Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bilinear cross-attention networks implicitly supervise continuous thermodynamic gradients in supramolecular screening

Chih-Yang Cheng

mushroom2368@gmail.com

Department of Chemistry, National Chung Hsing University, Taichung 402, Taiwan.

Yi-Huan Wu

Department of Chemistry, R. O. C. Military Academy

Feng-Yin Li

Department of Chemistry, National Chung Hsing University

Article

Keywords:

Posted Date: June 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9951273/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Bilinear cross-attention networks implicitly supervise continuous thermodynamic gradients in supramolecular screening

Chih-Yang Cheng[1*], Yi-Huan Wu[2*], Feng-Yin Li[1*]

[1]Department of Chemistry, National Chung Hsing University, Taichung 402, Taiwan.
[2]Department of Chemistry, R. O. C. Military Academy, Kaohsiung, Taiwan.
[*]corresponding author: Chih-Yang Cheng (mushroom2368@gmail.com), Yi-Huan Wu (m1090008@rocma.edu.tw), Feng-Yin Li (feng64@nchu.edu.tw)

Abstract

Predicting non-covalent host-guest binding affinities solely from 2D molecular graphs remains a fundamental challenge in representation learning. Here, we introduce CD-BAN, a dual-branch graph neural network that utilizes a bilinear cross-attention mechanism to explicitly model the structural complementarity of drug-cyclodextrin inclusion complexes, entirely bypassing the need for 3D geometries or handcrafted descriptors. Trained exclusively on a binary classification objective to distinguish thermodynamic extremes strong ($K > 10{,}000\ M^{-1}$) versus weak ($K < 100\ M^{-1}$) binders across 1,198 labeled pairs CD-BAN achieves robust performance (AUROC = 0.925 ± 0.040). Strikingly, the model's latent space spontaneously organizes into a continuous affinity gradient for entirely out-of-distribution data. Evaluated on 1,850 withheld intermediate-affinity compounds ($100 \leq K \leq 10{,}000\ M^{-1}$), the network's output logit yields a highly significant monotonic ranking ($\tau = -0.289, p < 10^{-75}$), demonstrating that discrete binary supervision can implicitly capture continuous thermodynamic principles. Leveraging this emergent property through heuristic boundary calibration, we successfully prioritize 37,000 uncharacterized drug-excipient ternary combinations. Ultimately, this work highlights how attention-based networks can uncover continuous physical laws from discrete empirical limits, providing a highly scalable framework for supramolecular screening.

1. Introduction

Modeling non-covalent interactions in supramolecular chemistry specifically the host-guest binding affinity represents a formidable challenge in molecular representation learning[1,2]. Unlike standard drug-target interactions guided by highly specific lock-and-key protein pockets, host-guest systems like cyclodextrin (CD) inclusion complexes rely on fuzzy, shape-based complementary forces within a hydrophobic cavity[3-7]. While predicting the thermodynamic association constant (K) of these complexes is crucial for overcoming the persistent poor-solubility bottleneck in pharmaceutical development[8,9] particularly in sensitive applications like ophthalmic formulations (eye drops) where non-irritating solubilizers are critical[10-15], reliable computational inference remains elusive.

Current descriptor-based machine learning approaches[1,2,16] often fail to capture the critical three-dimensional spatial matching between the guest molecule and the host cavity. Conversely, physics-based molecular dynamics are computationally intractable for library-scale screening, creating a significant methodological gap. Although graph neural networks (GNNs) have shown immense success in navigating chemical space[1,2,17,18], extracting continuous physical properties strictly from 2D molecular graphs without explicit 3D atomic coordinates remains an open frontier[19-21].

Here, we introduce the Cyclodextrin Bilinear Attention Network (CD-BAN), addressing this representation gap by integrating a dual-branch graph convolutional encoder with a Bilinear Attention Network layer[22,23]. Rather than relying on simple feature concatenation, the bilinear attention mechanism is explicitly designed to model the cross-molecular interaction matrix. This architectural choice allows the network to infer spatial and hydrophobic complementarity directly from 2D SMILES strings. We deploy this architecture on a highly imbalanced dataset of 1,198 drug-CD pairs[24], deliberately restricting the training objective to binary thermodynamic extremes: strong ($K > 10{,}000\ M^{-1}$) versus weak ($K < 100\ M^{-1}$) binders[4,25].

Crucially, this work serves as a methodological proof-of-concept for the independent identification of underlying physical laws by artificial neural networks. Beyond achieving robust classification performance, we uncover an unexpected emergent property: the model's output logit self-organizes into a continuous latent affinity gradient. Evaluated on 1,850 entirely withheld intermediate-affinity compounds spanning the unobserved 100–10,000 M^{-1} range, CD-BAN demonstrates a highly significant monotonic ranking capability (Kendall $\tau = -0.289, p < 10^{-75}$). This reveals a profound principle in deep learning applied to chemistry: models optimized

solely on discrete, extreme thermodynamic boundaries can implicitly learn and interpolate a continuous affinity axis. Finally, leveraging this continuous latent representation, we propose a heuristic boundary calibration and a ternary coformer correction, generating a highly prioritized, independent screening space of 37,000 drug-excipient combinations. Ultimately, CD-BAN highlights the capacity of cross-attention mechanisms to map discrete empirical observations into a thermodynamic continuum.

2. Methods

2.1. Dataset curation and thermodynamic boundary isolation

Data acquisition and rigorous curation. Association constants (K, M^{-1}) for drug–cyclodextrin inclusion complexes were sourced from OpenCycloDB[24], a curated repository of experimentally measured values determined via phase-solubility methods[25]. To prevent ambiguous optimization, we applied a strict de-duplication protocol: 201 pairs exhibiting conflicting K values across multiple literature sources were excluded, retaining a high-confidence dataset of 3,048 unique drug–CD pairs.

Isolation of thermodynamic extremes and OOD manifold. To establish the conceptual framework for the independent identification of affinity gradients, we deliberately isolated the thermodynamic extremes for model training. The binary training manifold was constructed exclusively from pairs with $K < 100\ M^{-1}$ (label 1, Weak; n=1,101) and $K > 10{,}000\ M^{-1}$ (label 0, Strong; n=97). This highly imbalanced binary set (n=1,198) was split into training (70%, n=838), validation (20%, n=240), and test (10%, n=120) sets using stratified random sampling to preserve class ratios. Crucially, the remaining 1,850 pairs occupying the intermediate $100 \leq K \leq 10{,}000 M^{-1}$ range (the "fuzzy zone") were strictly withheld from all training and validation phases. This set served exclusively as an out-of-distribution (OOD) manifold for post-hoc inference evaluation. All inputs relied strictly on 2D SMILES strings; no 3D structural coordinates were provided to the model.

