PHASE SYNCHRONIZATION DYNAMICS FOR MOLECULAR PROPERTY PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate molecular property prediction is predominantly addressed through graph representation learning. However, conventional spatial message-passing neural networks encounter inherent structural bottlenecks when modeling long-range dependencies and complex global topologies. We propose **MolSync** (**Mol**ecular **Sync**hronization Networks), which represents molecular graphs as networks of nonlinear phase oscillators and maps structural priors into a dynamic phase space. The resulting synchronization process provides an intrinsic feature-binding mechanism that organizes local chemical environments and higher-order structural context through learned oscillator interactions. Theoretically, we show that continuous synchronization removes the hard finite-hop dependency constraint of fixed-depth message passing, heterogeneous intrinsic dynamics sustain non-trivial representation variation at phase-locked states, and spectral evolution can distinguish a class of 1-WL-indistinguishable graphs with different weighted Laplacian spectra. Extensive experiments show strong predictive performance, while energy-based attribution and oscillator analyses reveal chemically meaningful, task-dependent patterns in the learned dynamics.

1 INTRODUCTION

Accurate molecular property prediction is a fundamental challenge in computational chemistry and machine learning-driven drug discovery Deng et al. (2023); Li et al. (2022); Wieder et al. (2020); Xia et al. (2023a). To address this, molecular representation learning, particularly through the lens of Graph Neural Networks (GNNs), has emerged as a dominant and highly effective approach Klicpera et al. (2020); Fang et al. (2022); Li et al. (2025). By explicitly leveraging topological structures, GNNs establish a robust mapping from discrete chemical compositions to continuous, expressive latent spaces, fundamentally transforming how molecular properties are predicted and understood Kipf & Welling (2017); Satorras et al. (2021).

Despite their widespread success, standard spatial Message-Passing Neural Networks (MPNNs) face critical structural limitations in molecular property prediction Gilmer et al. (2017). They inherently rely on static, strictly localized message-passing mechanisms, which impose severe bottlenecks when attempting to capture long-range dependencies Ying et al. (2021) and complex global topologies Morris et al. (2019). Furthermore, while advanced architectures attempt to provide structural importance or physical attribution via attention mechanisms Xiong et al. (2020), recent studies reveal a significant flaw: these dense attention weights frequently capture statistical artifacts rather than physically meaningful chemical interactions Jain & Wallace (2019); Miao et al. (2022). Consequently, existing models struggle to provide reliable, intrinsic physical attribution for predictions.

To overcome these bottlenecks, a physics-grounded mechanism is required for *feature binding*, *i.e.*, the process of grouping discrete atoms into meaningful functional subgraphs (*e.g.*, pharmacophores) without relying on computationally heavy post-hoc explainers. Our preliminary empirical analyses reveal a striking physical phenomenon: when molecular topological interactions are formulated through the lens of oscillatory synchronization Miyato et al. (2025), atoms undergoing the largest interaction-energy changes are enriched in chemically meaningful substructures and exert a disproportionate impact on predictive fidelity. Inspired by this alignment between synchronization dynamics and critical chemical interactions, we propose shifting molecular representation learning from static message passing to dynamic phase synchronization.

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

Driven by these empirical insights, this paper introduces **MolSync** (**Mol**ecular **Sync**hronization Networks), a framework based on non-linear phase synchronization dynamics. Instead of relying purely on discrete aggregation layers Gilmer et al. (2017), MolSync conceptualizes molecular graphs as dynamic networks of phase oscillators. Geometric Scattering Transform (GST) features Gao et al. (2019) provide multi-scale structural priors, while chemical bond attributes parameterize oscillator interactions. The resulting driven Kuramoto dynamics reorganize atom representations through synchronization, while the associated interaction energy provides intrinsic feature attribution.

Beyond this physical intuition, we analyze MolSync from three complementary perspectives. First, continuous synchronization removes the hard finite-hop dependency constraint of fixed-depth local message passing, while GST conditioning introduces an additional pathway for input sensitivity. Second, heterogeneous intrinsic dynamics sustain non-trivial representation variation at phase-locked states, counteracting global representation collapse. Third, spectral evolution can separate a class of 1-WL-indistinguishable graphs with different weighted Laplacian spectra. Together, these results characterize how synchronization dynamics expand information support, preserve phase diversity, and exploit global spectral structure.

The main contributions of this paper are summarized as follows:

- **Physics-Grounded Problem Discovery:** We empirically demonstrate that synchronization-associated energy changes align with critical chemical substructures.
- **Paradigm Shift in Molecular Modeling:** We propose MolSync, a molecular property prediction framework that models molecular graphs as dynamic phase synchronization systems.
- **Theoretical Characterization:** We establish finite-hop decoupling, a heterogeneity-induced non-collapse bound, and spectral separation for a class of 1-WL-indistinguishable graphs.
- **Superior Performance and Intrinsic Interpretability:** MolSync achieves state-of-the-art predictive performance, with learned dynamics aligning with task-specific molecular mechanisms.

2 PRELIMINARIES

Problem Formulation. A molecule is represented as an attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}_{atom}, \mathbf{X}_e)$, where $\mathcal{V}$ and $\mathcal{E}$ denote atoms and bonds and $\mathbf{X}_{atom}, \mathbf{X}_e$ contain their attributes. Given $\mathcal{D} = \{(\mathcal{G}_i, \mathbf{y}_i)\}_{i=1}^N$, a graph encoder h_θ produces $\mathbf{h}_\mathcal{G} = h_\theta(\mathcal{G})$ and a predictor p_ϕ outputs $\hat{\mathbf{y}} = p_\phi(\mathbf{h}_\mathcal{G})$.

Message Passing Neural Networks. Most molecular encoders are Message Passing Neural Networks (MPNNs) (see App. B). Their local aggregation induces a hard finite-hop dependency range, can attenuate long-range signals through graph bottlenecks (over-squashing) Alon & Yahav (2021); Topping et al. (2022), may homogenize node representations (over-smoothing) Li et al. (2018); Oono & Suzuki (2020), and inherits expressivity limitations related to the 1-Weisfeiler–Lehman test Morris et al. (2019). These limitations motivate a synchronization-based representation mechanism.

Kuramoto Synchronization on the Sphere. The classical Kuramoto model Kuramoto (1984) describes coupled scalar phases as

$$\dot{\theta}_v(t) = \omega_v + \sum_{u \in \mathcal{N}(v)} \kappa_{vu} \sin\big(\theta_u(t) - \theta_v(t)\big), \tag{1}$$

where ω_v is the intrinsic frequency and $\kappa_{vu} \in \mathbb{R}$ induces attractive ($\kappa_{vu} > 0$) or repulsive ($\kappa_{vu} < 0$) interactions. Following AKOrN Miyato et al. (2025), MolSync generalizes scalar phases to unit-vector states $\mathbf{x}_v(t) \in \mathbb{S}^{d-1}$:

$$\dot{\mathbf{x}}_v(t) = \mathbf{\Omega}_v \mathbf{x}_v(t) + \mathrm{Proj}_{\mathbf{x}_v(t)} \left(\mathbf{c}_v + \sum_{u \in \mathcal{N}(v)} \kappa_{vu} \mathbf{x}_u(t) \right), \tag{2}$$

where $\mathbf{\Omega}_v^\top = -\mathbf{\Omega}_v$ is the natural rotation, $\mathbf{c}_v$ is a contextual stimulus, $\kappa_{vu} \in \mathbb{R}$ is the signed coupling, and $\mathrm{Proj}_\mathbf{x}(\mathbf{y}) = \mathbf{y} - \langle \mathbf{x}, \mathbf{y} \rangle \mathbf{x}$ denotes tangent-space projection.

Signed Interaction Potential. For fixed stimuli and symmetric couplings $\kappa_{vu} = \kappa_{uv}$, define

$$E(\mathbf{X}) = -\frac{1}{2}\sum_{v\in\mathcal{V}}\sum_{u\in\mathcal{N}(v)}\kappa_{vu}\langle \mathbf{x}_v, \mathbf{x}_u\rangle - \sum_{v\in\mathcal{V}}\langle \mathbf{c}_v, \mathbf{x}_v\rangle. \tag{3}$$

The sign of κ_{vu} is retained in E, with positive and negative couplings favoring alignment and separation. The projected coupling/stimulus term in Eq. 2 is the negative Riemannian gradient of E; natural rotation adds tangential drift, so E need not be globally monotone. Sec. 5.2 separately uses the non-negative absolute-coupling Dirichlet variation $\mathcal{D}_{|\kappa|}$ to quantify representation variation.

3 Motivation: Physics-Grounded Feature Binding

As discussed in the Introduction, conventional spatial message passing encounters inherent structural bottlenecks, fundamentally lacking an intrinsic physical mechanism for *feature binding*. Formulating molecular graphs as dynamical systems governed by generalized Kuramoto dynamics provides a natural alternative: strongly interacting atoms can organize into coherent functional subsystems through learned synchronization dynamics.

To empirically validate this physical hypothesis, Fig. 1 illustrates a preliminary Energy-Based Attribution (EBA) analysis on representative benchmarks; comprehensive results are provided in Appendix D.3. As depicted, atoms exhibiting the most significant energy drop (ΔE) during dynamic updates tend to constitute critical predictive motifs. The results support two complementary properties: **(1) Chemical Plausibility**, evidenced by the significant enrichment of specific pharmacophoric features (e.g., H-bond acceptors) within high-ΔE nodes (Fig. 1a); and **(2) Predictive Fidelity**, demonstrated by a disproportionate performance degradation upon perturbing high-ΔE nodes, even when controlling for out-of-distribution artifacts via random baselines (Fig. 1b).

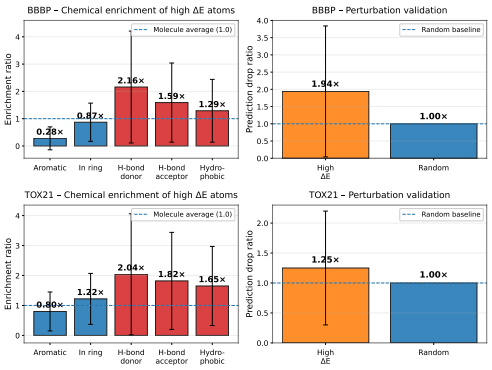

Figure 1: EBA validation of synchronization-driven feature binding. (a) Chemical enrichment among high-ΔE atoms. (b) Predictive fidelity versus random masking.

In summary, these empirical and physical observations motivate MolSync as a synchronization-based molecular encoder in which feature organization emerges from learned oscillator interactions rather than solely from repeated local aggregation.

4 MolSync Architecture

MolSync (**Mol**ecular **Sync**hronization Networks) realizes the graph encoder $h_\theta(\cdot)$ as a dynamical system on the unit hypersphere. As illustrated in Fig. 2, atoms are modeled as phase oscillators and bonds as interaction pathways. The model consists of three stages: **Multi-Scale Initialization**, **Driven Kuramoto Synchronization Updates**, and **Phase-Invariant Readout**.

