GSF-χ: GLOBAL STEREOCHEMICAL FIELDS FOR CHIRAL GRAPH TRANSFORMERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Enantiomers share atoms, bonds, and pairwise distances yet can behave differently in chiral environments, so molecular encoders must respect atom relabelings and proper rotations without becoming blind to reflection. We introduce GSF-χ, a graph transformer in which stereogenic units modulate all pairwise interactions rather than single out one atom as special. Each central or axial stereogenic unit creates a reflection-even phase field over all atoms, a handedness pseudoscalar χ sets the direction of a relative rotation on latent query–key blocks, giving a **Chiral-RoPE** that reflection inverts rather than leaves fixed. A C_2 projection separates mirror-even ECD peak counts and positions from mirror-odd peak signs. We prove the operator's even–odd decomposition and its annotation-inversion, permutation, and unit-order identities under explicit canonical-role conditions; property tests and a coordinate-reflection audit verify the laws end to end. GSF-χ leads every central-ECD output and improves axial Rotation and Symbol by 12.6% and 7.9% over the strongest baseline. Equal-budget controls attribute the Rotation advantage to global signed support rather than parameter or edge count; the C_2 projection yields exact enantiomer-pair consistency at a small raw-accuracy cost under complete supervision and becomes predictive when mirror supervision is scarce.

1 INTRODUCTION

An enantiomer pair exposes a basic failure mode in molecular representation learning: the molecules can agree in atom identities, bonds, and pairwise distances yet behave differently in a chiral environment. For example, (S)-thalidomide binds cereblon about tenfold more strongly than its (R)-enantiomer (Mori et al., 2018), and drug-development guidance calls for separate characterization when stereoisomers differ in pharmacology or toxicity (U.S. Food and Drug Administration, 1992). An encoder must preserve atom relabelings, translations, and proper rotations without identifying a molecule with its mirror image.

Distances and unsigned angles are unchanged by reflection (Gilmer et al., 2017; Schütt et al., 2018; Gasteiger et al., 2020b); an encoder invariant to the full Euclidean group therefore removes the orientation needed to distinguish enantiomers (Dumitrescu et al., 2025). The desired output law is also observable dependent. Configuration, optical rotation, and ECD peak signs reverse under reflection, whereas ECD peak number and position do not. Signed input alone is insufficient unless the readout preserves this parity.

Prior models introduce orientation through ordered neighborhoods, torsions, stereochemical graph features, pretraining, or determinant kernels (Pattanaik et al., 2020; Adams et al., 2021; Gaiński et al., 2023; Du et al., 2024; Yan et al., 2025; Shi et al., 2026b). This leaves open where chirality should enter the attention computation. Center-query constructions make one stereogenic anchor a privileged observer, and axial or multiple units need not provide a unique center. Selecting centers by chemical structure can equally satisfy permutation equivariance, so we treat the all-atom alternative as a design choice backed by the ablations of Section 6.1, keeping every atom in the same architectural role while stereogenic units shape the interactions between them.

GSF-χ realizes this requirement as a global stereochemical field (Figure 1). A central or axial unit produces a reflection-even phase field over all atoms, and one pseudoscalar χ selects the direction of a relative latent rotation. We call that signed rotary operator **Chiral-RoPE**; GSF-χ is the encoder that

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.

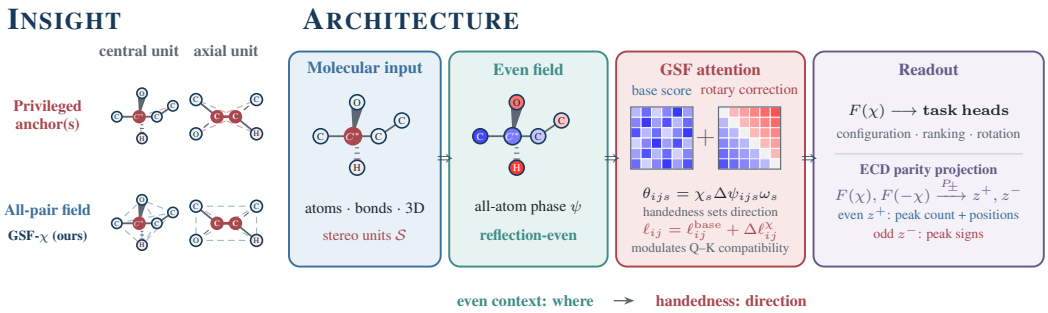


Figure 1: Insight and architecture of GSF-χ. **Left:** stereogenic units modulate all pairwise interactions while every atom keeps ordinary query, key, and value roles. **Right:** parity-even context yields a global phase field, handedness directs the rotary query–key modulation, and the ECD readout splits $F(\pm\chi)$ into even and odd streams (Appendix Figure 6).

drives it with the even field and reads it out through a parity projection. Even context thus determines *where* chirality acts; χ determines *how* the interaction reverses.

Our contributions are: (i) Chiral-RoPE, a relative rotary attention operator whose angle is an atom–unit phase difference and whose sign is a pseudoscalar rather than a sequence position, together with the reflection-even field that drives it over all atom pairs, so that every atom keeps the same architectural role and one interface covers central, axial, and multiple stereogenic units; (ii) exact mirror-symmetry guarantees: relative latent rotations with order-invariant unit aggregation and a C_2-projected ECD readout whose Number and Position are shared and whose Symbol predictions are exactly complementary across an enantiomer pair, verified by 85 property tests and a coordinate-level reflection audit; and (iii) strong empirical results, with GSF-χ achieving state-of-the-art accuracy on standard chirality classification benchmarks and outperforming all baselines on both central and axial ECD prediction.

2 Related Work

Molecular chirality representations. Reflection-even neighborhoods, distances, and unsigned angles cannot alone separate enantiomers (Gilmer et al., 2017; Schütt et al., 2018; Gasteiger et al., 2020b). Chirality-aware networks instead use ordered tetrahedral neighborhoods, permutation-sensitive aggregation, torsions, stereochemical graph features, or determinant kernels (Pattanaik et al., 2020; Gaiński et al., 2023; Adams et al., 2021; Yan et al., 2025; Shi et al., 2026b). Related models learn interpretable chiral kernels (Liu et al., 2023), pretrain Graph Transformers for handedness (Du et al., 2024), or diagnose poor chiral-token recognition in sequence Transformers (Yoshikai et al., 2024). GSF-χ instead asks where a signed signal should enter graph attention; every atom pair remains eligible, and each output receives its required parity.

Rotary and geometric positional encodings. Graphormer and Uni-Mol inject graph or three-dimensional pair structure into attention (Ying et al., 2021; Zhou et al., 2023), but these encodings remain reflection even. RoPE makes compatibility depend on a relative operator (Su et al., 2024); algebraic positional encodings and LieRE extend this principle beyond sequences (Kogkalidis et al., 2024; Ostmeier et al., 2025). Our Chiral-RoPE replaces sequence position with an atom–stereogenic-unit phase and lets a pseudoscalar reverse the relative operator. Unlike descriptor injection in the Dual Graph Transformer (Zhang & Lapkin, 2026), this rotation is defined for every atom pair and obeys an exact inversion identity.

Chiroptical prediction. ECDFormer predicts decoupled peak entities with molecule-conditioned queries (Li et al., 2025); ChiDeK evaluates central and axial ECD with learned chirality-dependent representations (Shi et al., 2026b). Their multitask heads learn the mirror relation from data. GSF-χ projects a shared representation into explicit even and odd parts, making peak number and position invariant and peak signs complementary under inversion of χ.

3 PRELIMINARIES

3.1 MOLECULAR GRAPHS AND STEREOCHEMICAL PREDICTION

We represent a conformer as

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{H}, \boldsymbol{X}), \qquad \boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^\top \in \mathbb{R}^{N\times 3}, \tag{1}$$

where $\mathcal{V}, \mathcal{E}, \boldsymbol{H}$, and $\boldsymbol{X}$ are atoms, bonds, chemical features, and coordinates. Bold symbols denote vectors or matrices, while calligraphic symbols denote sets. An encoder maps $\mathcal{G}$ to atom states $\boldsymbol{Z}$ and task predictions. Atom relabeling must permute $\boldsymbol{Z}$ consistently and leave molecular outputs unchanged.

Let $\mathcal{S}$ denote its stereogenic units. Central units have one anchor; axial units have multiple anchors. Each carries handedness $\chi_s \in \{-1, +1\}$; Section 4.1 defines the full unit.

3.2 RIGID-MOTION SYMMETRY AND PARITY

Coordinates transform as $\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{Q}^\top + \mathbf{1}\boldsymbol{t}^\top$, $\boldsymbol{Q} \in \mathrm{O}(3)$. Proper motions have $\det(\boldsymbol{Q}) = +1$; reflections have $\det(\boldsymbol{Q}) = -1$. Enantiomers are related by the latter but cannot be superposed by an atom relabeling and a proper motion. Scalar observables have two parity types (suppressing the unchanged features $\boldsymbol{H}$):

$$\underbrace{f^+(\boldsymbol{X}') = f^+(\boldsymbol{X})}_{\text{mirror-even}} \qquad \underbrace{f^-(\boldsymbol{X}') = \det(\boldsymbol{Q}) f^-(\boldsymbol{X})}_{\text{mirror-odd}}. \tag{2}$$

Distances and unsigned angles are even; oriented volumes and χ_s are odd. An encoder built only from $\mathrm{O}(3)$-invariant scalars cannot distinguish enantiomers (Dumitrescu et al., 2025); Appendix B gives the paired error bound. ECD peak number and position are even, whereas optical rotation and ECD peak signs are odd; reflection should exchange logits for an odd binary target.

3.3 RELATIVE ROTARY ATTENTION

RoPE applies orthogonal operators to queries and keys, making compatibility depend on $\boldsymbol{R}_i^\top \boldsymbol{R}_j$ rather than absolute positions (Su et al., 2024). A molecular graph has no canonical sequence, and ordinary graph position does not determine reflection parity. We retain the relative-operator principle, replace sequence position by an atom–unit phase, and let χ_s direct a content-dependent rotary modulation. We call the resulting construction *Chiral-RoPE*, since reflection inverts it rather than leaving it fixed. Equation 5 keeps the ordinary query–key path explicit.

4 METHOD

GSF-χ adds signed all-pair compatibility to a molecular Graph Transformer (Figure 1). Reflection-even chemistry defines a field, one pseudoscalar orients it, and the resulting relative rotation modifies query–key compatibility. No atom receives a special attention role.

4.1 A GLOBAL REFLECTION-EVEN FIELD

Let $\mathcal{S}$ contain the valid stereogenic units. A unit is represented by $s = (\tau_s, \mathcal{A}_s, \chi_s, \rho_s)$, where τ_s identifies central or axial chirality, $\mathcal{A}_s$ contains its anchor atom(s), $\chi_s \in \{-1, +1\}$ is handedness, and $\rho_s \in [0, 1]$ is annotation confidence. Each atom–unit relation $\boldsymbol{r}_{is}$ and unit descriptor $\boldsymbol{u}_s$ contain only reflection-even chemical, graph, and geometric quantities.

Each layer retains the initial atom embedding $\boldsymbol{e}_i$ as a separate even stream. Suppressing the layer index, the field is

$$\begin{aligned}(\delta_{is}, \eta_{is}) &= \mathrm{Field}([\boldsymbol{e}_i; \boldsymbol{u}_s; \boldsymbol{r}_{is}]), \qquad n_{is} = \sigma(\eta_{is}),\\ \psi_{is} &= \psi_{is}^{(0)} + \pi \tanh \delta_{is} - \frac{1}{N}\sum_k \left(\psi_{ks}^{(0)} + \pi \tanh \delta_{ks}\right).\end{aligned} \tag{3}$$

The even initialization $\psi^{(0)}$ uses branch/side roles and graph distance (Appendix A); centering fixes its arbitrary origin. The field is evaluated for every atom and independently for every unit.

Canonical stereochemical gauge. Role construction excludes χ_s: CIP information and chirality-free graph ranks align central branches and axial endpoints. Reflection therefore preserves role tensors and reverses only the pseudoscalar. Appendix B.2 gives the algorithm, tie policy, and audit.

4.2 CHIRAL-ROPE

Selected heads are split into three-channel blocks; the even unit descriptor conditions a rotation axis and a positive frequency, with shared and unit-conditioned variants deferred to Appendix A. For pair (i, j),

$$\Delta\psi_{ijs} = \psi_{js} - \psi_{is}, \qquad \theta_{ijshm} = \chi_s \Delta\psi_{ijs}\omega_{shm}, \qquad \boldsymbol{R}_{ijshm} = \exp(\theta_{ijshm}[\boldsymbol{a}_{shm}]_\times) \in \mathrm{SO}(3). \tag{4}$$

The axis rotates latent channels, not physical coordinates.

We apply this Chiral-RoPE rotation as a residual, leaving ordinary attention intact and all values unrotated. For selected head h,

$$\ell^h_{ij} = \underbrace{\frac{(\boldsymbol{q}^h_i)^\top \boldsymbol{k}^h_j}{\sqrt{d_h}} + b^h(\boldsymbol{p}_{ij})}_{\text{ordinary Graph Transformer}} + \underbrace{\frac{\lambda_h}{\sqrt{|\mathcal{S}|\,3M}} \sum_{s\in\mathcal{S}} \rho_s g_{ijs} \sum_{m=1}^{M} (\boldsymbol{q}^h_{im})^\top (\boldsymbol{R}_{ijshm} - \boldsymbol{I}_3)\boldsymbol{k}^h_{jm}}_{\text{global stereochemical correction}}. \tag{5}$$

Here $\boldsymbol{p}_{ij}$ and the relevance gate g_{ijs} are reflection even, and the sum runs over the $M{=}\lfloor d_h/3 \rfloor$ latent channel triples of the selected head with axes $\boldsymbol{a}_{shm}$ (Eq. 14). The second term is not a content-independent bias: $\boldsymbol{q}^\top(\boldsymbol{R} - \boldsymbol{I}_3)\boldsymbol{k} = \boldsymbol{q}^\top \boldsymbol{R}\boldsymbol{k} - \boldsymbol{q}^\top \boldsymbol{k}$ bilinearly modulates current query–key compatibility. At coefficient one, $\boldsymbol{q}^\top \boldsymbol{k} + \boldsymbol{q}^\top(\boldsymbol{R} - \boldsymbol{I}_3)\boldsymbol{k} = \boldsymbol{q}^\top \boldsymbol{R}\boldsymbol{k}$ recovers direct rotary replacement; learned coefficients make it a residual. Any pair, including a remote–remote pair, may receive the correction. Each unit becomes a scalar before summation, so non-commuting rotations require no multiplication order.

Figure 2 visualizes the two obstacles: even geometry cannot orient an odd response, and an anchor query restricts where the sign can enter attention. GSF-χ keeps context even while making the reversible operator available to every pair.

4.3 STRUCTURAL GUARANTEES

Proposition 1 (first-order parity separation). For one rotary block, let $c(\theta) = \boldsymbol{q}^\top(\boldsymbol{R}(\theta) - \boldsymbol{I}_3)\boldsymbol{k}$ with $\theta = \chi_s\omega(\psi_j - \psi_i)$. If the field, axis, frequency, and gate are reflection even, then

$$\frac{c(\theta) - c(-\theta)}{2} = \sin\theta\, \boldsymbol{q}^\top(\boldsymbol{a} \times \boldsymbol{k}) = \theta\, \boldsymbol{q}^\top(\boldsymbol{a} \times \boldsymbol{k}) + \mathcal{O}(\theta^3),$$

whereas the even correction begins at $\mathcal{O}(\theta^2)$. Thus handedness creates a first-order parity-odd compatibility signal without discarding ordinary parity-even attention. By contrast, a reflection-blind encoder has paired squared error at least y^2 on targets $(y, -y)$. The complete decomposition and proof are in Appendix B.

Only χ_s is odd. Handedness inversion and pair reversal therefore invert each block, while achiral input reduces exactly to the ordinary backbone:

$$\boldsymbol{R}_{ijshm}(-\chi_s) = \boldsymbol{R}_{ijshm}(\chi_s)^\top = \boldsymbol{R}_{jishm}(\chi_s), \tag{6}$$

$$\mathcal{S} = \varnothing \text{ or } \rho_s = 0\ \forall s \Longrightarrow \Delta\ell^h_{ij} = 0. \tag{7}$$

Proper-motion-invariant geometric inputs make molecular readout $\mathrm{SE}(3)$ invariant; shared maps give unit-order invariance and atom-permutation equivariance after canonical alignment. Appendix B proves the claims, states the tie qualification, and derives the full decomposition.