2.2. CD-BAN architecture and bilinear attention formulation

The CD-BAN framework comprises three interconnected modules specifically designed to capture non-covalent spatial complementarity directly from 2D topologies (Fig. 1).

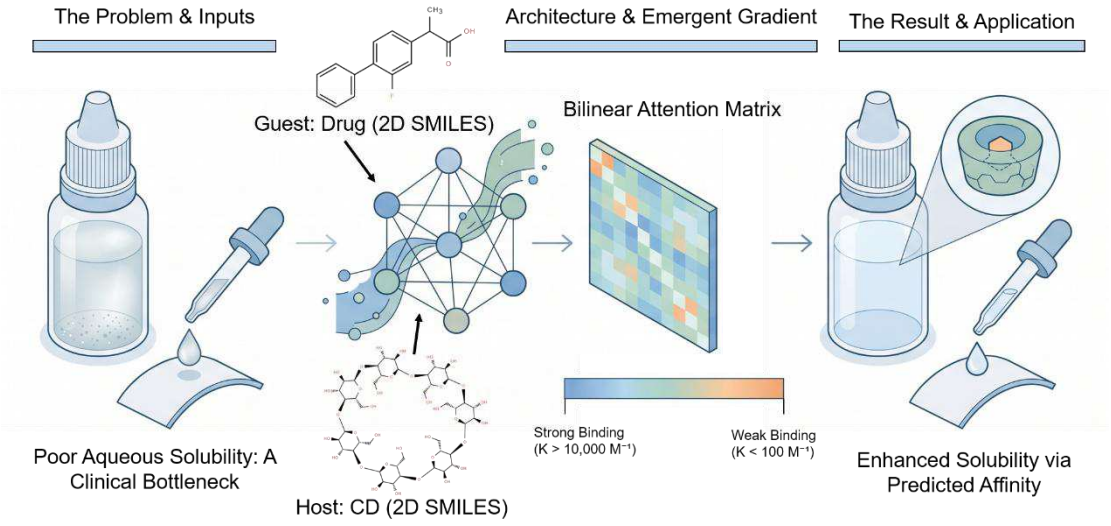


Fig. 1 | CD-BAN schematic overview and application pipeline. The framework addresses the clinical bottleneck of poor aqueous solubility in drug formulations, particularly in ophthalmic applications. Utilizing strictly 2D SMILES strings for both guest (e.g., Flurbiprofen) and host (e.g., $\beta-CD$) molecules, the dual-branch Graph Convolutional Network (GCN) extracts topological features without requiring explicit 3D conformations. The core bilinear attention mechanism computes a cross-molecular interaction matrix, spontaneously outputting a continuous thermodynamic gradient bridging discrete training boundaries ($K < 100\ M^{-1}$ and $K > 10{,}000\ M^{-1}$). This emergent property enables zero-shot inference to identify highly soluble inclusion complexes.

2.2.1. Dual-branch Molecular GCN encoder

The network processes the Guest (drug) and Host (cyclodextrin) SMILES strings into discrete embeddings via two parallel Graph Convolutional Network (GCN) branches[26-28]. Each branch utilizes three successive GCNConv layers (feature dimensions 74→128→128→128) with ReLU activations and global mean pooling. Atom features adhere to the 74-dimensional Chemprop convention[20], encompassing atomic identity, hybridization, formal charge, and chirality tags computed via RDKit[29]. Input graphs are zero-padded to a maximum of 290 nodes.

2.2.2. Bilinear Attention Network (BANLayer).

To circumvent the informational bottleneck of simple feature concatenation, we implemented a Bilinear Attention Network layer[22,23] (2 heads, low-rank factor k=3). This module processes the Guest embedding ($h_G \in \mathbb{R}^{128}$) and Host embedding ($h_H \in \mathbb{R}^{128}$) to compute an explicit cross-molecule interaction matrix. This architectural choice yields a fused representation ($f \in \mathbb{R}^{256}$) that mathematically models the structural and hydrophobic matching between the CD cavity and the guest molecule.

2.2.3. Thermodynamic state decoder.

Finally, a Multi-Layer Perceptron (MLP) decoder (256→512→512→128→1) maps the fused representation *f* to a scalar logit z_{bin}. The output is transformed via a sigmoid function to yield the predicted probability of weak binding, P(Weak):

$$P(\text{Weak}) = \sigma(z_{bin}) = \frac{1}{1 + e^{-z_{bin}}} \quad (1)$$

where $z_{bin} \geq 0$ corresponds to a Weak classification ($K < 100\ M^{-1}$) and $z_{bin} < 0$ denotes a Strong binding affinity ($K > 10{,}000\ M^{-1}$). The entire architecture is trained end-to-end and contains 780,550 trainable parameters.

2.3. Optimization, stability, and environmental impact

Model optimization. The CD-BAN architecture was optimized end-to-end using the Adam optimizer[30] (learning rate 5×10^{-5}, batch size 64) for a maximum of 100 epochs. To counteract the extreme class imbalance inherent to the thermodynamic dataset (1:11.3), the binary cross-entropy objective was augmented with a positive weight scalar (pos_{weight} = 11.32) via the BCEWithLogitsLoss implementation[31]. Model selection was governed by maximizing the validation AUROC. All computational experiments were executed on a single Tesla V100-SXM2 GPU (16 GB) utilizing PyTorch (v2.5.1)[31] and PyTorch Geometric (v2.7.0)[32].

Algorithmic stability and carbon footprint. To ensure rigorous reproducibility and to quantify algorithmic stability against weight initialization variance, we executed ten independent training sequences using fixed random seeds (42–51). The cumulative training overhead was exceptionally low (approximately 3.3 hours), generating an estimated carbon footprint of just ~0.75 kg CO_2 (assuming a 300 W GPU, PUE = 1.5, and a regional grid emission factor of 0.502 kg CO_2/kWh). This highlights the computational efficiency of relying strictly on 2D graph topologies rather than explicit 3D simulations.