4.1 Multi-Scale Initialization

For each atom i, the raw molecular input combines atomic and incident-bond features,

$$\mathbf{h}_i^{\text{raw}} = \left[\mathbf{x}_{atom,i} \| \text{mean}_{j\in\mathcal{N}(i)}\mathbf{x}_{e,ij}\right]. \tag{4}$$

A Geometric Scattering Transform (GST) Gao et al. (2019) then yields multi-scale structural features

$$\mathbf{S} = \text{GST}(G, \mathbf{H}^{\text{raw}}), \qquad \mathbf{s}_i = \mathbf{S}_{i,:}, \tag{5}$$

where the dyadic scattering construction follows the standard graph-scattering formulation. GST supplies a structural prior beyond immediate neighborhoods and induces graph-conditioned heterogeneity for the subsequent synchronization dynamics.

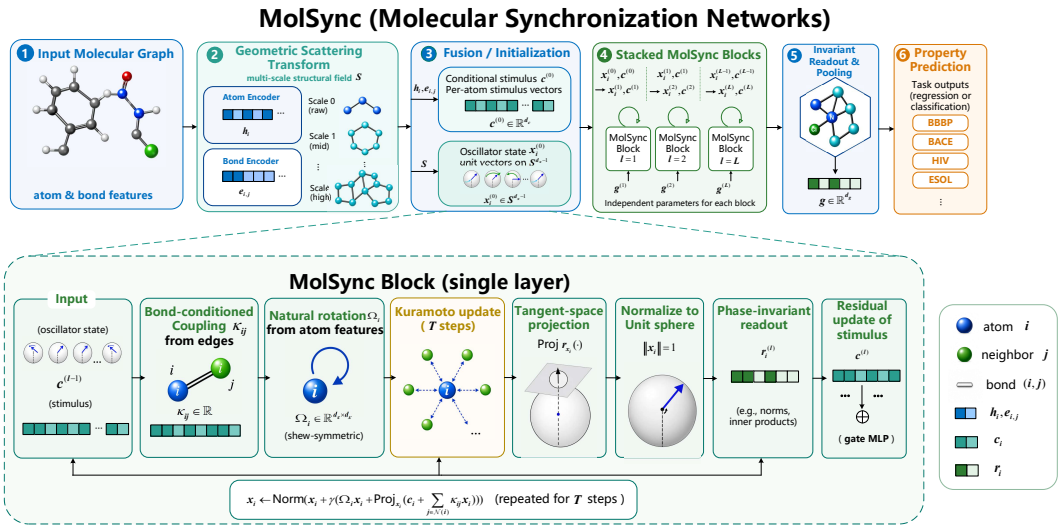

Figure 2: Overview of MOLSYNC. GST provides multi-scale structural priors; stacked MolSync blocks evolve oscillator states through signed bond-conditioned couplings and natural rotations; phase-invariant readout refines node representations for molecular prediction.

MolSync initializes two coupled fields:

$$\mathbf{c}_i^{(0)} = \mathrm{MLP}_{stim}\left([\mathbf{h}_i^{\mathrm{raw}} \| \mathbf{s}_i]\right), \tag{6}$$

$$\mathbf{x}_i^{(0)}(0) = \frac{\mathbf{W}_{gst}\mathbf{s}_i + \mathbf{W}_x\mathbf{h}_i^{\mathrm{raw}} + \boldsymbol{\epsilon}}{\|\mathbf{W}_{gst}\mathbf{s}_i + \mathbf{W}_x\mathbf{h}_i^{\mathrm{raw}} + \boldsymbol{\epsilon}\|_2}, \tag{7}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. Here, $\mathbf{x}_i^{(0)}(0) \in \mathcal{S}^{d-1}$ is the oscillator state and $\mathbf{c}_i^{(0)} \in \mathbb{R}^d$ is the graph-conditioned stimulus. The Gaussian perturbation acts as a stochastic symmetry-breaking term.

4.2 DRIVEN KURAMOTO SYNCHRONIZATION UPDATES

MolSync contains L synchronization blocks, each with T internal evolution steps. Within block ℓ, the stimulus $\mathbf{c}_i^{(\ell-1)}$ remains static while the oscillator state $\mathbf{x}_i^{(\ell)}(t) \in \mathcal{S}^{d-1}$ evolves. For $\ell > 0$, the block is initialized by $\mathbf{x}_i^{(\ell)}(0) = \mathbf{x}_i^{(\ell-1)}(T)$.

Bond-conditioned couplings and atom-dependent natural rotations are defined by

$$\begin{aligned} \kappa_{ij}^{(\ell)} &= \tanh(\mathrm{MLP}_{edge}(\mathbf{x}_{e,ij})), \\ \mathbf{W}_i^{(\ell)} &= \mathrm{MLP}_{atom}\left(\mathbf{c}_i^{(\ell-1)}\right), \qquad \boldsymbol{\Omega}_i^{(\ell)} = \frac{1}{2}\left(\mathbf{W}_i^{(\ell)} - \mathbf{W}_i^{(\ell)\top}\right). \end{aligned} \tag{8}$$

Thus, $\kappa_{ij}^{(\ell)} \in (-1, 1)$ permits both attractive ($\kappa_{ij} > 0$) and repulsive ($\kappa_{ij} < 0$) interactions. For undirected molecular graphs, $\kappa_{ij}^{(\ell)} = \kappa_{ji}^{(\ell)}$. The skew-symmetric $\boldsymbol{\Omega}_i^{(\ell)}$ generates tangential natural rotation on the sphere. Couplings encode chemically conditioned attraction or separation, whereas $\boldsymbol{\Omega}_i^{(\ell)}$ supplies atom-specific intrinsic motion, supporting heterogeneous phase patterns.

Following the global continuous energy defined in Eq. 3, the local energy for atom i at block ℓ and time t is tractably evaluated as $E_i^{(\ell)}(t) = -\frac{1}{2}\sum_{j\in\mathcal{N}(i)} \kappa_{ij}^{(\ell)} \langle \mathbf{x}_i^{(\ell)}(t), \mathbf{x}_j^{(\ell)}(t)\rangle - \langle \mathbf{c}_i^{(\ell-1)}, \mathbf{x}_i^{(\ell)}(t)\rangle$. Here $E_i^{(\ell)}(t)$ is a local contribution to the global interaction potential and is computed only for EBA. The coupling/stimulus term follows the negative Riemannian gradient of the global potential, while natural rotation adds tangential drift; we apply T unit-norm Riemannian updates. For each internal step t, the state is updated as:

$$\begin{aligned} \mathbf{d}_i^{(\ell)}(t) &= \mathbf{c}_i^{(\ell-1)} + \sum_{j\in\mathcal{N}(i)} \kappa_{ij}^{(\ell)} \mathbf{x}_j^{(\ell)}(t), \\ \mathbf{x}_i^{(\ell)}(t+1) &= \mathrm{Retract}\left(\mathbf{x}_i^{(\ell)}(t) + \gamma\left[\boldsymbol{\Omega}_i^{(\ell)}\mathbf{x}_i^{(\ell)}(t) + \mathrm{Proj}_{\mathbf{x}_i^{(\ell)}(t)}\left(\mathbf{d}_i^{(\ell)}(t)\right)\right]\right), \end{aligned} \tag{9}$$

where $\mathrm{Proj}_{\mathbf{x}}(\mathbf{y}) = \mathbf{y} - \langle \mathbf{x}, \mathbf{y} \rangle \mathbf{x}$ and $\mathrm{Retract}(\mathbf{z}) = \mathbf{z}/\|\mathbf{z}\|_2$. After T steps, $\mathbf{x}_i^{(\ell)}(T)$ is passed to the readout to update the stimulus field for the next block.

4.3 Phase-Invariant Readout and Residual Update

The readout converts the final oscillator states into K coherence channels over $\tilde{\mathcal{N}}(i) = \mathcal{N}(i) \cup \{i\}$. Following the compact formulation in the original architecture, the attention coefficient and channel-wise coherence are written in one line:

$$\alpha_{ij,k} = \frac{\exp\Big(\mathbf{w}_k^\top [\mathbf{c}_i^{(\ell-1)} \| \mathbf{c}_j^{(\ell-1)}]\Big)}{\sum_{w \in \tilde{\mathcal{N}}(i)} \exp\Big(\mathbf{w}_k^\top [\mathbf{c}_i^{(\ell-1)} \| \mathbf{c}_w^{(\ell-1)}]\Big)}, \quad z_{i,k} = \left\| \sum_{j \in \tilde{\mathcal{N}}(i)} \alpha_{ij,k} \mathbf{x}_j^{(\ell)}(T) \right\|_2. \tag{10}$$

Because the Euclidean norm is invariant to a common orthogonal transformation, $\mathbf{z}_i = [z_{i,1}, \ldots, z_{i,K}]^\top$ depends on relative synchronization coherence rather than the arbitrary global orientation of the oscillator coordinates. The readout maps this dynamical coherence back into the stimulus field, which then parameterizes the intrinsic dynamics of the next block and forms an iterative synchronization-to-stimulus feedback.

The coherence representation is projected back to the stimulus space through a residual update,

$$\mathbf{c}_i^{(\ell)} = \mathbf{c}_i^{(\ell-1)} + \mathrm{MLP}_{update}(\mathbf{z}_i). \tag{11}$$

After L blocks, global mean pooling gives $\mathbf{h}_{\mathcal{G}} = |\mathcal{V}|^{-1} \sum_{i \in \mathcal{V}} \mathbf{c}_i^{(L)}$, which is passed to a task-specific prediction head for end-to-end supervised learning.

5 Theoretical Analysis

Standard spatial MPNNs are known to suffer from information over-squashing Alon & Yahav (2021); Topping et al. (2022), representation over-smoothing Li et al. (2018); Oono & Suzuki (2020), and expressivity limitations related to the 1-Weisfeiler–Lehman (1-WL) test Morris et al. (2019). The purpose of this section is not to claim that one dynamical construction universally eliminates all three phenomena. Instead, we isolate the precise mechanisms that can be established for MolSync and state their assumptions explicitly. Theorem 1 analyzes the continuous synchronization core and its information support; Theorem 2 characterizes exact collapse in the canonical heterogeneous phase-locked regime; and Theorem 3 gives a sufficient spectral condition under which the continuous dynamics have separating capacity beyond 1-WL. The implemented model uses sphere-valued states, tangent projection, retraction, finite integration steps, learned signed couplings, and GST-conditioned stimuli. Accordingly, the scalar/linearized results below should be read as analyses of the corresponding synchronization mechanisms rather than as complete convergence guarantees for every component of the training pipeline. Proofs are provided in Section F.

5.1 Long-Range Information Flow: Removing the Hard Finite-Hop Barrier

A K-layer local MPNN has a strict support constraint: if $d(i, j) > K$, then the representation at node i cannot depend on the initial representation of node j. Over-squashing is a stronger phenomenon than this support limitation: even when a dependency exists, its magnitude may still decay rapidly through a graph bottleneck. We therefore distinguish *global reachability* from *sensitivity preservation*.