4.4 PARITY-COMPATIBLE ECD READOUT

R/S, optical Rotation, and Ranking read the pooled graph state. ECD mixes parities: Number and Position are even; Symbol is odd. We evaluate the shared encoder F at observed and inverted signs

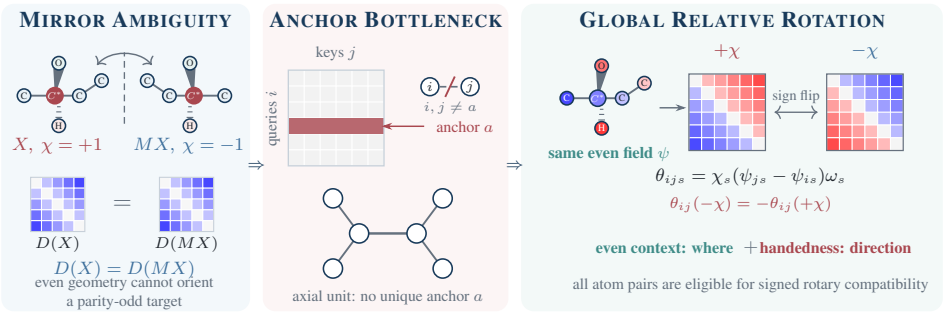


Figure 2: Why chirality must enter all interactions. A reflected input $\boldsymbol{MX}$ has $\boldsymbol{D}(\boldsymbol{X}) = \boldsymbol{D}(\boldsymbol{MX})$ but opposite handedness. An anchor query confines signed compatibility to one row and supplies no unique center for an axial unit. Heatmaps show θ_{ijs}, not attention weights; inversion gives $\theta_{ij}(-\chi) = -\theta_{ij}(+\chi)$.

under identical stochastic masks:

$$\underbrace{\boldsymbol{z}^{+} = \tfrac{1}{2}[F(\chi) + F(-\chi)]}_{\text{even}}, \qquad \underbrace{\boldsymbol{z}^{-} = \tfrac{1}{2}[F(\chi) - F(-\chi)]}_{\text{odd}}. \tag{8}$$

Number and Position read $\boldsymbol{z}^{+}$. A bias-free Symbol score is linear in $\boldsymbol{z}^{-}$, conditioned on $\boldsymbol{z}^{+}$, and produces logits $[-s_k, s_k]$. Consequently,

$$\widehat{N}(-\chi) = \widehat{N}(\chi), \qquad \widehat{\boldsymbol{P}}(-\chi) = \widehat{\boldsymbol{P}}(\chi), \qquad \boldsymbol{\ell}_k^{\text{Symbol}}(-\chi) = \boldsymbol{J}\boldsymbol{\ell}_k^{\text{Symbol}}(\chi), \tag{9}$$

where $\boldsymbol{J}$ swaps Symbol classes. All three ECD metrics come from one encoder and checkpoint, not separate spectrum models.

4.5 Training and Complexity

Training uses task-specific classification, ranking, or structured ECD losses. The correction adds one term per stereogenic unit to each attention entry, so direct evaluation costs $\mathcal{O}(L|\mathcal{S}|N^2d)$ against $\mathcal{O}(LN^2d)$ for the ordinary graph transformer it extends; the asymptotic penalty is the unit count $|\mathcal{S}|$, which averages 1.07 per molecule on ACMP. The measured gap is therefore dominated by constant factors rather than by $|\mathcal{S}|$; with the field disabled the same backbone runs $2.36\times$ faster (Table 17), reflecting the per-pair rotary block and field network. ECD evaluates both signs as one concatenated batch, doubling encoder arithmetic but not asymptotic order. Appendix A provides features, objectives, and settings.

5 Experiments

5.1 Experimental Setup

Datasets. The central tasks use the official ChIRo partitions (Adams et al., 2021): 466,683 conformers for absolute R/S classification and 34,560 enantiomer pairs for Ranking. Central ECD uses the 20,662-molecule public CMCDS reconstruction (Li et al., 2025); enantiomer partners remain in the same 80/10/10 partition. Axial experiments use all 1,192 ACMP molecules released with ChiDeK (Shi et al., 2026b), split into 952/120/120 molecules for training/validation/test. We audit base-identifier overlap and pair preservation before caching features; exact counts are in Appendix Table 5.

Baselines. We compare with 3D geometric networks (DimeNet++ (Gasteiger et al., 2020a) and SphereNet (Liu et al., 2022)), stereochemistry-aware message-passing models (Tetra-DMPNN (Pattanaik et al., 2020), ChIRo (Adams et al., 2021), ChiGNN (Yan et al., 2025), and SPMS (Zhang et al., 2021)), and the ECD/attention models ECDFormer (Li et al., 2025) and ChiDeK (Shi et al., 2026b). Broad benchmark rows follow ChiDeK's comparison (Shi et al., 2026b); the controlled central-ECD

and axial studies retrain ECDFormer, ChiDeK, and GSF-χ on identical molecules, splits, budgets, and evaluators.

Implementation and evaluation. Models are trained from scratch. Unless stated otherwise, means and standard deviations use three runs; the axial comparison averages five runs for the baselines and three for GSF-χ. We report accuracy for R/S, Ranking, axial Rotation, and ECD Symbol, and RMSE for ECD peak Number and Position (Number-RMSE, Position-RMSE, Symbol-Acc, and Rotation accuracy; definitions in Appendix C.2). Appendix Table 3 gives architecture dimensions, optimization, batch sizes, and decoding rules; further experimental details and the split audit appear in Appendices C and C.1, with reporting and comparison regimes in Appendix F.

5.2 Overall Performance

Central chirality. Table 1 gives one row per method. GSF-χ reaches 99.9% R/S and 72.9% Ranking accuracy and leads every central-ECD output against all baselines, raising the parity-sensitive Symbol accuracy from 50.3% to 53.8% (our best prespecified configuration; the like-for-like default attains 53.1%, Appendix C.3). R/S is already nearly saturated, whereas Ranking is a more sensitive test of whether the representation distinguishes two otherwise identical configurations. The ECD numbers in each row come from one jointly trained checkpoint rather than separate models for the three outputs, so the joint improvement cannot be attributed to metric-specific model selection. The upper-block ECD entries were measured on ChiDeK's unreleased processed archive and follow a different data contract, which is why ECD bolding is scoped to the lower block, and the gaps are consistent across runs and statistically significant under two-sided Welch tests on the run statistics (Appendix C.3). The appendix split matrix also reports the historical sample-level split, in which enantiomer partners may cross partitions; ECDFormer then attains the best even metrics, consistent with pair leakage, since Position and Number are shared within an enantiomer pair, whereas the parity-sensitive Symbol still favors GSF-χ. We prespecify the pair-preserving split as primary because it withholds both members of every test pair. The advantage therefore appears in peak location, count, and sign under one data contract, rather than arising from a single favorable output.

Axial chirality. GSF-χ reaches 77.8% Rotation and 74.5% Symbol and leads or ties all four axial outputs, with the largest margins on Rotation and Symbol, the two outputs whose target changes under handedness reversal (Table 1). Several central-chirality models remain near chance on parity-sensitive axial outputs. Appendix Table 24 confirms the ordering under identical splits, with statistically significant gaps on Rotation, Number, and Symbol, and a paired bootstrap over the 60 complete enantiomer pairs resolves both parity-odd outputs with confidence intervals excluding zero (Appendix E.7). This pattern is consistent with the intended bias, as the largest gains occur when orientation, rather than peak location alone, is the missing information. Appendix G makes the paired output law visible qualitatively; GSF-χ preserves enantiomer-paired peak counts and signs where the baselines make visible count or sign errors, and the appendix adds multi-unit and failure cases under a deterministic selection rule (Appendix G.1).

6 Ablation Studies

We organize the ablations around three questions; the extended studies and the axial comparison appear in Appendices D and F.1. In brief, the gains follow from *where* chirality enters attention rather than from how much correction is retained (Q1, Q2), and the C_2 projection converts this into exact mirror consistency at a small, supervision-dependent accuracy cost (Q3).

Q1: Does chirality need to condition all interactions? Replacing the global field with anchor queries, anchor-local injection, or a pooled chirality token degrades every parity-sensitive output on both central and axial tasks, and removing the signed input returns Rotation and Symbol to chance while parity-even targets barely move. The effect sits in the interaction topology rather than in a fragile combination of tuned parts; one learned component, the pair gate, dominates, and every other component swap moves parity-sensitive outputs only marginally (Section 6.1; Figure 3; Table 8).

Q2: Does global placement help beyond support budget? At an equal correction budget, random-global and distance-matched global support improve Rotation over query-anchor support by more

Table 1: Central and axial chirality results. R/S, Ranking, Rotation, and Symbol are accuracies in percent; Pos. and Num. are Position and Number RMSEs; smaller subscripts denote standard deviations. **Upper block**: published and inherited numbers on the official ChIRo splits and the published ACMP evaluation; the central-ECD entries in this block come from ChiDeK's unreleased archive, a different data contract not ranked against the block below. **Lower block**: models retrained on identical public molecules, splits, evaluators, and budgets (central n=3; axial n=5 for the baselines and n=3 for GSF-χ); †reproduces published values, and ECD bolding is scoped to this block (Appendix C.3, Table 24).

	Central chirality					Axial chirality (ACMP)			
Method	R/S ↑	Rank. ↑	Pos. ↓	Num. ↓	Sym. ↑	Rot. ↑	Pos. ↓	Num. ↓	Sym. ↑
DimeNet++	$65.7_{\pm 2.9}$	$58.4_{\pm 0.2}$	$2.12_{\pm 0.19}$	$1.04_{\pm 0.15}$	$50.8_{\pm 0.1}$	$50.0_{\pm 0.0}$	$3.62_{\pm 0.17}$	$1.13_{\pm 0.08}$	$50.0_{\pm 0.1}$
Tetra-DMPNN (c)	$99.7_{\pm 0.1}$	$70.1_{\pm 0.5}$	$2.38_{\pm 0.12}$	$1.25_{\pm 0.09}$	$50.0_{\pm 0.1}$	$50.0_{\pm 0.1}$	$3.15_{\pm 0.12}$	$1.20_{\pm 0.15}$	$52.8_{\pm 0.2}$
Tetra-DMPNN (p)	$99.7_{\pm 0.1}$	$67.6_{\pm 0.6}$	$2.36_{\pm 0.16}$	$1.28_{\pm 0.08}$	$50.0_{\pm 0.1}$	$50.0_{\pm 0.1}$	$3.16_{\pm 0.10}$	$1.05_{\pm 0.12}$	$52.8_{\pm 0.2}$
SphereNet	$98.2_{\pm 0.2}$	$68.6_{\pm 0.3}$	$2.36_{\pm 0.15}$	$1.02_{\pm 0.11}$	$51.9_{\pm 0.3}$	$52.5_{\pm 0.2}$	$3.36_{\pm 0.22}$	$1.08_{\pm 0.09}$	$52.4_{\pm 0.4}$
ChIRo	$98.5_{\pm 0.2}$	$72.0_{\pm 0.5}$	$2.67_{\pm 0.13}$	$1.22_{\pm 0.13}$	$51.0_{\pm 0.4}$	$50.0_{\pm 0.1}$	$3.78_{\pm 0.18}$	$1.11_{\pm 0.11}$	$51.1_{\pm 0.3}$
ECDFormer	$92.3_{\pm 1.2}$	$58.6_{\pm 0.3}$	$2.02_{\pm 0.10}$	$1.01_{\pm 0.09}$	$50.3_{\pm 0.3}$	$53.5_{\pm 0.3}$	$3.89_{\pm 0.25}$	$1.16_{\pm 0.14}$	$51.5_{\pm 0.5}$
ChiGNN	$84.5_{\pm 0.9}$	$59.6_{\pm 0.4}$	$2.39_{\pm 0.18}$	$1.24_{\pm 0.10}$	$49.9_{\pm 0.5}$	$50.2_{\pm 0.1}$	$2.89_{\pm 0.12}$	$1.06_{\pm 0.13}$	$50.8_{\pm 0.8}$
SPMS	$81.4_{\pm 0.7}$	$60.4_{\pm 0.4}$	$2.58_{\pm 0.17}$	$1.22_{\pm 0.12}$	$50.9_{\pm 0.3}$	$65.0_{\pm 0.6}$	$3.69_{\pm 0.21}$	$1.16_{\pm 0.15}$	$60.4_{\pm 0.3}$
ChiDeK	$99.8_{\pm 0.1}$	$72.8_{\pm 0.2}$	$2.20_{\pm 0.14}$	$1.18_{\pm 0.09}$	$53.3_{\pm 0.6}$	$69.2_{\pm 0.5}$	$3.24_{\pm 0.14}$	$1.05_{\pm 0.12}$	$71.2_{\pm 0.6}$
ECDFormer†	$92.3_{\pm 1.2}$	$58.6_{\pm 0.3}$	$2.72_{\pm 0.07}$	$1.32_{\pm 0.02}$	$50.4_{\pm 0.2}$	$50.0_{\pm 0.6}$	$3.70_{\pm 0.24}$	$1.14_{\pm 0.09}$	$52.8_{\pm 0.1}$
ChiDeK†	$99.8_{\pm 0.1}$	$72.8_{\pm 0.2}$	$2.62_{\pm 0.07}$	$1.24_{\pm 0.01}$	$50.3_{\pm 0.2}$	$65.2_{\pm 1.6}$	$\mathbf{2.77}_{\pm 0.10}$	$1.12_{\pm 0.06}$	$66.6_{\pm 1.4}$
GSF-χ (ours)	$\mathbf{99.9}_{\pm 0.0}$	$\mathbf{72.9}_{\pm 0.1}$	$2.19_{\pm 0.02}$	$1.17_{\pm 0.004}$	$\mathbf{53.8}_{\pm 0.3}$	$\mathbf{77.8}_{\pm 1.7}$	$\mathbf{2.77}_{\pm 0.09}$	$\mathbf{1.01}_{\pm 0.03}$	$\mathbf{74.5}_{\pm 1.3}$

than 10%. Equal counts isolate placement from support size, and the ordering is specific to Rotation (Section 6.2; Figure 4a; Appendix E.10).

Q3: What does the C_2-projected readout contribute? With complete paired supervision, projection trades raw Symbol accuracy for an exact enantiomer-pair complement. When mirror supervision is scarce it becomes predictive, improving pair-disjoint and withheld-mirror Symbol substantially (Sections 6.1 and 6.2; Figure 4b,c; Tables 10 and 2).

6.1 Why Does the Global Field Work?

Figure 3 tests three central tasks. Every restriction of the global field (anchor queries, anchor-incident pairs, local injection, or a pooled token) degrades Ranking and all three joint-ECD metrics, with the coarsest restrictions costing the most. This task-level consistency matters because a local sign can solve an almost saturated configuration label, yet it is less effective when a spectrum must coordinate peak count, location, and sign across the molecule.

The axial ablation (Appendix Table 8) repeats the same separation on all four targets. Removing the signed input returns both parity-sensitive outputs to chance while the even targets barely move, and the reflection-even control even attains the lowest Position RMSE, so the field is not a generic capacity increase. Restricting support to anchor queries, or replacing the field with local injection or a pooled token, removes most of the Rotation and Symbol gains, and the effect concentrates in the interaction topology rather than in any single tuned component. Complete values appear in Appendix D.2.

Separating interaction support from output projection. Appendix Table 10 crosses global versus anchor-local interaction with unconstrained versus C_2-projected readout: global support helps under both heads, and projection lowers raw Symbol accuracy by about 5% but makes the enantiomer-pair complement exact, so that a deployed model can never assign physically inconsistent signs to an enantiomer pair. Projection is a structural guarantee, not a universal regularizer; the unconstrained head remains more accurate with complete pairs, whereas projection becomes predictive when mirror supervision is scarce (Figure 4b,c), including after one-enantiomer supervision (Appendix E.11).

