ChemHyperMag: Physics-informed magnetic hypergraph learning improves molecular ADMET prediction

Hexiao Ding [*1] **Hongzhao Chen** [*1] **Jing Lan** [*1] **Yufeng Jiang** [1] **Zihong Luo** [2] **Zehua Xiong** [3] **Tianlong Ruan** [4] **Yunlin Mao** [1] **Nga Chun Ng** [1] **Gwing Kei Yip** [1] **Gerald W.Y. Cheng** [1] **Kate Inyoung Oh** [1] **Jing Cai** [1] **Liang-Ting Lin** [1] **Jung Sun Yoo** [#1]

Abstract

Accurate prediction of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) is important for drug discovery. Most predictors use undirected molecular graphs and pairwise edges. This choice misses asymmetric interactions, nonreversible dynamics, and motif level effects from functional groups and ring systems. We propose ChemHyperMag for multitask ADMET prediction under missing labels. ChemHyperMag builds a functional group hypergraph from rings, BRICS fragments, Bemis-Murcko scaffolds, and bonds. It also defines a potential driven nonreversible flow guided by electronegativity and Gasteiger partial charges. The resulting circulation is encoded by a Hermitian magnetic Laplacian and processed with a magnetic Chebyshev encoder. We perturb magnetic phases to form stochastic views and train with an InfoNCE objective. Experiments on multiple ADMET benchmarks show improvements over recent methods with fewer labeled samples and no conformers. ChemHyperMag is scalable and provides interpretable directional signals through its magnetic phases.

1. Introduction

The prediction of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) is a key question in the discovery of small-molecule drugs. It determines which candidates justify costly synthesis and assays. ADMET benchmarks in machine learning encompass multi-task learning, heterogeneous label semantics, and pervasive missingness (Wu et al., 2018; Huang et al., 2022). Currently, many workflows begin with SMILES (Weininger, 1988), which are modeled as sequences or transformed into two-dimensional atom-bond graphs by RDKit (Bento et al., 2020). Molecular encoders underperform in this regime because their representations lack comprehensive structural detail and fail to capture the physical and chemical nature of molecules.

Sequence encoders built on SMILES provide only indirect access to molecular structure, so functional regularities must be inferred from sparse and heterogeneous endpoints (Weininger, 1988). Message passing on bond graphs exposes topology, yet it is typically defined through undirected pairwise interactions. This formulation ensures invariance to edge reversal but neglects bond polarity and donor–acceptor roles, thus higher-order motifs are merely approximated via local aggregation (Yang et al., 2019). However, the abstraction of these representations is misaligned with the physical drivers behind many ADMET endpoints. Transport and partitioning are shaped by local polarity fields induced by electronegativity imbalance and charge separation (Pauling, 1960; Gasteiger & Marsili, 1980), while conjugated substructures exhibit directional redistribution and cyclic electronic circulation that cannot be expressed by undirected diffusion. Geometry-based and multimodal models can recover part of this signal. However, they rely on conformer generation and large-scale three-dimensional supervision, which raises data and compute demands and often conflicts with high-throughput ADMET settings (Zhou et al., 2023; Lu et al., 2024; Wang et al., 2024). The challenge is to design a SMILES-based pipeline that captures higher-order motif structures. At the same time, it should embed physically and chemically grounded directionality to enhance learning processes.

We propose ChemHyperMag to address this issue. It introduces directionality by integrating physically and chemically grounded energy-structure coupling. We start from the standard 2D molecular graph extracted from SMILES, which fixes the atom set and covalent connectivity. We then lift this graph to a functional group hypergraph that

[*]Equal contribution [#]Corresponding author. [1]Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hong Kong SAR, China [2]University of Liverpool, Liverpool, United Kingdom [3]Shanghai Jiao Tong University, Shanghai, China [4]Fudan University, Shanghai, China. Correspondence to: Jung Sun Yoo <jungsun.yoo@polyu.edu.hk>.

Proceedings of the AI4Physics Workshop at the 43rd International Conference on Machine Learning (AI4Physics@ICML 2026), Seoul, South Korea. 2026. Copyright 2026 by the author(s).

augments pairwise bonds with motif-defined higher-order neighborhoods, so that propagation is supported not only on covalent edges but on atom pairs that co-occur within chemically meaningful motifs. On this hypergraph-supported neighborhood we define a potential-driven, nonreversible diffusion. Each atom is assigned a chemical potential computed from its electronegativity and partial charge, and transitions are biased along potential gradients. The resulting Markov operator induces a stationary flow that decomposes into a symmetric component encoding diffusion strength and an antisymmetric component encoding circulation. We encode this circulation with a magnetic Laplacian, where the complex phase represents cyclic currents while the operator remains Hermitian, enabling stable learning while retaining directional information (Fanuel et al., 2017; Zhang et al., 2021; Benko et al., 2025). To improve robustness under missing labels, we construct stochastic views by perturbing the magnetic phase while keeping the hypergraph fixed. We then optimize an InfoNCE objective to align these views at the operator level, following directed contrastive learning (Tong et al., 2021). This design removes the need for conformers, but it introduces sparse hypergraph construction and complex-valued propagation. ***Figure 1*** illustrates the concept and workflow of ChemHyperMag.

The main contributions of this work are summarized as follows. First, we build a SMILES-derived functional group hypergraph that encodes motif-level neighborhoods beyond pairwise bonds. Second, we introduce a physicochemically-grounded nonreversible Markov flow on this hypergraph and encode its directional bias and circulation with a Hermitian magnetic Laplacian through complex phases. Third, we propose magnetic phase perturbation as operator-level view generation for InfoNCE regularization in multitask ADMET with missing labels.

2. Related Work

ADMET prediction is framed as multi-endpoint learning with heterogeneous labels and strong emphasis on standardized splits and metrics, as in MoleculeNet (Wu et al., 2018) and the Therapeutics Data Commons (Huang et al., 2022).

2.1. Statistical Learning Methods

Quantitative Structure-Activity Relationship (QSAR) baselines remain central for ADMET. Extended-connectivity fingerprints used physicochemical descriptors and fingerprints with linear models or tree ensembles (Rogers & Hahn, 2010). The other QSAR models such as pkCSM (Pires et al., 2015), SwissADME (Daina et al., 2017), vNN-ADMET (Schyman et al., 2017), and ADMETboost (Tian et al., 2022) serve as rule- and descriptor-based screening tools.

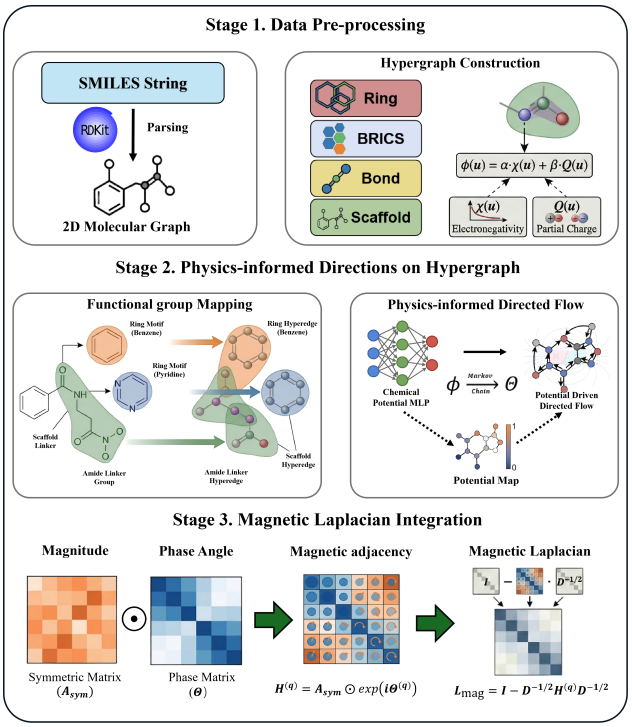

Figure 1. Workflow of ChemHyperMag. ***Stage 1*** parses SMILES into a 2D molecular graph with RDKit and computes atom electronegativity and Gasteiger partial charges together with structural annotations. ***Stage 2*** constructs a functional group hypergraph from rings, BRICS fragments, scaffolds, and bonds, then defines a chemical potential that induces directed flow under the hypergraph constraints. ***Stage 3*** encodes this directionality with a magnetic phase, forms the magnetic adjacency, and builds the normalized magnetic Laplacian for spectral propagation.