Let L_κ denote the weighted Laplacian induced by symmetric couplings $\kappa_{ij} \in \mathbb{R}$, which may be positive (attractive) or negative (repulsive). Consider the local scalar linearization,

$$\dot{X}(t) = \Omega - L_\kappa X(t). \tag{12}$$

For nodes i and j, let $d = d(i, j)$ and let $\mathcal{P}_d(j \to i)$ be the set of shortest paths of length d. Define the signed shortest-path coefficient

$$a_{ij} = \sum_{p \in \mathcal{P}_d(j \to i)} \prod_{e \in p} \kappa_e. \tag{13}$$

Theorem 1 (Removal of the Hard Finite-Hop Support Constraint). *For the linearized dynamics in Eq. 12, the state Jacobian is $\mathcal{J}(t) = \partial X(t)/\partial X(0) = e^{-L_\kappa t}$. If $a_{ij} \neq 0$, then*

$$\mathcal{J}_{ij}(t) = \frac{a_{ij}}{d!} t^d + \mathcal{O}(t^{d+1}). \tag{14}$$

Hence $\mathcal{J}_{ij}(t)$ is not identically zero and, for sufficiently small $t > 0$, node j influences node i even when $d(i, j)$ exceeds any fixed discrete message-passing depth. If all active couplings are positive and the graph is connected, the stronger property $\mathcal{J}_{ij}(t) > 0$ holds for every $t > 0$. For signed couplings, cancellations can occur at special parameter values or times; the theorem therefore establishes removal of the hard finite-hop support barrier under a non-degeneracy condition, not a uniform lower bound on long-range sensitivity.

Implementation-aware sensitivity decomposition. Within each MolSync block, the implemented dynamics contain an additional pathway that is absent from Eq. 12. Let H denote the input node features to the block (with graph topology and bond features fixed), let $X(0) = g(H)$ denote the initialized oscillator state, and let $C = h_{\mathcal{G}}(H)$ denote the GST-conditioned stimulus held fixed during the internal continuous evolution. Write the block-level continuous vector field abstractly as $\dot{X} = F(X, C)$; the stacked architecture composes these block-level sensitivities through the chain rule. Along a trajectory, define

$$A(t) = \frac{\partial F}{\partial X}, \qquad B(t) = \frac{\partial F}{\partial C}, \tag{15}$$

and let $\Phi(t, \tau)$ be the state-transition operator generated by $A(t)$. Differentiating the dynamics with respect to the original input gives the exact variational decomposition

$$\frac{\partial X(t)}{\partial H} = \Phi(t, 0)\frac{\partial g(H)}{\partial H} + \int_0^t \Phi(t, \tau)B(\tau)\frac{\partial h_{\mathcal{G}}(H)}{\partial H}\, d\tau. \tag{16}$$

The first term is the state-propagation pathway, whereas the second term is specific to the graph-conditioned stimulus and captures how GST-derived structural information modulates the subsequent synchronization dynamics. In the simplified specialization $F(X, C) = \Omega(C) - L_\kappa X$ with $g(H) = H$, Eq. 16 reduces to

$$\frac{\partial X(t)}{\partial X(0)} = e^{-L_\kappa t} + \int_0^t e^{-L_\kappa (t-\tau)}\frac{\partial \Omega}{\partial C}\frac{\partial C}{\partial X(0)}\, d\tau. \tag{17}$$

This decomposition explains why the full architecture is not governed solely by repeated local diffusion. It does *not*, by itself, guarantee preservation of the magnitude of distant-node sensitivities, which may still depend on the learned dynamics and graph structure. Accordingly, our theoretical result concerns the removal of the hard finite-hop support barrier rather than a universal elimination of sensitivity attenuation.

5.2 Avoiding Exact Global Collapse under Heterogeneous Phase Locking

Over-smoothing is commonly associated with the collapse of node representations toward an indistinguishable state. Because MolSync permits signed couplings, the signed interaction potential and a non-negative measure of representation variation must be distinguished. The coupling sign determines whether an interaction is attractive or repulsive, whereas representation variation should remain non-negative. We therefore define the signed-weight-aware Dirichlet variation

$$\mathcal{D}_{|\kappa|}(X) = \frac{1}{2}\sum_{(i,j)\in\mathcal{E}} |\kappa_{ij}|\, \|x_i - x_j\|_2^2. \tag{18}$$

For scalar phases, Eq. 18 reduces to $\frac{1}{2}\sum_{(i,j)} |\kappa_{ij}|(\theta_i - \theta_j)^2$. This quantity is used only to measure representational variation; it is not identified with the signed Kuramoto interaction potential.

Theorem 2 (Non-Collapse Bound for Heterogeneous Phase-Locked Kuramoto Dynamics). *Consider the canonical scalar Kuramoto dynamics on a connected active-coupling graph,*

$$\dot{\theta}_i = \omega_i + \sum_{j\in\mathcal{N}(i)} \kappa_{ij}\sin(\theta_j - \theta_i), \tag{19}$$

with symmetric signed couplings $\kappa_{ij} \in \mathbb{R} \setminus \{0\}$. Suppose the system reaches a phase-locked state θ^* with common angular velocity $\bar{\omega}$. Let B be an oriented incidence matrix, $K = \mathrm{diag}(\kappa_e)$, $L_{|\kappa|} = B|K|B^\top$, and $\delta\omega = \omega - \bar{\omega}\mathbf{1}$. Then

$$\mathcal{D}_{|\kappa|}(\theta^*) \geq \frac{\|\delta\omega\|_2^2}{2\,\lambda_{\max}(L_{|\kappa|})}. \tag{20}$$

Consequently, if the intrinsic frequencies are not all identical, $\delta\omega \neq 0$ and exact global phase collapse is impossible at a phase-locked solution: $\mathcal{D}_{|\kappa|}(\theta^*) > 0$.

Theorem 2 formalizes the specific synchronization mechanism used in our over-smoothing discussion. It does not assert that every finite-step state of the complete sphere-valued MolSync architecture satisfies the scalar bound in Eq. 20. Instead, it shows that heterogeneous intrinsic dynamics are incompatible with an exactly homogeneous phase-locked solution in the canonical Kuramoto system, while Eq. 18 gives the natural vector-state diagnostic. This distinction is important for two edge cases. First, if all intrinsic frequencies are identical, the lower bound becomes zero and the theorem makes no non-collapse guarantee. Second, the isotropic Gaussian perturbation used in initialization acts as a practical symmetry-breaking mechanism, but it is not a substitute for the heterogeneity assumption in the theorem. Thus, localized or cluster synchronization can be used for feature binding without equating synchronization itself with pathological global over-smoothing.

5.3 Conditional Spectral Separability beyond 1-WL

The expressive power of standard spatial MPNNs is bounded by the 1-WL test Morris et al. (2019). Spectral dynamics provide a different source of graph sensitivity, but a spectral argument does not imply a universal strict separation from 1-WL: co-spectral non-isomorphic graphs remain an immediate limitation, and even for graphs with different spectra a degenerate input signal can fail to excite discriminative modes. We therefore state the result as a sufficient separation condition.

Theorem 3 (Spectral Separation on a Sufficient Graph Class). *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two non-isomorphic graphs with the same number of nodes that are indistinguishable by 1-WL. Let L_1 and L_2 be their real symmetric weighted Laplacians. If L_1 and L_2 have different spectra, then for every $t > 0$ the continuous transfer operators*

$$H_r(t) = e^{-L_r t}, \qquad r \in \{1, 2\}, \tag{21}$$

have different spectra and cannot be related by a node permutation. Consequently, there exists an input graph signal q such that

$$H_1(t)q \neq P^\top H_2(t)Pq \tag{22}$$

for every permutation matrix P. Therefore, the linearized synchronization dynamics have the capacity to separate this class of 1-WL-indistinguishable graph pairs when the input/stimulus excites the discriminative spectral modes. The theorem does not cover co-spectral non-isomorphic graphs and does not claim that every learned initialization or readout necessarily realizes the separation.

Theorem 3 gives a precise interpretation of the spectral contribution of MolSync. The continuous dynamics apply graph-dependent spectral transfer functions to atom-level signals; when two 1-WL-indistinguishable graphs possess different weighted Laplacian spectra, these transfer operators are mathematically distinct and can induce different temporal responses. Molecular rings and other global topological motifs can influence these spectral responses, but the Laplacian spectrum is not a complete graph invariant. We therefore use the theorem as a sufficient mechanism for additional topological sensitivity rather than as a universal statement that MolSync strictly dominates 1-WL on all graph pairs.

6 Experiments

This section empirically evaluates MolSync from multiple perspectives. First, we assess its predictive performance on MoleculeNet classification tasks. Next, we analyze functional-group synchronization to examine whether the oscillator dynamics induce chemically meaningful structural organization. Finally, we conduct architectural ablation studies to validate the contributions of key components, including GST initialization, Kuramoto dynamics, natural rotations, and phase-invariant readout. Additional experimental details and supplementary results, including hyperparameter settings, regression results, extended EBA analyses, dynamical hyperparameter sensitivity, oscillator trajectory visualizations, and algorithmic complexity analysis, are provided in the Appendix.

Datasets. Models are evaluated on multiple molecular property prediction benchmarks from the MoleculeNet suite Wu et al. (2018), covering both classification and regression tasks. The classification setting includes eight classification datasets, such as BBBP, Tox21, and ClinTox, spanning diverse

Table 1: Performance on classification tasks. Mean±std ROC-AUC (%) over 5 random trials.

Methods	BBBP ↑	Tox21 ↑	ToxCast ↑	SIDER ↑	ClinTox ↑	MUV ↑	HIV ↑	BACE ↑	Avg ↑
AttrMask	65.0±2.3	74.8±0.2	62.9±0.1	61.2±0.1	87.7±1.1	73.4±2.0	76.8±0.5	79.7±0.3	72.68
ContextPred	65.7±0.6	74.2±0.0	62.5±0.3	62.2±0.5	77.2±0.8	75.3±1.5	77.1±0.8	76.0±2.0	71.28
GraphCL	69.7±0.6	73.9±0.6	62.4±0.5	60.5±0.8	76.0±2.6	69.8±2.6	78.5±1.2	75.4±1.4	70.78
InfoGraph	67.5±0.1	73.2±0.4	63.7±0.5	59.9±0.3	76.5±1.0	74.1±0.7	75.1±0.9	77.8±0.8	70.98
GROVER	70.0±0.1	74.3±0.1	65.4±0.4	64.8±0.6	81.2±3.0	67.3±1.8	62.5±0.9	82.6±0.7	71.01
MolCLR	66.6±1.8	73.0±0.1	62.9±0.3	57.5±1.7	86.1±0.9	72.5±2.3	76.2±1.5	71.5±3.1	70.79
GraphMAE	72.0±0.6	75.5±0.6	64.1±0.3	60.3±1.1	82.3±1.2	76.3±2.4	77.2±1.0	83.1±0.9	73.85
Mole-BERT	71.9±1.6	76.8±0.5	64.3±0.2	62.8±1.1	78.9±3.0	78.6±1.8	78.2±0.8	80.8±1.4	74.04
MoleculeSDE	71.8±0.7	76.8±0.3	65.0±0.2	60.8±0.3	87.0±0.5	80.9±0.3	78.8±0.9	79.5±2.1	75.07
3D InfoMax	70.4±1.0	75.5±0.5	63.1±0.7	64.1±0.1	89.8±1.2	72.8±1.0	74.9±0.3	80.7±0.6	73.91
Galformer	71.6±0.9	75.7±0.8	64.0±0.4	64.3±0.5	84.7±0.9	73.4±1.1	75.6±0.4	80.9±1.0	74.02
GraphMVP	71.5±1.3	76.1±0.9	64.3±0.6	64.7±0.7	85.4±0.8	74.9±1.2	76.0±0.6	81.5±1.2	74.86
MoleBlend	73.0±0.8	77.8±0.8	66.1±0.0	64.9±0.3	87.6±0.7	77.2±2.3	79.0±0.8	83.7±1.4	76.16
LEMON	73.7±1.1	77.5±0.6	65.1±0.5	64.3±0.9	85.9±3.2	79.4±4.3	79.3±1.1	87.8±1.4	76.62
GEM	72.4±0.4	78.1±0.1	69.2±0.4	67.2±0.4	90.1±1.3	81.7±0.5	80.6±0.9	85.6±1.1	78.18
Uni-Mol	72.9±0.6	79.6±0.5	**69.6±0.1**	65.9±1.3	91.9±1.8	82.1±1.3	80.8±0.3	85.7±0.2	78.63
DeMol	75.1±0.6	80.9±0.4	69.3±0.5	68.4±0.3	92.6±0.7	82.5±1.4	**81.2±0.9**	89.0±1.1	79.96
MolSync (Ours)	**78.3±1.2**	**81.4±0.3**	67.8±0.2	**68.6±0.2**	**92.9±0.2**	**82.6±0.4**	80.9±1.2	**89.1±1.2**	**80.20**

biomedical and toxicological endpoints. The regression setting includes six continuous prediction datasets: ESOL, FreeSolv, and Lipophilicity are used to evaluate physicochemical properties, while QM7, QM8, and QM9 are used to evaluate quantum-mechanical targets. To assess out-of-distribution structural generalization, standard scaffold splitting is applied across all tasks.