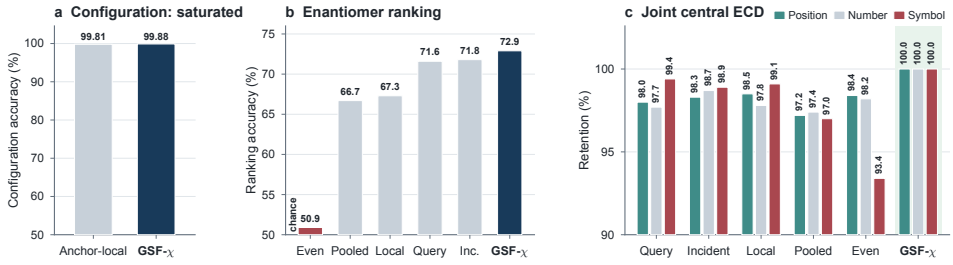


Figure 3: Central-chirality mechanism analysis. **(c)** Joint central-ECD retention relative to the complete model (100%): RMSE uses the full/variant ratio and Symbol the variant/full ratio, so higher is uniformly better. Means over three runs; standard deviations stay below 0.72 (Ranking) and 1.45 (Symbol) percentage points.

6.2 Support Interventions and Structural Evidence

At the same correction budget, random-global and distance-matched global support improve Rotation over query-anchor support by 11.11% and 10.28% (Figure 4a). Equal counts isolate placement from support size, so the gain follows from where the correction may act rather than from how much of it is retained. The ordering is specific to Rotation; Appendix E.10 gives the complete table for all four targets, including a hard top-k variant that does not help.

Corruption, subtype, capacity, efficiency, and same-checkpoint controls appear in Appendix E (Appendices E.1–E.6). All 85 operator and readout tests pass. A separate audit reflects coordinates and reruns preprocessing on all 1,192 ACMP molecules; its field residuals are at machine precision and all output-parity residuals are zero, so the algebraic guarantee survives stereochemical unit extraction and canonical gauge construction.

6.3 Beyond Curated Units and Conformations

Automatic stereogenic units. Handedness is computed from the conformer rather than read from a label (Appendix B.2), so the unit set is the only external input. Replacing it with an external automatic extractor (Shi et al., 2026a) and evaluating the frozen checkpoints without retraining costs about 6% of Rotation and 9% of Symbol accuracy while leaving parity-even targets unchanged, yet the automatic setting still exceeds what ChiDeK attains *with* curated annotations, locating the limitation in upstream detection rather than the operator (Appendix E.8).

Alternative conformations. RotA is ACMP's source corpus, not an independent test; we use it to vary conformation at fixed molecular identity. Table 2 evaluates the same frozen checkpoints on 283 physical mirror pairs from 59 held-out molecules. GSF-χ remains the strongest method under the conformational shift, and its margin over ChiDeK widens from $+12.6$ and $+7.9$ points on ACMP to $+15.4$ and $+10.7$ points here. An advantage that grows on unseen conformations is what one expects if the signal is geometric rather than memorized.

What enforces the mirror law. The right-hand block of Table 2 asks how predictions respond to physical reflection. The C_2-projected Symbol head complements every one of the 283 pairs with zero variance across the three GSF-χ runs, whereas the unprojected Rotation head flips on most of them, indistinguishable from ChiDeK. Both heads read one representation and one checkpoint and differ only in the projection, so exact mirror consistency comes from the projection rather than from learning.

6.4 Representations under Axial Rotation

Following ChiDeK, we rotate one side of an annotated axis through 18 increments of 20°, re-encoding every conformer. Figure 5 visualizes the parity split $z^{\pm} = \frac{1}{2}(F(\chi) \pm F(-\chi))$ used by the ECD readout; the even context stays essentially fixed while the odd trajectory varies with torsion, separating

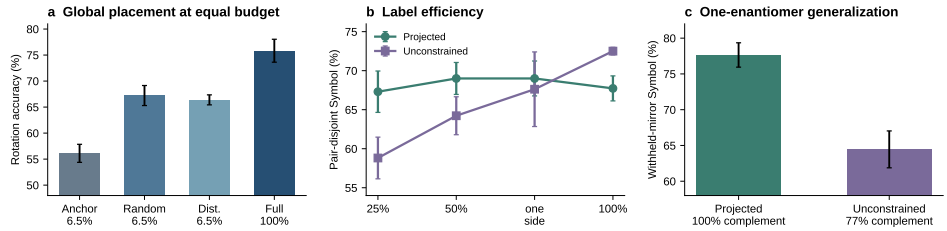


Figure 4: Controls for global placement and parity projection. **(a)** Rotation under equal 6.5% support budgets; full support is a reference. **(b)** Pair-disjoint Symbol under reduced supervision. **(c)** Withheld-mirror Symbol after one-enantiomer supervision. Means ± standard deviations over three runs.

Table 2: RotA conformer evaluation. Mean ± sample standard deviation over frozen ACMP-trained runs (five per baseline, three for GSF-χ) on the same 283 physical mirror pairs; no RotA result is used for model selection. Rotation flip: the predicted class changes between mirror copies; Symbol complement is evaluated at the true peak slots; only the Symbol head is C_2 projected (Appendix E.9).

	Accuracy on RotA conformers				Response to physical reflection	
Method	Rotation ↑	Position ↓	Number ↓	Symbol ↑	Rotation flip ↑	Symbol complement ↑
ECDFormer	50.0 ± 0.0	3.60 ± 0.29	1.19 ± 0.08	53.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
ChiDeK	57.7 ± 2.6	2.80 ± 0.09	1.08 ± 0.04	62.1 ± 2.5	78.7 ± 10.0	73.4 ± 9.3
GSF-χ (ours)	$\mathbf{73.1 \pm 2.0}$	$\mathbf{2.78 \pm 0.03}$	$\mathbf{1.01 \pm 0.05}$	$\mathbf{72.8 \pm 1.3}$	$\mathbf{78.8 \pm 6.5}$	$\mathbf{100.0 \pm 0.0}$

the nine $+\chi$ from the nine $-\chi$ conformers. UMAP serves only the two-dimensional view; polar panels use the original space (McInnes et al., 2018); construction in Appendix F.2.

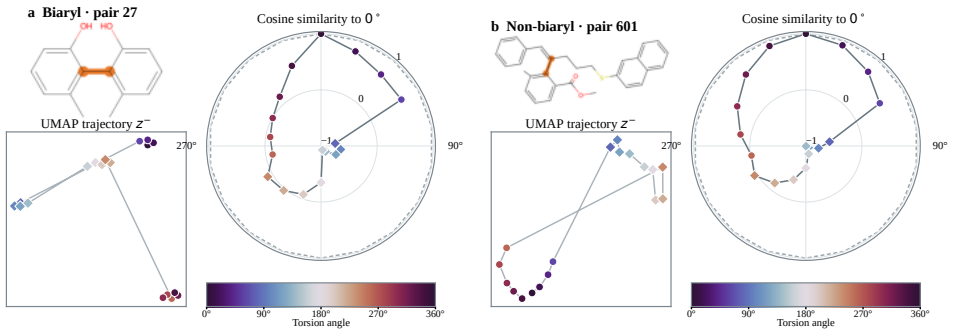


Figure 5: GSF-χ representations during a full axial rotation. Molecule above the UMAP trajectory of z^-; polar plots give cosine similarity to the 0° conformer for z^- (colored) and z^+ (dashed). Color encodes torsion; markers distinguish opposite axial configurations.

7 Conclusion

We presented GSF-χ, a graph transformer that resolves a basic failure mode of molecular representation learning, distinguishing enantiomers that agree in atoms, bonds, and pairwise distances without giving up invariance to relabeling and proper rotation. GSF-χ lets an even context field determine where signed interactions act and a handedness pseudoscalar χ, the Chiral-RoPE operator, determine their direction. It preserves exact mirror-symmetry guarantees by construction and reduces to a standard achiral encoder without stereogenic units. It outperforms all baselines on chirality classification and ECD prediction, and controlled studies attribute the gains to global signed interactions rather than model size or edge count. More broadly, our results offer a general representation that integrates chirality into graph transformers.

AI USE STATEMENT

In this work, we used generative AI writing assistants for proofreading and polishing the English of the manuscript. We have not used generative AI tools for generating research ideas, hypotheses, or experimental designs, and no reported number was produced by a generative AI tool. We have reviewed and revised all AI-assisted text and take responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

All models are trained from scratch; architecture dimensions, optimization, batch sizes, and decoding rules are listed in Appendix Table 3. Dataset construction and the base-identifier overlap audit are described in Appendix C.1 and Table 5. The central-ECD and axial comparisons (Appendices C.3 and F.1) use prespecified run sets, splits, budgets, and evaluators. Per-run values behind every comparison table are serialized in the machine-readable manifests released in the anonymous code supplement at https://anonymous.4open.science/r/GSF-chi-FF2A.

REFERENCES

Keir Adams, Lagnajit Pattanaik, and Connor W Coley. Learning 3d representations of molecular chirality with invariance to bond rotations. *arXiv preprint arXiv:2110.04383*, 2021.

Yifan Du, Junwu Chen, and Philippe Schwaller. Pretrained graph transformers for chirality-related tasks. In *AI for Accelerated Materials Design-Vienna 2024*, 2024.

Alexandru Dumitrescu, Dani Korpela, Markus Heinonen, Yogesh Verma, Valerii Iakovlev, Vikas Garg, and Harri Lähdesmäki. E(3)-equivariant models cannot learn chirality: Field-based molecular generation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=mXHTifc1Fn.

Piotr Gaiński, Michał Koziarski, Jacek Tabor, and Marek Śmieja. Chienn: embracing molecular chirality with graph neural networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 36–52. Springer, 2023.

Johannes Gasteiger, Shankari Giri, Johannes T Margraf, and Stephan Günnemann. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020a.

Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020b. URL https://openreview.net/forum?id=B1eWbxStPH.

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

Konstantinos Kogkalidis, Jean-Philippe Bernardy, and Vikas Garg. Algebraic positional encodings. *Advances in Neural Information Processing Systems*, 37:34824–34845, 2024.

Greg Landrum et al. RDKit: Open-source cheminformatics. https://www.rdkit.org, 2024.

Hao Li, Da Long, Li Yuan, Yu Wang, Yonghong Tian, Xinchang Wang, and Fanyang Mo. Decoupled peak property learning for efficient and interpretable electronic circular dichroism spectrum prediction. *Nature Computational Science*, 5(3):234–244, 2025.

Yi Liu, Limei Wang, Meng Liu, Yuchao Lin, Xuan Zhang, Bora Oztekin, and Shuiwang Ji. Spherical message passing for 3d molecular graphs. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=givsRXsOt9r.

Yunchao Lance Liu, Yu Wang, Oanh Vu, Rocco Moretti, Bobby Bodenheimer, Jens Meiler, and Tyler Derr. Interpretable chirality-aware graph neural network for quantitative structure activity relationship modeling in drug discovery. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 14356–14364, 2023.

Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018. doi: 10.21105/joss.00861.

Tomoyuki Mori, Takumi Ito, Shujie Liu, Hideki Ando, Satoshi Sakamoto, Yuki Yamaguchi, Etsuko Tokunaga, Norio Shibata, Hiroshi Handa, and Toshio Hakoshima. Structural basis of thalidomide enantiomer binding to cereblon. *Scientific reports*, 8(1):1294, 2018.

Sophie Ostmeier, Brian Axelrod, Maya Varma, Michael Moseley, Akshay Chaudhari, and Curtis Langlotz. Liere: lie rotational positional encodings. In *Proceedings of the 42nd International Conference on Machine Learning*, ICML'25. JMLR.org, 2025.

Lagnajit Pattanaik, Octavian-Eugen Ganea, Ian Coley, Klavs F Jensen, William H Green, and Connor W Coley. Message passing networks for molecules with tetrahedral chirality. *arXiv preprint arXiv:2012.00094*, 2020.

Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet–a deep learning architecture for molecules and materials. *The Journal of chemical physics*, 148(24), 2018.

Runhan Shi, Chi Zhang, Gufeng Yu, Xiaohong Huo, and Yang Yang. A unifying geometric framework for computational representation of stereoisomers based on mixed product. *Cell Reports Physical Science*, 7(1), 2026a.

Runhan Shi, Zhicheng Zhang, Letian Chen, Gufeng Yu, and Yang Yang. Learning molecular chirality via chiral determinant kernels. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=E7mrnBuMFd.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

U.S. Food and Drug Administration. Development of new stereoisomeric drugs. Guidance for Industry, Center for Drug Evaluation and Research, 1992. URL https://www.fda.gov/regulatory-information/search-fda-guidance-documents/development-new-stereoisomeric-drugs. Accessed 2026-08-19.

Jiaxin Yan, Haiyuan Wang, Wensheng Yang, Xiaonan Ma, Yajing Sun, and Wenping Hu. Chignn: Interpretable algorithm framework of molecular chiral knowledge-embedding and stereosensitive property prediction. *Journal of Chemical Information and Modeling*, 65(7):3239–3247, 2025.

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? *Advances in neural information processing systems*, 34:28877–28888, 2021.

Yasuhiro Yoshikai, Tadahaya Mizuno, Shumpei Nemoto, and Hiroyuki Kusuhara. Difficulty in chirality recognition for transformer architectures learning chemical structures from string representations. *Nature Communications*, 15(1):1197, 2024.

Shuo-Qing Zhang, Xin Hong, Li-Cheng Xu, Xin Li, Miao-Jiong Tang, Luo-Tian Yuan, and Jia-Yu Zheng. A molecular stereostructure descriptor based on spherical projection. *Synlett*, 32(18): 1837–1842, 2021.

Shuyuan Zhang and Alexei A Lapkin. Enhancing molecular property prediction of transformer models with dual graph representation. *Nature Communications*, 2026.

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

A Full GSF-χ Parameterization

This appendix specifies the implementation summarized in Section 4. These details are required for reproduction.

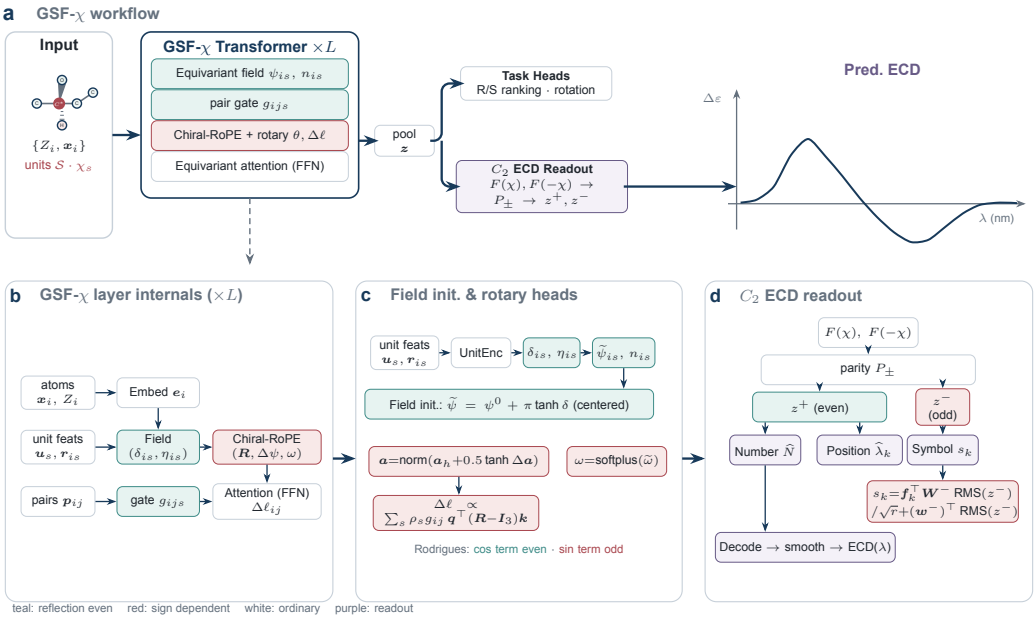


Figure 6: GSF-χ pipeline. **(a)** A chiral molecule supplies parity-even inputs and one handedness sign χ_s; the even field and Chiral-RoPE blocks produce signed attention, stacked L times with ordinary transformer updates. Task heads read the pooled representation directly, while the C_2-projected ECD readout decodes peak counts, positions, and signs into a predicted spectrum. **(b)** Per-layer data flow. **(c)** Field, gate, and rotary-head parameterizations. **(d)** Parity-split ECD readout. Exact definitions are given in Section 4 and Appendix A.