2.2. Deep Learning Methods

The next-generation ADMET platforms combine these ingredients with learned-models and curated endpoints at scale. ADMETlab 2.0 (Xiong et al., 2021) and 3.0 (Fu et al., 2024) and ADMET-AI (Swanson et al., 2024) are pioneer works. Besides, end-to-end representation learning largely follows message passing on atom-bond graphs, with PotentialNet (Feinberg et al., 2018) and the D-MPNN family implemented in Chemprop (Yang et al., 2019; Heid et al., 2023). Notably, Chemprop v2 is now serving as a widely adopted ADMET backbone (Graff et al., 2025). Moreover, recent advances in transformer architectures and 3D pre-training frameworks, such as Graphormer (Ying et al., 2021) and Uni-Mol (Zhou et al., 2023), have shown improved capacity to address the long-context problem and capture geometric effects in molecular ADMET prediction. Current deep learning models increasingly leverage endpoint structure and analogue comparisons in ADMET prediction. Notable strategies include adaptive auxiliary task selection in MTGL-ADMET (Du et al., 2023a) and pairwise difference learning in DeepDelta (Fralish et al., 2023), while OmniMol emphasizes multi-task robustness under imperfect labeling

conditions (Wang et al., 2025a).

2.3. Physics-informed Methods

Physics-informed ADMET models incorporate physicochemical priors by supervising molecular representations with electronic or thermodynamic features beyond connectivity. Recent work uses quantum chemistry supervision to pretrain graph transformers on atom-in-a-molecule properties and transfers the resulting encoders to ADMET endpoints (Kim et al., 2024; Fallani et al., 2025). Physics-informed methods improve model performance and sample efficiency when labels are sparse or missing.

2.4. Hypergraph and Magnetic Laplacian

2.4.1. Hypergraph Neural Networks

Hypergraph neural networks generalize graph message passing by defining transmission over node-hyperedge incidence structures (Feng et al., 2019). Surveys have organized existing approaches into convolutional, attention, and spectral formulations (Yang & Xu, 2025). For undirected hypergraphs, multiset-based encoders such as AllSet learn permutation-invariant node-hyperedge aggregation and unify several incidence variants (Chien et al., 2022). In parallel, sheaf hypergraph networks enrich hyperedge states and produce more expressive Laplacians for modeling higher-order diffusion (Duta et al., 2023). Directed hypergraphs assign each hyperedge a tail and a head, thereby inducing nonreversible dynamics and asymmetric Laplacians. DHGNN introduces a directed-hypergraph Laplacian for convolution and link prediction (Ma et al., 2024), while DSHN unifies directed and undirected learning through a complex-valued directed sheaf Laplacian (Mule et al., 2025). In molecular learning, hyperedges are used to represent rings, fragments, and reaction environments (Kajino, 2019; Chen & Schwaller, 2024). Functional-group hypergraphs have been shown to improve toxicity prediction and related ADMET endpoints in MolHGCN (Chen et al., 2021), while molecular hypergraph message passing captures irregular higher-order conjugation beyond conventional bond graphs (Chen & Schwaller, 2024).

2.4.2. Magnetic Laplacian

Magnetic Laplacians encode edge orientation in complex phases while remaining Hermitian, which enables spectral learning under nonreversible dynamics (Fanuel et al., 2017). MagNet introduces a charge-parameterized magnetic Laplacian for directed representation learning (Zhang et al., 2021), and MSGNN generalizes it to signed digraphs through a magnetic signed Laplacian with Chebyshev polynomial filters (He et al., 2022b). SD-GCN derives a related Hermitian magnetic Laplacian for signed directed graphs and uses it for spectral convolution (Ko et al., 2023b). HyperMagNet extends magnetic operators to hypergraphs by defining an edge-dependent random walk on the incidence bipartite graph and constructing a magnetic hypergraph Laplacian with learnable charge (Benko et al., 2025).

2.5. Graph Contrastive Learning (CL)

2.5.1. Graph CL without Directions

Graph contrastive learning (GCL) learns robust graph representations by promoting representation scattering, often via agreement between augmented views or distribution discrimination (He et al., 2024). Representative frameworks such as GraphCL (You et al., 2020) have standardized objectives and augmentation pipelines. Follow-up research improves view construction and loss formulation using adversarial and prototypical strategies (Suresh et al., 2021; Lin et al., 2022). Empirical analyses also identify scenarios where augmentations enhance representation learning or cause collapse (Zhu et al., 2021). For molecular ADMET predictions, labeled endpoints are heterogeneous and often limited. Recent GCL frameworks enhance data efficiency by tailoring chemically-valid perturbations and enforcing cross-view alignment. MolCLR (Wang et al., 2022), GraphMVP (Liu et al., 2022b), and 3DGCL (Moon et al., 2023) advance 2D-3D contrast; ATMOL (Liu et al., 2022a) introduces attention-guided masking; and ACML (Wang et al., 2025b) develops multimodal asymmetric alignment. Collectively, these methods achieve consistent gains on drug-like benchmarks with ADMET endpoints.

2.5.2. Graph CL with Directions

Directed Graph CL avoids orientation loss by perturbing graph operators. DiGCL leverages directed Laplacian perturbations with dynamic view sampling (Tong et al., 2021), and UGCL extends this framework to unify Laplacian perturbations across undirected, directed, and signed graphs (Ko et al., 2023c). For signed directed structure, SDGCL (Ko et al., 2023a) introduces Laplacian augmentation in magnetic space and S2-DiGCL further couples path-based spatial views with operator-level perturbations (Su et al., 2025). However, molecular representations are generally treated as undirected graphs, and directed graph contrastive learning remains unexplored in ADMET prediction.

3. Preliminaries

3.1. Problem Definition

Let $\mathcal{G}$ be the set of molecular graphs and let $T \in \mathbb{N}$ be the number of prediction tasks. The training set is

$$\mathcal{D} = \left\{(G_i, y_i, m_i)\right\}_{i=1}^{N} \subseteq \mathcal{G} \times \mathbb{R}^T \times \{0,1\}^T, \quad (1)$$

where $G_i \in \mathcal{G}$ is the ith molecule, $y_i = (y_{i1}, \ldots, y_{iT}) \in \mathbb{R}^T$ is its multitask label vector with y_{it} the ground truth target for task t, and $m_i = (m_{i1}, \ldots, m_{iT}) \in \{0,1\}^T$ is the observation mask with $m_{it} = 1$ if y_{it} is observed and $m_{it} = 0$ otherwise. A model $f_\theta : \mathcal{G} \to \mathbb{R}^T$ maps an input G to a prediction vector $f_\theta(G)$, whose tth component $f_\theta(G)_t$ predicts task t.

On a mini batch $\mathcal{B} \subset \mathcal{D}$ the supervised multitask loss is

$$\mathcal{L}_{\text{sup}} = \frac{1}{|\mathcal{B}|} \sum_{(G,y,m)\in\mathcal{B}} \sum_{t=1}^{T} m_t \, \ell_t\big(f_\theta(G)_t,\, y_t\big), \tag{2}$$

where ℓ_t is the loss for task t. ChemHyperMag implements f_θ as a shared magnetic hypergraph encoder with task specific heads and combines (2) with a contrastive term.

3.2. Functional Group Hypergraphs

3.2.1. Molecular Graphs

A two dimensional molecular graph is $G = (V, B, X)$, where $V = \{1, \ldots, n\}$ is the set of atoms, $B \subseteq V \times V$ is the set of undirected covalent bonds, and $X \in \mathbb{R}^{n\times d}$ is the node feature matrix with rows $x_u \in \mathbb{R}^d$ for $u \in V$.

The bond graph (V, B) encodes connectivity and bond distance. For $u, v \in V$, a bond path of length k from u to v is a sequence $(u_0, \ldots, u_k)$ with $u_0 = u$, $u_k = v$, and $(u_{r-1}, u_r) \in B$ for all $r \in \{1, \ldots, k\}$. Define the bond distance

$$\begin{aligned}\text{dist}_B(u,v) = \min\big\{k \in \mathbb{N}_0 : \exists\,(u_0,\ldots,u_k) \text{ s.t.} \\ u_0 = u,\ u_k = v,\ (u_{r-1},u_r) \in B, \forall r \in [k]\big\},\end{aligned} \tag{3}$$

with $\text{dist}_B(u,v) = +\infty$ if no such path exists. In particular, u and v are connected if and only if $\text{dist}_B(u,v) < +\infty$.

3.2.2. Hypergraphs

To represent functional groups and structural motifs, we lift the molecule to a hypergraph $\mathcal{H} = (V, E)$ with the same node set V and hyperedges $E = \{e_1, \ldots, e_m\}$, where each $e_j \subseteq V$. The incidence matrix is denoted by $\mathbf{M} \in \{0,1\}^{n\times m}$ with entries

$$\mathbf{M}_{u,j} = \begin{cases} 1, & u \in e_j, \\ 0, & u \notin e_j, \end{cases} \tag{4}$$

for $u \in V$ and $j \in \{1, \ldots, m\}$. In ChemHyperMag, each hyperedge e_j corresponds to a ring, a BRICS fragment, a Bemis-Murcko scaffold, or a bond hyperedge.