Baselines. MolSync is systematically evaluated against 17 state-of-the-art baselines across four distinct paradigms: (1) *self-supervised 2D pre-training* (AttrMask Hu et al. (2020), ContextPred Hu et al. (2020), InfoGraph Sun et al. (2019)); (2) *contrastive graph learning* (GraphCL You et al. (2020), MolCLR Wang et al. (2022), LEMON); (3) *masked modeling and Transformers* (GROVER Rong et al. (2020), GraphMAE Hou et al. (2022), Mole-BERT Xia et al. (2023b)); and (4) *geometry-aware 2D/3D multi-modal and foundation models* (3D InfoMax Stärk et al. (2022), GraphMVP Liu et al. (2022), MoleculeSDE Liu et al. (2023), Galformer Bai et al. (2023), MoleBlend Yu et al. (2024), GEM Fang et al. (2022), Uni-Mol Zhou et al. (2023), DeMol Liu et al. (2026)).

Implementation Details. MolSync uses 4 synchronization blocks with hidden dimension 32. For all classification benchmarks, we adopt an 80/10/10 Bemis–Murcko scaffold split, with hyperparameters and checkpoints selected exclusively on the validation set. Molecular graphs are constructed from SMILES using RDKit, and missing labels in multi-task datasets are masked during training and evaluation. We tune the Kuramoto steps $T \in \{4, 8\}$, phase-coherence readout channels $K \in \{8, 16\}$, and batch size $\in \{64, 128, 256\}$. AdamW is used throughout, with BCE for classification, L_2 loss for ESOL, FreeSolv, and Lipophilicity, and Smooth L_1 loss for QM7–9. We employ a 5% linear warm-up followed by cosine learning-rate decay, and compute ROC-AUC independently over valid tasks before macro-averaging. MolSync results report mean±std over 5 independent runs. Dataset-specific optimization settings and selected hyperparameters are provided in Table 3.

6.1 Predictive Performance on MoleculeNet Classification Tasks

As detailed in Tab. 1, MolSync establishes highly competitive performance across the eight MoleculeNet classification benchmarks, consistently outperforming state-of-the-art baselines on six out of eight datasets. Unlike conventional static message-passing architectures, MolSync effectively captures both local chemical environments and higher-order structural motifs. For complex endpoints such as toxicity and target inhibition, the Kuramoto dynamics capture coordinated interactions among functional groups and scaffolds through task-conditioned phase synchronization, supporting predictive generalization and the subsequent energy-based interpretability analysis. The gains are distributed across toxicity-, bioactivity-, and permeability-related endpoints rather than concentrated on a single task, indicating that the synchronization representation remains effective across distinct molecular property regimes.

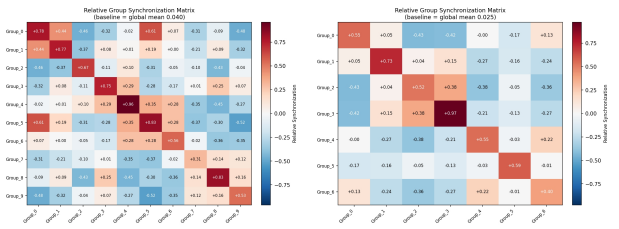

Figure 3: Group synchronization patterns. Positive and negative entries denote local coherence vs. phase separation.

Table 2: Ablation study on the BBBP dataset. Best result is highlighted in bold.

Variant	ROC-AUC ↑
w/o GST	73.9±1.4
w/o Kuramoto	73.8±0.9
w/o Ω	74.7±0.3
w/o phase-invariant readout	76.8±1.2
MolSync	**78.3±1.2**

6.2 Functional-Group Synchronization and Intrinsic Feature Binding

To investigate whether MolSync captures meaningful chemical motifs, the emergent synchronization patterns among distinct molecular substructures are analyzed. For any two functional groups $\mathcal{G}_p$ and $\mathcal{G}_q$, their interaction is quantified via the mean centered cosine similarity of their final oscillator states: $S_{pq} = \mathbb{E}_{v\in\mathcal{G}_p, u\in\mathcal{G}_q}[\langle \tilde{\mathbf{x}}_v, \tilde{\mathbf{x}}_u \rangle]$, where $\tilde{\mathbf{x}}_v$ denotes the L_2-normalized deviation from the global mean phase. This metric naturally isolates relative phase alignments, with positive and negative scalar values denoting cooperative local synchronization and competitive phase separation, respectively.

Fig. 3 is consistent with the non-collapse mechanism characterized in Theorem 2. The learned dynamics organize distinct structural motifs rather than a globally homogeneous state. Positive diagonal blocks confirm local coherence within functional groups, whereas negative off-diagonal entries reveal phase separation across substructures. This structured organization suggests that MolSync binds task-relevant motifs without collapsing the global molecular representation. This representation-level evidence complements the EBA results in Appendix D.3: group synchronization describes how motifs are organized in latent phase space, whereas EBA identifies molecular regions that most strongly influence the prediction.

6.3 Ablation Study

Tab. 2 shows the architectural ablation on the BBBP dataset. The full MolSync outperforms all isolated variants. Disabling the continuous-time Kuramoto dynamics ($T = 0$) or omitting GST initialization causes the largest degradation ($\sim 4.5\%$ drop), supporting the combination of dynamic synchronization and multi-scale structural priors. Removing the natural rotation term ($\mathbf{\Omega}_v^{(\ell)}$) causes a 3.6% drop, indicating that atom-specific motion helps maintain phase diversity, while replacing the phase-invariant readout with direct state pooling over $\mathbf{X}$ reduces performance by 1.5%, confirming the importance of rotation-invariant structural coherence. These ablations also show that the components are complementary rather than interchangeable. GST supplies multi-scale structural context, the dynamical terms govern how that context is reorganized, and the invariant readout converts the resulting coherence into task-relevant features. The consistent degradation across these removals indicates that performance is not attributable to a single isolated module.

7 Conclusions

This paper introduces MolSync, a physics-grounded framework that reformulates molecular property prediction through phase synchronization dynamics. By combining multi-scale structural initialization with vectorized Kuramoto-style dynamics on a spherical manifold, MolSync provides an intrinsic mechanism for feature binding. Our theory characterizes this mechanism under explicit assumptions: continuous synchronization removes the hard finite-hop support constraint, heterogeneous phase locking prevents exact global collapse, and spectral evolution provides conditional separation capacity for certain 1-WL-indistinguishable graphs. Extensive experiments further demonstrate strong predictive performance and chemically meaningful intrinsic attribution. Overall, MolSync offers a unified dynamical perspective for molecular representation learning that connects structural modeling, synchronization, and interpretability.

AI USE STATEMENT

Generative AI tools were used to assist with language editing and manuscript writing, including improving grammar, clarity, phrasing, and overall presentation, and with minor visual editing of a scientific figure. They were not used to develop theoretical or conceptual frameworks, formulate mathematical claims, provide ingredients for or assist with proofs, propose or refine research hypotheses, design the methodology or experiments, implement the method, generate or process datasets or numerical results, or interpret experimental results. All technical content, theoretical analyses, experimental results, scientific claims, and the scientific content of the edited figure were specified, developed, and verified by the authors. The authors reviewed all AI-assisted edits and take full responsibility for the final manuscript.

ETHICS STATEMENT

This work follows the ICLR Code of Ethics and does not involve human subjects, personal or sensitive data, or interventions on individuals. The experiments use established molecular property prediction benchmarks for research purposes. We nevertheless acknowledge a potential dual-use concern: improved molecular property prediction may support beneficial applications such as drug discovery and materials design, but related techniques could also be misused to assist the identification or design of harmful chemical compounds. MolSync is presented as a research model for molecular representation and property prediction rather than as a clinical, regulatory, or safety-critical decision system. The energy-based attribution results are intended as model-level scientific analyses and should not be interpreted as experimentally validated biochemical or clinical explanations. We report these limitations and potential broader impacts to support responsible use of the method.

REPRODUCIBILITY STATEMENT

We provide the information needed to reproduce the main results throughout the paper and supplementary material. The MolSync architecture and synchronization updates are described in Section 4, while the theoretical assumptions and claims are stated in Section 5 with complete proofs in Appendix F. Experimental protocols, datasets, optimization settings, and evaluation procedures are described in Section 6 and Appendix D; dataset-specific hyperparameters are summarized in Table 3. Additional regression results, attribution analyses, dynamical sensitivity studies, oscillator visualizations, and complexity details are also included in the appendix. An anonymous implementation is available at `https://anonymous.4open.science/r/MolSync-demo-3E77/`.

REFERENCES

Selahattin Akkas and Ariful Azad. Gnnshap: Scalable and accurate GNN explanation using shapley values. In Tat-Seng Chua, Chong-Wah Ngo, Ravi Kumar, Hady W. Lauw, and Roy Ka-Wei Lee (eds.), *WWW*, pp. 827–838, 2024.

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *ICLR*, 2021.

Burouj Armgaan, Manthan Dalmia, Sourav Medya, and Sayan Ranu. Graphtrail: Translating GNN predictions into human-interpretable logical rules. In *NeurIPS*, 2024.

Steve Azzolin, Antonio Longa, Stefano Teso, and Andrea Passerini. Reconsidering faithfulness in regular, self-explainable and domain invariant gnns. In *ICLR*, 2025.

Peizhen Bai, Xianyuan Liu, and Haiping Lu. Geometry-aware line graph transformer pre-training for molecular property prediction. *CoRR*, abs/2309.00483, 2023.