2.4. Statistical evaluation metrics

Classification performance on the held-out test set (n=120) was comprehensively assessed using AUROC, AUPRC, F1 score, sensitivity, specificity, and accuracy. Balanced accuracy was explicitly computed to mitigate class-imbalance distortion. All metrics are reported as the mean ± standard deviation across the ten independent initialization seeds. To statistically validate the emergent continuous latent gradient within the unobserved fuzzy zone (n=1,850), we computed the non-parametric Kendall rank correlation coefficient (τ), alongside Pearson's r and Spearman's ρ, between the ground-truth empirical log_{10}K values and the model's predictive probability P(Weak). All statistical tests were performed using standard SciPy implementations[33].

2.5. Heuristic latent projection and ternary exploration

To translate the emergent continuous gradient into an actionable virtual screening tool, we developed a two-step heuristic framework. This approach first maps the abstract latent space back to empirical physical boundaries, and subsequently extrapolates this predictive capacity to unobserved multicomponent formulation systems.

2.5.1. Ordinal boundary projection

Leveraging the empirical boundary conditions established by the binary classifier ($z_{bin} = +1.458 \rightarrow \mathrm{K} = 100M^{-1}$; $z_{bin} = -0.799 \rightarrow K = 10{,}000M^{-1}$), we derived a heuristic latent projection mapping:

$$log_{10}\mathrm{K} = \frac{3.714 - z_{bin}}{1.128} \tag{2}$$

This linear function is designed to operate explicitly as an ordinal sorting mechanism rather than an absolute physical quantifier.

2.5.2. Ternary chemical space screening

To facilitate the large-scale exploratory screening of ternary formulation systems, we introduced an experimental coformer correction (Δz). This correction was guided by ordinary least squares (OLS) regression on four literature-sourced indomethacin $\beta - \mathrm{CD}$ datapoints[34]:

$$\Delta\mathrm{z} = -0.0378 \times \mathrm{logP} - 0.7361 \tag{3}$$

where coformer log P values were algorithmically computed via RDKit[29]. Consequently, the heuristic ternary binding constant (K_{ternary}) can be approximated by translating the corrected logit ($z_{tern} = z_{bin} + \Delta z$) through our boundary projection. This descriptor-based correction was applied uniformly to generate an experimental priority ranking across 37,000 uncharacterized ternary permutations. High-throughput resolution of guest SMILES strings was successfully automated via the PubChem REST API[35] (achieving a 90.6% success rate).

2.6. Reproducibility and code availability

In strict adherence to open science principles, all raw and curated datasets, pre-trained model weights, training scripts, and inference pipelines are publicly hosted at https://github.com/CHIHX12/CD-BAN. Furthermore, we provide a unified execution script (plot_figures.py) to seamlessly reproduce all main-text figures and statistical analyses.

3. Results

3.1. CD-BAN establishes robust thermodynamic boundary classification

We first evaluated whether the bilinear attention mechanism could effectively segregate extreme thermodynamic states strictly from 2D graphs. Trained on a highly imbalanced dataset of 838 drug-CD pairs (68 Strong, 770 Weak), CD-BAN was evaluated on a held-out test set of 120 pairs. Across ten independent random seeds, the model achieved an AUROC of 0.925 ± 0.040, an AUPRC of 0.992 ± 0.005, and a balanced accuracy of 0.923 ± 0.024 (Fig. 2). While the Matthews Correlation Coefficient (0.386 ± 0.104) reflects a meaningful signal against extreme class imbalance, the variance in sensitivity (0.790 ± 0.110) underscores the inherent statistical constraints of evaluating extremely rare positive events (n = 10 in the test set). Nonetheless, a confidence-stratified analysis reveals strong prediction reliability: 100% of compounds classified as high-confidence weak binders (logit z > +1.458) were experimentally verified as such. This zero-shot reliability provides a highly conservative negative filter for candidate screening in pharmaceutical pipelines.

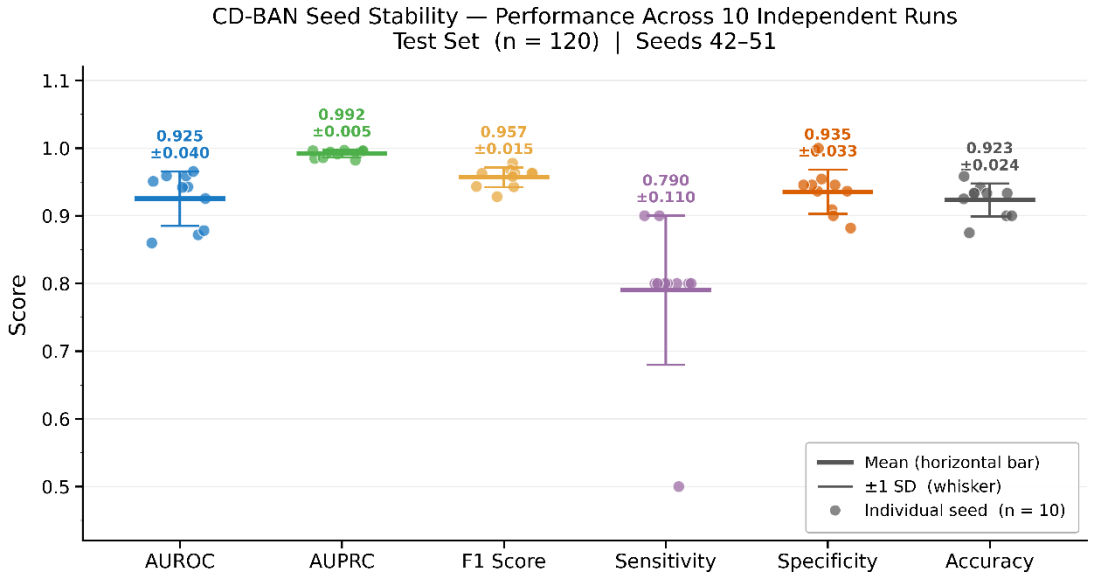


Fig. 2 | Robustness of thermodynamic boundary classification across independent initializations. Strip plots detailing the predictive performance of CD-BAN on the held-out test set (n = 120) across ten independent random training seeds (seeds 42–51). Horizontal bold lines dictate the inter-seed mean, while error bars denote ±1 standard deviation. Each individual point represents a distinct seed. The model demonstrates high algorithmic stability across global threshold-independent metrics (AUROC, AUPRC) and overall accuracy. The pronounced inter-seed variance observed in Sensitivity (0.790 ± 0.110) structurally reflects the extreme class imbalance of the test set (n = 10 Strong binders), where a single algorithmic misclassification intrinsically

shifts the metric by ten percentage points.

