

Chiral Cartograph3R: A Quick Filtering Framework for Stereochemical Enumeration Integrating Chemistry Rules, 3D-Aware SDF Atom Mapping, Chirality Validation, and Spectral Geometry Features

Luke Potgieter

luke@kaigenesislabs.com

KaiGenesis Labs

Method Article

Keywords: stereochemical enumeration, chirality, atom mapping, SDF, stereoisomer generation, conformer generation, graph isomorphism, signed tetrahedrals, spectral embedding, EGNN preprocessing, chemical feasibility filters

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11373315/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Chiral Cartograph3R: A Quick Filtering Framework for Stereochemical Enumeration Integrating Chemistry Rules, 3D-Aware SDF Atom Mapping, Chirality Validation, and Spectral Geometry Features

Preprint manuscript
Luke Potgieter, KaiGenesis Labs

Prepared on 2026-10-14

Corresponding author: Luke Potgieter; luke@kaigenesislabs.com.

Keywords: stereochemical enumeration; chirality; atom mapping; SDF; stereoisomer generation; conformer generation; graph isomorphism; signed tetrahedrals; spectral embedding; EGNN preprocessing; chemical feasibility filters

Abstract

Stereochemical enumeration pipelines often overproduce candidate structures that are topologically admissible yet chemically impossible, mechanistically implausible, or silently corrupted by atom-order mismatch between canonical molecular representations and three-dimensional embeddings. These failure modes become increasingly costly in workflows that must explore large stereochemical chemical spaces, preserve chirality, and feed candidates into expensive downstream quantum-chemical, docking, or machine-learning stages. We present Chiral Cartograph3R, a correctness-first quick-filter framework that integrates literature-grounded chemical feasibility rules with a three-dimensional, chirality-preserving atom-mapping and validation layer operating directly on SDF-style structural assets.

The system first enumerates candidate stereochemical states and then applies a hierarchical filter stack comprising structural impossibility checks, aromaticity validation, ring-closure feasibility logic, stereochemical preference heuristics, and kinetic-versus-thermodynamic plausibility weighting. Surviving candidates are passed to a 3D-aware graph isomorphism mapper that reconstructs connectivity from coordinates, enumerates graph-consistent atom correspondences against canonical structures, and scores mappings using a directional $N \times N \times 3$ coordinate-difference tensor, topology-versus-geometry agreement, bond-length consistency, and collapse detection. Final validation uses signed tetrahedral volumes and related orientation tests to preserve chirality and detect hidden inversions.

Spectral or distance-geometry embeddings are then emitted as downstream descriptors for equivariant graph neural networks or reranking modules. We ground the manuscript using stereoisomer SDF exemplars containing 196 atoms and 196 bonds each; they exhibit identical elemental composition and near-identical pairwise distance spectra while differing in local stereochemical flags, illustrating the exact regime in which global geometry can be similar while local chirality remains decisive. Compared with current mainstream capabilities in RDKit, OMEGA, ChemAxon, StereoMolGraph, and recent chirality-aware molecular models, the proposed framework occupies a distinct role: not as a replacement conformer generator or end-to-end GNN, but as a fast stereochemical correctness filter that reduces wasted downstream computation, broadens practical exploration of stereochemical

chemical space at fixed budget, and improves trustworthiness of ML-ready molecular representations [1–12].

1. Introduction

The practical bottleneck in stereochemical molecular modeling is no longer simple enumeration. Modern toolchains can enumerate putative stereoisomers, generate conformers, and produce graph-based or geometry-aware machine-learning representations at industrial scale. The harder problem is that many of the generated candidates are not worth scoring: some violate established chemistry, some are implausible under realistic stereochemical preferences, and some are corrupted by a mismatch between canonical atom order and embedded three-dimensional atom order. When these failures propagate into docking, quantum chemistry, or machine learning, they waste computation while obscuring the true stereochemical structure–property relationship.