Throughout, $u, v \in V$ denote atoms, and an ordered pair (u, v) denotes a directed transition from u to v. For $e \in E$ and $u \in e$, define

$$N_e(u) = \{v \in e \mid v \neq u\}, \tag{5}$$

and the hypergraph neighborhood

$$N(u) = \bigcup_{e\in E\,:\,u\in e} N_e(u). \tag{6}$$

Thus $v \in N(u)$ means that u and v co-occur in at least one hyperedge. All Markov transitions in Section 3.3 are supported on ordered pairs (u, v) with $v \in N(u)$.

Each atom $u \in V$ carries a scalar chemical potential

$$\phi(u) = \alpha\,\chi(u) + \beta\,Q(u), \tag{7}$$

where $\chi(u)$ is the Pauling electronegativity of the atom at u and $Q(u)$ is its Gasteiger partial charge. The scalars α and β are learned parameters, and ϕ biases the nonreversible Markov dynamics.

3.3. Magnetic Markov Operators

3.3.1. Ergodic Markov Chains

Let $V = \{1, \ldots, n\}$ be the atom set of Section 3.2, which we also use as the finite state space of the Markov chain. A row stochastic matrix is $P \in [0,1]^{n\times n}$ with nonnegative entries and

$$P\mathbf{1} = \mathbf{1}, \tag{8}$$

where $\mathbf{1}$ is the all ones vector. We interpret P_{uv} as the one step transition probability from state u to state v. A distribution $\pi \in \mathbb{R}^n$ with $\pi_u \geq 0$ and $\mathbf{1}^\top \pi = 1$ is stationary for P if $\pi^\top P = \pi^\top$.

To guarantee ergodicity, we add a teleportation term. The same teleporting random walk is used in directed graph spectral learning (Tong et al., 2021) and in magnetic hypergraph operators (Benko et al., 2025). Given a base transition matrix P and a distribution $\nu \in \mathbb{R}^n$ with full support, meaning $\nu_u > 0$ for all $u \in V$, for $\eta \in (0,1)$ define

$$P^{(\eta)} = (1-\eta)P + \eta\,\mathbf{1}\,\nu^\top. \tag{9}$$

Since $(P^{(\eta)})_{uv} \geq \eta\nu_v > 0$ for all $u, v \in V$, the chain is irreducible and aperiodic, hence admits a unique stationary distribution. We denote it by $\pi^{(\eta)}$, and when η is fixed we write π.

3.3.2. Phase Map and Magnetic Adjacency

Let $A \in \mathbb{R}^{n\times n}$ be a real matrix supported on the allowed ordered pairs, meaning $A_{uv} = 0$ whenever $v \notin N(u)$. Define its symmetric and skew symmetric parts

$$A^{\text{sym}} = \frac{A + A^\top}{2}, \qquad A^{\text{skew}} = \frac{A - A^\top}{2}. \tag{10}$$

For $u, v \in V$ with $A^{\text{sym}}_{uv} > 0$ define

$$R_{uv} = \frac{A^{\text{skew}}_{uv}}{A^{\text{sym}}_{uv} + \varepsilon}, \qquad \Theta^{(q)}_{uv} = 2\pi q\,\tanh(R_{uv}), \tag{11}$$

with $\varepsilon > 0$ and $q \geq 0$. Then $\Theta^{(q)}_{uv} = -\Theta^{(q)}_{vu}$ and $\Theta^{(q)}_{uu} = 0$. Define $\Xi^{(q)} \in \mathbb{C}^{n\times n}$ by

$$\Xi^{(q)}_{uv} = \begin{cases} \exp\big(\mathrm{i}\,\Theta^{(q)}_{uv}\big), & A^{\mathrm{sym}}_{uv} > 0, \\ 1, & A^{\mathrm{sym}}_{uv} = 0, \end{cases} \tag{12}$$

where the choice on $\{(u,v) \mid A^{\mathrm{sym}}_{uv} = 0\}$ is irrelevant since $H^{(q)}_{uv} = 0$ there. For $A^{\mathrm{sym}}_{uv} > 0$, $|\Xi^{(q)}_{uv}| = 1$ and $\Xi^{(q)}_{uv} = \overline{\Xi^{(q)}_{vu}}$, hence $\Xi^{(q)}$ is Hermitian and generally not real.

The magnetic adjacency is

$$H^{(q)} = A^{\mathrm{sym}} \odot \Xi^{(q)}, \qquad (A \odot B)_{uv} = A_{uv}B_{uv}. \tag{13}$$

3.3.3. Magnetic Laplacian

Let $d_u = \sum_{v\in V} A^{\mathrm{sym}}_{uv}$ and $D = \mathrm{diag}(d_1, \ldots, d_n)$. Assume $d_u > 0$ for all $u \in V$. The normalized magnetic Laplacian is

$$L_{\mathrm{mag}} = I - D^{-1/2} H^{(q)} D^{-1/2}. \tag{14}$$

Proposition 1. L_{mag} is Hermitian and positive semidefinite.
Proof. Since A^{sym} is real symmetric and $\Xi^{(q)}$ is Hermitian, the Hadamard product $H^{(q)} = A^{\mathrm{sym}} \odot \Xi^{(q)}$ satisfies $H^{(q)}_{uv} = \overline{H^{(q)}_{vu}}$, hence $H^{(q)}$ is Hermitian. The matrix $D^{-1/2}$ is real diagonal, hence Hermitian, so L_{mag} is Hermitian. For $z \in \mathbb{C}^n$ define $\widehat{z}_u = z_u/\sqrt{d_u}$. Using $H^{(q)}_{uv} = A^{\mathrm{sym}}_{uv}\Xi^{(q)}_{uv}$ and $|\Xi^{(q)}_{uv}| = 1$, one obtains

$$z^* L_{\mathrm{mag}} z = \frac{1}{2} \sum_{u,v\in V} A^{\mathrm{sym}}_{uv} \left|\widehat{z}_u - \Xi^{(q)}_{uv}\widehat{z}_v\right|^2 \geq 0, \tag{15}$$

which proves positive semidefiniteness.

3.4. Contrastive Objectives

An InfoNCE objective is applied to pairs of stochastic views of the same molecule (Oord et al., 2018). Let $\mathcal{X}$ be the space of molecular inputs. A stochastic augmentation kernel $\mathcal{T}$ maps $x \in \mathcal{X}$ to a random view $\widetilde{x} \sim \mathcal{T}(\cdot \mid x)$. An encoder $f_\theta : \mathcal{X} \to \mathbb{R}^p$ produces embeddings $h = f_\theta(\widetilde{x})$. For each x, sample two views $\widetilde{x}^{(1)}, \widetilde{x}^{(2)} \sim \mathcal{T}(\cdot \mid x)$ and set $h^{(k)} = f_\theta(\widetilde{x}^{(k)})$.

For a mini batch of size B, write $(h^{(1)}_i, h^{(2)}_i)$ for the pair from the ith molecule. With cosine similarity

$$\mathrm{sim}(u,v) = \frac{\langle u, v\rangle}{\|u\|_2\,\|v\|_2}, \tag{16}$$

and temperature $\tau > 0$, define the batchwise scores

$$s^{12}_{ij} = \mathrm{sim}\big(h^{(1)}_i, h^{(2)}_j\big), \qquad s^{21}_{ij} = \mathrm{sim}\big(h^{(2)}_i, h^{(1)}_j\big). \tag{17}$$

The symmetric InfoNCE loss is

$$\mathcal{L}_{\mathrm{NCE}} = \frac{1}{2B} \sum_{i=1}^{B} \big(\ell^{(1)}_i + \ell^{(2)}_i\big), \tag{18}$$

where

$$\ell^{(1)}_i = -\log \frac{\exp\big(s^{12}_{ii}/\tau\big)}{\sum_{j=1}^{B} \exp\big(s^{12}_{ij}/\tau\big)}, \tag{19}$$

$$\ell^{(2)}_i = -\log \frac{\exp\big(s^{21}_{ii}/\tau\big)}{\sum_{j=1}^{B} \exp\big(s^{21}_{ij}/\tau\big)}. \tag{20}$$

In ChemHyperMag, $\mathcal{T}$ perturbs $\Xi^{(q)}$ while keeping the molecular hypergraph $\mathcal{H} = (V, E)$ fixed, and $\mathcal{L}_{\mathrm{NCE}}$ is combined with $\mathcal{L}_{\mathrm{sup}}$ in ***Equation*** (2).

4. Methodology

ChemHyperMag combines a functional group hypergraph representation with magnetic spectral propagation for multitask molecular prediction. Hypergraph construction and notation follow Section 3.2. The encoder instantiates the magnetic operator of Section 3.3 from atom level chemical potential in ***Equation*** (7), applies a Chebyshev spectral filter on the resulting complex magnetic Laplacian in ***Equation*** (14), and produces real node embeddings used by task specific pooling and predictors. Training minimizes the masked supervised loss in ***Equation*** (2) and the contrastive objective in ***Equation*** (18). The model framework of ChemHyperMag is shown in ***Figure 2***.