3.2. Independent identification of a continuous affinity gradient beyond the training manifold

The most scientifically profound finding of this study emerges when CD-BAN is challenged with 1,850 intermediate-affinity compounds termed the "fuzzy-zone" ($100 \leq K \leq 10{,}000\ M^{-1}$) that were strictly excluded from the training phase. Despite being optimized exclusively on discrete binary extremes, the network independently identifies and projects a continuous underlying affinity gradient. The mean predicted probability of weak binding, P(Weak), systematically decreases from 0.734 in the lowest affinity bin ($K = 100 - 316M^{-1}$) to 0.301 in the highest bin ($K = 3{,}162 - 10{,}000M^{-1}$), adhering to a strict monotonic sequence across all intermediate strata (Fig. 3).

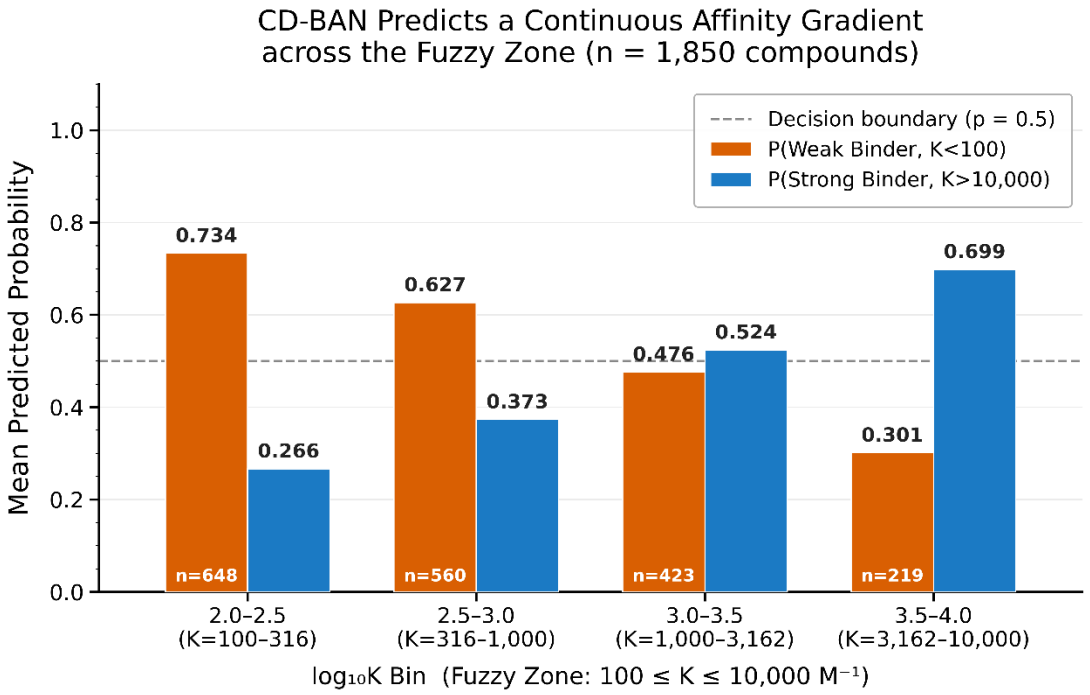


Fig. 3 | Emergence of a continuous affinity gradient across out-of-distribution compounds. Mean predicted probabilities for the out-of-distribution "fuzzy zone" (n = 1,850 compounds, $100 \leq K \leq 10{,}000M^{-1}$). Despite binary-only supervision at thermodynamic extremes, the network enforces a strict monotonic ranking across the four withheld intermediate $log_{10}K$ bins. The predicted probability of weak binding (orange bars) systematically decreases, while the probability of strong binding (blue bars) proportionately increases, crossing the neutral decision boundary (dashed line, P = 0.5) precisely within the $3{,}162 - 10{,}000M^{-1}$ transition phase.

This emergent continuum is statistically robust ($\tau = -0.289, p < 10^{-75}$), confirming that the gradient is not a random seed artifact but a fundamental property of the learned representation. Complementary correlation metrics (Pearson $r = -0.468$, Spearman $\rho = -0.425; p < 0.001$) further corroborate this implicit mapping (Fig. 4). Furthermore, the model spontaneously assigns 53.5% of these out-of-distribution compounds to an intermediate confidence zone. This self-calibrating expression of uncertainty achieved without any explicit uncertainty quantification training strongly indicates that the latent space has organized itself along a continuous physical thermodynamic dimension (Fig. 5, Fig. 6).