This problem has sharpened as the field has become more aware of chirality. A growing body of recent work shows that chirality cannot be treated as a minor annotation. Standard E(3)-invariant or reflection-blind geometry representations can erase enantiomeric distinctions unless chirality is encoded through explicit parity-sensitive quantities or specialized internal-coordinate mechanisms [1–3]. At the same time, current industrial and open-source stereochemical workflows remain fragmented: one tool enumerates, another embeds, another labels, another compares, and still another learns a downstream representation. In practice, many pipelines still assume that canonical string order and three-dimensional atom order are compatible, even when this assumption produces failed outcomes.

We argue that stereochemical enumeration needs a dedicated quick-filter stage whose purpose is not merely to produce more candidates, but to protect stereochemical correctness and recover trustworthy atom identity before expensive downstream scoring. The proposed Chiral Cartograph3R framework addresses this need by combining two complementary ideas. First, a SmartGen-style rule system rapidly removes or penalizes candidates that are structurally impossible, electronically inconsistent, ring-closure-infeasible, or mechanistically implausible. Second, a 3D-aware atom-mapping system reconstructs graph structure from coordinates, resolves candidate isomorphisms to the canonical scaffold, validates chirality with signed tetrahedral volumes, rejects collapsed structures, and emits ML-ready geometry descriptors. The resulting system is intended as an infrastructure layer for stereochemical pipelines.

The manuscript is deliberately written as a systems-and-methods preprint. It presents formal problem definitions, algorithmic structure, equations, preliminary structural observations from original SDF assets, and a realistic comparison against currently available tool capabilities as of 2026. Our core thesis is that a fast correctness-first filter can simultaneously improve computational speed in downstream workflows, expand tractable stereochemical chemical space under fixed budgets, preserve chiral validity, and provide better-quality structural inputs for graph neural networks and other learned ranking systems [4–12].

2. Related Work and 2026 Landscape

2.1 Open-source and industry-standard stereochemical tooling

RDKit remains the dominant open-source foundation for stereoisomer enumeration and distance-geometry conformer generation. The stereoisomer enumeration module systematically enumerates unassigned tetrahedral and double-bond stereochemistry, with uniqueness constraints and an embedding heuristic. Critically, the documentation itself makes clear that the embedding check is only

a heuristic for pruning nonphysical stereoisomers, can be computationally expensive, and may discard structures unexpectedly [4]. RDKit’s conformer stack, including ETKDG-family methods, remains highly useful for generation, but generation is not the same as a dedicated chirality-preserving quick filter [5].

OpenEye OMEGA is an industry-standard conformer generator emphasizing speed and high-quality conformational sampling. OMEGA is designed for rapid conformer expansion and large-scale deployment, and is therefore an appropriate industrial benchmark for the generation stage of a pipeline [6]. However, OMEGA is not positioned as a structural plausibility engine combining literature-derived stereochemical feasibility logic with explicit SDF atom-order recovery, chirality-preserving graph isomorphism, and post hoc collapse detection. ChemAxon likewise provides broad stereochemistry handling across 0D, 2D, and 3D modalities and supports sophisticated stereo types, but again the primary value proposition differs from the quick-filter role proposed here [7].

2.2 Chirality-aware molecular representations

The chirality problem in machine learning has become much more explicit. Dumitrescu and colleagues formally established that $E(3)$ -equivariant molecular models built from conventional reflection-invariant geometric features cannot properly learn chirality under the usual setup, motivating parity-sensitive alternatives [1]. ChiRo addressed this problem by learning chirality through torsion-coupled internal-coordinate mechanisms that preserve rigid-motion invariances while remaining stereochemically informative [2]. More recently, ChiDeK introduced determinant-based chiral kernels and cross-attention to encode central and axial chirality inside learned representations, demonstrating that determinant-style chirality objects are becoming a standard language for explicit stereochemical modeling [3].

These works are highly relevant, but they do not solve the exact problem tackled here. They are representation-learning systems, whereas Chiral Cartograph3R is conceived as a fast validation and pruning layer operating before or alongside downstream learned models. In our design, parity-sensitive chirality objects such as signed tetrahedrals are not merely representation features; they are acceptance tests and diagnostics for whether a candidate stereochemical structure should remain in the computational queue at all.