Zhuomin Chen, Jiaxing Zhang, Jingchao Ni, Xiaoting Li, Yuchen Bian, Md Mezbahul Islam, Ananda Mondal, Hua Wei, and Dongsheng Luo. Generating in-distribution proxy graphs for explaining graph neural networks. In *ICML*, pp. 7712–7730, 2024.

Jianyuan Deng, Zhibo Yang, Hehe Wang, Iwao Ojima, Dimitris Samaras, and Fusheng Wang. A systematic study of key elements underlying molecular property prediction. *Nature Communications*, 14(1):6395, 2023.

Joseph L. Durant, Burton A. Leland, Douglas R. Henry, and James G. Nourse. Reoptimization of MDL keys for use in drug discovery. *Journal of Chemical Information and Computer Sciences*, 42 (6):1273–1280, 2002.

Xiaomin Fang, Lihang Liu, Jieqiong Lei, Donglong He, Shanzhuo Zhang, Jingbo Zhou, Fan Wang, Hua Wu, and Haifeng Wang. Geometry-enhanced molecular representation learning for property prediction. *Nat. Mach. Intell.*, 4(2):127–134, 2022.

Fabian Fumagalli, Maximilian Muschalik, Paolo Frazzetto, Janine Strotherm, Luca Hermes, Alessandro Sperduti, Eyke Hüllermeier, and Barbara Hammer. Exact computation of any-order shapley interactions for graph neural networks. *CoRR*, abs/2501.16944, 2025.

Feng Gao, Guy Wolf, and Matthew J. Hirn. Geometric scattering for graph data analysis. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *ICML*, pp. 2122–2131, 2019.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In Doina Precup and Yee Whye Teh (eds.), *ICML*, pp. 1263–1272, 2017.

Zhichun Guo, Kehan Guo, Bozhao Nan, Yijun Tian, Roshni G. Iyer, Yihong Ma, Olaf Wiest, Xiangliang Zhang, Wei Wang, Chuxu Zhang, and Nitesh V. Chawla. Graph-based molecular representation learning. In *IJCAI*, pp. 6638–6646, 2023.

Yonatan Harnik and Anat Milo. A focus on molecular representation learning for the prediction of chemical properties. *Chemical Science*, 15(14):5052–5055, 2024.

Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders. In *KDD*, pp. 594–604, 2022.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay S. Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *ICLR*, 2020.

Sarthak Jain and Byron C. Wallace. Attention is not explanation. In *NAACL-HLT*, pp. 3543–3556, 2019.

Apakorn Kengkanna and Masahito Ohue. Enhancing property and activity prediction and interpretation using multiple molecular graph representations with MMGX. *Communications Chemistry*, 7 (1):74, 2024.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *ICLR*, 2020.

Yoshiki Kuramoto. Chemical turbulence. In *Chemical oscillations, waves, and turbulence*, pp. 111–140. Springer, 1984.

Longlong Li, Yipeng Zhang, Guanghui Wang, and Kelin Xia. Kolmogorov–arnold graph neural networks for molecular property prediction. *Nature Machine Intelligence*, 7(8):1346–1354, 2025.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In Sheila A. McIlraith and Kilian Q. Weinberger (eds.), *AAAI*, pp. 3538–3545, 2018.

Zhen Li, Mingjian Jiang, Shuang Wang, and Shugang Zhang. Deep learning methods for molecular representation and property prediction. *Drug Discovery Today*, 27(12):103373, 2022.

Shengchao Liu, Hanchen Wang, Weiyang Liu, Joan Lasenby, Hongyu Guo, and Jian Tang. Pre-training molecular graph representation with 3d geometry. In *ICLR*, 2022.

Shengchao Liu, Weitao Du, Zhi-Ming Ma, Hongyu Guo, and Jian Tang. A group symmetric stochastic differential equation model for molecule multi-modal pretraining. In *ICML*, pp. 21497–21526, 2023.

Yunqing Liu, Yi Zhou, and Wenqi Fan. Enhancing molecular property predictions by learning from bond modelling and interactions. *CoRR*, abs/2603.00568, 2026.

Ziyang Liu, Chaokun Wang, Shuwen Zheng, Cheng Wu, Hao Feng, Li Xu, Yue Zheng, Liang Rong, and Peng Li. Molecular motif learning as a pretraining objective for molecular property prediction. *Nature Communications*, 16(1):11676, 2025.

Antonio Longa, Steve Azzolin, Gabriele Santin, Giulia Cencetti, Pietro Lio, Bruno Lepri, and Andrea Passerini. Explaining the explainers in graph neural networks: a comparative study. *ACM Comput. Surv.*, 57(5):120:1–120:37, 2025.

Siqi Miao, Mia Liu, and Pan Li. Interpretable and generalizable graph learning via stochastic attention mechanism. In *ICML*, pp. 15524–15543, 2022.

Takeru Miyato, Sindy Löwe, Andreas Geiger, and Max Welling. Artificial kuramoto oscillatory neurons. In *ICLR*, 2025.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *AAAI*, pp. 4602–4609, 2019.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *ICLR*, 2020.

Jianbo Qiao, Junru Jin, Ding Wang, Saisai Teng, Junyu Zhang, Xuetong Yang, Yuhang Liu, Yu Wang, Lizhen Cui, Quan Zou, Ran Su, and Leyi Wei. A self-conformation-aware pre-training framework for molecular property prediction with substructure interpretability. *Nature Communications*, 16 (1):4382, 2025.

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

Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *NeurIPS*, 2020.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In Marina Meila and Tong Zhang (eds.), *ICML*, pp. 9323–9332, 2021.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NIPS*, pp. 991–1001, 2017.

Hannes Stärk, Dominique Beaini, Gabriele Corso, Prudencio Tossou, Christian Dallago, Stephan Günnemann, and Pietro Lió. 3d infomax improves gnns for molecular property prediction. In *ICML*, pp. 20479–20502, 2022.

Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. *CoRR*, abs/1908.01000, 2019.

Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. In *ICLR*, 2022.

Bowen Wang, Junyou Li, Donghao Zhou, Lanqing Li, Jinpeng Li, Ercheng Wang, Jianye Hao, Liang Shi, Chengqiang Lu, Jiezhong Qiu, Tingjun Hou, Dongsheng Cao, Guangyong Chen, and Pheng Ann Heng. Unified and explainable molecular representation learning for imperfectly annotated data from the hypergraph view. *Nature Communications*, 16(1):8717, 2025.

Yuyang Wang, Jianren Wang, Zhonglin Cao, and Amir Barati Farimani. Molecular contrastive learning of representations via graph neural networks. *Nat. Mach. Intell.*, 4(3):279–287, 2022.

Oliver Wieder, Stefan Kohlbacher, Mélaine Kuenemann, Arthur Garon, Pierre Ducrot, Thomas Seidel, and Thierry Langer. A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies*, 37:1–12, 2020.

Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: a benchmark for molecular machine learning. *Chemical Science*, 9(2):513–530, 2018.

Zhenxing Wu, Jike Wang, Hongyan Du, Dejun Jiang, Yu Kang, Dan Li, Peichen Pan, Yafeng Deng, Dongsheng Cao, Chang-Yu Hsieh, and Tingjun Hou. Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking. *Nature Communications*, 14(1):2585, 2023.

Jun Xia, Lecheng Zhang, Xiao Zhu, Yue Liu, Zhangyang Gao, Bozhen Hu, Cheng Tan, Jiangbin Zheng, Siyuan Li, and Stan Z. Li. Understanding the limitations of deep models for molecular property prediction: Insights and solutions. In *NeurIPS*, 2023a.

Jun Xia, Chengshuai Zhao, Bozhen Hu, Zhangyang Gao, Cheng Tan, Yue Liu, Siyuan Li, and Stan Z. Li. Mole-bert: Rethinking pre-training graph neural networks for molecules. In *ICLR*, 2023b.

Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, and Mingyue Zheng. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of Medicinal Chemistry*, 63(16):8749–8760, 2020.

Kevin Yang, Kyle Swanson, Wengong Jin, Connor W. Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian P. Kelley, Miriam Mathea, Andrew Palmer, Volker Settels, Tommi Jaakkola, Klavs F. Jensen, and Regina Barzilay. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform badly for graph representation? In *NeurIPS*, pp. 28877–28888, 2021.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. In *NeurIPS*, 2020.

Qiying Yu, Yudi Zhang, Yuyan Ni, Shikun Feng, Yanyan Lan, Hao Zhou, and Jingjing Liu. Multi-modal molecular pretraining via modality blending. In *ICLR*, 2024.

Zhaoning Yu and Hongyang Gao. MAGE: model-level graph neural networks explanations via motif-based graph generation. In *ICLR*, 2025.

Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, and Guolin Ke. Uni-mol: A universal 3d molecular representation learning framework. In *ICLR*, 2023.

Abdeljalil Zoubir and Badr Missaoui. GeoScatt-GNN: a geometric scattering transform-based graph neural network model for Ames mutagenicity prediction. *Neural Networks*, 195:108214, 2026.

A Related Work

Molecular Representation Learning Different featurization schemes lead to different predictive behavior across benchmarks, making molecular representation a central factor in molecular property prediction Guo et al. (2023); Harnik & Milo (2024); Wang et al. (2025). Early methods used fingerprints and hand-crafted descriptors to encode chemical structure, and they remain strong baselines on MoleculeNet-style benchmarks Durant et al. (2002); Rogers & Hahn (2010); Wu et al. (2018). Neural representation learning then shifted the focus from fixed descriptors to learnable molecular states. Message-passing models learn atom- and bond-level features from molecular graphs, while directed propagation and attention-based mechanisms improve the modeling of local chemical environments and task-relevant interactions Gilmer et al. (2017); Yang et al. (2019); Xiong et al. (2020). Geometry-aware models further capture three-dimensional molecular structure beyond 2D connectivity by encoding interatomic distances, bond angles, molecular conformers, and equivariant constraints Schütt et al. (2017); Klicpera et al. (2020); Zhou et al. (2023). Recent pretraining frameworks further improve transferability by learning molecular regularities from large-scale data and auxiliary chemical objectives Rong et al. (2020); Xia et al. (2023b); Qiao et al. (2025); Liu et al. (2025). In parallel, geometric scattering provides stable multiscale graph descriptors for molecular toxicity prediction, offering a complementary route to neural molecular representations Zoubir & Missaoui (2026).

Task-Specific Chemical Rationales Molecular property prediction is not a uniform task. Different endpoints arise from different assays and chemical mechanisms, so a high benchmark score does not always imply that a model has learned meaningful chemical evidence. Recent studies show that model comparison can change due to scaffold splits, distribution shift between training and test molecules, or the dominance of activity cliffs in difficult cases Deng et al. (2023); Xia et al. (2023a). This motivates rationale analysis that matches chemical reasoning. Recent molecular explanation methods therefore shift from atom-level saliency to chemically coherent structural evidence Wu et al. (2023); Kengkanna & Ohue (2024); Qiao et al. (2025); Yu & Gao (2025); Chen et al. (2024). Broader graph explanation studies also show that post-hoc rationales are useful but sensitive to methodological choices and evaluation protocols Armgaan et al. (2024); Akkas & Azad (2024); Longa et al. (2025); Azzolin et al. (2025); Fumagalli et al. (2025). Despite progress in extracting predictive substructures and chemically meaningful rationales, most methods still rely on externally optimized attribution mechanisms or predefined chemical units, and rarely examine task specificity for the same molecule across different prediction tasks. Mechanistically grounded molecular explanations thus require intrinsic model quantities that can reveal chemically distinct rationales across task contexts.