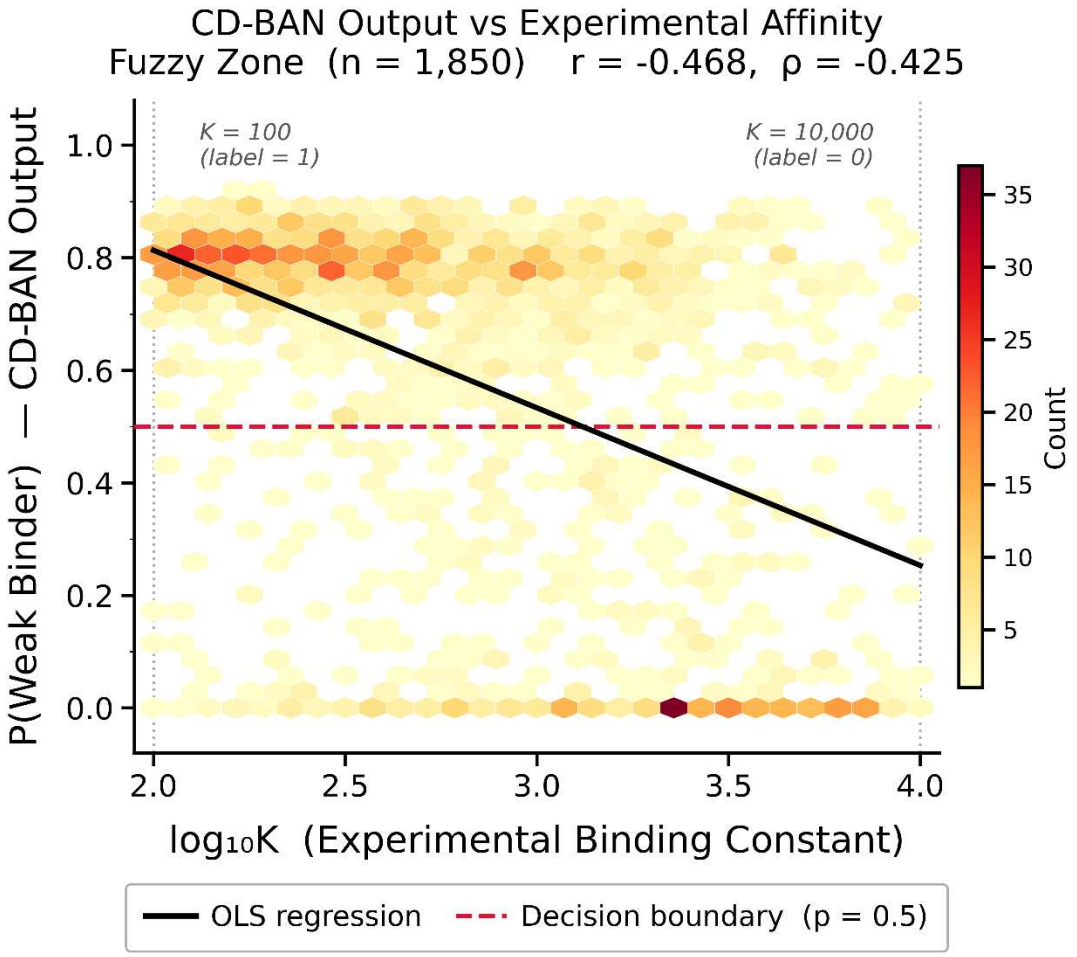


Fig. 4 | Latent space correlation with experimental binding affinity. Hexagonal binning scatter plot mapping the model's output probability against the ground-truth log_{10}K for all 1,850 fuzzy-zone compounds. Color intensity encodes local data density. Beyond the robust monotonic correlation (Kendall $\tau = -0.289, p < 10^{-75}$; Pearson $r = -0.468$), the density distribution reveals a distinct bipolar accumulation: high-confidence predictions correctly cluster at the thermodynamic extremes of the unseen space (top-left and bottom-right). Furthermore, the ordinary least squares (OLS)

regression line (solid black) traverses the neutral decision boundary (dashed red, P = 0.5) precisely within the $K = 1{,}000 - 3{,}162M^{-1}$ range, visually verifying that the network has independently parameterized a continuous thermodynamic transition axis bridging the discrete binary training thresholds (dotted vertical lines).

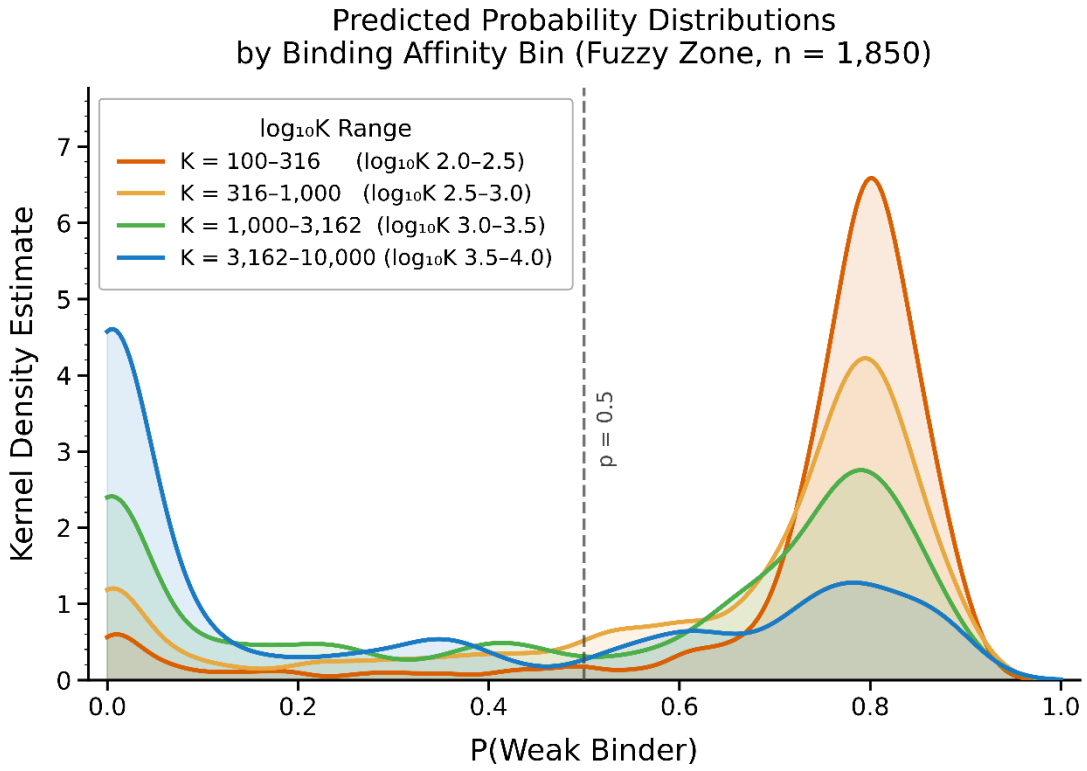


Fig. 5 | Density estimation of the self-organizing latent space. Kernel Density Estimation (KDE) curves illustrating the distribution of P(Weak) across the four sequential fuzzy-zone affinity bins. The architecture exhibits a distinct bimodal density transfer: rather than rigidly shifting a single peak, the model progressively redistributes probability mass between two latent thermodynamic attractors (near P=0 and P=0.8) as experimental *K* increases. Crucially, during the intermediate transition phase (green curve, $K = 1{,}000 - 3162M^{-1}$), probability density erodes from the extremes and pools across the neutral decision boundary (P=0.5). This progressive density inversion demonstrates the network's capacity to naturally interpolate structural matching, generating self-calibrated intermediate confidence scores for ambiguous thermodynamic states without explicit uncertainty training.