A.1 Stereochemical Units and Even Inputs

For $s = (\tau_s, \mathcal{A}_s, \chi_s, \rho_s)$, each atom–unit pair has a 26-dimensional relation vector $\boldsymbol{r}_{is}$ and each unit has a 342-dimensional descriptor $\boldsymbol{u}_s$. Under $\boldsymbol{x}'_i = \boldsymbol{Q}\boldsymbol{x}_i + \boldsymbol{t}$, $\boldsymbol{Q} \in \mathrm{O}(3)$,

$$\chi'_s = \det(\boldsymbol{Q})\chi_s, \qquad \boldsymbol{r}'_{is} = \boldsymbol{r}_{is}, \qquad \boldsymbol{u}'_s = \boldsymbol{u}_s. \tag{10}$$

For a center, $\boldsymbol{r}_{is}$ contains graph and Euclidean distance, shortest-path bond statistics, a center indicator, chirality-free first-hop branch identity, and even non-planarity summaries. For an axis, it contains distances to canonically ordered endpoints, radial and longitudinal relations, side and branch roles, path statistics, axis length, $|\sin\phi_s|$, and $\cos\phi_s$. The unit descriptor collects anchor/substituent features and even geometric summaries. The ordinary pair stream is

$$\boldsymbol{p}_{ij} = \left[\frac{\min\{d_G(i,j), 15\}}{15}, \frac{\min\{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2, 20\}}{20}, \mathbf{1}[(i,j) \in \mathcal{E}]\right]. \tag{11}$$

A.2 Field Network and Initialization

At layer ℓ, an even atom stream $\boldsymbol{e}_i = \mathrm{Embed}(\boldsymbol{h}_i^{\mathrm{raw}})$ is held separate from the chirality-modulated hidden state. We compute

$$\begin{aligned}
\boldsymbol{u}_s^{(\ell)} &= \mathrm{UnitEnc}^{(\ell)}(\boldsymbol{u}_s), \\
(\delta_{is}^{(\ell)}, \eta_{is}^{(\ell)}) &= \mathrm{Field}^{(\ell)}([\boldsymbol{e}_i; \boldsymbol{u}_s^{(\ell)}; \boldsymbol{r}_{is}]), \\
\widetilde{\psi}_{is}^{(\ell)} &= \psi_{is}^{(0)} + \pi \tanh \delta_{is}^{(\ell)}, \qquad n_{is}^{(\ell)} = \sigma(\eta_{is}^{(\ell)}), \\
\psi_{is}^{(\ell)} &= \widetilde{\psi}_{is}^{(\ell)} - N^{-1} \sum_{k=1}^{N} \widetilde{\psi}_{ks}^{(\ell)}.
\end{aligned} \tag{12}$$

The chemistry-informed initial phase is

$$\psi_{is}^{(0)} = \begin{cases} \alpha_{b_c(i)} + 0.12\, d_G(c,i), & s \text{ is center } c, \\ \frac{1}{2}\sigma_{is}\,\mathrm{atan2}(|\sin\phi_s|, \cos\phi_s) + 0.12\,\min\{d_G(i,a), d_G(i,b)\}, & s \text{ is axis } (a,b), \end{cases} \tag{13}$$

with $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 2\pi/3, 4\pi/3, 0)$. Branch and side terms separate chemical regions; graph distance propagates the initialization beyond the first shell. Handedness is never used in this initialization.

A.3 Conditioned Rotations and Pair Gates

For head width d_h, the first $3M$ channels of a selected head form $M = \lfloor d_h/3 \rfloor$ blocks. The implementation uses unit-conditioned axes and supports both shared and unit-conditioned frequencies:

$$\boldsymbol{a}_{shm} = \mathrm{norm}(\boldsymbol{a}_{hm} + 0.5\tanh\Delta\boldsymbol{a}_{shm})\,, \qquad \omega_{hm} = \mathrm{softplus}(\widetilde{\omega}_{hm}), \tag{14}$$

where the axis residual is generated from $\boldsymbol{u}_s^{(\ell)}$ and initialized at zero. Frequencies are initialized logarithmically from 1 to 10^{-2} and learned during training. The central tasks and the axial comparison use the parity-even conditioned form $\omega_{shm} = \omega_{hm}\exp(0.5\tanh\Delta\omega_{shm})$; our best axial configuration uses the shared special case $\Delta\omega_{shm} = 0$.

For $\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{a} \in \mathbb{R}^3$, Rodrigues' identity evaluates the rotary correction without a generic matrix exponential:

$$\boldsymbol{q}^\top(\boldsymbol{R} - \boldsymbol{I}_3)\boldsymbol{k} = (\cos\theta - 1)[\boldsymbol{q}^\top\boldsymbol{k} - (\boldsymbol{q}^\top\boldsymbol{a})(\boldsymbol{k}^\top\boldsymbol{a})] + \sin\theta\,\boldsymbol{q}^\top(\boldsymbol{a}\times\boldsymbol{k}). \tag{15}$$

The cosine term is even and the orientation-sensitive sine term is odd in θ. Pair relevance is

$$g_{ijs} = \sqrt{n_{is}n_{js}}\,\sigma(\mathrm{MLP}_g([\boldsymbol{r}_{is}; \boldsymbol{r}_{js}; \boldsymbol{p}_{ij}]))\,. \tag{16}$$

The complete correction is

$$\Delta\ell_{ij}^h = \frac{\lambda_h}{\sqrt{|\mathcal{S}|}\sqrt{3M}} \sum_{s\in\mathcal{S}} \rho_s g_{ijs} \sum_{m=1}^{M} (\boldsymbol{q}_{im}^h)^\top(\boldsymbol{R}_{ijshm} - \boldsymbol{I}_3)\boldsymbol{k}_{jm}^h. \tag{17}$$

This is not a conventional additive attention bias; it depends bilinearly on the current $\boldsymbol{q}_i^h$ and $\boldsymbol{k}_j^h$. Indeed, $\boldsymbol{q}^\top(\boldsymbol{R} - \boldsymbol{I}_3)\boldsymbol{k} = \boldsymbol{q}^\top\boldsymbol{R}\boldsymbol{k} - \boldsymbol{q}^\top\boldsymbol{k}$, whereas the query–key-independent pair bias is the separate term $b^h(\boldsymbol{p}_{ij})$. The layer then applies the standard softmax over $\boldsymbol{q}_i^\top\boldsymbol{k}_j/\sqrt{d_h} + b^h(\boldsymbol{p}_{ij}) + \Delta\ell_{ij}^h$ and aggregates unrotated values.

A.4 Heads, Objectives, and Regularization

After mean pooling, a LayerNorm–MLP predicts configuration and optical rotation. Ranking uses an MSE term plus a pairwise margin loss,

$$\mathcal{L}_{\mathrm{rank}} = \mathcal{L}_{\mathrm{MSE}} + \beta\max(0, m - y_{ab}(\widehat{y}_a - \widehat{y}_b)). \tag{18}$$

For ECD, Number and Position heads read $\boldsymbol{z}^+$. The Symbol score

$$s_k = \frac{\boldsymbol{f}_k(\boldsymbol{z}^+)^\top\boldsymbol{W}_k^-\,\mathrm{RMS}(\boldsymbol{z}^-)}{\sqrt{r}} + (\boldsymbol{w}_k^-)^\top\,\mathrm{RMS}(\boldsymbol{z}^-) \tag{19}$$

is bias free in the odd input and produces logits $[-s_k, s_k]$. A private even position context may condition $\boldsymbol{f}_k$ without sending Symbol gradients into the formal Position head.

The ECD objective is

$$\mathcal{L}_{\mathrm{ECD}} = w_N\mathcal{L}_N + w_P\mathcal{L}_P + 2\mathcal{L}_{\mathrm{Symbol}} + \mathcal{L}_{\mathrm{structured}} + \gamma_\ell\mathcal{L}_{\mathrm{logit}} + \gamma_\psi\mathcal{L}_{\mathrm{smooth}}, \tag{20}$$

$$\mathcal{L}_{\mathrm{logit}} = L^{-1}\sum_{\ell=1}^{L}\mathrm{mean}[(\Delta\ell^{(\ell)})^2],$$

$$\mathcal{L}_{\mathrm{smooth}} = \frac{\sum_s\sum_{(i,j)\in\mathcal{E}}(\psi_{is} - \psi_{js})^2}{\sum_s\sum_{(i,j)\in\mathcal{E}}1}. \tag{21}$$

Expected squared-distance risk is used for ordered Number and Position classes. Whole-unit dropout replaces ρ_s by $\rho_s m_s/(1-p)$, $m_s \sim \mathrm{Bernoulli}(1-p)$, with the same mask in both mirror forwards.

A.5 Task-Specific Training Configurations

Table 3 records the configurations behind the reported models. All models are initialized from scratch; full command-line arguments and histories are included in the released machine-readable manifests.

Table 3: Implementation configurations. H_χ/H is the number of GSF heads over total attention heads; p_u is stereogenic-unit dropout.

Task	d	L	H_χ/H	Epochs	Batch	Optimizer	LR / WD	p_u / scale
Central R/S	48	2	2/4	10	128	AdamW	$5\times10^{-4}/10^{-4}$	$0.2/5.0$
Central Ranking	384	12	9/12	200	128	AdamW + EMA	$10^{-4}/3\times10^{-4}$	$0.2/4.6$
Central ECD	48	2	4/4	50	128	AdamW	$5\times10^{-4}/10^{-4}$	$0.2/5.0$
Axial Rotation	48	2	2/4	25	24	AdamW	$5\times10^{-4}/10^{-4}$	$0.0/5.0$
Axial ECD	48	2	2/4	25	32	Adam	$3\times10^{-4}/10^{-5}$	$0.2/5.0$

Central Ranking additionally uses dropout 0.15, margin 0.3, and EMA decay 0.9993. Its best configuration shares only the GSF unit encoder, latent axes, and frequencies across the 12 layers; phase fields, attention projections, and feed-forward blocks remain layer-specific. It has $28{,}895{,}233$ trainable parameters. Central ECD uses parity-split readout, categorical Number, posterior-mean Position decoding, and risk weights 0.15 for Number and Position. Axial ECD uses ordinal Number. The central tasks and the axial replication use unit-conditioned frequencies, whereas our best axial configuration uses the shared-frequency special case. The remaining configurations use dropout 0.1.

A.6 Implementation Cost and Scope

For L layers, N atoms, width d, and $|\mathcal{S}|$ units, direct evaluation costs $\mathcal{O}(L|\mathcal{S}|N^2d)$ time and stores $\mathcal{O}(B|\mathcal{S}||\mathcal{H}_\chi|N^2M)$ rotary intermediates. Units, heads, blocks, and atom pairs are vectorized, and Eq. 15 is applied only to selected heads. "Field" denotes a learned field over atom tokens.

B Proofs of the Structural Guarantees

This section expands the proof sketches in Section 4.3. The argument has three layers. First, a no-go result explains why a reflection-blind representation cannot predict a nonzero parity-odd observable. Second, we characterize the symmetry of the GSF-χ rotary operator, including the exact even and odd parts of its logit correction. Third, we show that the paired readout is the canonical projection onto the two representations of the reflection group C_2. Operator inversion and output parity are deliberately kept separate; the former creates a signed interaction, while only the latter enforces an exact task-level law.

B.1 Assumptions and Notation

For a stereogenic unit s, let $\boldsymbol{u}_s$ collect its parity-even descriptor and let $\boldsymbol{r}_{is}$ collect the parity-even atom–unit relation. We use the following conditions; the released preprocessing is checked by both property tests and the dataset-wide audit below.

1. The tensors $\boldsymbol{u}_s$, $\boldsymbol{r}_{is}$, $\boldsymbol{p}_{ij}$, and the confidence ρ_s are unchanged by a proper rigid motion. Among these constructed inputs, reflection changes only the explicit handedness sign χ_s.
2. Atom and unit relabelings permute the corresponding tensor axes; all learned maps share parameters over those axes.
3. Axial endpoints and substituent roles are put in the operational canonical gauge of Appendix B.2 before $\boldsymbol{u}_s$, $\boldsymbol{r}_{is}$, ψ_{is}, and χ_s are constructed. The physical-reflection guarantee requires this preprocessing map to preserve all even tensors and reverse the supplied pseudoscalar sign; this condition is audited directly rather than inferred from the network algebra.
4. The two forwards used by the ECD parity projection share weights and, during stochastic training, share every random mask.

The channel-space axis $\boldsymbol{a}_{shm}$ has unit norm, and $\boldsymbol{A}_{shm} = [\boldsymbol{a}_{shm}]_\times$ is its skew-symmetric cross-product matrix. Throughout, $S = |\mathcal{S}|$ denotes the number of valid stereogenic units.

B.2 Operational Canonicalization and Reflection Audit

The implementation first computes RDKit canonical graph ranks with chirality disabled. External substituents on each side of an axial unit are sorted by the tuple

$$(\text{CIPRank}, Z, \text{isotope}, \text{formal charge}, \text{graph rank}). \tag{22}$$

The endpoint signature concatenates its sorted substituent keys with the same parity-even atom attributes. The lexicographically larger signature defines the first endpoint; if the signatures agree, RDKit's unique canonical rank, again computed without chirality, fixes the endpoint direction. Handedness is computed only after these roles are fixed. Singleton spiral-unit records use the same neighbor key. Equal singleton-neighbor keys retain their stable serialized order; hence the exact atom-permutation statement is conditional on canonical alignment for these rare records, whereas the physical-reflection law is unaffected because reflection does not change atom serialization.

We test the preprocessing condition by applying $\boldsymbol{Q} = \text{diag}(-1, 1, 1)$, with $\det(\boldsymbol{Q}) = -1$, to every RDKit conformer and to every serialized geometric stereochemical frame. We recompute the frame determinants and rerun the standard sample builder rather than manually negating a network tensor. Table 4 reports the result. Endpoint signatures are chemically equivalent for 552 of 1,258 axial-unit instances, and every such case is resolved by the chirality-free unique canonical rank. The 22 singleton instances contain 24 equal neighbor equivalence classes. Two zero determinants occur only among unused alternate frame candidates; the selected frame remains nondegenerate.

This audit establishes the transformation law for the released ACMP preprocessing path and the tested annotation records. It does not imply that arbitrary external conformer generators or uncertain stereochemical annotations will satisfy the same contract.

B.3 Why Parity-Sensitive Information Is Necessary

Let $\mathcal{M}$ denote a molecular reflection, including the induced sign change $\chi_s \mapsto -\chi_s$ for every valid stereogenic unit.

Proposition (no-go for reflection-blind encoders). Suppose an encoder Φ is reflection blind, $\Phi(\mathcal{M}\mathcal{G}) = \Phi(\mathcal{G})$. Then no deterministic head $g \circ \Phi$ can represent a nonzero mirror-odd target y^- satisfying $y^-(\mathcal{M}\mathcal{G}) = -y^-(\mathcal{G})$ on both members of an enantiomeric pair.

Indeed, invariance forces the two predictions to be the same value c. For a scalar target with $y^-(\mathcal{G}) = y \neq 0$, the mean paired squared error obeys

$$\frac{(c-y)^2 + (c+y)^2}{2} = c^2 + y^2 \geq y^2. \tag{23}$$

For complementary binary labels, identical logits likewise cannot classify both members correctly. The proposition does not require that handedness be encoded specifically by GSF-χ; it says that some parity-sensitive information is necessary. Our construction isolates that information in the pseudoscalar χ_s, while every field and gate used to distribute it remains reflection even.

B.4 Inverse Response to Handedness and Pair Direction

The block rotation is

$$\boldsymbol{R}_{ijshm}(\chi_s) = \exp(\theta_{ijshm}\boldsymbol{A}_{shm}), \qquad \theta_{ijshm} = \chi_s \omega_{shm}(\psi_{js} - \psi_{is}). \tag{24}$$

Because $\boldsymbol{A}_{shm}^\top = -\boldsymbol{A}_{shm}$ and the exponential of a matrix commutes with its own inverse,

$$\boldsymbol{R}_{ijshm}(\chi_s)^{-1} = \exp(-\theta_{ijshm}\boldsymbol{A}_{shm}) = \boldsymbol{R}_{ijshm}(\chi_s)^\top. \tag{25}$$

All factors other than χ_s are parity even, so stereochemical inversion maps θ to $-\theta$. Pair reversal also maps θ to $-\theta$ because

$$\psi_{is} - \psi_{js} = -(\psi_{js} - \psi_{is}). \tag{26}$$

Substitution into Eq. 25 proves Eq. 6. The statement holds independently for every unit, head, and complete three-channel block; residual one- or two-dimensional channels remain ordinary attention channels.