B Preliminaries

The predominant paradigm for instantiating the graph encoder $h_\theta(\cdot)$ is the Message Passing Neural Networks (MPNNs) Gilmer et al. (2017); Kipf & Welling (2017). In this framework, the hidden state $\mathbf{h}_v^{(l)}$ of atom v at the l-th layer is iteratively updated by aggregating features from its topological neighborhood $\mathcal{N}(v)$, that is,

$$\mathbf{m}_v^{(l+1)} = \sum_{u \in \mathcal{N}(v)} \mathrm{Message}^{(l)} \left(\mathbf{h}_v^{(l)}, \mathbf{h}_u^{(l)}, \mathbf{x}_{e,uv} \right), \quad \mathbf{h}_v^{(l+1)} = \mathrm{Update}^{(l)} \left(\mathbf{h}_v^{(l)}, \mathbf{m}_v^{(l+1)} \right), \tag{23}$$

where $\mathrm{Message}(\cdot)$ and $\mathrm{Update}(\cdot)$ represent layer-wise, differentiable message and update functions, $\mathcal{N}(v)$ denotes the neighbor set of node v, and $\mathbf{h}_v^{(0)}$ is initialized with the atom feature $\mathbf{x}_{atom,v}$.

C Introduction of Datasets

MoleculeNet datasets. The eight MoleculeNet Wu et al. (2018) classification datasets considered in this work span different experimental settings, assay scales, and biomedical endpoints. They include permeability prediction, toxicity profiling, side-effect classification, antiviral activity, target inhibition, and virtual screening tasks. Such diversity makes them suitable for evaluating whether a molecular graph model can generalize across heterogeneous chemical and biological prediction scenarios.

- **BBBP** (Blood–Brain Barrier Penetration) contains 2,039 small molecules and evaluates whether a compound can penetrate the blood–brain barrier. It is formulated as a binary classification task that distinguishes permeable and non-permeable molecules, which is important for assessing central nervous system drug delivery.
- **Tox21** is derived from high-throughput screening assays conducted by the U.S. Environmental Protection Agency and the National Institutes of Health. It contains over 8,000 compounds and evaluates toxicity against 12 biological targets, including nuclear receptor signaling and stress response pathways.
- **ToxCast** extends the toxicity evaluation setting of Tox21 by providing a broader toxicological profiling benchmark. It contains approximately 8,000 to over 10,000 molecules and includes results from more than 600 in vitro high-throughput screening assays covering diverse biochemical pathways.
- **SIDER** records adverse drug reactions across 27 system organ classes. It contains molecules with diverse structural and pharmacological properties, ranging from conventional small-molecule drugs to larger peptide-like compounds.
- **ClinTox** focuses on clinical toxicity and drug approval outcomes. It contains 1,478 drugs and provides binary labels related to clinical toxicity and FDA approval status, making it useful for evaluating safety-related molecular prediction.
- **MUV** is derived from PubChem BioAssay and is designed as a challenging benchmark for virtual screening. It contains around 90,000 molecules and 17 bioassay tasks, aiming to distinguish active compounds from carefully selected decoys.
- **HIV** evaluates whether compounds can inhibit the replication of the Human Immunodeficiency Virus. It contains 41,127 molecules labeled as active or inactive according to their antiviral activity.
- **BACE** is constructed for Alzheimer's disease-related drug discovery and measures the binding and inhibitory effects of compounds against human β-secretase 1 (BACE-1). It contains 1,513 molecules labeled as active inhibitors or inactive compounds.

Regression benchmarks. Six continuous benchmarks evaluate the framework across macroscopic physicochemical properties and microscopic quantum mechanical targets.

- **ESOL** is an aqueous solubility prediction dataset, where the goal is to predict the experimentally measured water solubility of small molecules. This task is closely related to molecular polarity, hydrophobicity, hydrogen-bonding capacity, and overall topology, making it a standard benchmark for physicochemical property prediction.
- **FreeSolv** is used to predict hydration free energies of small molecules in water. Although relatively small in scale, the dataset directly reflects solvation interactions between molecules and the aqueous environment, requiring the model to capture local functional groups, charge distribution, and polar interactions.
- **Lipophilicity** focuses on molecular lipophilicity, usually characterized by octanol/water partition or distribution behavior. This property is closely related to drug absorption, membrane permeability, and in vivo distribution, requiring the model to jointly capture hydrophobic scaffolds, aromatic structures, polar groups, and the overall molecular architecture.
- **QM7** is a quantum-chemical regression dataset mainly used for predicting molecular atomization energies. It consists of small organic molecules with energy labels computed from quantum-chemical calculations, making it suitable for evaluating whether a model can represent chemical bond stability and basic electronic-structure information.
- **QM8** is a quantum-chemical dataset for predicting electronic spectra and excited-state properties. It contains multiple regression targets related to electronic transitions, making it more suitable for evaluating multi-task learning of quantum-chemical response variables.
- **QM9** is a large-scale quantum-chemical benchmark containing small organic molecules with computed molecular properties. It covers geometric, energetic, electronic, and thermodynamic targets, and is widely used to evaluate a model's ability to learn complex relationships between molecular structure, electronic distribution, and quantum-chemical properties.

D Experimental Details and Additional Results

D.1 Hyperparameter Settings

Table 3 summarizes the selected hyperparameter settings of MolSync on the MoleculeNet classification datasets. All reported models use four MolSync blocks with hidden dimension 32. Hyperparameters are selected exclusively by validation ROC-AUC from predefined search spaces: learning rate $\in \{5 \times 10^{-5}, 1 \times 10^{-4}, 3 \times 10^{-4}\}$, batch size $\in \{64, 128, 256\}$, epochs $\in \{300, 400\}$, Kuramoto steps $T \in \{4, 8\}$, and phase-coherence readout channels $K \in \{8, 16\}$. AdamW is used with a 5% linear warm-up followed by cosine learning-rate decay to 10^{-6}. The selected configurations are reported in Table 3.

Table 3: Hyperparameter settings of MolSync on MoleculeNet classification datasets.

Dataset	Learning Rate	Batch Size	Epochs	T	K
BBBP	1×10^{-4}	128	400	4	8
Tox21	3×10^{-4}	64	400	4	8
ToxCast	1×10^{-4}	64	400	4	8
SIDER	1×10^{-4}	128	400	4	8
ClinTox	1×10^{-4}	128	400	4	8
MUV	5×10^{-5}	256	400	8	16
HIV	5×10^{-5}	256	400	8	16
BACE	1×10^{-4}	128	300	4	8

D.2 Predictive Performance on MoleculeNet Regression Tasks

Table 4 reports the results of MolSync on six MoleculeNet regression benchmarks. Overall, MolSync demonstrates favorable continuous property modeling ability on both physicochemical and quantum-chemical prediction tasks. For QM7–9, interatomic distances computed from the coordinates provided by the datasets are concatenated with the original bond features to parameterize the edge-conditioned coupling κ_{ij}, allowing three-dimensional geometry to modulate synchronization while preserving invariance to global rotations and translations. For QM9, we follow the same three-target evaluation setting as Uni-Mol. On ESOL, FreeSolv, and Lipophilicity, MolSync remains competitive with strong molecular representation learning baselines, while stable performance is also observed on QM7–9. These results indicate that the proposed oscillator synchronization mechanism can be applied to both classification and continuous molecular property prediction.

Table 4: Regression performance on molecular property prediction benchmarks. RMSE is reported for ESOL, FreeSolv, and Lipophilicity, while MAE is reported for QM7, QM8, and QM9. Results are shown as mean±std where available; lower is better.

Regression (lower is better ↓)						
	RMSE			**MAE**		
Datasets	**ESOL**	**FreeSolv**	**Lipo**	**QM7**	**QM8**	**QM9**
# Molecules	1128	642	4200	6830	21786	133885
# Tasks	1	1	1	1	12	3
D-MPNN	1.050±0.008	2.082±0.082	0.683±0.016	103.5±8.6	0.0190±0.0001	0.00814±0.00001
AttentiveFP	0.877±0.029	2.073±0.183	0.721±0.001	72.0±2.7	0.0179±0.001	0.00812±0.00001
N-Gram	1.074±0.107	2.688±0.085	0.812±0.028	92.8±4.0	0.0236±0.0006	0.01037±0.00016
$\text{N-Gram}_{\text{XGB}}$	1.083±0.082	5.061±0.744	2.072±0.030	81.9±1.9	0.0215±0.0005	0.00964±0.00031
PretrainGNN	1.100±0.006	2.764±0.002	0.739±0.003	113.2±0.6	0.0200±0.0001	0.00922±0.00004
$\text{GROVER}_{\text{base}}$	0.983±0.090	2.176±0.052	0.817±0.008	94.5±3.8	0.0218±0.0004	0.00984±0.00055
$\text{GROVER}_{\text{large}}$	0.895±0.017	2.272±0.051	0.823±0.010	92.0±0.9	0.0224±0.0003	0.00986±0.00025
GraphMVP	1.029±0.033	–	0.681±0.010	–	–	–
MolCLR	1.271±0.040	2.594±0.249	0.691±0.004	66.8±2.3	0.0178±0.0003	–
GEM	**0.798±0.029**	1.877±0.094	0.660±0.008	58.9±0.8	**0.0171±0.0001**	0.00746±0.00001
MolSync	0.816±0.033	**1.780±0.103**	**0.653±0.008**	**51.8±1.9**	0.0173±0.0001	**0.00722±0.00007**

D.3 Additional Results of Energy-Based Attribution

To complement the preliminary Energy-Based Attribution (EBA) observations in the main text, we report additional results on the full MoleculeNet classification benchmarks. These results correspond to the two claims illustrated in Fig. 1: chemical plausibility and predictive fidelity.

Chemical enrichment analysis. As shown in Fig. 4, high-ΔE atoms exhibit clear task-dependent chemical enrichment patterns. BBBP, ClinTox, SIDER, and Tox21 are mainly enriched in polar hydrogen-bond donor/acceptor sites, indicating that MolSync focuses on hydrogen-bonding and polar interaction regions for ADMET-related endpoints. In contrast, BACE, HIV, and MUV show stronger enrichment in aromatic, ring, and hydrophobic regions, suggesting greater emphasis on molecular scaffolds and hydrophobic interactions in binding- and screening-related tasks. These results suggest that EBA captures task-dependent chemical patterns rather than being biased toward a fixed atom type.

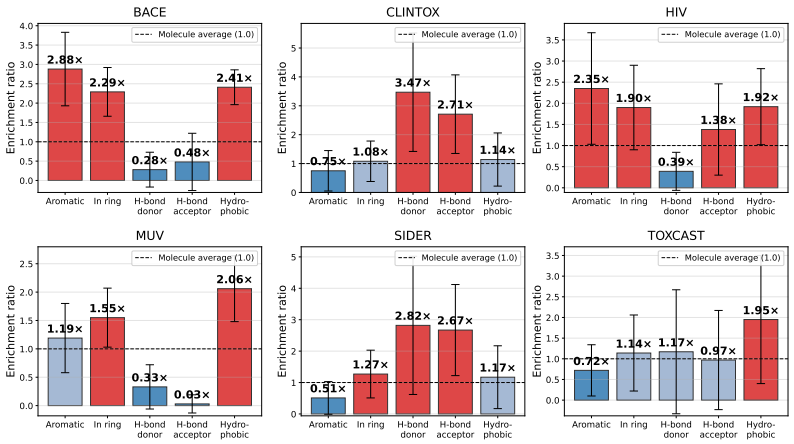

Figure 4: Chemical enrichment analysis of high-ΔE atoms across MoleculeNet datasets.

