through ion-dipole interactions, while the remaining candidates are neutral hydrophobic cages that bind through cavity filling and the hydrophobic effect alone, as adamantane itself does. Groups unfavorable for CB[7] like anionic carboxylates are rare. Together with the scaffold analysis, this profile substantiates that the top-ranked guests incorporate a rigid hydrophobic cage with, in roughly half of the cases, a portal-anchoring cationic group.

C.3 Comparison with Heuristic Virtual Screening

The physical and chemical scores defined in Table 2 are used for post-training corpus filtering, and they are also cheap and heuristic proxies for CB[7]-guest binding affinity. This raises a natural concern: has CB7Mol merely internalized these heuristics, or has it acquired a more comprehensive understanding of CB[7]-guest complexes from additional supervisions? We answer this by contrasting the CB7Mol ranked top scaffolds with those ranked by heuristic physical and chemical scores. We use the same generic scaffold definition in Appendix A.9, and the top-10 scaffolds extracted from top-300 molecules ranked by heuristic scores are listed in Table 6. The top-ranked molecules from heuristic scoring are dominated by flat and fused aromatic (hetero)cycles, especially benzene and benzimidazole. However, the flat aromatic plate is not complementary with CB[7] cavity in

Table 5: Peripheral functional groups and physicochemical profiles of the 349 GEOM-Drugs strong candidates with predicted $\log K_a > 8$. Under *Functional groups*, we organize the peripheral chemical motifs into three ionization classes based on the charge they confer. Each candidate is then assigned to one of four classes based on chemical properties (cationic, anionic, neutral with polar groups, or apolar). Under *Physicochemical properties*, we report median values of all involved candidates with inter-quartile range for each property.

Descriptor	Value
Functional groups (% of candidates)	
Cationic / basic (portal ion–dipole anchor)	**51%**
protonated ammonium ($-\mathrm{NH}_n^+$)	28%
tertiary amine	19%
secondary amine	7%
primary amine	5%
guanidine / amidine	3%
Anionic (unfavorable for CB[7])	**3%**
carboxylic acid / carboxylate	3%
Neutral (no ionizable center)	**47%**
Non-ionizable polar (across all classes)	
amide	53%
hydroxyl	13%
ether	5%
Physicochemical properties (median [IQR])	
molecular weight (Da)	265 [226–298]
heavy atoms	19 [16–21]
rings	4 [3–5]
rotatable bonds	2 [1–4]
fraction of sp^3 carbons	0.90 [0.72–0.94]
cLogP	2.6 [1.8–3.4]

Table 6: Top-10 generic scaffolds among the top-300 GEOM molecules ranked by the physical and chemical heuristic scores respectively. n: number of molecules.

Physical score			**Chemical score**		
n	Generic Scaffold	SMILES	n	Generic Scaffold	SMILES
55	Benzene	`c1ccccc1`	20	Benzene	`c1ccccc1`
16	Pyridine	`c1ccncc1`	9	Quinoline	`c1ccc2ncccc2c1`
10	Benzimidazole	`c1ccc2ncnc2c1`	6	Benzimidazole	`c1ccc2ncnc2c1`
8	Quinoline	`c1ccc2ncccc2c1`	6	Benzimidazo[1,2-a]pyrimidine	`c1ccc2c(c1)NC1=NCCCN12`
4	2,1,3-Benzoxadiazole	`c1ccc2nonc2c1`	5	1,2,3,4-Tetrahydroquinoline	`c1ccc2c(c1)CCCN2`
3	Acridine	`c1ccc2nc3ccccc3cc2c1`	4	Pyridine	`c1ccncc1`
3	Pyrimidine	`c1cncnc1`	4	Thiadiazino[3,2-a]benzimidazole	`C1=Nn2c(nc3ccccc32)SC1`
3	Quinoxaline	`c1ccc2nccnc2c1`	4	Benzothiazole	`c1ccc2scnc2c1`
3	Pyrrolo[1,2-a]quinoxaline	`c1ccc2c(c1)NCC1CCCN21`	4	5,6-Dihydroimidazo[2,1-b]thiazole	`C1=CN2CCN=C2S1`
3	2,3-Dihydroperimidine	`c1cc2c3c(cccc3c1)NCN2`	4	Piperazine	`C1CNCCN1`

shape. In contrast, the top-ranked molecules by CB7Mol are dominated by adamantane and its derivatives, which have the optimal hydrophobic core to fit in CB[7] cavity. This supports that the transferred supervision helps CB7Mol prioritize candidates consistent with established CB[7] recognition determinants beyond the input heuristics alone.