Table 4: Coordinate-reflection audit over all 1,192 ACMP molecules. Feature entries are compared after rebuilding the complete released model input; output logits use one locked checkpoint on the test split.

Quantity	Expected transformation	Maximum absolute residual
node, pair, unit descriptor/relation, confidence	even	0
initialized phase field	even	1.19×10^{-7}
handedness and local signed input	odd	0
CIP ranks	even	0
Number logits	even	0
Position logits	even	0
Symbol logits	class complement	0

B.5 Exact Even–Odd Decomposition of a Rotary Block

The inverse identity also exposes where the chiral signal enters attention. For fixed $\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{a}$ and $\boldsymbol{R}(\theta) = \exp(\theta[\boldsymbol{a}]_\times)$, define the added block score $c(\theta) = \boldsymbol{q}^\top(\boldsymbol{R}(\theta) - \boldsymbol{I}_3)\boldsymbol{k}$. Rodrigues' identity gives the two exact components

$$c_{\mathrm{even}}(\theta) = \frac{c(\theta) + c(-\theta)}{2} = (\cos\theta - 1)\big[\boldsymbol{q}^\top\boldsymbol{k} - (\boldsymbol{q}^\top\boldsymbol{a})(\boldsymbol{k}^\top\boldsymbol{a})\big], \tag{27}$$

$$c_{\mathrm{odd}}(\theta) = \frac{c(\theta) - c(-\theta)}{2} = \sin\theta\, \boldsymbol{q}^\top(\boldsymbol{a} \times \boldsymbol{k}). \tag{28}$$

Because $\theta = \chi_s\omega_{shm}(\psi_{js} - \psi_{is})$ and the conditioned frequency is parity even, reflection reverses the second term exactly. Locally, $c_{\mathrm{odd}}(\theta) = \theta\boldsymbol{q}^\top(\boldsymbol{a} \times \boldsymbol{k}) + \mathcal{O}(\theta^3)$, whereas the even term begins at order θ^2. Thus the signed phase difference creates a first-order orientation-sensitive compatibility signal while retaining an even geometric correction.

This decomposition is conditional on the current $\boldsymbol{q}$ and $\boldsymbol{k}$ block. At deeper layers those states may themselves depend on earlier signed interactions, so it does not claim that every hidden channel has a fixed parity. This is precisely why operator inversion and the exact output projection proved below serve different roles.

B.6 Gauge, Endpoint, and Unit-Order Independence

For an arbitrary constant c_s,

$$(\psi_{js} + c_s) - (\psi_{is} + c_s) = \psi_{js} - \psi_{is}, \tag{29}$$

so neither θ_{ijshm} nor the attention correction depends on the zero of the phase field. Endpoint-order independence is a separate canonicalization property; Assumption 3 maps both axial endpoint tuple orders to the same canonical endpoint and branch roles, after which every tensor entering Eq. 24 is identical.

A naive multi-unit construction could multiply the unit rotations. This would introduce an unphysical enumeration choice because $\mathrm{SO}(3)$ is non-commutative: for rotations about non-collinear axes, in general

$$\boldsymbol{R}_s\boldsymbol{R}_t \neq \boldsymbol{R}_t\boldsymbol{R}_s. \tag{30}$$

GSF-χ instead maps each unit rotation to a scalar logit contribution before combining units.

Let C^h_{ijs} be the scalar contribution of unit s after its confidence and pair gates. For any permutation π of the S units,

$$\frac{1}{\sqrt{S}}\sum_{s=1}^{S} C^h_{ijs} = \frac{1}{\sqrt{S}}\sum_{s=1}^{S} C^h_{ij,\pi(s)}. \tag{31}$$

This proves unit-order invariance. Notice that the construction sums scalar bilinear query–key modulations, rather than multiplying unit-specific $\mathrm{SO}(3)$ matrices; therefore no arbitrary order is imposed on generally non-commuting rotations. The symmetric $S^{-1/2}$ factor controls scale and does not change this result.

B.7 Interaction-Support Hierarchy

The only difference among the three scope interventions is the binary mask

$$m_{ijs}^{\text{query}} = a_{is}, \qquad m_{ijs}^{\text{incident}} = \max(a_{is}, a_{js}), \quad m_{ijs}^{\text{global}} = 1,$$
$$\operatorname{supp}(m^{\text{query}}) \subseteq \operatorname{supp}(m^{\text{incident}}) \subseteq \operatorname{supp}(m^{\text{global}}). \tag{32}$$

If $a_{is} = 1$, then $\max(a_{is}, a_{js}) = 1$ for every j; hence every query-anchor pair is also an anchor-incident pair. Every binary pair is, trivially, contained in the all-one global mask. This proves Eq. 32. The containments are strict whenever the molecule contains both anchor and non-anchor queries and contains a non-anchor–non-anchor pair; in degenerate cases they remain valid but may coincide.

The result is architectural rather than statistical. Under query-anchor scope, $a_{is} = 0$ implies $C_{ijs}^h = 0$ for all learned weights. Under anchor-incident scope, $a_{is} = a_{js} = 0$ implies the same. In the global model these entries are not hard-coded to zero; whether they are used is learned by the continuous node and pair gates. The scope intervention thus removes a well-defined set of paths without changing any parameter.

B.8 Proper Rigid Motions, Relabeling, and Achiral Reduction

Let $\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{Q}^\top + \mathbf{1}\boldsymbol{t}^\top$ for $\boldsymbol{Q} \in \mathrm{SO}(3)$. Translation cancels in all coordinate differences. For any difference vectors v, w,

$$\|\boldsymbol{v}\boldsymbol{Q}^\top\|_2 = \|\boldsymbol{v}\|_2, \qquad (\boldsymbol{v}\boldsymbol{Q}^\top)(\boldsymbol{w}\boldsymbol{Q}^\top)^\top = \boldsymbol{v}\boldsymbol{w}^\top, \qquad \det(\boldsymbol{Q}) = 1. \tag{33}$$

Distances, dot products, unsigned torsional magnitudes, and the canonical proper-orientation sign used by the preprocessor are therefore unchanged. The tensors supplied to the network are identical, which proves rigid-motion invariance.

For an atom permutation matrix $\boldsymbol{\Pi}$, node tensors map to $\boldsymbol{\Pi}\boldsymbol{H}$, pair tensors to $\boldsymbol{\Pi}\boldsymbol{P}\boldsymbol{\Pi}^\top$, and atom–unit tensors to $\boldsymbol{\Pi}\boldsymbol{R}$. Shared pointwise maps commute with $\boldsymbol{\Pi}$, pair attention contracts the relabeled key index, and softmax commutes with a common permutation of its entries. Induction over layers gives

$$\boldsymbol{Z}(\boldsymbol{\Pi}\boldsymbol{H}, \boldsymbol{\Pi}\boldsymbol{X}, \boldsymbol{\Pi}\mathcal{S}) = \boldsymbol{\Pi}\boldsymbol{Z}(\boldsymbol{H}, \boldsymbol{X}, \mathcal{S}). \tag{34}$$

Masked sum or mean pooling removes the remaining atom axis, yielding a permutation-invariant molecular representation.

Finally, every chiral correction contains ρ_s and the validity mask. If there is no valid unit, every C_{ijs}^h is exactly zero and each layer is the ordinary graph-attention layer with the same weights. This proves the achiral reduction in Eq. 7.

B.9 Reflection-Group Projection and Exact ECD Output Parity

Let $\mathcal{T}$ act on an encoder response by $(\mathcal{T}F)(\chi) = F(-\chi)$. Reflection is an involution, so $\mathcal{T}^2 = \mathcal{I}$. The two group-averaging operators

$$\mathcal{P}_+ = \frac{\mathcal{I} + \mathcal{T}}{2}, \qquad \mathcal{P}_- = \frac{\mathcal{I} - \mathcal{T}}{2} \tag{35}$$

are complementary idempotent projectors:

$$\mathcal{P}_\pm^2 = \mathcal{P}_\pm, \qquad \mathcal{P}_+\mathcal{P}_- = 0, \qquad \mathcal{T}\mathcal{P}_\pm = \pm\mathcal{P}_\pm, \qquad \mathcal{P}_+ + \mathcal{P}_- = \mathcal{I}. \tag{36}$$

Consequently every paired response has the unique decomposition $F = \mathcal{P}_+F + \mathcal{P}_-F$ into the trivial (even) and sign (odd) representations of C_2. Evaluating these projectors pointwise gives

$$\boldsymbol{z}^+(\chi) = \frac{F(\chi) + F(-\chi)}{2}, \qquad \boldsymbol{z}^-(\chi) = \frac{F(\chi) - F(-\chi)}{2}. \tag{37}$$

Direct substitution gives

$$\boldsymbol{z}^+(-\chi) = \frac{F(-\chi) + F(\chi)}{2} = \boldsymbol{z}^+(\chi), \tag{38}$$

$$\boldsymbol{z}^-(-\chi) = \frac{F(-\chi) - F(\chi)}{2} = -\boldsymbol{z}^-(\chi). \tag{39}$$

This argument does not assume that F is linear or that its intermediate channels carry a fixed parity.

Number and Position are functions only of $\boldsymbol{z}^+$ and are therefore exactly even. For a Symbol slot, the score used by the parity head has the form

$$s_k(\boldsymbol{z}^+, \boldsymbol{z}^-) = \frac{\langle \boldsymbol{u}_k(\boldsymbol{z}^+), \boldsymbol{W}^- \operatorname{RMS}(\boldsymbol{z}^-)\rangle}{\sqrt{r}} + \langle \boldsymbol{w}_k^-, \operatorname{RMS}(\boldsymbol{z}^-)\rangle, \tag{40}$$

where $\boldsymbol{W}^-$ and $\boldsymbol{w}_k^-$ are bias free. The implemented parameter-free normalization is

$$\operatorname{RMS}(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\sqrt{d^{-1}\|\boldsymbol{x}\|_2^2 + \epsilon}}, \tag{41}$$

so $\operatorname{RMS}(-\boldsymbol{z}^-) = -\operatorname{RMS}(\boldsymbol{z}^-)$. The coefficient $\boldsymbol{u}_k(\boldsymbol{z}^+)$ is unchanged by reflection. Hence $s_k(\boldsymbol{z}^+, -\boldsymbol{z}^-) = -s_k(\boldsymbol{z}^+, \boldsymbol{z}^-)$, and the binary logits $[-s_k, s_k]$ exchange under reflection. Shared stochastic masks make these identities sample-wise during training as well as evaluation. This completes the proof of Eq. 9.

B.10 Combined Structural Statement

Theorem (structural guarantees of GSF-χ). Under the assumptions above, GSF-χ is equivariant to atom relabeling and invariant to unit enumeration and phase-origin shifts. Its molecular readout is invariant to proper rigid motions. Handedness inversion and atom-pair reversal invert every complete latent rotary block; with no valid stereogenic unit, the model reduces exactly to its reflection-even graph-attention backbone. Finally, the C_2-projected ECD readout is exactly mirror even for Number and Position and exchanges the two Symbol logits under reflection.

Each clause was proved independently above. Together they separate the minimal parity-sensitive input required by the no-go proposition, the symmetry-compatible mechanism that distributes it over all atom pairs, and the task-level projector that enforces the requested output law.

Scope of the guarantees. These proofs constrain the operator, data flow, and declared task outputs. They do not imply that an arbitrary hidden channel has physical meaning, that the learned field is a continuous electromagnetic field, or that algebraic parity alone guarantees predictive accuracy.

C Additional Experimental Details

C.1 Dataset Audit

We audit splits before constructing any model cache. Enantiomer partners must share a partition, base identifiers must have zero overlap across partitions, and paired parity-even tensors must agree after canonical alignment. The central public-ECD reconstruction additionally requires complementary R/S labels and complementary peak signs. Table 5 records the audited split sizes used by our code.

Table 5: Pair-preserving split audit.

Dataset	Train	Validation	Test
ChIRo R/S conformers	326,865	70,099	69,719
ChIRo R/S base pairs	27,542	5,874	5,840
ChIRo Ranking pairs	24,192	5,184	5,184
public CMCDS molecules	16,528	2,066	2,068
ACMP molecules	952	120	120

For central ECD, the public archive contains fewer usable spectra than stated in the published processed benchmark. We do not impute missing examples or silently mix splits. All central-ECD tables therefore state the exact public molecule counts they use.

C.2 Evaluation Metrics

Each ECD spectrum is discretized into a sequence of peaks, each carrying a position and a sign (Cotton effect). Prediction quality is then scored along three axes, all computed with the released evaluator on identical discretizations:

- **Number-RMSE** (Num. in the tables) scores count fidelity: $\sqrt{\frac{1}{|\mathcal{D}|}\sum_{i\in\mathcal{D}}(\widehat{N}_i - N_i)^2}$, the root-mean-square discrepancy between the reference peak count N_i and the predicted count $\widehat{N}_i$, i.e., whether the model invents or misses intensity-carrying transitions.
- **Position-RMSE** (Pos. in the tables) scores localization: peaks are sorted by position and paired in order over the first $\min(N_i, \widehat{N}_i)$ slots, and the score is the root-mean-square displacement of the paired peaks, deliberately decoupled from count errors.
- **Symbol-Acc** (Sym. in the tables) scores sign fidelity: the fraction of real peaks whose predicted sign agrees with the reference at the paired slots; molecules with no real peaks contribute the evaluator's fixed correct slots.

Number and Position are mirror-even targets shared within an enantiomer pair, whereas Symbol is parity sensitive: enantiomers share peak positions with complementary signs. The three scores thus probe complementary failure modes—spurious or missing peaks, mislocated peaks, and flipped signs—and all come from one joint-spectrum checkpoint. **Rotation** accuracy is the remaining task output: the model classifies the binary sign of the optical rotation, which on the axial tasks is tied to the handedness of the stereogenic unit. Its target is mirror-odd, so chance level is 50% and reflection-even encoders cannot exceed chance (Section 6).

C.3 Controlled Central-ECD Retraining

To separate architecture from split construction, we retrain ECDFormer, ChiDeK, and GSF-χ from scratch under the same preprocessing, budget, and three-run setting. Table 7 reports both the pair-preserving split used for our main comparison and the historical sample-level split, where members of an enantiomer pair may cross partitions. All three outputs in a row come from the same joint-spectrum checkpoint. Table 6 summarizes the headline pair-preserving numbers, including both the like-for-like GSF-χ default and the best prespecified configuration quoted in the main-text Table 1 row.

The pair-preserving split has zero train–test pair overlap and is the primary analysis. GSF-χ leads all three outputs there. Under this contract every mean gap exceeds the sum of the two reported standard deviations, and two-sided Welch tests on the run statistics (n=3 per method) give $p < 0.01$ for Position and Symbol against both baselines and for Number against ECDFormer; the remaining comparison, Number against ChiDeK, gives t=4.1 (p=0.034). The sample-level split changes the ordering of the even metrics, illustrating why numbers from different split contracts should not be pooled into one comparison.

D Extended Ablation Studies

D.1 Full Central-Chirality Mechanism Results

Figure 3 summarizes the main mechanism trends, with values printed on the bars. Each variant changes exactly one component of the complete model; architecture, training budget, the three-run set, and the checkpoint rule are held fixed, so the remaining differences are attributable to the ablated component alone. The Number, Position, and Symbol columns are produced by one joint-ECD checkpoint shared by all central rows, not by separately tuned task-specific models.

D.2 Axial Ablation Setup

Table 8 ablates our best configuration in Table 1. All rows use the same ACMP split, $d = 48$, $L = 2$, $H_\chi/H = 2/4$, optimizer, number of epochs, three retained runs, and within-run checkpoint rule. No ablation variant receives its own run, checkpoint rule, or decoder. Disabled modules remain instantiated and have their output replaced by the stated fixed quantity; parameter count therefore cannot explain the differences.