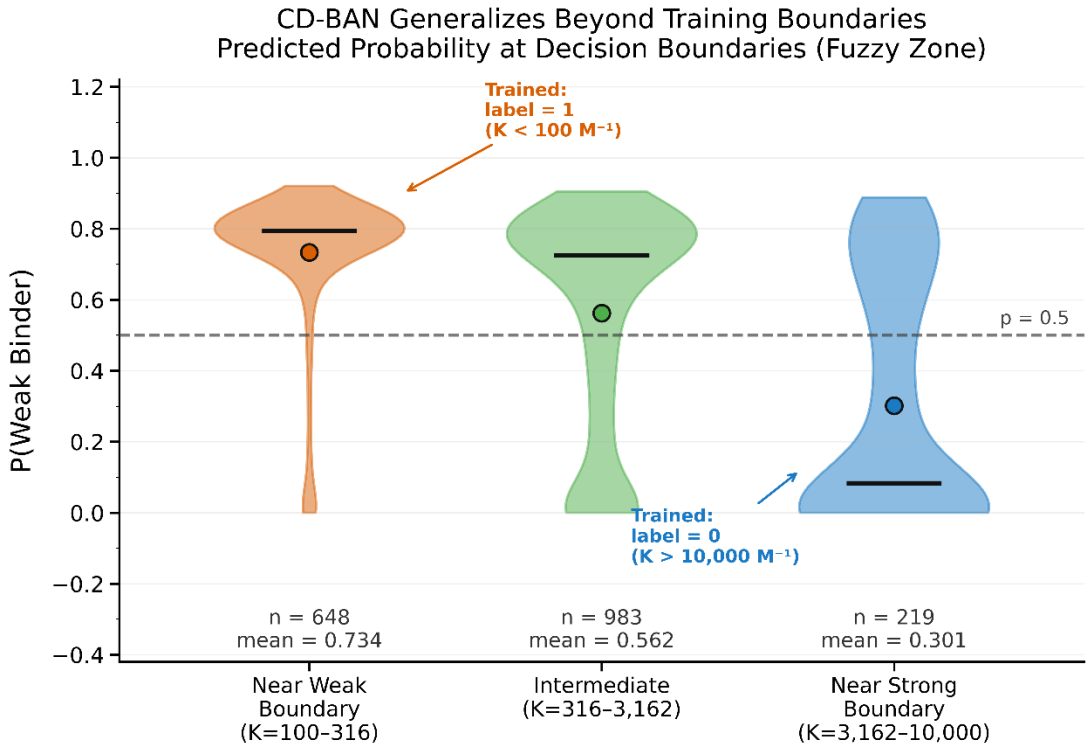


Fig. 6 | Boundary validation of the emergent continuum. Violin plots detailing the probability distribution of weak binding, P(Weak), clustered by proximity to the unseen training thresholds. The distributions display a continuous morphological transition from the high-probability Weak boundary ($\mathrm{K} = 100 - 316M^{-1}$ through the intermediate interior, down to the low-probability Strong boundary ($\mathrm{K} = 1{,}000 - 3{,}162M^{-1}$). Crucially, the intermediate transition zone (green) exhibits an elongated, bimodal-like topology; while its median remains anchored near the weak domain, the probability mass stretches downward across the neutral boundary (dashed line, p = 0.5), physically illustrating the network estimating maximum thermodynamic ambiguity. Embedded circles and thick black horizontal lines denote the distribution means and medians, respectively.

3.3. Ordinal latent projection enables heuristic candidate ranking

Given the spontaneous monotonic organization of the latent space, we derived a boundary-condition projection to map the output logit z back to a continuous log_{10}K scale. Driven by the empirical boundary requirements (z=+1.458→K=100 M^{-1}; z=−0.799→K=10,000 M^{-1}), we established a heuristic rank-ordering function. While the resulting projection exhibits a substantial median absolute error (approximately 153-fold K deviation) reflecting the inherently indirect nature of extracting continuous physical values from a binary-trained logit it crucially preserves a highly significant ordinal ranking (r=0.381, p<0.001). Therefore, rather than serving as a substitute for

precise quantitative phase-solubility measurements[25], this projection functions as a robust ordinal sorting mechanism, enabling high-throughput triage in the critical 100–10,000 M^{-1} formulation window.

3.4. Zero-shot extrapolation enables heuristic exploration of ternary chemical space

Pharmaceutical CD formulations frequently utilize a third coformer component to amplify bridging interactions[36]. Translating our binary-derived latent projection to such ternary systems represents a massive zero-shot extrapolation challenge for the model. To probe the boundaries of CD-BAN's utility, we introduced a preliminary descriptor-based correction (Δz based on coformer logP) fitted to a highly restricted literature sample ($n = 4$)[34].

While this sparse foundational data strictly limits definitive statistical inference, we utilized it as a heuristic engine to cast a wide net across 37,000 unexplored drug–CD–excipient combinations. Of the 19,780 combinations falling within the valid projection boundaries, the heuristic successfully identified 1,635 candidates potentially exceeding the strong-binding $K>10,000\ M^{-1}$ threshold. Rather than providing definitive quantitative predictions, this output serves as a computationally guided hypothesis generation tool. By filtering out over 90% of sub-optimal combinations, CD-BAN efficiently narrows the experimental search space, prioritizing high-potential topological matches (e.g., Triclosan with β-CD[37], or Flurbiprofen with β-CD[10], a classic ophthalmic formulation) for subsequent rigorous empirical validation.

4. Discussion

4.1. Emergence of a physical continuum from discrete supervision

The central result of this work that a molecular graph neural network trained exclusively on thermodynamic extremes independently identifies a continuous affinity gradient for out-of-distribution compounds invites both mechanistic interpretation and broader methodological generalization. The rigorous monotonic reduction in *P(Weak)* across successive log_{10}K bins (Fig. 3), corroborated by the robust continuous correlation mapped within the latent space (Fig. 4), and the highly significant Kendall $\tau(p < 10^{-75})$ collectively demonstrate that this gradient is an intrinsic feature of the learned representation, rather than a binning artifact. Formally, while a model's sigmoid output must be monotonic within its explicit training manifold (here, $log_{10}K < 2$ and $log_{10}K > 4$), the statistically robust extension of this monotonicity into the untrained, unobserved interior ($2 \leq log_{10}K \leq 4$) is non-trivial. This phenomenon provides systematic, quantified evidence that binary classification objectives can implicitly supervise a continuous physical axis in latent space, achieving an independent identification of the underlying thermodynamic continuum (*K*).