C.4 Detailed Analysis for Case Study

Adamantane-1,3-diyldimethanamine dihydrochloride. This guest was chosen for experimental testing to examine how well CB7Mol could predict a well known scaffold with unfavorable aspects field experts expected to diminish binding affinity. CB7Mol predicts $\log K_a$ of 10.8 and ranks it as the top-1 binder, which is fairly close to the experimental average binding affinity of 12.6. This guest's high affinity can be attributed to the combination of a rigid adamantane core and two protonatable amines, which are favorable motifs for CB[7] recognition. However, this guest adopts a non-axial 1,3-disubstitution pattern, in which the two aminomethyl arms project from different di-

rections of the rigid adamantane framework. When the adamantane core is deeply included, both ammonium groups may therefore be unable to occupy optimal, low-strain positions at the CB[7] portals simultaneously. The second arm may instead crowd the cavity entrance, reducing packing complementarity and partially offsetting the favorable contribution of the additional cationic group. This explains why the guest binds strongly but does not reach the affinity of optimally fitted mono-substituted 1-adamantylammonium Liu et al. (2005).

3-Azaspiro[5.5]undecane hydrochloride. This guest was chosen for experimental testing to examine if CB7Mol could successfully identify a novel scaffold with high binding affinity. For 3-azaspiro[5.5]undecane hydrochloride, CB7Mol predicts $\log K_a$ of 10.6, very close to the experimental average of 10.4. The guest combines a hydrophobic bicyclic framework with a protonatable ring nitrogen, supporting cavity occupation and portal anchoring respectively. However, its elongated spirocyclic framework is less compact and isotropic than privileged cage-like guests such as adamantane, resulting in less optimal shape complementarity with the CB[7] cavity. CB7Mol nevertheless recognizes that the favorable hydrophobic and electrostatic contributions are sufficient to produce strong binding while assigning a lower affinity than would be expected for an optimally packed cage guest.

2-(4-Methylpiperidin-1-yl)ethan-1-amine dihydrochloride. This guest was chosen for experimental testing to examine if CB7Mol could predict a medium affinity guest rather than just very high affinity binders. CB7Mol predicts $\log K_a$ of 7.4 for this guest, reasonably close to the experimental average of 8.3. Although the piperidine ring provides a potential hydrophobic core, the guest remains substantially more flexible than the rigid cage-like binders. Moreover, both the piperidine nitrogen and the terminal primary amine can be protonated, while their short ethylene separation prevents them from simultaneously anchoring the two opposite CB[7] portals. In a binding-compatible geometry, the two cationic centers may instead remain close to the same portal, where they compete for favorable interactions with the portal carbonyls and experience intramolecular electrostatic repulsion. This would introduce an unfavorable desolvation penalty. Together, these effects limit the benefit of its protonatable groups and explain why CB7Mol correctly assigns it a substantially lower affinity than the rigid, compact strong binders.

Bromantane. This guest was chosen for experimental testing because CB7Mol predicted a far lower affinity than field experts would predict, suggesting there may be a limitation of the model. CB7Mol predicts $\log K_a$ of 7.2, substantially below the experimental value of 12.9. The model identifies its shape-complementary adamantane core and overall hydrophobic character, but its exceptionally strong binding appears to involve features that are uncommon in the affinity labels. In particular, the secondary amino linker may participate in a portal interaction based on hydrogen bonds, while the brominated aromatic group is poorly represented among known CB[7] guests. CB7Mol therefore captures the familiar adamantane-based determinant but fails to translate the unusual binding mode and functional group into the corresponding affinity gain.

Maxacalcitol. CB7Mol overestimates Maxacalcitol, predicting a $\log K_a$ of 4.2 for an experimentally validated non-binder of CB[7]. The molecule contains several hydrophobic substructures that may locally resemble favorable cavity-binding motifs, and the model partially offsets these signals by recognizing its high flexibility. However, Maxacalcitol is excessively large, elongated, and conformationally complex relative to the CB[7] cavity. Its hydrophobic groups therefore cannot form a productive inclusion complex, regardless of their local chemical favorability. This case shows that CB7Mol does not fully enforce cavity compatibility as a prerequisite that must be satisfied before other favorable binding determinants can contribute to affinity.

C.5 Wet Lab Details for Case Study

We validate the predicted binding affinities using ^{1}H competition NMR in D_2O, as shown in Figure 7a. For each candidate guest G, we first compare its free-state spectrum with that obtained after adding CB[7]. Changes in the guest resonances provide evidence of complex formation. We then introduce a reference guest R, whose CB[7] association constant is known, such as *p*-xylylenediamine (benzene-1,4-dimethanamine) for the lower affinity guest and ferrocene-ammonium ((ferrocenylmethyl)trimethylammonium chloride) for the higher affinity guests so as to have a distribution of