Restricting support to stereogenic-anchor queries lowers Rotation from 77.78% to 69.17%, a decrease of 8.61%. The representation controls are more consistent on Symbol; anchor-local attention and a pooled chirality token lower it by 7.85% and 10.61%, respectively. Together with the separately disclosed same-checkpoint support intervention in Table 13, these results show that no single restricted

Table 6: Central-ECD retraining on the same 20,662 public CMCDS molecules and pair-preserving split. All three metrics come from one joint-spectrum checkpoint. The final row is our best configuration.

Method	Position ↓	Number ↓	Symbol ↑
ECDFormer	$2.718_{\pm 0.065}$	$1.316_{\pm 0.023}$	$50.420_{\pm 0.230}$
ChiDeK	$2.623_{\pm 0.068}$	$1.235_{\pm 0.006}$	$50.264_{\pm 0.208}$
GSF-χ (ours)	$\mathbf{2.280}_{\pm 0.022}$	$\mathbf{1.197}_{\pm 0.015}$	$\mathbf{53.102}_{\pm 0.086}$
GSF-χ (best)	$2.186_{\pm 0.015}$	$1.174_{\pm 0.004}$	$53.757_{\pm 0.315}$

Table 7: Central-ECD split matrix under controlled retraining. Bold marks the best mean within each split and metric.

	Pair-preserving 80/10/10			Sample-level 90/5/5		
Model	Position ↓	Number ↓	Symbol ↑	Position ↓	Number ↓	Symbol ↑
ECDFormer	2.718 ± 0.065	1.316 ± 0.023	50.420 ± 0.230	$\mathbf{1.022 \pm 0.027}$	$\mathbf{0.848 \pm 0.029}$	41.669 ± 1.756
ChiDeK	2.623 ± 0.068	1.235 ± 0.006	50.264 ± 0.208	2.527 ± 0.183	1.318 ± 0.131	50.233 ± 0.770
GSF-χ	$\mathbf{2.280 \pm 0.022}$	$\mathbf{1.197 \pm 0.015}$	$\mathbf{53.102 \pm 0.086}$	2.058 ± 0.018	1.153 ± 0.003	$\mathbf{58.314 \pm 0.842}$

Table 8: Axial mechanism ablation around the exact GSF-χ headline configuration in Table 1. Rotation and Symbol are parity-sensitive; Position and Number are parity-even diagnostics. Every row uses the same three runs, training budget, and evaluation procedure as the complete model.

Variant	Rotation ↑	Position ↓	Number ↓	Symbol ↑
GSF-χ (full)	$\mathbf{77.8 \pm 1.7}$	2.77 ± 0.09	1.01 ± 0.03	74.5 ± 1.3
Anchor queries only	69.2 ± 3.0	2.80 ± 0.06	0.99 ± 0.03	67.3 ± 2.6
w/o unit-conditioned axes	75.3 ± 1.3	2.78 ± 0.05	0.99 ± 0.02	73.9 ± 1.3
+ unit-conditioned frequencies	75.6 ± 1.7	2.77 ± 0.09	1.00 ± 0.02	$\mathbf{74.9 \pm 2.1}$
w/o learned pair gate	76.9 ± 4.1	2.83 ± 0.04	1.01 ± 0.05	66.5 ± 5.8
w/o geometric phase initialization	77.5 ± 0.8	2.81 ± 0.11	0.99 ± 0.02	74.5 ± 3.8
Anchor-local chirality	71.1 ± 4.9	2.81 ± 0.07	1.01 ± 0.03	66.7 ± 3.5
Pooled chirality token	71.7 ± 1.7	2.80 ± 0.04	1.01 ± 0.02	63.9 ± 2.6
Reflection-even backbone	49.7 ± 0.5	$\mathbf{2.69 \pm 0.07}$	1.00 ± 0.02	52.2 ± 0.0

topology dominates the complete model across the parity-sensitive outputs. All nine rows instantiate 154,440 parameters.

The internal controls are likewise metric dependent. Removing unit-conditioned axes lowers Rotation by 2.50% and Symbol by 0.63%; adding unit-conditioned frequencies raises Symbol by 0.43% but lowers Rotation by 2.22%. Removing the learned pair gate lowers Symbol by 8.07%, whereas removing geometric phase initialization leaves the rounded Symbol mean unchanged. Position and Number change much less and do not consistently favor the full model, as expected for parity-even targets. The reflection-even control remains at chance on Rotation; its 52.23% official Symbol score is exactly the 50% real-peak baseline plus the released evaluator's fixed correct slots for zero-peak molecules.

E Robustness and Fine-Grained Evaluation

Each subsection below pairs an experimental procedure with its results: annotation corruption (Appendix E.1), subtype and multi-unit breakdowns (Appendix E.2), distance-resolved field reach (Appendix E.3), head allocation and capacity (Appendix E.4), and efficiency and scope (Appendix E.5). Further controls (paired bootstrap, automatic unit detection, RotA conformer robustness, and low-data mirror generalization) follow as separate subsections with their own tables.

Table 9: Per-run values for the headline-base axial ablation. Rotation and ECD checkpoints use their corresponding within-run selection rules. The last row uses sample standard deviation.

Run	Rotation epoch	ECD epoch	Rotation ↑	Position ↓	Number ↓	Symbol ↑
A	20	25	78.33	2.730	0.975	74.52
B	24	25	75.83	2.703	1.041	75.80
C	24	25	79.17	2.876	1.017	73.25
Mean ± std	–	–	77.78 ± 1.73	2.770 ± 0.093	1.011 ± 0.033	74.52 ± 1.27

Table 10: Factorial separation of interaction support and ECD output projection on ACMP. All configurations use the same three prespecified runs and checkpoint rule; the test split is evaluated only after restoring the selected checkpoint. Accuracies are percentages.

Interaction	Readout	Parameters	Position ↓	Number ↓	Symbol ↑	Pair complement ↑
Global GSF-χ	C_2-projected	154,440	2.844 ± 0.054	1.014 ± 0.021	67.73 ± 1.60	100.00 ± 0.00
Anchor-local	C_2-projected	154,440	2.765 ± 0.070	0.994 ± 0.035	65.82 ± 3.51	100.00 ± 0.00
Global GSF-χ	Unconstrained	134,714	2.785 ± 0.036	0.997 ± 0.038	72.51 ± 0.49	92.22 ± 5.00
Anchor-local	Unconstrained	134,714	2.810 ± 0.018	1.002 ± 0.035	65.07 ± 2.17	82.00 ± 2.00

E.1 Annotation Corruption

For corruption rate r, we draw a fixed pair-level mask from the evaluation seed. Half of the selected units are removed ($\rho_s = 0$) and half have their sign flipped. The same corruption is applied to both members of a pair so that the experiment measures annotation quality rather than accidental pair leakage. Rates are 2.5, 5, 7.5, and 10%, matching the range used by ChiDeK. We report both task performance and exact-parity residuals; the latter should remain at numerical precision even as accuracy degrades, because the readout law is independent of annotation correctness.

At 10% pair-level annotation corruption, official Symbol decreases by 0.64% while Number and Position remain exactly unchanged. The curve is not monotone at every intermediate rate; we therefore interpret the endpoint and per-seed values rather than fitting a trend. All mirror residuals remain exactly zero, confirming that an incorrect annotation can hurt accuracy without breaking the algebraic output law.

E.2 Axial Subtypes and Number of Units

ACMP contains chiral-atom pairs, allene-like structures, spiral atoms/chains, biaryls, C–N, C–B, and rare C–C heterobiaryls, and nonbiaryls. We report sample count, real-peak count, and all four task metrics for every subtype. A second stratification groups molecules by one, two, or three valid stereogenic units. This is particularly diagnostic for GSF-χ because, unlike a single-center query, its unit contributions are defined as an order-independent sum.

Subtype and scaling rows are descriptive. The dominant biaryl group contains 30 test pairs and reaches 81.44% real-peak Symbol accuracy; all other subtypes contain at most nine pairs. In particular, the perfect multi-unit row contains only two pairs. Molecular-size bins are likewise confounded with subtype.

E.3 Distance-Resolved Field Reach

For every atom pair, we record the minimum graph distance of either atom from a stereogenic anchor and aggregate $|\Delta\ell_{ij}|$ by distance. We then mask the correction to query-anchor, one-hop, two-hop, or global pairs without changing parameters. This produces a fixed-model support-sensitivity curve: performance changes are tied directly to which atom-pair corrections remain active.

The post-training intervention probes support without retraining; keeping only anchor-query corrections lowers Symbol by 5.31%, and keeping all anchor-incident corrections lowers it by 5.73%. Aggregate Number and Position metrics are unchanged across the three masks. Together with the comparable correction magnitudes at graph distances 0 through 4+, this shows that the trained checkpoint relies on remote–remote corrections.

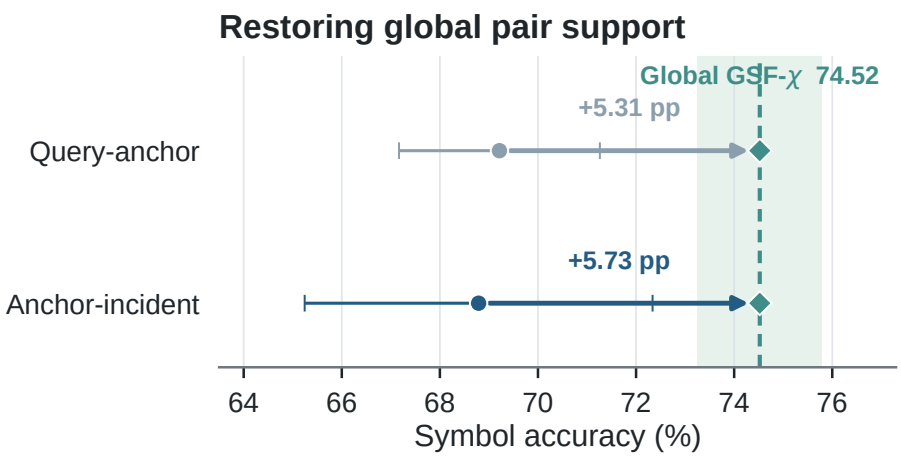


Figure 7: Same-checkpoint support intervention. Arrows connect each restricted-mask mean to the all-pair GSF-χ mean. Horizontal whiskers and the green band show standard deviations over three runs.

E.4 Head Allocation and Capacity

Table 11 keeps the same runs and checkpoint rule while changing head allocation or model size. Making every head chiral reduces Symbol by 8.28%, supporting a mixed even/odd attention pathway. The 64-times larger $d = 256, L = 8$ model slightly improves Position but worsens Number and loses 5.31% Symbol accuracy. The axial gain therefore cannot be attributed to parameter count alone.

E.5 Efficiency and Scope of the Evaluation

The global field has a measurable computational cost; at batch size 32 it processes 6,370 molecules/s and peaks at 754 MiB, compared with 15,033 molecules/s and 173 MiB for the reflection-even control, a $2.36\times$ latency and $4.36\times$ memory overhead of the direct $|\mathcal{S}|N^2$ implementation, while staying below 1 GiB at 154,440 parameters. Every row in Table 17 uses the C_2-projected ECD readout, so these measurements include the concatenated $F(\chi)/F(-\chi)$ encoder batch; the comparison therefore isolates field topology while measuring end-to-end projected inference. On achiral molecules the operator reduces exactly to that backbone (Eq. 7), so the cost is paid only where stereogenic units exist. Beyond cost, the empirical scope has two boundaries: central ECD uses the reproducible public subset rather than ChiDeK's unreleased processed set, and the implementation assumes a supplied stereogenic unit set, though handedness within a unit is computed from the conformer (Appendix B.2) and Section 6.3 measures the cost of removing that remaining input. These boundaries affect empirical scope and uncertainty, not the exact parity identities.

E.6 End-to-End Cost of the C_2 Projection

The projected implementation concatenates χ and $-\chi$ into one accelerator batch. We benchmark the complete ECD forward, including Number, Position, and Symbol heads, against the one-sign unconstrained model. Timing uses 20 warm-up iterations followed by seven synchronized repetitions.

At batch size one, projection changes latency by $1.18\times$; at batch size 32, its peak allocation is $1.93\times$ the one-sign readout and throughput decreases from 8264 to 6370 molecules/s. These measurements include the concatenated two-sign encoder evaluation and therefore quantify the deployment cost omitted by an encoder-only parameter count.

E.7 Pair-Level Uncertainty for Parity-Odd Targets

The dispersion reported elsewhere in this paper is taken across training runs and describes optimization variability at a fixed test set. We additionally quantify uncertainty with respect to sampling of the test set itself, by jointly resampling the same 60 complete ACMP enantiomer pairs for GSF-χ and

Table 11: Head allocation and capacity around our best configuration. Every row uses the same runs and within-run checkpoint rule.

Configuration	Parameters	Position ↓	Number ↓	Symbol ↑
$d = 48, L = 2$, 2/4 chiral heads	154,440	2.770 ± 0.093	$\mathbf{1.011 \pm 0.033}$	$\mathbf{74.52 \pm 1.27}$
$d = 48, L = 2$, 4/4 chiral heads	157,644	2.820 ± 0.012	1.035 ± 0.017	66.24 ± 2.78
$d = 256, L = 8$, 4/8 chiral heads	9,898,102	$\mathbf{2.734 \pm 0.047}$	1.037 ± 0.047	69.21 ± 0.37

Table 12: Annotation-corruption analysis on ACMP using our best checkpoint set. Mean ± sample standard deviation over three runs.

Corruption	Position ↓	Number ↓	Symbol ↑	Real-peak Symbol ↑
0.0%	2.770 ± 0.093	1.011 ± 0.033	74.52 ± 1.27	73.33 ± 1.33
2.5%	2.770 ± 0.093	1.011 ± 0.033	74.95 ± 0.74	73.78 ± 0.77
5.0%	2.770 ± 0.093	1.011 ± 0.033	75.58 ± 0.74	74.44 ± 0.77
7.5%	2.770 ± 0.093	1.011 ± 0.033	73.89 ± 1.27	72.67 ± 1.33
10.0%	2.770 ± 0.093	1.011 ± 0.033	73.89 ± 1.69	72.67 ± 1.76

Table 13: Post-training support-mask intervention over our best checkpoint set. The full model is trained globally; only the correction mask is changed at evaluation.

Retained correction support	Position ↓	Number ↓	Symbol ↑	Real-peak Symbol ↑
anchor queries only	2.773 ± 0.098	1.011 ± 0.033	69.21 ± 2.05	67.78 ± 2.14
anchor-incident pairs	2.773 ± 0.098	1.011 ± 0.033	68.79 ± 3.55	67.33 ± 3.71
all atom pairs	2.770 ± 0.093	1.011 ± 0.033	74.52 ± 1.27	73.33 ± 1.33

ChiDeK over 10,000 bootstrap replicates. Resampling whole enantiomer pairs rather than individual molecules keeps each pair's two configurations together, and the complete reported run sets are aggregated before each paired difference is computed. Unlike a comparison between marginal confidence intervals, this procedure removes the pair-specific difficulty shared by the two methods, which is why we report the paired difference itself rather than two separate intervals.

Both intervals exclude zero. For Rotation, a two-sided paired bootstrap test gives $p = 0.0154$ and a pair-cluster sign-flip randomization test gives $p = 0.0251$. This comparison therefore statistically resolves the advantage of GSF-χ on both parity-odd outputs, molecular Rotation and ECD Symbol, which are exactly the two targets whose value reverses under handedness inversion. The parity-even targets are unaffected, as expected. This complements the exact architectural parity guarantees of Section 4.3 with finite-test-set empirical evidence.

E.8 Automatic Stereogenic-Unit Detection

Handedness within a unit is already computed from the conformer rather than read from an annotation (Appendix B.2), so the remaining external input is the stereogenic unit set itself. We therefore evaluate whether GSF-χ can operate with units extracted automatically from molecular structure. We apply the ChiralFinder/RDKit extraction pipeline (Shi et al., 2026a; Landrum et al., 2024), an external tool developed independently of this work, and match each detected unit against the curated ACMP unit by its complete atom-index set. Optical-rotation and ECD labels are never used during extraction.

Only the atom set of each unit is taken from the extractor. Handedness is recomputed from the conformer under our own canonical gauge exactly as in the curated setting, so the extractor's chirality descriptor enters neither the model input nor the evaluation. This keeps the substitution a test of unit localization alone, and prevents the detector from supplying the very pseudoscalar whose effect is being measured.