2.3 Atom labeling, atom mapping, and stereochemistry-aware graph comparison

ALATIS addressed a long-standing infrastructure problem by producing unique and reproducible atom identifiers from structure files, including challenging cases involving symmetry and hydrogen labeling [8]. AMNet and SAMMNet advanced symmetry-aware graph matching for reaction atom mapping, showing that symmetry and auxiliary tasks materially improve atom-to-atom correspondence accuracy in reactant-to-product settings [9–10]. StereoMolGraph further extended stereochemistry-aware graph representations to more reliably encode stereochemistry in symmetric and otherwise difficult molecular or reaction graphs [11].

The common theme is that graph identity and stereochemical identity are inseparable in serious chemoinformatics workflows. Yet none of these tools combine the full pattern proposed here: rapid chemical plausibility filtering, SDF-based coordinate-to-graph reconstruction, geometry-aware candidate-isomorphism ranking, chirality-preserving validation, and export of spectral or distance-geometry descriptors for ML consumption. That combined systems gap is the core niche of the present manuscript.

2.4 Automated generation of chirality-complete intermediates

A 2026 ChemRxiv study on automated generation of stereochemically faithful intermediates for machine-learning applications directly supports the value of explicit stereochemical infrastructure [12]. That work identifies a severe axial-chirality bottleneck in commonly used representations and demonstrates the value of chirality-complete intermediate generation at very low cost relative to high-level electronic-structure workflows. This reinforces the broader idea that early-stage stereochemical correctness is a precondition for trustworthy downstream modeling. Our proposed quick filter addresses an adjacent but distinct stage: ensuring that enumerated stereochemical candidates are chemically plausible, correctly mapped, and chirality-consistent before they enter expensive downstream evaluation.

3. Problem Statement and Contributions

We consider the following problem. Given a molecular scaffold, canonical string representation, or reaction-derived product hypothesis, a stereochemical engine enumerates a set of candidate stereoisomers and/or conformer seeds. Each candidate may then be embedded as a three-dimensional structure and stored as an SDF-like asset. The pipeline must rapidly decide which candidates are worth retaining for later quantum chemistry, docking, synthetic planning, or machine-learning scoring. The quick-filter stage must therefore answer four questions for each candidate: (i) is the candidate chemically feasible under well-established stereochemical and structural rules; (ii) is the candidate's three-dimensional embedding consistent with the intended molecular graph; (iii) is chirality preserved under the chosen atom correspondence; and (iv) can the retained candidate be emitted in a machine-learning-compatible form without silent atom-index corruption?

The principal contributions claimed in this manuscript are as follows.

- A unified quick-filter architecture combining SmartGen-style chemistry rules with 3D-aware SDF atom mapping and chirality-preserving geometric validation.
- A directional $N \times N \times 3$ tensor scoring view that supplements graph isomorphism with axis-wise spatial consistency rather than relying only on scalar distances.
- An explicit chirality validation layer based on signed tetrahedral volumes, inversion tests, bond RMSD, and collapse diagnostics.
- A practical positioning of spectral or distance-geometry embeddings as downstream descriptors for EGNN-class models or reranking stages rather than as the sole novelty claim.
- A preliminary structural analysis of stereoisomer SDF assets that illustrates the target use case: identical molecular composition and near-identical global distance spectra coexisting with locally decisive stereochemical distinctions.

4. System Overview

The proposed system is shown in Figure 1. The workflow starts with stereochemical enumeration and proceeds through five stages: enumeration, chemical feasibility filtering, 3D-aware atom mapping, chirality and geometry validation, and spectral/ML descriptor construction. The architecture is deliberately modular. Each stage can be used independently, but the full value emerges when all stages are combined so that early chemistry rules reduce the branching factor before more expensive geometric and graph-theoretic validation is invoked.