4.1. Functional Group Hypergraph

For each molecule $G = (V, B, X)$, we construct the hypergraph $\mathcal{H} = (V, E)$ as in Section 3.2, with hyperedges from SSSR rings (Figueras, 1996), BRICS fragments (Degen et al., 2008), the Bemis-Murcko scaffold (Bemis & Murcko, 1996), and bond hyperedges $\{u, v\}$ for all $(u, v) \in B$. We implement $\mathcal{H}$ as an incidence graph whose node set augments atoms with one node per hyperedge, and whose directed edges connect each hyperedge node to its incident atom nodes in both directions. The heterogeneous input projector maps both node types to a shared width. For atom nodes, we apply a linear map to the first d_{atom} input channels. For hyperedge nodes, we apply a linear map to the first d_{he} channels. We restore the original node order, then apply LayerNorm and ReLU. This yields real initial states $H^{(0)} \in \mathbb{R}^{n\times d_h}$ on the incidence graph.

4.2. Magnetic Operator from Chemical Potential

When using magnetic propagation, each node u is assigned electronegativity $\chi(u)$ and Gasteiger charge $Q(u)$, and the chemical potential is computed by ***Equation*** (7). In the incidence graph, $\chi(u) = Q(u) = 0$ for hyperedge nodes, while

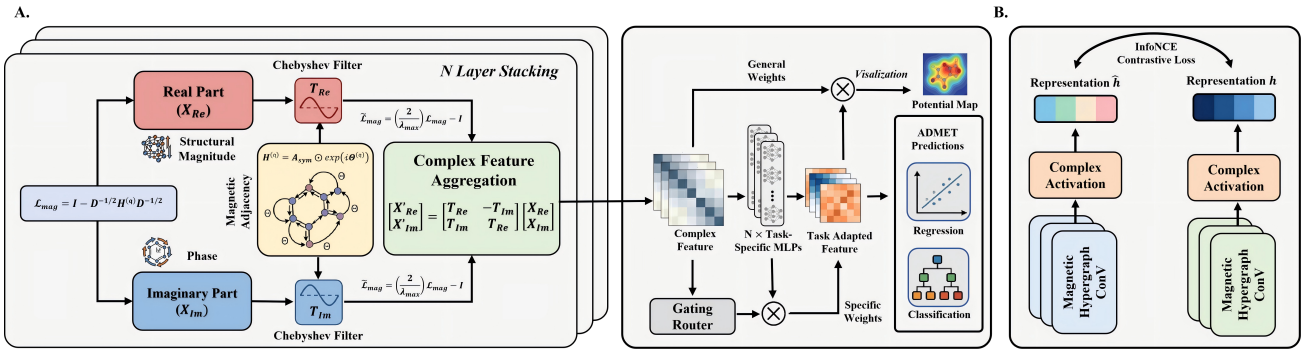

Figure 2. The model framework of ChemHyperMag. **A.** Internal structure of the Magnetic Chebyshev Encoder and subsequent gating mechanism; **B.** Brief overview of contrastive learning.

atom nodes use the stored descriptors, with electronegativity rescaled by a constant factor 4.0 in the implementation. Let $\mathrm{Adj} \in \{0,1\}^{n\times n}$ be the directed adjacency matrix of the incidence graph. We construct a biased transition matrix P by assigning unnormalized weights

$$W_{uv} = \mathrm{Adj}_{uv} \exp\big(\phi(v) - \phi(u)\big), P_{uv} = \frac{W_{uv}}{\sum_w W_{uw}}, \quad (21)$$

and setting $P_{u\cdot}$ to the uniform distribution when $\sum_w W_{uw}$ is numerically zero. We then apply teleportation with uniform ν and rate $\eta \in (0,1)$ using ***Equation*** (9), implemented as $\nu_u = 1/n$ and $\eta = \tau$ with $\tau = 0.15$. The stationary distribution π is computed by a fixed number of power iterations, initialized uniformly and iterated as $\pi \leftarrow (P^{(\eta)})^\top \pi$ for 20 steps with renormalization.

Given $P^{(\eta)}$ and π, we form the stationary flow matrix

$$\widetilde{A} = \mathrm{diag}(\pi)\, P^{(\eta)}, \quad (22)$$

and decompose it into $\widetilde{A}^{\mathrm{sym}}$ and $\widetilde{A}^{\mathrm{skew}}$ via ***Equation*** (10). We then compute the phase ratio and phase map exactly as in ***Equation*** (11) with $A^{\mathrm{sym}} = \widetilde{A}^{\mathrm{sym}}$, $A^{\mathrm{skew}} = \widetilde{A}^{\mathrm{skew}}$, $\varepsilon = 10^{-7}$, and phase strength q. The phase factors $\Xi^{(q)}$ and magnetic adjacency $H^{(q)}$ are computed by ***Equation*** (12) and ***Equation*** (13), and the normalized magnetic Laplacian L_{mag} follows ***Equation*** (14) with degrees computed from $\widetilde{A}^{\mathrm{sym}}$.

4.3. Magnetic Chebyshev Encoder

Each magnetic layer applies a Chebyshev filter on the normalized operator $\widetilde{L} = -D^{-1/2} H^{(q)} D^{-1/2}$, which equals $L_{\mathrm{mag}} - I$. Given real input features $h \in \mathbb{R}^{n\times d_{\mathrm{in}}}$, we first apply a real linear map $x = Wh$ and lift it to complex by $x_{\mathbb{C}} = x + \mathrm{i}0$. We then compute Chebyshev iterates by

$$\begin{aligned} T_0 &= x_{\mathbb{C}}, \quad T_1 = \widetilde{L}\, x_{\mathbb{C}}, \\ T_k &= 2\widetilde{L}\, T_{k-1} - T_{k-2}, \quad (k \geq 2), \end{aligned} \quad (23)$$

where each multiplication by $\widetilde{L}$ is implemented as $u \mapsto -D^{-1/2} H^{(q)} D^{-1/2} u$ using degrees from $\widetilde{A}^{\mathrm{sym}}$. The layer output before nonlinearity follows the ChebNetII reparametrization (He et al., 2022a), which learns the filter response at Chebyshev interpolation nodes and builds on the Chebyshev spectral filtering of ChebNet (Defferrard et al., 2016). We write the expansion as

$$z = \frac{c_0}{2} T_0 + \sum_{k=1}^{K} c_k T_k, \quad (24)$$

with coefficients c_k derived from nonnegative parameters $\tilde{c}_j = \mathrm{ReLU}(t_j)$. Let $x_j = \cos\big((K - j + 0.5)\pi/(K+1)\big)$ for $j = 0, \ldots, K$. The coefficients are

$$c_k = \frac{2}{K+1} \sum_{j=0}^{K} \tilde{c}_j\, T_k(x_j), \quad (25)$$

where $T_k(\cdot)$ denotes the scalar Chebyshev polynomial.

To return physically interpretable real representations while still using phase during propagation, we apply a magnitude based activation that does not separate real and imaginary parts. Concretely, we map z to real features by $|z| = \sqrt{\Re(z)^2 + \Im(z)^2}$ elementwise and apply ReLU on this magnitude,

$$\tilde{h} = \mathrm{ReLU}(|z| + b)\,, \quad (26)$$

with a learned bias b. We do not apply ReLU to $\Re(z)$ and $\Im(z)$ separately, since independent truncation modifies phase in an uncontrolled manner and conflicts with the interpretation of phase as directed bias. In ***Equation*** (26) the phase influences message mixing through $H^{(q)}$, while the stored features remain real and nonnegative. Each layer adds a residual term $\mathrm{ReLU}(W_{\mathrm{res}} h)$ and applies batch normalization to the resulting real features. The encoder stacks three such layers.

4.4. Gating Mechanism

From the final node states $H^{(L)}$, we compute one molecular embedding per task by gated sum pooling. For task t, a learned scalar gate $w_{u,t} = \sigma(a_t^\top h_u^{(L)})$ is applied to each node, and the task embedding is $z_t = \sum_u w_{u,t} h_u^{(L)}$. Each task uses its own two layer MLP head with dropout, ReLU, and batch normalization, followed by a linear output.

When a primary task index p is specified, we fuse auxiliary task embeddings into the primary pathway using a graph level gate conditioned on the mean node embedding $h_G = \frac{1}{n}\sum_u h_u^{(L)}$. For each auxiliary task $t \neq p$, we compute the batch averaged cosine similarity between z_p and z_t, sort tasks by this score, then select a subset $\mathcal{A}_p$ by first taking tasks with similarity above a fixed threshold and, if fewer than a prescribed minimum are selected, adding the top scoring tasks until the minimum is reached. For each $t \in \mathcal{A}_p$, a linear map produces a two dimensional logit vector from h_G, a softmax yields $\gamma_t \in \mathbb{R}^2$, and we form a convex combination $\tilde{z}_{p,t} = \gamma_{t,1} z_t + \gamma_{t,2} z_p$. The fused representation is $z_p^{\text{fused}} = \sum_{t \in \mathcal{A}_p} \tilde{z}_{p,t}$, and when $\mathcal{A}_p$ is empty, we set $z_p^{\text{fused}} = z_p$. Only the primary head consumes z_p^{fused}, while all other heads use their own z_t.