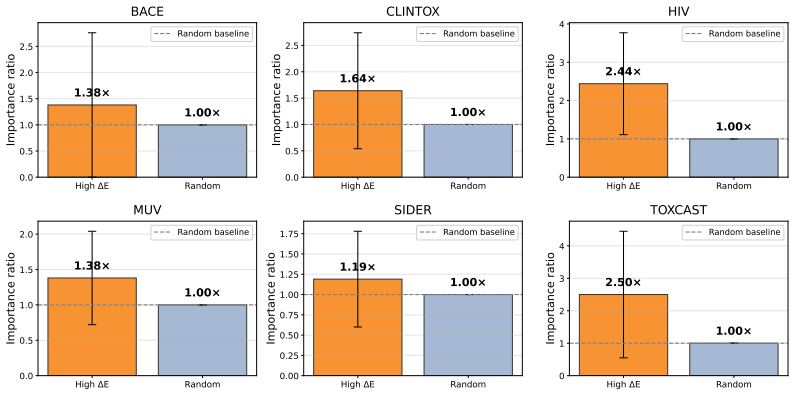

Figure 5: Perturbation validation comparing high-ΔE atom masking with random atom masking.

Perturbation validation. As shown in Fig. 5, perturbing high-ΔE atoms induces larger prediction changes than random perturbation across the evaluated datasets. This suggests that high-ΔE regions are more influential under the adopted perturbation protocol. Therefore, the realized local interaction-potential decrease provides a dynamics-derived atom-level attribution signal for identifying synchronization-relevant molecular regions associated with MolSync's predictions.

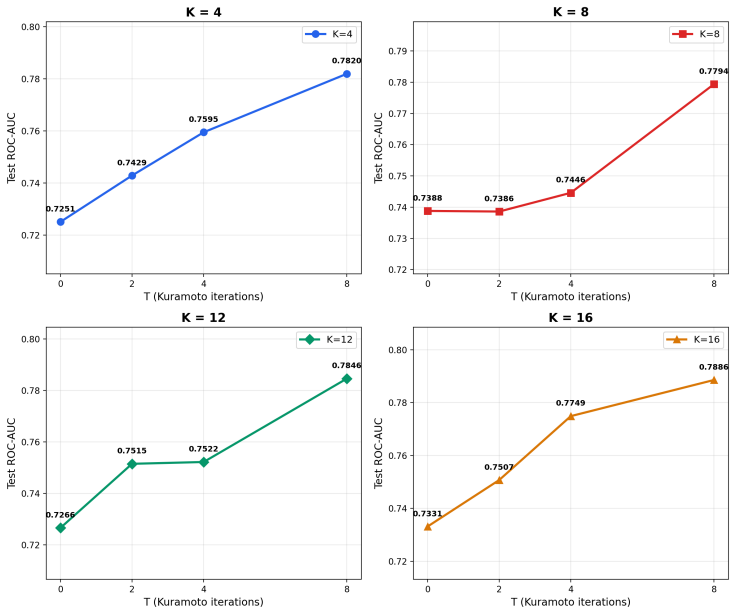

Figure 6: Sensitivity analysis of Kuramoto iteration steps T and phase-coherence readout channels K on BBBP.

D.4 Hyperparameter Study

Analysis of Kuramoto iteration steps and phase-coherence readout channels. To evaluate the contribution of the dynamical module, we conduct a sensitivity study on BBBP by varying the Kuramoto iteration steps T and the number of phase-coherence readout channels K, while keeping other hyperparameters fixed. Here, $T = 0$ removes the synchronization process and serves as a static baseline. As shown in Fig. 6, using Kuramoto evolution consistently improves performance over $T = 0$, indicating that MolSync benefits from explicit multi-step synchronization rather than only from the initialized molecular features. Increasing T generally leads to better performance, while a larger K provides a richer set of coherence channels for summarizing synchronization patterns across neighboring oscillator states. The best results are obtained when sufficient synchronization depth is combined with a relatively large number of phase-coherence readout channels. These results confirm that T and K jointly control the expressive capacity of the synchronization and readout process in MolSync.

D.5 Computational Efficiency

Empirical runtime and memory. We compare practical efficiency on HIV under identical hardware, data split, batch size, and measurement settings. Timing measurements are repeated over three runs, and Table 5 reports the mean values.

Table 5: Runtime and peak GPU memory usage on HIV.

Model	Train (s/epoch)	Train Mem. (MB)	Infer (s/test)	Infer Mem. (MB)
GIN	8.72	271.0	1.07	76.2
GraphGPS	21.82	4176.7	1.67	1430.2
Transformer	13.92	6055.9	0.87	1525.0
MolSync	17.01	2625.5	1.13	748.8

Compared with GIN, MolSync requires $1.95\times$ the training time but only $1.06\times$ the full-test-set inference time. Its peak memory usage is higher than GIN but substantially lower than GraphGPS and the Transformer, indicating moderate training overhead while retaining practical inference efficiency.

D.6 PCA VISUALIZATION OF OSCILLATOR TRAJECTORIES.

To visualize how MolSync reorganizes atom representations during Kuramoto evolution, we extract oscillator states from the first MolSync block at different internal time steps and perform a joint PCA over all states. In Fig. 7, each column corresponds to one Kuramoto step; the molecular structure is colored by mapping the first three principal components to RGB values, while the scatter plot shows the atom states in the two-dimensional PCA space. The initial states are mainly determined by atom features and local chemical environments, whereas subsequent Kuramoto iterations progressively reorganize them. Some atoms move closer and form compact local clusters, while others become separated, indicating both synchronization and anti-synchronization. The resulting clusters do not strictly match manually defined functional groups, suggesting that MolSync does not simply reproduce predefined fragments or uniformly smooth atom representations. Instead, it learns a task-conditioned dynamic organization of atoms through multi-step oscillator evolution.

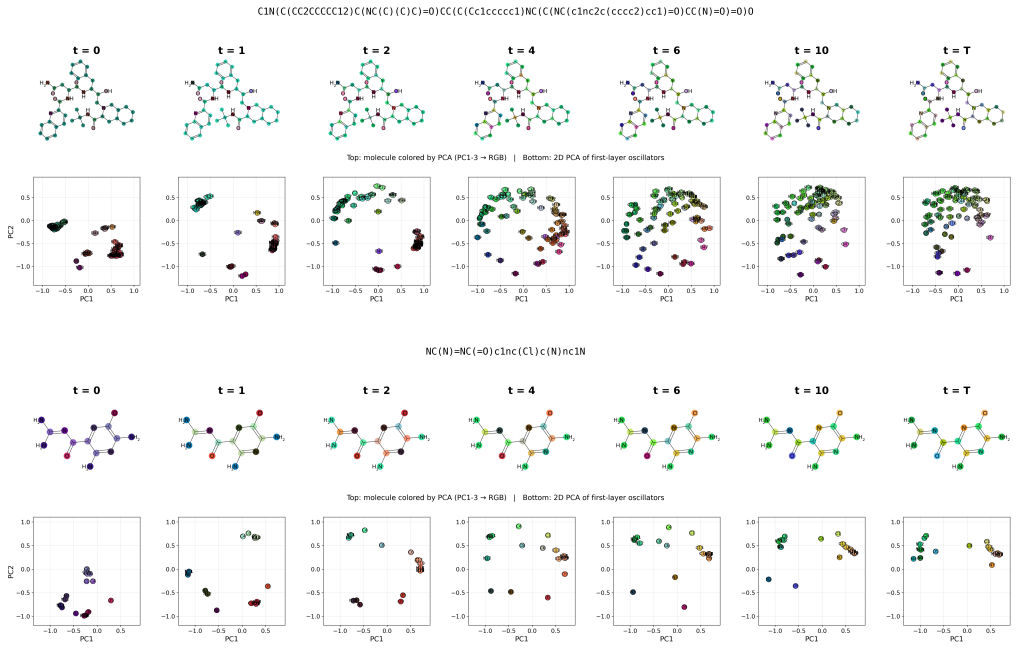

Figure 7: PCA visualization of oscillator trajectories across Kuramoto iteration steps.

E ALGORITHM AND COMPLEXITY ANALYSIS

The forward pass of the proposed MolSync is detailed in Algorithm 1. The architecture numerically integrates the vectorized synchronization dynamics while maintaining sparse edge-wise computation.

The MolSync framework achieves strict linear scalability with respect to graph size, ensuring efficiency comparable to standard message-passing neural networks.

Time Complexity. The computational overhead is dominated by neighborhood aggregation within the internal integration steps and the phase-invariant readout. For a graph with n nodes and $|\mathcal{E}|$ edges, the L-block architecture with T integration steps dictates a complexity of $\mathcal{O}(L(T|\mathcal{E}|d + K|\mathcal{E}|d))$, where K is the number of attention heads. Given that T and K are small constants, the total execution time scales linearly with the number of edges.

Space Complexity. Memory requirements are primarily bounded by the storage of phase states, static stimuli, and multi-scale scattering features, yielding a complexity of $\mathcal{O}(n(d + d_{gst}) + |\mathcal{E}|d_{edge})$. Notably, the discrete-time Riemannian update avoids storing intermediate Jacobian matrices, maintaining a compact memory footprint.

Algorithm 1 Forward Pass of the MOLSYNC Framework

Input: Molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, atom features $\mathbf{X}$, bond features $\mathbf{X}_e$, GST features $\mathbf{S}$.
Parameters: Block count L, internal steps T, step size γ.
Output: Property prediction logits $\hat{y}$.

Stage 1: Bifurcated Phase-Stimulus Initialization

1. $[\mathbf{c}_v^{(0)}, \mathbf{x}_v^{(0,0)}] \leftarrow \mathrm{BifurcatedInit}(\mathbf{x}_v, \mathbf{s}_v)$

Stage 2: Repeated MOLSYNC Blocks (Iterative Evolution & Readout)

for $\ell = 1, \ldots, L$ **do**

// (a) Continuous-Time Kuramoto Dynamics

2. $\kappa_{vu}^{(\ell)} \leftarrow \tanh\big(\mathrm{MLP}_{edge}(\mathbf{x}_{e,vu})\big)$.
3. $\mathbf{\Omega}_v^{(\ell)} \leftarrow \frac{1}{2}(\mathbf{W}_v - \mathbf{W}_v^\top)$, where $\mathbf{W}_v = \mathrm{MLP}_{atom}(\mathbf{c}_v^{(\ell-1)})$.
4. **for** $t = 0, \ldots, T-1$ **do**
5. $\mathbf{d}_v^{(\ell,t)} \leftarrow \mathbf{c}_v^{(\ell-1)} + \sum_{u \in \mathcal{N}(v)} \kappa_{vu}^{(\ell)} \mathbf{x}_u^{(\ell,t)}$.
6. $\mathbf{x}_v^{(\ell,t+1)} \leftarrow \mathrm{Retract}\Big(\mathbf{x}_v^{(\ell,t)} + \gamma\big[\mathbf{\Omega}_v^{(\ell)} \mathbf{x}_v^{(\ell,t)} + \mathrm{Proj}_{\mathbf{x}_v^{(\ell,t)}}(\mathbf{d}_v^{(\ell,t)})\big]\Big)$.
7. **end for**

// (b) Phase-Invariant Readout and Stimulus Update

8. Extract coherence vector $\mathbf{z}_v$ via attention over final states $\mathbf{x}_u^{(\ell,T)}$.
9. $\mathbf{c}_v^{(\ell)} \leftarrow \mathbf{c}_v^{(\ell-1)} + \mathrm{MLP}_{update}(\mathbf{z}_v)$.

end for

Stage 3: Node-to-Graph Readout and Prediction

10. $\mathbf{h}_\mathcal{G} \leftarrow \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \mathbf{c}_v^{(L)}$. *// Global Mean Pool*

return $\hat{y} = \mathrm{MLP}_{pred}(\mathbf{h}_\mathcal{G})$.