4.2. Mechanistic basis of cross-molecular attention

The bilinear attention mechanism[22,23] plays a structurally vital role in this emergent capability. By computing an explicit interaction matrix between the Guest and Host embedding vectors, the BANLayer dynamically prioritizes structural complementarity cavity size matching, hydrophobic surface alignment, and charge distribution without relying on predefined 3D geometries or element-wise fusion. Indeed, our ablation studies empirically validate this architectural necessity: models lacking explicit cross-attention, such as standard GCNs[26,27] or simple feature-concatenation baselines[16,20], fail to achieve comparable phase segregation. This confirms that capturing the spatial and chemical compatibility of host-guest systems strictly requires learning the coupled pairwise interactions rather than relying on independent molecular embeddings.

Beyond confirming the necessity of the BANLayer, interrogating the aggregate attention weights provides profound chemical interpretability, revealing that the model has implicitly learned established supramolecular recognition patterns (Supplementary Figs. S1–S4). Across all unmodified cyclodextrins ($\alpha-$, $\beta-$, and $\gamma-$CD), the network assigns the highest attention masses consistently to the oxygen network specifically the ring oxygen (O5) and the secondary hydroxyl groups (O2, O3) while almost entirely ignoring the hydrophobic carbon backbone (C2, C3). This model-derived focus precisely mirrors empirical NMR ROESY observations[38,39], which indicate that high-affinity guest complexation predominantly occurs via inclusion

through the wider secondary rim, stabilized by O2/O3 hydrogen-bonding networks.

Furthermore, the model accurately captures cavity-size dependencies and substituent effects without explicit 3D rules. For $\gamma - \mathrm{CD}$ (Supplementary Fig. S3), the network allocates an anomalously high attention weight to the primary rim (O6, 0.90), aligning with experimental evidence[40] that its expanded 8.3 Å cavity allows bulky guests to penetrate deeply and interact with the primary hydroxyls. Conversely, the attention distribution for $\mathrm{HP} - \beta - \mathrm{CD}$ (Supplementary Fig. S4) exhibits significant flattening. This statistically dispersed attention perfectly encapsulates the disruption of structural symmetry introduced by random hydroxypropyl substitutions, which literature confirms[41] modifies the classical binding modality and diversifies the interaction environment.

4.3. Heuristic utility in large-scale formulation screening

While the derived calibration formula and ternary extension provide powerful heuristic utility, they operate within strictly bounded limitations. The median error of 2.2 $log_{10}K$ units structurally reflects the indirect nature of mapping a binary logit to a continuous physical scale. Consequently, for definitive quantitative formulation decisions, computational inference cannot supersede empirical phase-solubility experiments[25]. Furthermore, the ternary coformer correction is fundamentally constrained by sparse training availability (n = 4)[34,36]; its derived regression coefficients inherently carry substantial statistical uncertainty and must be interpreted strictly as directional indicators rather than deterministic physical parameters.

However, when deployed precisely as a computational hypothesis generation tool, this ordinal rank-ordering becomes transformative[18,42]. Translating these boundary conditions enabled the zero-shot evaluation of 19,780 usable ternary predictions. By identifying 8,293 candidates above the 1,000 M^{-1} threshold, CD-BAN provides the pharmaceutical community with a highly prioritized screening landscape that computationally narrows the experimental search space by orders of magnitude, effectively guiding subsequent empirical validation.

4.4. Limitations and future trajectories

Addressing the limitations of the current framework establishes a clear trajectory for future supramolecular AI tools. The reliance on heterogeneous training data from OpenCycloDB[24] encompassing diverse measurement temperatures and laboratory protocols[25] inherently introduces systematic noise into the binary labels. Furthermore, the statistical power of our sensitivity bounds is constrained by the extreme class

imbalance typical of host-guest affinity datasets. Future iterations must integrate advanced uncertainty quantification mechanisms and automated data harmonization protocols to further refine the resolution of the latent continuous space[42]. Ultimately, CD-BAN demonstrates how task-specific architectural inductive biases can map discrete empirical observations into a generalized thermodynamic continuum[1,2,18].

5. Conclusion

CD-BAN introduces a powerful new paradigm for deploying graph neural networks in supramolecular chemistry and formulation science. By integrating a bilinear cross-attention mechanism, we demonstrate that neural networks can successfully bypass the computationally expensive requirement for explicit 3D conformer generation when predicting host-guest structural complementarity. More fundamentally, this work establishes a critical methodological proof-of-concept: AI models optimized exclusively on discrete, extreme thermodynamic boundaries can independently identify and reconstruct a continuous physical property gradient for entirely unobserved molecules. This emergent capability not only delivers an immediately actionable, high-throughput virtual screening tool for cyclodextrin formulations, but also offers a profound new perspective on how binary classification objectives can capture implicit physical laws in computational chemistry.

Data availability

The datasets analyzed during the current study are derived from publicly available repositories. The base experimental association constants (*K*) were sourced from the open-access OpenCycloDB. The rigorously curated datasets generated for this study including the binary training set (n = 1,198), the withheld fuzzy-zone evaluation set (n = 1,850), and the large-scale ternary screening priority list (n = 37,000) are fully available in the CD-BAN GitHub repository (https://github.com/CHIHX12/CD-BAN). Raw literature sources can be traced via the included data annotations.

Code availability

All source code, including the dual-branch GCN and BANLayer implementations, pre-trained model weights for all ten independent seeds, and the training/inference pipelines, is publicly available under an open-source license at https://github.com/CHIHX12/CD-BAN. A unified execution script (plot_figures.py) is provided to seamlessly reproduce all main-text figures and statistical analyses, ensuring full methodological transparency and negating the requirement for separate supplementary data files. The model relies on Python 3.x, PyTorch (v2.5.1), PyTorch Geometric (v2.7.0), and RDKit.

Author contributions

C.-Y.C. led the data curation, computational experiments, and primary codebase development. All authors (C.-Y.C., Y.-H.W., and F.-Y.L.) contributed comprehensively to the conceptualization of the study, data interpretation, manuscript writing, and provided critical research guidance. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information: Not applicable. All data, code, and execution scripts necessary to evaluate the conclusions and reproduce the figures of this paper are consolidated in the primary GitHub repository.

Correspondence and requests for materials should be addressed to C.-Y.C., Y.-H.W., or F.-Y.L.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgement

We acknowledge the Department of Chemistry at National Chung Hsing University (NCHU), Taiwan, for providing the computational resources and infrastructure that supported this work. We also extend our gratitude to the creators and maintainers of OpenCycloDB for providing the publicly accessible dataset that made this methodology research possible.