4.5. Contrastive Learning

Contrastive learning uses two stochastic views that share Adj, ϕ, $P^{(\eta)}$, π, and $\widetilde{A}^{\text{sym}}$, and differ only in the magnetic phases. Starting from $\Theta^{(q)}$ computed by ***Equation*** (11), we sample an elementwise noise matrix Δ with independent entries $\Delta_{uv} \sim \mathrm{Unif}(-\rho, \rho)$, and define the perturbed phase

$$\widetilde{\Theta}^{(q)} = \Theta^{(q)} + \Delta. \tag{27}$$

This perturbs only the phase $\Theta^{(q)}$, leaving the hypergraph $\mathcal{H} = (V, E)$ and magnitude $\widetilde{A}^{\text{sym}}$ fixed, so atoms, bonds, motifs, and diffusion strength are unchanged. Since ϕ is only an estimate from electronegativity and Gasteiger charges, a small zero-mean $\Delta_{uv} \sim \mathrm{Unif}(-\rho, \rho)$ yields a plausible directional variant of the same molecule, forming a valid positive pair as in operator-level directed graph contrastive learning (Tong et al., 2021).

We then construct the perturbed phase factors entrywise by

$$\widetilde{\Xi}_{uv}^{(q)} = \cos\big(\widetilde{\Theta}_{uv}^{(q)}\big) + \mathrm{i}\sin\big(\widetilde{\Theta}_{uv}^{(q)}\big), \tag{28}$$

and the perturbed magnetic adjacency is $\widetilde{H}^{(q)} = \widetilde{A}^{\text{sym}} \odot \widetilde{\Xi}^{(q)}$. This is the only augmentation applied inside the magnetic operator, with ρ implemented as a scalar perturbation passed to the adjacency constructor. For supervised prediction, we set $\rho = 0$; and for contrastive views, we sample two independent noise matrices to obtain two independent phase perturbations. The two views are encoded and trained with the symmetric InfoNCE loss in ***Equation*** (18). Supervision uses the masked multitask loss in ***Equation*** (2), with MSE for regression tasks and binary cross entropy for classification tasks. The final objective is $\mathcal{L} = \mathcal{L}_{\text{sup}} + \lambda\, \mathcal{L}_{\text{NCE}}$ with $\lambda > 0$.

Table 1. Model performance on selected MTGL-ADMET datasets.

Endpoint	Metric	[ST]-GCN	[ST]-MGA	[MT]-GCN	MGA	MTGL	ChemBFN	Uni-Mol	Ours
Distribution									
PPB	R^2	.577$_{.028}$	.585$_{.004}$	.589$_{.036}$	.568$_{.038}$	<u>.626</u>$_{.029}$	.537$_{.030}$	.562$_{.028}$	**.687**$_{.012}$
Toxicity									
Hepatotoxicity	ROC-AUC	.653$_{.040}$	.669$_{.022}$	.612$_{.039}$	**.713**$_{.053}$	<u>.701</u>$_{.036}$	.697$_{.041}$	.660$_{.037}$	.664$_{.019}$
Respiratory tox.	ROC-AUC	.842$_{.018}$	**.872**$_{.013}$	.810$_{.014}$	.828$_{.021}$	.859$_{.010}$	.818$_{.028}$	.865$_{.015}$	**.872**$_{.007}$
Cardiotoxicity-1	ROC-AUC	.707$_{.026}$	.703$_{.020}$	.683$_{.028}$	.684$_{.023}$	.740$_{.023}$	.745$_{.027}$	<u>.761</u>$_{.024}$	**.816**$_{.010}$
Physicochemical									
IGC50	R^2	.703$_{.055}$	.818$_{.021}$	.618$_{.027}$	.772$_{.021}$	.819$_{.008}$	.797$_{.014}$	**.832**$_{.012}$	<u>.830</u>$_{.010}$
ESOL	R^2	.814$_{.030}$	.896$_{.013}$	.824$_{.030}$	.866$_{.020}$	<u>.931</u>$_{.038}$	.907$_{.042}$	.920$_{.039}$	**.936**$_{.016}$

Note: [ST] denotes single-task training and **[MT]** denotes multi-task training with a shared encoder and task-specific heads, while **GCN** refers to graph convolutional networks and **MGA** refers to multitask graph attention. Results are reported as mean±std under a random $80/10/10$ split. Baseline numbers are taken from the original papers with 10 repetitions for MTGL-ADMET baselines (Du et al., 2023b) and MTGIB-UNet baselines (Li et al., 2025). We evaluated ChemHyperMag under the same protocol with 10 runs. The best result(s) are in **bold**, and the second-best result is <u>underlined</u>. For tied best results, all tied entries are highlighted in **bold**.

5. Experiments

5.1. Experimental Setup

We evaluated ChemHyperMag on six ADMET endpoints: PPB, IGC50, ESOL, Hepatotoxicity, Respiratory toxicity, and Cardiotoxicity-1, including three regression tasks and three classification tasks (based on MTGL-ADMET (Du et al., 2023b) and MoleculeNet Leaderboard (Wu et al., 2018) for fair comparisons). Following the baseline setting, each dataset was randomly split into training, validation, and test sets in an 8:1:1 ratio. Results were reported as mean $\pm$ standard deviation over ten runs with different random seeds. All methods used the same RDKit atom featurization. ChemHyperMag further augmented the atom features with Pauling electronegativity and RDKit Gasteiger partial charges, and built the corresponding hypergraph incidence structure for the encoder. We adopted ROC-AUC for classification and R^2 for regression, where higher is better.

5.2. Baselines

We compared against single task and multitask baselines in ADMET prediction. Single task baselines trained one model per endpoint with a GCN (Kipf, 2016) or a graph attention model (Veličković et al., 2017). Multitask graph baselines used a shared encoder with task-specific heads, matching the supervision setting in ***Equation*** (2). We additionally included recent multitask frameworks and pretrained molecular models reported in the benchmark, including MTGL (Du et al., 2023a), ChemBERTa (Chithrananda et al., 2020), Uni-Mol (Zhou et al., 2023), ADMET-AI (Swanson et al., 2024), and Chemprop v2 (Graff et al., 2025).

5.3. Implementation Details

ChemHyperMag was implemented in PyTorch and DGL and trained on NVIDIA A100 GPUs. Hyperedges were constructed from SSSR rings, BRICS fragments, Bemis-Murcko scaffolds, and bond hyperedges as described in Section 3.2. The magnetic operator was computed from the teleported Markov chain in ***Equation*** (9) and the magnetic adjacency in ***Equation*** (13). We set the magnetic strength $q = 0.1$ and the teleportation rate $\eta = 0.15$. The encoder used three magnetic Chebyshev layers with hidden width 64, intermediate molecular width 128, and output width 64, with Chebyshev order $K = 2$. We trained with Adam using learning rate 10^{-3} and weight decay 10^{-5}, batch size 128, and up to 200 epochs. We applied early stopping based on the validation score aggregated over tasks. For contrastive training, we used the symmetric InfoNCE loss in ***Equation*** (18) with temperature $\tau = 0.07$.

5.4. Main Results

Table 1 reports results on the MTGL-ADMET datasets. ChemHyperMag achieves the best or tied-best result on four of the six selected endpoints and ranks second on IGC50. It improves the strongest baseline by +0.061 on PPB (0.687 vs. 0.626) and by +0.055 on Cardiotoxicity-1 (0.816 vs. 0.761), matches the best on Respiratory toxicity (0.872), and attains the highest ESOL score (0.936). On IGC50 it stays within 0.002 of the top result (0.830 vs. 0.832). It trails the leading baselines only on Hepatotoxicity, where MGA reaches 0.713 and ChemHyperMag obtains 0.664.

The ablation in ***Table 2*** uses MoleculeNet. The full model is strongest on five of the six endpoints, showing that ChebNet, the magnetic Laplacian, hypergraph construction, and contrastive learning contribute complementary gains. Cardiotoxicity-1 is the most sensitive endpoint. Removing the hypergraph, magnetic Laplacian, contrastive objective, or ChebNet lowers it from 0.851 to 0.720, 0.734, 0.749, and 0.767. PPB drops sharply without ChebNet (0.439) or the hypergraph (0.479). Respiratory toxicity is the only endpoint where an ablated variant beats the full model, indicating a task-specific trade-off rather than a uniform gain.

6. Limitations

Random splits and masked supervision may overestimate scaffold generalization and bias multitask effects under missing not at random labels. The directional signal relies on a heuristic chemical potential from electronegativity and Gasteiger charges, so outcomes can be sensitive to charge and protonation protocols and remain weakly mechanistically grounded for ADMET. Fixed step stationary estimation and complex spectral propagation introduce approximation and stability risks, and magnitude projection removes phase information. Phase-only contrastive learning regularizes sparse supervision but imposes a narrow invariance that may not be chemically faithful. Future work should use scaffold and distribution shift splits, stress test charge and protonation settings, and adopt convergence-controlled solvers with scaling analyses.