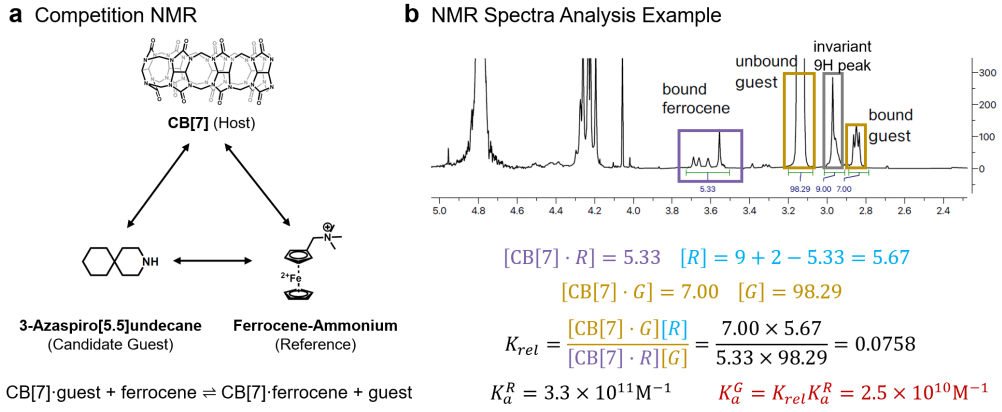


Figure 7: **Competition NMR Demonstration.** (a) A candidate guest G and a reference guest R with known association constant compete for the same CB[7] cavity, allowing their relative affinity to be determined from the equilibrium populations. (b) Representative analysis using 3-azaspiro[5.5]undecane as the candidate and ferrocene-ammonium as the reference. The highlighted resonances are used to estimate the four required relative populations for candidate association constant derivation.

bound and unbound guests for both the guest of interest and the reference guest. Because the candidate and reference guest compete for the same CB[7] cavity, their relative affinity is given by:

$$K_{\text{rel}} = \frac{K_a^G}{K_a^R} = \frac{[\text{CB}[7]\cdot G][R]}{[\text{CB}[7]\cdot R][G]}. \tag{7}$$

where $[\text{CB}[7]\cdot G], [\text{CB}[7]\cdot R], [G], [R]$ represent the estimated relative population of bound guest, bound reference, unbound guest and unbound reference respectively. Figure 7b illustrates how the four relative populations in Equation 7 are obtained in a representative competition experiment. Where the corresponding resonances are sufficiently resolved, these populations are estimated directly from their relative proton-peak integrations. If a resonance overlaps with another signal, the corresponding population is instead inferred from the total population and its complementary resolved population. For example, the free-reference resonance in our example overlaps with the CB[7] signals and is therefore recovered from the total reference population. The invariant 9-proton signal provides an internal scale for interpreting the peak integrations as relative populations. On this scale, the reference signals used to distinguish the free and bound states represent 11 proton equivalents in total, including nine protons from the ferrocene rings and two from the methylene linker. After population estimation, the association constant of the candidate is obtained as:

$$K_a^G = K_{\text{rel}} K_a^R. \tag{8}$$

We repeat the measurement at several candidate-to-reference ratios and report an average affinity as the estimates vary slightly across mixtures. When exchange broadening prevents reliable separation of the free and bound resonances, the ratio of bound to unbound of the guest can be estimated by the quantity of known guest bound relative to total host concentration, assuming guests are strong enough for near quantitative binding of available hosts. In the case of maxacalcitol, no appreciable shifts in the NMR spectrum were observed upon addition of CB[7] and it was unable to displace a competitive guest before hitting the limit of its solubility, leaving no evidence of substantial binding to CB[7].

D Broader Implications of Physicochemical Grounding

CB[7]-based host-guest affinity prediction represents a broader class of scientific discovery problems in which direct task labels are costly or scarce, while relevant physicochemical principles and indirect supervision are comparatively accessible. Similar low-data conditions arise in catalyst and materials discovery, where obtaining target-specific labels often requires costly synthesis, testing,

and characterization (Zahrt et al., 2019; Raccuglia et al., 2016). In such settings, generic model or data scaling alone may be insufficient: broad pre-training objectives optimize coverage of the overall data distribution, whereas specialized scientific targets may depend on narrow physicochemical factors that are weakly represented in generic training data. CB7Mol provides one example of how such scientific knowledge can instead be operationalized through complementary transferable supervision.

Operationalize scientific principles through transferable surrogate supervision. Scientific principles are often too abstract to supervise directly, while a complete physical simulator may be too expensive to integrate into learning. A practical alternative is to decompose the target task into distinct knowledge requirements and construct learnable surrogate objectives for each requirement using richer auxiliary supervision. A useful surrogate should have a clear physicochemical interpretation, an independently measurable objective, and supervision that complements rather than duplicates the target labels. CB7Mol follows this principle by separately acquiring molecular semantics, binding geometry and conformational information, and affinity supervision from different compatible sources.

Match transferable supervision to its physicochemical validity range. Different forms of scientific knowledge transfer across different domains. Generic geometric or molecular knowledge may be learned from broad data distributions, whereas supervision closer to the final prediction target may require source systems that share more specific physicochemical determinants with the target. In CB7Mol, enlarging the generic unlabeled post-training corpus provides no consistent improvement (Appendix B.3), while the contribution of cross-host affinity supervision varies across source hosts. These observations suggest that auxiliary data should be selected not only by scale, but also by whether the transferred supervision remains scientifically valid for the target system.

Physicochemical grounding does not imply causal completeness. Transferable supervision encodes selected determinants of the underlying scientific process rather than the complete causal chain. CB7Mol can therefore fail when a molecule exhibits an unusual interaction pattern not represented by its supervision, or when locally favorable determinants are present but an unmodeled prerequisite such as binding feasibility is violated. Physicochemical grounding should therefore be further enhanced by a complete physical or causal model indicating underlying reasoning process.