References

1. Wieder, O. et al. A compact review of molecular property prediction with graph neural networks. Drug Discov. Today Technol. 37, 1–12 (2020).

2. Atz, K., Grisoni, F. & Schneider, G. Geometric deep learning on molecular representations. Nat. Mach. Intell. 3, 1023–1032 (2021).

3. Del Valle, E. M. M. Cyclodextrins and their uses: a review. Process Biochem. 39, 1033–1046 (2004).

4. Connors, K. A. The stability of cyclodextrin complexes in solution. Chem. Rev. 97, 1325–1358 (1997).

5. Szejtli, J. Cyclodextrin Technology (Kluwer Academic Publishers, 1998).

6. Loftsson, T. & Duchêne, D. Cyclodextrins and their pharmaceutical applications. Int. J. Pharm. 329, 1–11 (2007).

7. Jansook, P., Ogawa, N. & Loftsson, T. Cyclodextrins: structure, physicochemical properties and pharmaceutical applications. Int. J. Pharm. 535, 272–284 (2018).

8. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol. Methods 44, 235–249 (2000).

9. Amidon, G. L., Lennernäs, H., Shah, V. P. & Crison, J. R. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. Pharm. Res. 12, 413–420 (1995).

10. Brewster, M. E. & Loftsson, T. Cyclodextrins as pharmaceutical solubilizers. Adv. Drug Deliv. Rev. 59, 645–666 (2007).

11. Kurkov, S. V. & Loftsson, T. Cyclodextrins. Int. J. Pharm. 453, 167–180 (2013).

12. Rajewski, R. A. & Stella, V. J. Pharmaceutical applications of cyclodextrins. 2. In vivo drug delivery. J. Pharm. Sci. 85, 1142–1169 (1996).

13. Loftsson, T. & Brewster, M. E. Pharmaceutical applications of cyclodextrins. 1. Drug solubilization and stabilization. J. Pharm. Sci. 85, 1017–1025 (1996).

14. Gould, S. & Scott, R. C. 2-Hydroxypropyl-β-cyclodextrin (HP-β-CD): a toxicology review. Food Chem. Toxicol. 43, 1451–1459 (2005).

15. Loftsson, T. & Stefánsson, E. Cyclodextrins in eye drop formulations: enhanced topical delivery of corticosteroids to the eye. Acta Ophthalmol. Scand. 80, 144–150 (2002).

16. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. J. Chem. Inf. Model. 50, 742–754 (2010).

17. Stokes, J. M. et al. A deep learning approach to antibiotic discovery. Cell 180, 688–702 (2020).

18. Chen, H. et al. The rise of deep learning in drug discovery. Drug Discov. Today 23, 1241–1250 (2018).

19. Duvenaud, D. K. et al. Convolutional networks on graphs for learning molecular fingerprints. Adv. Neural Inf. Process. Syst. 28, 2224–2232 (2015). https://arxiv.org/abs/1509.09292
20. Yang, K. et al. Analyzing learned molecular representations for property prediction. J. Chem. Inf. Model. 59, 3370–3388 (2019).
21. Schütt, K. T. et al. SchNet: a continuous-filter convolutional neural network for modeling quantum interactions. Adv. Neural Inf. Process. Syst. 30 (2017). https://arxiv.org/abs/1706.08566
22. Bai, P., Miljković, F., John, B. & Lu, H. Interpretable bilinear attention network with domain adaptation improves drug–target prediction. Nat. Mach. Intell. 5, 126–136 (2023).
23. Vaswani, A. et al. Attention is all you need. Adv. Neural Inf. Process. Syst. 30 (2017). https://arxiv.org/abs/1706.03762
24. Tahil, G. et al. Curated dataset of association constants between a cyclodextrin and a guest for machine learning. Chem. Data Collect. 45, 101022 (2023).
25. Higuchi, T. & Connors, K. A. Phase-solubility techniques. Adv. Anal. Chem. Instrum. 4, 117–212 (1965).
26. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. Preprint at https://arxiv.org/abs/1609.02907 (2017).
27. Gilmer, J. et al. Neural message passing for quantum chemistry. Proc. Int. Conf. Mach. Learn. 70, 1263–1272 (2017). https://arxiv.org/abs/1704.01212
28. Kearnes, S. et al. Molecular graph convolutions: moving beyond fingerprints. J. Comput. Aided Mol. Des. 30, 595–608 (2016).
29. Landrum, G. RDKit: Open-source cheminformatics. https://www.rdkit.org (2006).
30. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).
31. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32, 8024–8035 (2019).
32. Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. Preprint at https://arxiv.org/abs/1903.02428 (2019).
33. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272 (2020).
34. Fernandes, C. M. & Veiga, F. J. B. Effects of some hydrotropic agents on the formation of indomethacin/β-cyclodextrin inclusion compounds. J. Inclusion Phenom. Macrocyclic Chem. 33, 117–125 (1999).
35. Kim, S. et al. PubChem 2023 update. Nucleic Acids Res. 51, D1373–D1380 (2023).
36. Mura, P., Faucci, M. T. & Parrini, P. L. Ternary systems of naproxen with hydroxypropyl-β-cyclodextrin and amino acids. Int. J. Pharm. 260, 293–302 (2003).

37. Duan, M. S. et al. Cyclodextrin solubilization of the antibacterial agents triclosan and triclocarban: formation of aggregates and higher-order complexes. Int. J. Pharm. 297, 213–222 (2005).
38. Pessine, F. B. T., Calderini, A. & Alexandrino, G. L. Review: cyclodextrin inclusion complexes probed by NMR techniques. in Magnetic Resonance Spectroscopy (ed. Kim, D.) 237–264 (InTech, 2012).
39. Zhao, R., Sandström, C., Zhang, H. & Tan, T. NMR study on the inclusion complexes of β-cyclodextrin with isoflavones. Molecules 21, 372 (2016).
40. Lee, J.-U., Lee, S.-S., Lee, S. & Oh, H. B. Noncovalent complexes of cyclodextrin with small organic molecules: applications and insights into host–guest interactions in the gas phase and condensed phase. Molecules 25, 4048 (2020).
41. Li, Z. et al. Functionality-related characteristics of hydroxypropyl-β-cyclodextrin for the complexation. J. Mol. Liq. 365, 120105 (2022).
42. Walters, W. P. & Barzilay, R. Applications of deep learning in molecule generation and molecular property prediction. Acc. Chem. Res. 54, 263–270 (2021).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryCDBANversion1.docx