Table 2. Full ablation study results on MoleculeNet benchmark.

	Regression (R^2)			*Classification (ROC-AUC)*		
Variant	**PPB**	**IGC50**	**ESOL**	**Hepato.**	**Resp.**	**Cardio-1**
Base Model (All modules off)						
Base Model	$.376_{.161}$	$.709_{.078}$	$.829_{.083}$	$.547_{.060}$	$.843_{.038}$	$.745_{.053}$
Single Module Enabled						
Only ChebNet	$\mathbf{.418}_{.135}$	$\mathbf{.790}_{.049}$	$\underline{.892}_{.045}$	$\underline{.628}_{.103}$	$.794_{.059}$	$.730_{.064}$
Only Magnetic	$\underline{.376}_{.161}$	$.709_{.078}$	$.829_{.083}$	$.547_{.060}$	$\mathbf{.843}_{.038}$	$\mathbf{.745}_{.053}$
Only Hypergraph	$.363_{.250}$	$\underline{.742}_{.043}$	$\mathbf{.905}_{.019}$	$\mathbf{.640}_{.079}$	$\underline{.842}_{.038}$	$\underline{.741}_{.115}$
Two Modules Enabled						
ChebNet + CL	$.418_{.135}$	$\underline{.790}_{.049}$	$.892_{.045}$	$\underline{.628}_{.103}$	$.794_{.059}$	$\underline{.730}_{.064}$
Magnetic + CL	$.376_{.161}$	$.709_{.078}$	$.829_{.083}$	$.547_{.060}$	$\mathbf{.843}_{.038}$	$\mathbf{.745}_{.053}$
Magnetic + ChebNet	$.418_{.135}$	$\underline{.790}_{.049}$	$.892_{.045}$	$\underline{.628}_{.103}$	$.794_{.059}$	$\underline{.730}_{.064}$
Hypergraph + CL	$\underline{.428}_{.119}$	$\mathbf{.802}_{.042}$	$\mathbf{.905}_{.020}$	$\mathbf{.639}_{.083}$	$.808_{.032}$	$.705_{.151}$
Hypergraph + ChebNet	$.423_{.214}$	$.765_{.012}$	$.847_{.061}$	$.588_{.010}$	$\underline{.837}_{.005}$	$.723_{.107}$
Hypergraph + Magnetic	$\mathbf{.451}_{.093}$	$.756_{.097}$	$\underline{.902}_{.006}$	$.582_{.006}$	$.815_{.034}$	$.723_{.124}$
Leave-One-Out						
w/o ChebNet	$.439_{.000}\downarrow$	$.789_{.000}\downarrow$	$\underline{.915}_{.000}$	$\mathbf{.655}_{.000}$	$.813_{.000}$	$\mathbf{.767}_{.000}$
w/o Hyp	$.479_{.000}$	$\mathbf{.802}_{.000}$	$.912_{.000}$	$\underline{.647}_{.000}$	$\mathbf{.869}_{.000}$	$.720_{.000}\downarrow$
w/o MagLap	$\underline{.505}_{.020}$	$\underline{.798}_{.015}$	$\mathbf{.920}_{.015}$	$.606_{.020}\downarrow$	$.813_{.005}\downarrow$	$.734_{.017}$
w/o CL	$\mathbf{.523}_{.043}$	$.797_{.016}$	$.907_{.032}\downarrow$	$.641_{.029}$	$\underline{.829}_{.011}$	$\underline{.749}_{.031}$
Full Model	.544	.806	.925	.669	.835	.851

Note: Bold/underline/↓ indicate highest/second-highest performance among ablated models and the largest drop relative to the Full Model, respectively. Ties are marked for all corresponding entries. Subscripts denote the absolute difference between test and validation results.

7. Conclusion

We propose ChemHyperMag for multitask ADMET prediction with sparse labels. It builds a functional group hypergraph, applies magnetic spectral propagation to encode directional effects, and adds phase perturbation contrastive learning as regularization. ChemHyperMag shows strong performance, and ablation studies confirm that the hypergraph, magnetic operator, and contrastive learning each contribute to the gains.

Conflict of Interest Disclosure

There are no conflicts of interest among all authors.

Impact Statement

ChemHyperMag is a physics-informed, conformer-free, and lightweight model for small-molecule ADMET prediction. Our code can be found at `https://github.com/1anj/ChemHyperMag`.

Acknowledgements

This work was fully supported by a Collaborative Research Fund (Project No. C5082-25E) and a General Research Fund (Project No. 15101422) from the Research Grants Council of the Hong Kong Special Administrative Region, P. R. China, and an internal grant (P0051278) from The Hong Kong Polytechnic University.

References

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

Benko, T., Buck, M., Amburg, I., Young, S. J., and Aksoy, S. G. Hypermagnet: A magnetic laplacian based hypergraph neural network. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=Gdf4P7sEzE.

Bento, A. P., Hersey, A., Félix, E., Landrum, G., Gaulton, A., Atkinson, F., Bellis, L. J., De Veij, M., and Leach, A. R. An open source chemical structure curation pipeline using rdkit. *Journal of Cheminformatics*, 12(1):51, 2020.

Chen, F., Park, J., and Park, J. A hypergraph convolutional neural network for molecular properties prediction using functional group. *arXiv*, 2021.

Chen, J. and Schwaller, P. Molecular hypergraph neural networks. *The Journal of Chemical Physics*, 160(14), 2024.

Chien, E., Pan, C., Peng, J., and Milenkovic, O. You are allset: A multiset function framework for hypergraph neural networks. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=hpBTIv2uy_E.

Chithrananda, S., Grand, G., and Ramsundar, B. Chemberta: large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885*, 2020.

Daina, A., Michielin, O., and Zoete, V. Swissadme: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific reports*, 7(1):42717, 2017.

Defferrard, M., Bresson, X., and Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

Degen, J., Wegscheid-Gerlach, C., Zaliani, A., and Rarey, M. On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem*, 3(10):1503, 2008.

Du, B.-X., Xu, Y., Yiu, S.-M., Yu, H., and Shi, J.-Y. Admet property prediction via multi-task graph learning under adaptive auxiliary task selection. *Iscience*, 26(11), 2023a.

Du, B.-X., Xu, Y., Yiu, S.-M., Yu, H., and Shi, J.-Y. Mtgl-admet: a novel multi-task graph learning framework for admet prediction enhanced by status-theory and maximum flow. In *International Conference on Research in Computational Molecular Biology*, pp. 85–103. Springer, 2023b.

Duta, I., Cassarà, G., Silvestri, F., and Liò, P. Sheaf hypergraph networks. *Advances in Neural Information Processing Systems*, 36:12087–12099, 2023.

Fallani, A., Nugmanov, R., Arjona-Medina, J., Wegner, J. K., Tkatchenko, A., and Chernichenko, K. Pretraining graph transformers with atom-in-a-molecule quantum properties for improved admet modeling. *Journal of Cheminformatics*, 17(1):25, 2025.

Fanuel, M., Alaiz, C. M., and Suykens, J. A. Magnetic eigenmaps for community detection in directed networks. *Physical Review E*, 95(2):022302, 2017.

Feinberg, E. N., Sur, D., Wu, Z., Husic, B. E., Mai, H., Li, Y., Sun, S., Yang, J., Ramsundar, B., and Pande, V. S. Potentialnet for molecular property prediction. *ACS central science*, 4(11):1520–1530, 2018.

Feng, Y., You, H., Zhang, Z., Ji, R., and Gao, Y. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 3558–3565, 2019.

Figueras, J. Ring perception using breadth-first search. *Journal of chemical information and computer sciences*, 36 (5):986–991, 1996.

Fralish, Z., Chen, A., Skaluba, P., and Reker, D. Deepdelta: predicting admet improvements of molecular derivatives with deep learning. *Journal of cheminformatics*, 15(1): 101, 2023.

Fu, L., Shi, S., Yi, J., Wang, N., He, Y., Wu, Z., Peng, J., Deng, Y., Wang, W., Wu, C., et al. Admetlab 3.0: an updated comprehensive online admet prediction platform enhanced with broader coverage, improved performance, api functionality and decision support. *Nucleic acids research*, 52(W1):W422–W431, 2024.

Gasteiger, J. and Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*, 36(22):3219–3228, 1980.

Graff, D. E., Morgan, N. K., Burns, J. W., Doner, A. C., Li, B., Li, S.-C., Manu, J., Menon, A., Pang, H.-W., Wu, H., et al. Chemprop v2: An efficient, modular machine learning package for chemical property prediction. *Journal of Chemical Information and Modeling*, 2025.