The extractor therefore supplies a deliberately noisy, high-recall annotation setting, recovering 92.66% of curated units while proposing roughly twice as many candidates in total. We then evaluate the curated-trained checkpoints directly under these automatically extracted units, with no retraining, no adaptation, and no change to the operator.

Table 14: ACMP subtype analysis for our best global model over three runs. Groups with very few pairs are descriptive only.

Subtype	Pairs	Rotation ↑	Position ↓	Number ↓	Real-peak Symbol ↑
Allene-like structure	2	75.00 ± 25.00	4.151 ± 0.225	1.886 ± 0.408	66.67 ± 28.87
Biaryl structure	30	72.78 ± 1.93	2.215 ± 0.176	0.850 ± 0.022	81.44 ± 2.64
Chiral atom pair	5	80.00 ± 10.00	2.942 ± 0.053	0.632 ± 0.000	69.05 ± 4.12
Heterobiaryl (C-B)	4	87.50 ± 12.50	4.088 ± 0.260	0.500 ± 0.000	57.14 ± 14.29
Heterobiaryl (C-N)	6	94.44 ± 9.62	4.412 ± 0.096	1.225 ± 0.000	76.19 ± 21.82
Nonbiaryl	9	70.37 ± 6.42	2.859 ± 0.075	1.217 ± 0.026	56.94 ± 9.62
Spiral atom and chain	4	95.83 ± 7.22	2.040 ± 0.164	1.323 ± 0.000	69.23 ± 7.69

Table 15: Scaling analysis for our best model on ACMP. Multi-unit results contain only two test pairs.

Stratification	Group	Pairs	Rotation ↑	Position ↓	Number ↓	Real-peak Symbol ↑
unit count	multiple units	2	100.00 ± 0.00	2.005 ± 0.192	0.000 ± 0.000	100.00 ± 0.00
unit count	one unit	58	77.01 ± 1.80	2.796 ± 0.102	1.028 ± 0.034	72.60 ± 1.37
atom count	15–25 atoms	16	76.04 ± 1.80	2.593 ± 0.088	0.989 ± 0.036	68.42 ± 2.63
atom count	26–39 atoms	29	72.99 ± 1.99	3.065 ± 0.055	1.124 ± 0.050	71.23 ± 1.37
atom count	40+ atoms	15	88.89 ± 3.85	2.388 ± 0.224	0.775 ± 0.000	82.05 ± 4.44

Table 16: Magnitude of the learned correction by minimum graph distance from a stereogenic anchor for one representative run of our best configuration.

Distance	Atom-pair blocks	Mean $\lvert\Delta\ell_{ij}\rvert$
0	16048	0.05101
1	29860	0.04588
2	42248	0.05403
3	33832	0.05794
4+	47528	0.05597

Under this substitution Rotation is $68.61 \pm 3.76\%$ and Symbol is $62.00 \pm 3.51\%$, drops of 5.56 and 9.34 points from the same checkpoints evaluated with curated units; the parity-even targets are essentially unchanged at 2.823 ± 0.065 Position and 1.016 ± 0.049 Number, as expected for quantities the signed field does not orient. Two observations follow. First, both parity-odd outputs stay far above the 50% floor, so the operator degrades gracefully rather than collapsing when its unit set is wrong roughly half the time at the molecule level. Second, the Rotation accuracy obtained with fully automatic extraction still exceeds the 65.17% that ChiDeK attains with curated annotations.

We do not claim an annotation-free pipeline. Detection precision of 42.57% is the limiting factor, and axial-unit perception is an open problem in its own right, orthogonal to the operator studied here. What these results establish is that the parity-sensitive signal GSF-χ extracts is a property of the geometry and the interaction topology rather than of curated labels, and that improvements in upstream detection translate directly into end-to-end gains without any change to the model.

E.9 Conformer Robustness on RotA

Evaluation setting. RotA contains 3,140 conformers of 650 axially chiral molecules (Shi et al., 2026a). Because ACMP is derived from this collection, RotA is *not* an independent test corpus; we use it instead to test whether predictions remain stable across alternative conformations of the same held-out molecular identities. We freeze the ACMP-trained checkpoints of each method (five per baseline, three for GSF-χ) and evaluate every RotA conformer associated with the ACMP test set, without retraining, checkpoint selection, or parameter fitting on RotA. Each conformer is paired with a physically reflected copy, while its Rotation and ECD targets are inherited from the corresponding ACMP enantiomer pair. This gives 283 complete physical mirror pairs (566 samples) drawn from 59 held-out molecules, which remain the independent sampling unit. One additional test molecule is excluded because its singleton stereogenic-unit annotation does not define the two-endpoint geometric gauge needed for a physical axial reflection audit.

Table 17: Inference efficiency on one NVIDIA H100 80GB HBM3 with batch size 32. Latency is the median of five 50-batch measurements after ten warm-up batches. All controls retain the inactive modules, so parameter counts are exactly matched. The last row is the ordinary graph transformer this work extends, obtained by disabling the signed field, and is therefore the reference for the cost of the mechanism.

Mechanism	Parameters	Molecules/s ↑	Peak MiB ↓
global GSF-χ	154,440	6370.3	754.1
anchor-local χ	154,440	14744.2	172.8
pooled χ token	154,440	13967.6	172.8
reflection-even backbone	154,440	15032.5	172.8

Table 18: End-to-end inference cost on one NVIDIA H100 80GB. Mean $\pm$ sample standard deviation over three checkpoints.

Readout	Parameters	Batch-1 latency (ms)	Batch-1 peak MiB	Batch-32 molecules/s	Batch-32 peak MiB
one-sign unconstrained	134,714	3.432 ± 0.015	65.7 ± 0.0	8264.4 ± 18.1	391.6 ± 0.0
C_2-projected	154,440	4.056 ± 0.006	66.3 ± 0.0	6370.3 ± 6.6	754.1 ± 0.0

Table 19: Paired bootstrap on the two parity-odd ACMP targets against retrained ChiDeK. Differences are percentage points; intervals are 95% pair-bootstrap percentile intervals obtained by resampling the same complete enantiomer pairs for both methods. Positive values favor GSF-χ.

Target	Paired Δ	95% CI
Rotation	$+\mathbf{9.67}$	$[+1.83, +17.83]$
Symbol	$+\mathbf{7.96}$	$[+0.60, +15.24]$

Table 2 in the main text reports the accuracies. GSF-χ degrades gracefully under the conformational shift, reaching 73.1% Rotation and 72.8% Symbol on RotA conformers, against 77.8% and 74.5% on ACMP (Table 1). Its margin over the matched ChiDeK retraining meanwhile widens, from $+12.6$ and $+7.9$ points on ACMP to $+15.4$ and $+10.7$ points here, and Number RMSE improves by 0.07. GSF-χ also holds the lowest Position RMSE (2.78 vs. 2.80), a parity-even target where no signed advantage is expected. That the advantage grows rather than shrinks on unseen conformations is what one expects if the signal is carried by geometry rather than by conformer-specific memorization.

ECDFormer produces identical outputs for a molecule and its mirror image, which forces its Rotation accuracy to exactly chance and pins its Symbol score at the evaluator floor with zero variance across runs; the reflection diagnostics below confirm this directly.

Table 2 in the main text isolates the parity projection rather than the encoder. Both GSF-χ rows come from one shared representation and one checkpoint; the two heads differ only in whether the C_2 projection of Section 4.4 is applied. The projected Symbol head complements on every one of the 283 mirror pairs, with zero variance across all three GSF-χ runs, whereas the unprojected Rotation head flips on 78.8% of them, statistically indistinguishable from ChiDeK's 78.7%. A head trained without the projection therefore behaves like a conventional classifier even when it reads a parity-correct representation, which is direct evidence that exact mirror consistency comes from the projection and cannot be obtained by learning alone. Rotation flip should accordingly be read as a diagnostic, not as a guarantee the architecture makes.

This law is realized through the complete coordinate-reflection pipeline on every evaluated conformer, not only under an internal sign-flip intervention; across all 283 pairs, the parity-even GSF inputs match exactly and every supplied pseudoscalar changes sign after reflection.

Implementation audit. The released ChiralFinder annotations are recovered for 88.05% of the curated axes in this conformer set. Native ChiralFinder preprocessing completes for 556 of 566 samples; for the remaining ten samples (1.77%) we use the same deterministic coordinate-based fallback as in the ACMP evaluation adapter (Appendix E.8). All detected, nondegenerate mirror pairs reverse their determinant sign.

Table 20: Automatic stereogenic-unit detection on all 1,192 ACMP molecular configurations. Precision, recall, and F1 are micro-averaged over exact atom-index-set matches against the 1,280 curated units.

Detection statistic	Value
True positive units	1,186
False positive units	1,600
False negative units	94
Precision	42.57%
Recall	92.66%
F1 score	58.34%
Exact molecule-level unit set	50.00%
No detected unit	4.70%
Mean detected units per molecule	2.34

E.10 Training-Time Support-Budget Controls

We retrain every support rather than masking a dense checkpoint. Random and graph-distance-matched supports retain exactly the query-anchor entry count for each molecule and stereogenic unit. The masks are label independent, parity even, and shared by the two members of every enantiomer pair. The hard-top-k variants select the same parity-even pair gate at three support budgets; they are included as exploratory negative controls.

At the identical 6.5% budget, moving support from anchor queries to random-global and distance-matched global pairs improves Rotation by 11.11% and 10.28%, respectively. Full support adds another 8.61–9.44%. Thus the Rotation gain is not explained by retaining more entries alone; the ability to place signed interactions away from the anchor matters. Symbol does not share this ordering (query-anchor is comparable to random-global and higher than distance-matched global), so this control supports a target-specific global-placement claim rather than universal dominance on every output. The hard top-k selector collapses near chance at all three budgets. We treat this as an optimization failure of discrete support selection, not evidence against dense global interaction, and do not use it for the mechanism claim.

E.11 Low-Data Mirror Generalization

The pair-fraction setting retains complete pair identities. In contrast, one-side supervision labels one configuration per training and validation pair and withholds its mirror from both optimization and checkpoint selection. The member is selected by a target-blind hash of pair identity. The original ACMP test pairs remain pair-disjoint from training in every setting.

Projection is most useful when mirror supervision is missing; it improves withheld-mirror Symbol accuracy by 13.20% and enforces exact pair complement. On pair-disjoint test molecules, its gain is largest at 25% paired supervision (8.49%), decreases at 50% (4.78%), and reverses with complete paired supervision, where the unconstrained head has enough labels to learn the relation approximately. This is the expected inductive-bias trade-off rather than a universal raw-accuracy gain.

F Comparison and Analysis Procedures

Table 1 reports our best benchmark settings so that each task has one complete headline comparison; the ECD columns of ECDFormer and ChiDeK in Table 1 are matched retrainings, described in Section 5.2. The matched comparisons in Appendices C.3 and F.1 use prespecified run sets and hyperparameters, retraining the baselines on identical molecules, splits, budgets, and evaluators.

F.1 Controlled Axial Retraining

We additionally retrain ECDFormer and ChiDeK on the same 1,192 ACMP molecules, pair-preserving split, evaluator, and 25-epoch budget. Table 24 reports the retrained baselines alongside our best GSF-χ configuration; per-run values are serialized in the released machine-readable manifests. Two-sided

Table 21: Training-time support controls on ACMP. Accuracies are percentages; every row uses the same three runs and checkpoint rule.

Support	Density	Rotation ↑	Position ↓	Number ↓	Symbol ↑
Query-anchor	6.5%	56.11 ± 1.73	2.793 ± 0.043	1.019 ± 0.031	70.06 ± 2.30
Random-global	6.5%	67.22 ± 1.92	2.841 ± 0.013	1.027 ± 0.012	69.64 ± 6.99
Distance-matched global	6.5%	66.39 ± 0.96	2.817 ± 0.010	1.011 ± 0.010	65.82 ± 3.21
Hard top-k global	6.5%	50.00 ± 0.00	2.785 ± 0.060	1.026 ± 0.063	52.87 ± 0.64
Hard top-k global	15.0%	51.11 ± 0.48	2.780 ± 0.052	1.013 ± 0.062	53.08 ± 0.74
Hard top-k global	30.0%	51.39 ± 1.27	2.814 ± 0.013	0.991 ± 0.044	52.87 ± 0.64
Full global	100.0%	75.83 ± 2.20	2.817 ± 0.052	1.024 ± 0.038	68.37 ± 1.95

Table 22: Pair-disjoint ACMP test performance under reduced ECD supervision. Symbol and complement are percentages.

Labels	Readout	Position ↓	Number ↓	Symbol ↑	Complement ↑
25% pairs	projected	2.830 ± 0.111	1.065 ± 0.000	67.30 ± 2.65	100.00 ± 0.00
25% pairs	unconstrained	2.799 ± 0.104	1.065 ± 0.000	58.81 ± 2.67	41.78 ± 20.79
50% pairs	projected	2.826 ± 0.116	1.065 ± 0.000	69.00 ± 2.05	100.00 ± 0.00
50% pairs	unconstrained	2.790 ± 0.046	1.065 ± 0.000	64.23 ± 2.43	73.11 ± 4.73
one side	projected	2.840 ± 0.105	1.065 ± 0.000	69.00 ± 2.24	100.00 ± 0.00
one side	unconstrained	2.835 ± 0.033	1.054 ± 0.015	67.62 ± 4.78	70.89 ± 7.34
100% pairs	projected	2.844 ± 0.054	1.014 ± 0.021	67.73 ± 1.60	100.00 ± 0.00
100% pairs	unconstrained	2.785 ± 0.036	0.997 ± 0.038	72.51 ± 0.49	92.22 ± 5.00

Table 23: Mirror completion after one-enantiomer supervision. Withheld training mirrors are never used for optimization or checkpoint selection. Symbol and complement are percentages.

Readout	Mirror Symbol ↑	Complement ↑
projected	77.65 ± 1.70	100.00 ± 0.00
unconstrained	64.45 ± 2.58	77.34 ± 7.52

Welch tests give p=4.9×10^{-4} for Rotation, p=5.8×10^{-4} for Symbol, and p=0.014 for Number RMSE; Position is tied at the reported precision.

F.2 Continuous Axial-Rotation Construction

The analysis in Figure 5 uses ACMP pairs 27 and 601 because they already serve as the biaryl and non-biaryl qualitative examples; trajectory separation was not a selection criterion. For each molecule, we remove the canonically oriented axial bond, choose the connected component on its second endpoint, and rotate that component about the bond by $0°, 20°, \ldots, 340°$. Each of the 18 conformers then passes through the ordinary unit extraction, role canonicalization, feature builder, and one locked axial-ECD encoder. Both series contain nine conformers of each configuration. We compute z^+ and z^- before dimensionality reduction; UMAP uses cosine distance, five neighbors, minimum distance 0.08, and a fixed random state. The polar curves are not projected; they are raw cosine similarities to the $0°$ representation. Their machine-readable manifest contains every angle, extracted sign, embedding, and similarity.

G Case-Study Selection and Additional Examples

Figure 8 makes the paired output law visible on ACMP pair 601. Both configurations have $N = 2$ and $P = (3, 11)$, with Symbols $(-, +)$ and $(+, -)$. ECDFormer repeats the same three-peak prediction across the pair; ChiDeK predicts two peaks with $(+, +)$ in the first row but only one peak in the second; GSF-χ preserves the paired law, predicts both Symbols correctly, and places the peaks at $(3, 10)$, one bin from the reference.

Table 24: ACMP retraining with a common split, budget, and checkpoint rule; baselines average five runs. The final row is our best configuration. Rotation and Symbol are percentages.