F PROOF TO THEOREMS

This section proves the three results in Section 5 and makes explicit which statements concern the linearized/canonical synchronization mechanisms and which statements describe the implementation-aware sensitivity pathway.

F.1 PROOF OF THEOREM 1

Proof. The solution of Eq. 12 is

$$X(t) = e^{-L_\kappa t} X(0) + \int_0^t e^{-L_\kappa (t-\tau)} \Omega \, d\tau. \tag{24}$$

Since Ω is independent of $X(0)$ in this core linearization,

$$\mathcal{J}(t) = \frac{\partial X(t)}{\partial X(0)} = e^{-L_\kappa t} = \sum_{k=0}^{\infty} \frac{(-t)^k}{k!} L_\kappa^k. \tag{25}$$

Let $d = d(i,j)$. For every $k < d$, $(L_\kappa^k)_{ij} = 0$: any nonzero product contributing to the (i,j) entry must contain at least d off-diagonal transitions to move from j to i. At order $k = d$, no diagonal "stay" can occur, and the only contributions are shortest paths. Since $(L_\kappa)_{uv} = -\kappa_{uv}$ for $u \neq v$,

$$(L_\kappa^d)_{ij} = (-1)^d \sum_{p \in \mathcal{P}_d(j \to i)} \prod_{e \in p} \kappa_e = (-1)^d a_{ij}. \tag{26}$$

Substituting this into Eq. 25 gives

$$\mathcal{J}_{ij}(t) = \frac{(-t)^d}{d!} (L_\kappa^d)_{ij} + \mathcal{O}(t^{d+1}) = \frac{a_{ij}}{d!} t^d + \mathcal{O}(t^{d+1}), \tag{27}$$

which proves Eq. 14. If $a_{ij} \neq 0$, analyticity implies that $\mathcal{J}_{ij}(t)$ is not identically zero and is nonzero for sufficiently small positive t. This is the precise sense in which the continuous flow removes a hard finite-hop support cutoff.

If all active couplings are positive, then $-L_\kappa$ is an irreducible Metzler matrix on a connected graph. The semigroup $e^{-L_\kappa t}$ is therefore strictly positive for every $t > 0$, yielding $\mathcal{J}_{ij}(t) > 0$ for all node pairs. When couplings are signed, the Metzler argument no longer applies and path contributions can cancel. This is why the theorem uses the non-degeneracy condition $a_{ij} \neq 0$ and does not claim strict positivity for arbitrary signed couplings or arbitrary times. □

Derivation of the implementation-aware Jacobian. For one MolSync block, let the continuous vectorized dynamics be written abstractly as

$$\dot{X}(t) = F(X(t), C), \qquad X(0) = g(H), \qquad C = h_{\mathcal{G}}(H), \tag{28}$$

where H denotes the input node features to the block while graph topology and bond features are held fixed, and $h_{\mathcal{G}}$ contains the GST-conditioned stimulus construction. The stimulus is fixed during the internal integration of the block; sensitivities across stacked blocks are obtained by composition. Set $J_H(t) = \partial X(t)/\partial H$. Differentiating the ODE gives the non-homogeneous variational equation

$$\dot{J}_H(t) = A(t)J_H(t) + B(t)\frac{\partial h_{\mathcal{G}}(H)}{\partial H}, \qquad J_H(0) = \frac{\partial g(H)}{\partial H}, \tag{29}$$

where $A(t) = \partial F/\partial X$ and $B(t) = \partial F/\partial C$ are evaluated along the forward trajectory. Variation of constants yields

$$J_H(t) = \Phi(t,0)\frac{\partial g(H)}{\partial H} + \int_0^t \Phi(t,\tau)B(\tau)\frac{\partial h_{\mathcal{G}}(H)}{\partial H}\, d\tau, \tag{30}$$

which is Eq. 16. For the specialization $F(X, C) = \Omega(C) - L_\kappa X$, $g(H) = H$, and $\Phi(t, \tau) = e^{-L_\kappa(t-\tau)}$, we obtain Eq. 17. The first term is the usual state-propagation contribution; the second is the GST-conditioned sensitivity contribution. Neither formula supplies a universal positive lower bound on $\|J_{ij}(t)\|$, so the magnitude of long-range sensitivity is an empirical question rather than a consequence of nonzero reachability alone.

F.2 Proof of Theorem 2

Proof. Choose an arbitrary orientation of the m active edges and let $B \in \mathbb{R}^{n\times m}$ be the corresponding incidence matrix. Let $K = \mathrm{diag}(\kappa_e)$ and write $z = B^\top\theta^*$ for the vector of oriented phase differences. At a phase-locked state with common angular velocity $\bar{\omega}$, Eq. 19 can be written, up to the arbitrary edge orientation, as

$$\delta\omega := \omega - \bar{\omega}\mathbf{1} = BK\sin(z). \tag{31}$$

Let $S = \mathrm{diag}(\mathrm{sgn}(\kappa_e))$, so that $K = |K|^{1/2}S|K|^{1/2}$. Using the spectral norm, $\|S\|_2 = 1$, and the element-wise inequality $|\sin z_e| \leq |z_e|$, we obtain

$$\begin{aligned}
\|\delta\omega\|_2 &= \|B|K|^{1/2}S|K|^{1/2}\sin(z)\|_2 && (32)\\
&\leq \|B|K|^{1/2}\|_2\, \||K|^{1/2}\sin(z)\|_2 && (33)\\
&\leq \sqrt{\lambda_{\max}(B|K|B^\top)}\, \||K|^{1/2}z\|_2. && (34)
\end{aligned}$$

By definition,

$$\mathcal{D}_{|\kappa|}(\theta^*) = \frac{1}{2}\||K|^{1/2}B^\top\theta^*\|_2^2 = \frac{1}{2}\||K|^{1/2}z\|_2^2. \tag{35}$$

Therefore,

$$\|\delta\omega\|_2^2 \leq 2\lambda_{\max}(L_{|\kappa|})\,\mathcal{D}_{|\kappa|}(\theta^*), \tag{36}$$

which rearranges to Eq. 20.

If the intrinsic frequencies are heterogeneous, then they cannot all equal the common locked frequency $\bar{\omega}$, so $\delta\omega \neq 0$. Because the active-coupling graph is connected, $\lambda_{\max}(L_{|\kappa|}) > 0$, and the lower bound is strictly positive. Hence an exactly collapsed phase configuration cannot be a phase-locked solution under the stated assumptions. Notice that the use of $|K|$ in $\mathcal{D}_{|\kappa|}$ does not erase the physical distinction between attractive and repulsive interactions in Eq. 19; it only defines a non-negative diagnostic of representational variation. □

Scope relative to the implemented architecture. Theorem 2 analyzes the canonical scalar synchronization mechanism. MolSync adopts a vectorized sphere-valued realization with tangent projection, skew-symmetric natural rotations, retraction, finite integration steps, and residual stimulus updates. The theorem is therefore used to characterize why heterogeneous intrinsic dynamics are incompatible with exact homogeneous phase locking in the underlying Kuramoto mechanism; it is not presented as a complete convergence theorem for every intermediate state of the implemented network. In particular, if the GST-conditioned intrinsic dynamics are exactly identical, Eq. 20 becomes vacuous. The isotropic Gaussian perturbation in the practical initialization can break this degeneracy, but the theoretical guarantee itself remains conditional on heterogeneity.

F.3 Proof of Theorem 3

Proof. Because L_1 and L_2 are real symmetric, they admit orthogonal eigendecompositions $L_r = U_r \Lambda_r U_r^\top$. For every $t > 0$,

$$H_r(t) = e^{-L_r t} = U_r e^{-\Lambda_r t} U_r^\top. \tag{37}$$

Hence the eigenvalues of $H_r(t)$ are $\{e^{-\lambda t} : \lambda \in \mathrm{spec}(L_r)\}$. The map $\lambda \mapsto e^{-\lambda t}$ is injective on $\mathbb{R}$ for fixed $t > 0$. Therefore,

$$\mathrm{spec}(L_1) \neq \mathrm{spec}(L_2) \quad \Longrightarrow \quad \mathrm{spec}(H_1(t)) \neq \mathrm{spec}(H_2(t)). \tag{38}$$

If there existed a permutation matrix P such that $H_1(t) = P^\top H_2(t) P$, the two operators would be similar and would necessarily have identical spectra, a contradiction. Thus, for every permutation matrix P,

$$A_P := H_1(t) - P^\top H_2(t) P \neq 0. \tag{39}$$

Each null space $\ker(A_P)$ is therefore a proper subspace of $\mathbb{R}^n$. There are only finitely many permutation matrices, and a finite union of proper linear subspaces cannot cover $\mathbb{R}^n$. Consequently, there exists a vector q outside all these null spaces. For this q,

$$H_1(t) q \neq P^\top H_2(t) P q \tag{40}$$

for every permutation matrix P. Thus the continuous spectral transfer operators possess separating capacity for the stated graph pair. The homogeneous term already establishes this capacity; adding a graph-conditioned forcing term may enrich the dynamics but is not required for the result.

This conclusion is intentionally conditional. If the graphs are co-spectral, the spectral argument cannot separate them. Even when the spectra differ, a particular input signal may lie in a non-discriminative subspace, and a downstream readout may discard the difference. The theorem therefore establishes a sufficient mechanism available to MolSync rather than a universal strict expressivity ordering over all 1-WL-indistinguishable graphs. □

G Limitations

While MolSync provides mechanism-level theoretical guarantees under explicit assumptions, it currently faces two main practical limitations. First, the numerical ODE integration introduces higher computational overhead during training and inference compared to discrete MPNNs. Second, scaling these continuous-time dynamics from small drug-like molecules to massive macro-molecular graphs (e.g., proteins) remains challenging due to integration complexity. Future work will explore efficient solver approximations to address these bottlenecks.

H Broader Impacts

Positively, MolSync can significantly accelerate computer-aided drug discovery and the design of novel materials. However, we acknowledge a potential dual-use dilemma: highly accurate molecular property predictors could theoretically be misused to identify or design harmful chemical compounds, such as toxins.