He, D., Shan, L., Zhao, J., Zhang, H., Wang, Z., and Zhang, W. Exploitation of a latent mechanism in graph contrastive learning: Representation scattering. *Advances in Neural Information Processing Systems*, 37:115351–115376, 2024.

He, M., Wei, Z., and Wen, J.-R. Convolutional neural networks on graphs with chebyshev approximation, revisited. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022a. URL https://openreview.net/forum?id=jxPJ4QA0KAb.

He, Y., Perlmutter, M., Reinert, G., and Cucuringu, M. Msgnn: A spectral graph neural network based on a novel magnetic signed laplacian. In *Learning on graphs conference*, pp. 40–1. PMLR, 2022b.

Heid, E., Greenman, K. P., Chung, Y., Li, S.-C., Graff, D. E., Vermeire, F. H., Wu, H., Green, W. H., and McGill, C. J. Chemprop: a machine learning package for chemical property prediction. *Journal of Chemical Information and Modeling*, 64(1):9–17, 2023.

Huang, K., Fu, T., Gao, W., Zhao, Y., Roohani, Y., Leskovec, J., Coley, C. W., Xiao, C., Sun, J., and Zitnik, M. Artificial intelligence foundation for therapeutic science. *Nature chemical biology*, 18(10):1033–1036, 2022.

Kajino, H. Molecular hypergraph grammar with its application to molecular optimization. In *International Conference on Machine Learning*, pp. 3183–3191. PMLR, 2019.

Kim, J., Chang, W., Ji, H., and Joung, I. Quantum-informed molecular representation learning enhancing admet property prediction. *Journal of Chemical Information and Modeling*, 64(13):5028–5040, 2024.

Kipf, T. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Ko, T., Choi, Y., and Kim, C.-K. Signed directed graph contrastive learning with laplacian augmentation. *arXiv preprint arXiv:2301.05163*, 2023a.

Ko, T., Choi, Y., and Kim, C.-K. A spectral graph convolution for signed directed graphs via magnetic laplacian. *Neural Networks*, 164:562–574, 2023b.

Ko, T., Choi, Y., and Kim, C.-K. Universal graph contrastive learning with a novel laplacian perturbation. In *Uncertainty in Artificial Intelligence*, pp. 1098–1108. PMLR, 2023c.

Li, X., Du, W., Xia, J., Wang, J., Wang, X., Yang, Y., and Wang, Y. Mtgib-unet: A multi-task graph information bottleneck and uncertainty weighted network for admet prediction. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 7518–7526, 2025.

Lin, S., Liu, C., Zhou, P., Hu, Z.-Y., Wang, S., Zhao, R., Zheng, Y., Lin, L., Xing, E., and Liang, X. Prototypical graph contrastive learning. *IEEE transactions on neural networks and learning systems*, 35(2):2747–2758, 2022.

Liu, H., Huang, Y., Liu, X., and Deng, L. Attention-wise masked graph contrastive learning for predicting molecular property. *Briefings in bioinformatics*, 23(5):bbac303, 2022a.

Liu, S., Wang, H., Liu, W., Lasenby, J., Guo, H., and Tang, J. Pre-training molecular graph representation with 3d geometry. In *International Conference on Learning Representations*, 2022b. URL https://openreview.net/forum?id=xQUe1pOKPam.

Lu, S., Gao, Z., He, D., Zhang, L., and Ke, G. Data-driven quantum chemical property prediction leveraging 3d conformations with uni-mol+. *Nature communications*, 15 (1):7104, 2024.

Ma, Z., Zhao, W., and Yang, Z. Directed hypergraph representation learning for link prediction. In *International conference on artificial intelligence and statistics*, pp. 3268–3276. PMLR, 2024.

Montanari, F., Kuhnke, L., Ter Laak, A., and Clevert, D.-A. Modeling physico-chemical admet endpoints with multitask graph convolutional networks. *Molecules*, 25 (1):44, 2019.

Moon, K., Im, H.-J., and Kwon, S. 3d graph contrastive learning for molecular property prediction. *Bioinformatics*, 39(6):btad371, 2023.

Mule, E., Fiorini, S., Purificato, A., Siciliano, F., Coniglio, S., and Silvestri, F. Directional sheaf hypergraph networks: Unifying learning on directed and undirected hypergraphs. *arXiv preprint arXiv:2510.04727*, 2025.

Oord, A. v. d., Li, Y., and Vinyals, O. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Pauling, L. *The nature of the chemical bond and the structure of molecules and crystals: an introduction to modern structural chemistry*, volume 18. Cornell university press, 1960.

Pires, D. E., Blundell, T. L., and Ascher, D. B. pkcsm: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of medicinal chemistry*, 58(9):4066–4072, 2015.

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

Schyman, P., Liu, R., Desai, V., and Wallqvist, A. vnn web server for admet predictions. *Frontiers in pharmacology*, 8:889, 2017.

Su, D., Zhang, Y., Li, X., Li, R.-H., and Wang, G. Toward general digraph contrastive learning: A dual spatial perspective. *arXiv preprint arXiv:2510.16311*, 2025.

Suresh, S., Li, P., Hao, C., and Neville, J. Adversarial graph augmentation to improve graph contrastive learning. *Advances in Neural Information Processing Systems*, 34: 15920–15933, 2021.

Swanson, K., Walther, P., Leitz, J., Mukherjee, S., Wu, J. C., Shivnaraine, R. V., and Zou, J. Admet-ai: a machine learning admet platform for evaluation of large-scale chemical libraries. *Bioinformatics*, 40(7):btae416, 2024.

Tao, N. and Abe, M. Bayesian flow network framework for chemistry tasks. *Journal of Chemical Information and Modeling*, 65(3):1178–1187, 2025.

Tian, H., Ketkar, R., and Tao, P. Admetboost: a web server for accurate admet prediction. *Journal of molecular modeling*, 28(12):408, 2022.

Tong, Z., Liang, Y., Ding, H., Dai, Y., Li, X., and Wang, C. Directed graph contrastive learning. *Advances in neural information processing systems*, 34:19580–19593, 2021.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Wang, B., Li, J., Zhou, D., Li, L., Li, J., Wang, E., Hao, J., Shi, L., Lu, C., Qiu, J., et al. Unified and explainable molecular representation learning for imperfectly annotated data from the hypergraph view. *Nature Communications*, 16(1):1–18, 2025a.

Wang, Y., Wang, J., Cao, Z., and Barati Farimani, A. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4(3):279–287, 2022.

Wang, Y., Li, Y., Liu, L., Hong, P., and Xu, H. Advancing drug discovery with enhanced chemical understanding via asymmetric contrastive multimodal learning. *Journal of chemical information and modeling*, 2025b.

Wang, Z., Gao, Z., Zheng, H., Zhang, L., Ke, G., et al. Exploring molecular pretraining model at scale. *Advances in Neural Information Processing Systems*, 37:46956–46978, 2024.

Weininger, D. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K., and Pande, V. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., Yin, M., Zeng, X., Wu, C., Lu, A., et al. Admetlab 2.0: an integrated online platform for accurate and comprehensive predictions of admet properties. *Nucleic acids research*, 49(W1):W5–W14, 2021.

Yang, K., Swanson, K., Jin, W., Coley, C., Eiden, P., Gao, H., Guzman-Perez, A., Hopper, T., Kelley, B., Mathea, M., et al. Analyzing learned molecular representations for property prediction. *Journal of chemical information and modeling*, 59(8):3370–3388, 2019.

Yang, M.-R. and Xu, X.-J. Recent advances in hypergraph neural networks: M.-r. yang, x.-j. xu. *Journal of the Operations Research Society of China*, pp. 1–37, 2025.

Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., and Liu, T.-Y. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34:28877–28888, 2021.

You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., and Shen, Y. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33:5812–5823, 2020.

Zhang, X., He, Y., Brugnone, N., Perlmutter, M., and Hirn, M. Magnet: A neural network for directed graphs. *Advances in neural information processing systems*, 34: 27003–27015, 2021.

Zhou, G., Gao, Z., Ding, Q., Zheng, H., Xu, H., Wei, Z., Zhang, L., and Ke, G. Uni-mol: A universal 3d molecular representation learning framework. In *The eleventh international conference on learning representations*, 2023.

Zhu, Y., Xu, Y., Liu, Q., and Wu, S. An empirical study of graph contrastive learning. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021. URL https://openreview.net/forum?id=UuUbIYnHKO.