Method	Rotation ↑	Position ↓	Number ↓	Symbol ↑
ECDFormer	$50.00_{\pm 0.59}$	$3.696_{\pm 0.244}$	$1.139_{\pm 0.086}$	$52.77_{\pm 0.14}$
ChiDeK	$65.17_{\pm 1.60}$	$\mathbf{2.770}_{\pm 0.097}$	$1.118_{\pm 0.055}$	$66.62_{\pm 1.41}$
GSF-χ (best)	$\mathbf{77.78}_{\pm 1.73}$	$\mathbf{2.770}_{\pm 0.093}$	$\mathbf{1.011}_{\pm 0.033}$	$\mathbf{74.52}_{\pm 1.27}$

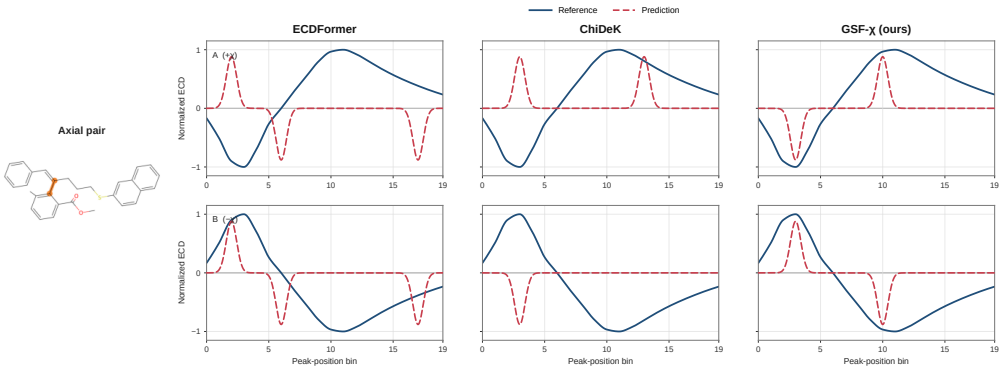


Figure 8: Axial-ECD prediction for an enantiomer pair. Rows are opposite configurations; solid and dashed curves are reference and prediction. GSF-χ recovers Number and both Symbols, with a one-bin displacement of the second peak; the baselines make visibly different count or sign errors.

The pair is chosen by a deterministic post-hoc rule. We first retain pairs for which the locked GSF-χ checkpoint predicts Number and Symbol exactly in both configurations, obeys the complete pair law, and misses no Position by more than one bin. We then maximize ChiDeK's combined Number and Symbol errors, breaking ties by total component error and base ID. This rule affects only the illustration, never model selection or aggregate metrics. This appendix also records the attribution procedure and adds multi-unit and failure cases.

Figure 9 opens the operator and shows that inversion preserves the parity-even field and gates while negating every signed latent angle, including for a two-unit molecule. This operator-level view is the visual counterpart of Section 4.3.

Every molecular panel is generated from serialized model outputs and an RDKit molecule; structures are not manually redrawn. Atom colors encode the centered phase, and green outlines identify stereogenic anchors. The compact case figure prioritizes direct paired Symbol predictions. The detailed diagnostic view in Figure 11 retains unit-specific phase curves and integrated-gradient attributions using the zero odd representation as baseline and 64 interpolation steps. It also contains the deterministically selected highest-loss failure, rather than hiding it in a success-only figure. The companion table summarizes even predictions and paired Symbol correctness, while the machine-readable manifest stores every $(\text{position}, \text{symbol})$ tuple. Table 14 complements these examples with exhaustive subtype-level statistics.

Pair 261 is the largest paired Symbol-NLL failure under the same locked checkpoint. GSF-χ, query-anchor, and no-chirality classify 2/6, 2/6, and 3/6 real-peak Symbols, respectively. We report the failure without assigning an unsupported chemical explanation that the experiment cannot identify; its broad field and attribution demonstrate that global reach alone does not guarantee correctness.

G.1 Additional Paired ECD Predictions

Figures 12–14 use the same plotting contract as the main paired example; each model column receives the identical reference sequence within a row, and dashed curves render only predicted peak number, position, and symbol at fixed amplitude. Pairs 27, 368, and 261 are the deterministic Symbol success, multi-unit example, and highest-loss failure under one representative visualization checkpoint.

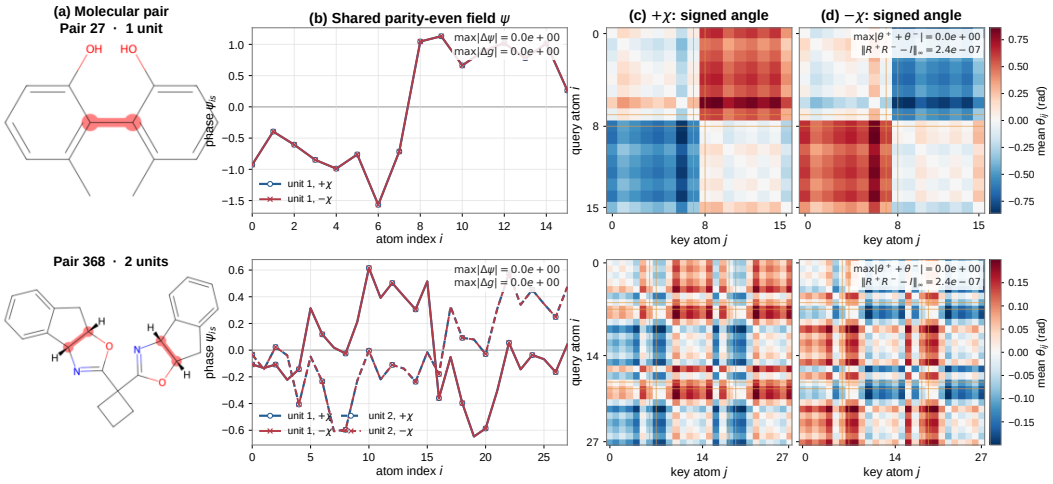


Figure 9: Configuration inversion inside GSF-χ. Rows show one- and two-unit molecules. **(a)** Molecular structure; **(b)** shared parity-even field; **(c,d)** signed-angle maps for opposite configurations. The residuals verify operator inversion to numerical precision; atom-pair indices are not physical rotation angles.

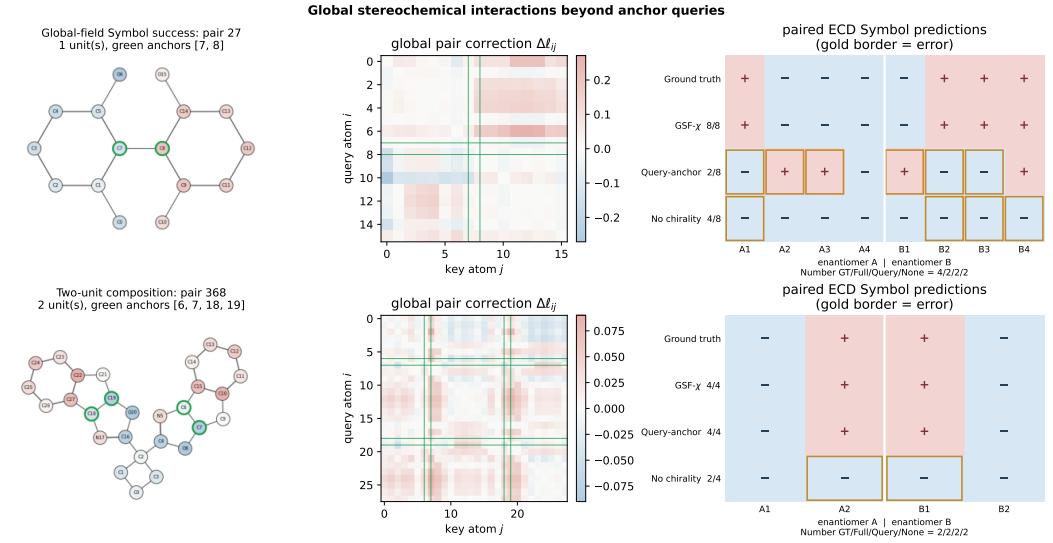


Figure 10: Axial-ECD Symbol cases. Rows show a control-separating Symbol success and a two-unit example. Columns show the phase-colored molecule, all-atom rotary correction, and paired Symbol predictions; gold borders mark errors.

Table 25: Predictions for the deterministic cases. Number and Position are exactly shared by each reflected pair; Symbol reports correctly classified real peaks across both members. P_{cmp} contains the first $\min(N, \widehat{N})$ positions used by the released evaluator; the manifest contains every individual signed tuple.

Pair / role	Ground truth	GSF-χ	Query-anchor	No chirality
27 / Symbol success	$N = 4,\ P = (3, 7, 11, 16)$	$N = 2,\ P_{\mathrm{cmp}} = (3, 11),\ 8/8$	$N = 2,\ P_{\mathrm{cmp}} = (3, 10),\ 2/8$	$N = 2,\ P_{\mathrm{cmp}} = (5, 11),\ 4/8$
368 / two units	$N = 2,\ P = (3, 9)$	$N = 2,\ P_{\mathrm{cmp}} = (3, 11),\ 4/4$	$N = 2,\ P_{\mathrm{cmp}} = (3, 11),\ 4/4$	$N = 2,\ P_{\mathrm{cmp}} = (5, 11),\ 2/4$
261 / highest loss	$N = 3,\ P = (1, 4, 12)$	$N = 3,\ P_{\mathrm{cmp}} = (3, 10, 13),\ 2/6$	$N = 3,\ P_{\mathrm{cmp}} = (3, 9, 13),\ 2/6$	$N = 3,\ P_{\mathrm{cmp}} = (3, 9, 13),\ 3/6$

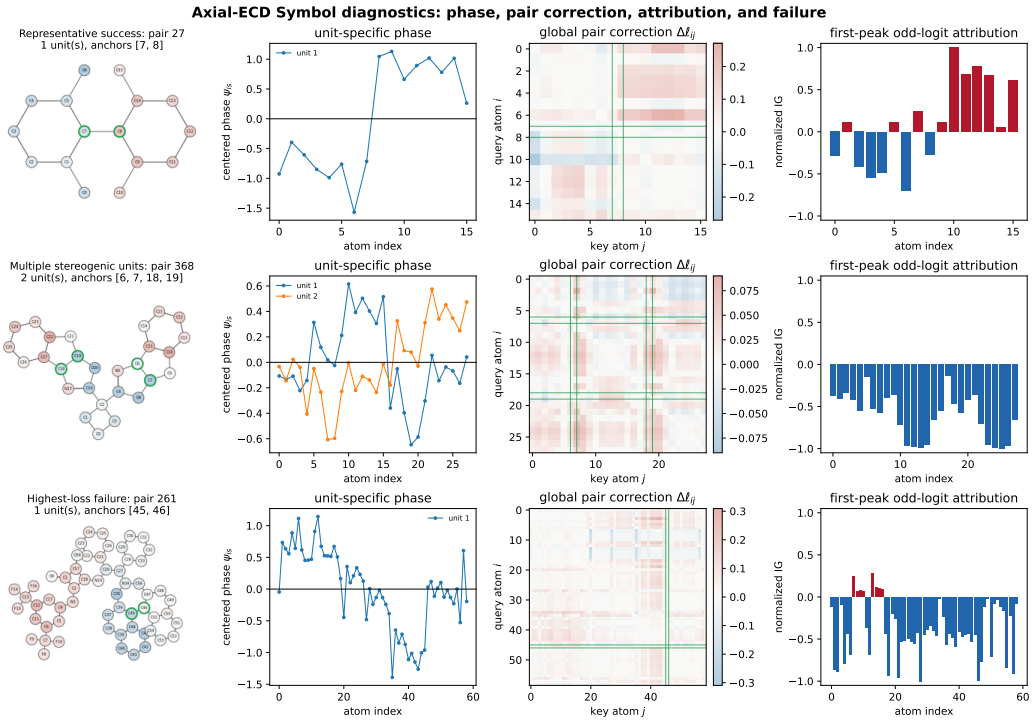


Figure 11: Detailed axial-ECD Symbol diagnostics. Rows contain the deterministic success (pair 27), multi-unit example (pair 368), and highest-loss failure (pair 261). Columns show the phase-colored structure, unit-specific phase curves, all-atom rotary correction, and integrated gradients for the first Symbol logit. These panels are qualitative diagnostics; the same-checkpoint mask intervention provides a separate support-sensitivity test.

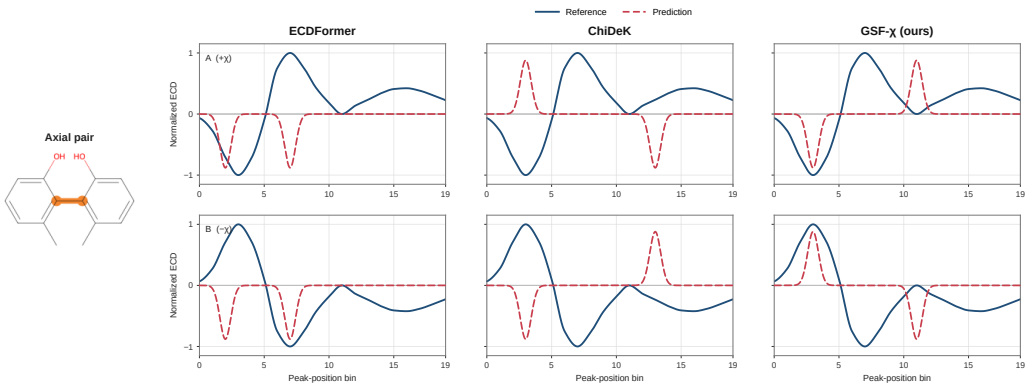


Figure 12: Control-separating Symbol case (ACMP pair 27). GSF-χ predicts complementary signs at Positions $(3, 11)$ and matches the corresponding released signs in both configurations. It predicts two rather than four peaks, so this is deliberately a Symbol-focused case rather than exact spectrum recovery; Table 25 evaluates all four real-peak slots and contrasts the scope controls.

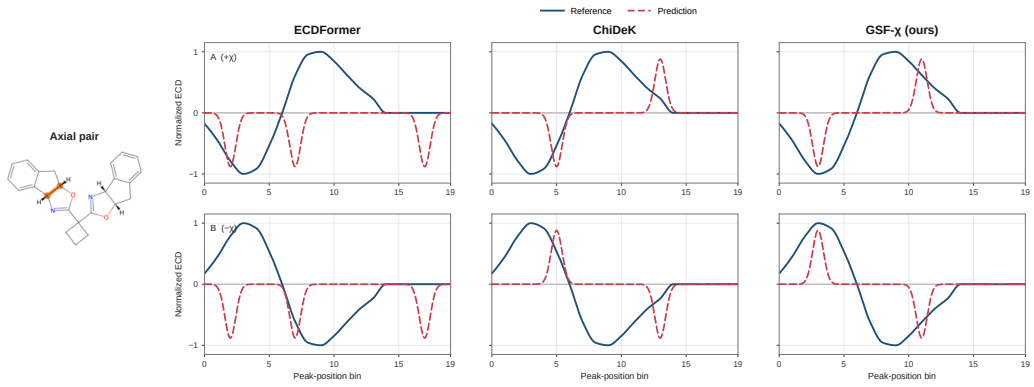


Figure 13: Two-unit case (ACMP pair 368). GSF-χ preserves Number and Position across the reflected pair and predicts both complementary Symbols; its second Position is shifted from 9 to 11. The example demonstrates that two stereogenic fields can contribute without imposing an order over units, not that every spectral component is exact.

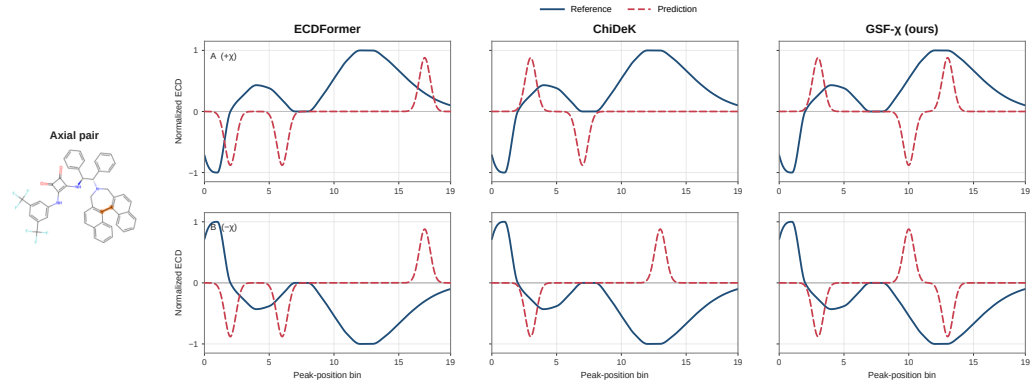


Figure 14: Locked highest-loss failure (ACMP pair 261). GSF-χ obeys its exact mirror law, with Number and Positions agreeing across configurations and the predicted Symbols complementing, but the prediction is inaccurate for both Positions and two of three Symbols. Algebraic consistency constrains the error; it does not guarantee chemical accuracy.