A. Additional Details

A.1. Notation

Table 3. Summary of notation

Symbol	Description
Problem Setup & Dataset	
$\mathcal{G}$	Set of molecular graphs
T	Number of prediction tasks
N	Number of training molecules
$\mathcal{D}$	Training set $\{(G_i, y_i, m_i)\}_{i=1}^{N}$
G_i	The i-th molecular graph
$y_i \in \mathbb{R}^T$	Multitask label vector
$m_i \in \{0,1\}^T$	Observation mask
f_θ	Model mapping $\mathcal{G} \to \mathbb{R}^T$
ℓ_t	Loss function for task t
$\mathcal{L}_{\text{sup}}, \mathcal{L}_{\text{NCE}}, \lambda$	Supervised / contrastive loss; weight
Molecular Graph	
$G = (V, B, X)$	2D molecular graph
$V = \{1, \ldots, n\}$	Set of atoms (nodes)
$B \subseteq V \times V$	Undirected covalent bonds
$X \in \mathbb{R}^{n \times d}$	Node feature matrix
Functional Group Hypergraph	
$\mathcal{H} = (V, E)$	Hypergraph over atoms
$E = \{e_1, \ldots, e_m\}$	Hyperedges
$\mathbf{M} \in \{0,1\}^{n \times m}$	Incidence matrix
$N(u)$	Hypergraph neighborhood
Adj	Adjacency of incidence graph
Chemical Potential	
$\chi(u), Q(u)$	Electronegativity; partial charge
α, β	Learnable scaling parameters
$\phi(u)$	Potential: $\alpha\chi(u) + \beta Q(u)$
Markov Chain & Flow	
W_{uv}, P	Transition weight; matrix
η, ν	Teleportation rate; distribution
π	Stationary distribution
$\widetilde{A}$	Flow: $\mathrm{diag}(\pi)P^{(\eta)}$
$\widetilde{A}^{\text{sym}}, \widetilde{A}^{\text{skew}}$	Symmetric / skew parts

Symbol	Description
Magnetic Laplacian	
q	Magnetic charge parameter
ε	Stability constant (10^{-7})
R_{uv}	Asymmetry ratio
$\Theta_{uv}^{(q)}$	Phase: $2\pi q \tanh(R_{uv})$
$\Xi^{(q)}$	Phase factor: $e^{\mathrm{i}\Theta_{uv}^{(q)}}$
$H^{(q)}$	Magnetic adjacency
D, d_u	Degree matrix; node degree
L_{mag}	Magnetic Laplacian
$\widetilde{L}$	Shifted: $L_{\text{mag}} - I$
$\odot$, i	Hadamard product; imaginary unit
Chebyshev Encoder	
K	Chebyshev order
T_k	k-th Chebyshev iterate
c_k, $\tilde{c}_j$	Expansion coefficients
$z \in \mathbb{C}^{n \times d_h}$	Complex layer output
$\lvert z \rvert$	Element-wise magnitude
$H^{(0)}, H^{(L)}$	Initial / final states
d_h, L	Hidden dim; num layers
Gating & Task Fusion	
a_t, $w_{u,t}$	Gate vector; weight
z_t	Task embedding
h_G	Graph mean embedding
$\mathcal{A}_p$, γ_t	Auxiliary tasks; gate
z_p^{fused}	Fused embedding
Contrastive Learning	
ρ, Δ	Perturbation magnitude; noise
$\widetilde{\Theta}^{(q)}$	Perturbed phase
$h^{(1)}, h^{(2)}$	Two-view embeddings
$\mathcal{B}$, $\lvert\mathcal{B}\rvert$	Mini-batch; size
$\mathrm{sim}(u, v)$	Cosine similarity
τ	InfoNCE temperature

A.2. Datasets

We evaluate ChemHyperMag on six selected ADMET endpoints drawn from the 24-endpoint benchmark, namely PPB, IGC50, ESOL, Hepatotoxicity, Respiratory toxicity, and Cardiotoxicity-1. These tasks cover distribution, toxicity, and physicochemical property prediction, and include three regression tasks and three binary classification tasks. PPB measures plasma protein binding and reflects the free fraction of a compound available for distribution, metabolism, and excretion. IGC50 measures 48-hour *Tetrahymena pyriformis* growth inhibition and is commonly used as an indicator of general cytotoxicity and environmental toxicity. ESOL measures aqueous solubility, which is directly related to dissolution-limited absorption and formulation feasibility. Hepatotoxicity predicts drug-induced liver injury risk, Respiratory toxicity captures pulmonary adverse effects such as fibrosis and bronchospasm, and Cardiotoxicity-1 corresponds to the strictest hERG-related threshold ($\mathrm{IC}_{50} < 1\,\mu\mathrm{M}$), identifying compounds with high cardiotoxic risk.

The selected datasets range from 1,128 to 1,830 molecules. Among the classification tasks, Hepatotoxicity contains 782 positive and 531 negative samples, Respiratory toxicity contains 844 positive and 555 negative samples, and Cardiotoxicity-1 contains 350 positive and 1,215 negative samples, indicating different degrees of class imbalance. The full statistics,

including data split sizes and task types, are summarized in Table 4.

Table 4. Statistics of the six selected ADMET endpoints.

Endpoint	#Mol	Pos/Neg	Train/Val/Test	Type
PPB	1,830	–	1,464/183/183	Reg
IGC50	1,787	–	1,429/179/179	Reg
ESOL	1,128	–	902/113/113	Reg
Hepatotoxicity	1,313	782/531	1,050/131/132	Cls
Respiratory toxicity	1,399	844/555	1,120/139/140	Cls
Cardiotoxicity-1	1,565	350/1,215	927/235/403	Cls

A.3. Baselines

We compare ChemHyperMag against representative methods spanning single-task models, multi-task architectures, and pretrained foundation models.

A.3.1. Single-Task Methods

ST-GCN (Montanari et al., 2019) applies graph convolutional networks to molecular property prediction. It treats atoms as nodes and bonds as edges, aggregating neighbor information through spectral convolutions on the normalized adjacency matrix. Node features are initialized with atom descriptors such as element type and formal charge. Graph-level representations are obtained via global pooling for endpoint prediction. ST-GCN is trained independently on each ADMET endpoint without cross-task parameter sharing.

ST-MGA (Xiong et al., 2021) extends graph attention networks to molecular graphs. Instead of fixed aggregation weights, ST-MGA computes attention coefficients that dynamically weight neighbor contributions based on their relevance. Multi-head attention enables the model to capture different structural patterns simultaneously. Originally developed for ADMETlab 2.0, ST-MGA is trained separately for each property as the single-task baseline.

Chemprop v2 (Graff et al., 2025) implements directed message passing neural networks (D-MPNN). Unlike node-centric GNNs, D-MPNN propagates messages along directed edges, distinguishing incoming and outgoing bond information to avoid redundant updates. Version 2 introduces significant engineering improvements including 10× faster training, native support for reaction inputs, and uncertainty quantification. Chemprop v2 represents the state-of-the-art single-task GNN baseline and serves as the backbone for several production ADMET tools.

A.3.2. Multi-Task Methods

MT-GCN (Montanari et al., 2019) extends GCN to multi-task learning by sharing the graph encoder across all endpoints while maintaining task-specific prediction heads. This hard parameter sharing provides regularization and enables knowledge transfer between related properties. The model handles missing labels through masked loss computation, updating only on observed endpoints for each molecule.

MGA (Xiong et al., 2021) is the multi-task graph attention architecture powering ADMETlab 2.0 and 3.0. Beyond basic parameter sharing, MGA incorporates bond attributes into attention computation and uses a virtual node connected to all atoms for improved global information flow. Attention weights often highlight pharmacologically relevant substructures such as toxicophores and metabolic soft spots.

MTGL-ADMET (Du et al., 2023b) introduces adaptive auxiliary task selection for multi-task ADMET prediction. The key insight is that not all auxiliary tasks benefit a given primary task; some may cause negative transfer. MTGL-ADMET constructs a task relationship graph and applies status theory with maximum flow optimization to select beneficial auxiliaries for each endpoint. This dynamic selection mechanism achieved state-of-the-art results among 2D methods prior to our work.

ADMET-AI (Swanson et al., 2024) combines Chemprop with 200 RDKit molecular descriptors including molecular weight, logP, and topological polar surface area. The learned and handcrafted features are concatenated before task-specific prediction heads. Trained on Therapeutics Data Commons datasets with careful hyperparameter tuning, ADMET-AI achieves the best average rank on the TDC ADMET Benchmark leaderboard and represents a strong ensemble baseline.

A.3.3. Pretrained and Foundation Models

ChemBFN (Tao & Abe, 2025) applies Bayesian Flow Networks to molecular representation learning. Operating on SMILES or SELFIES strings, ChemBFN learns to denoise corrupted molecular representations during pretraining, capturing chemical grammar and structural patterns from millions of unlabeled molecules. The pretrained encoder is then fine-tuned for ADMET prediction through task-specific output layers.

Uni-Mol (Zhou et al., 2023) is a 3D molecular foundation model with 209 million parameters pretrained on 209 million conformations. The SE(3)-invariant transformer processes atomic coordinates and types, capturing both local bonding patterns and global molecular shape. Pretraining combines masked atom prediction, coordinate denoising, and conformer contrastive learning. Uni-Mol requires 3D conformer generation as input, adding computational overhead but providing access to spatial information. It serves as a reference for methods with explicit geometric information.