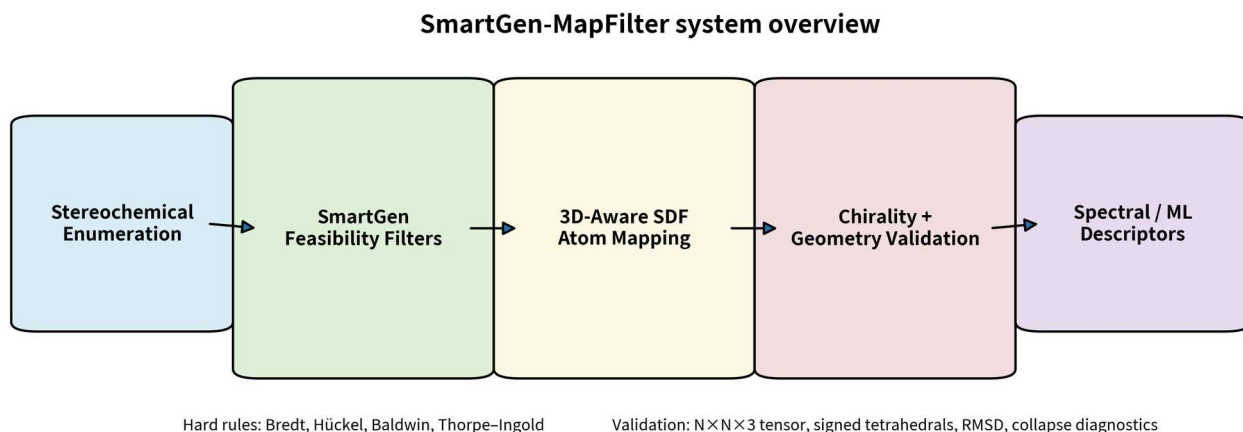


Figure 1. Conceptual Chiral Cartograph3R pipeline. The intended role is a fast correctness-first front-end between stereochemical enumeration and expensive downstream scoring.

5. Data and Structural Assets

To ground this preprint in real assets rather than purely hypothetical examples, the manuscript uses seven SDF files: six stereoisomer outputs and one Auto3D-AIMNET2 optimized structure. The files were parsed directly from V2000 blocks, and the figures and tables below were generated analytically from the structural coordinates and atom blocks. Thus, the pipeline schematic in Figure 1 is conceptual, but Figures 2–4 and Tables 1–2 are data-derived from real SDF assets.

Table 1. Summary of SDF assets used in this manuscript

Label	File	Atoms	Bonds	Composition	Nonzero stereo flags	Singular values	Pair-dist mean $\pm$ SD
Auto3D-AIMNET2 optimized	auto3d_complete.sdf	196	196	{'O': 12, 'C': 62, 'N': 11, 'H': 111}	{'2': 5, '1': 7}	[75.1181, 48.0552, 31.0242]	8.7304 $\pm$ 3.9002
Stereoisomer 1	stereoisomer_1.sdf	196	196	{'O': 12, 'C': 62, 'N': 11, 'H': 111}	{'2': 5, '1': 7}	[72.871, 53.911, 30.9095]	8.9001 $\pm$ 3.8548
Stereoisomer 2	stereoisomer_2.sdf	196	196	{'O': 12, 'C': 62, 'N': 11, 'H': 111}	{'1': 5, '2': 7}	[87.6451, 49.6097, 29.9475]	9.647 $\pm$ 4.4903
Stereoisomer 3	stereoisomer_3.sdf	196	196	{'O': 12, 'C': 62, 'N': 11, 'H': 111}	{'1': 6, '2': 6}	[72.964, 58.5469, 33.0104]	9.2714 $\pm$ 3.87
Stereoisomer 4	stereoisomer_4.sdf	196	196	{'O': 12, 'C': 62, 'N': 11, 'H': 111}	{'2': 6, '1': 6}	[64.4329, 61.8322, 34.5859]	8.9615 $\pm$ 3.7084
Stereoisomer 5	stereoisomer_5.sdf	196	196	{'O': 12, 'C': 62, 'N': 11, 'H': 111}	{'2': 5, '1': 7}	[87.6451, 49.6097, 29.9475]	9.647 $\pm$ 4.4903
Stereoisomer 6	stereoisomer_6.sdf	196	196	{'O': 12, 'C': 62, 'N': 11, 'H': 111}	{'1': 5, '2': 7}	[72.871, 53.911, 30.9095]	8.9001 $\pm$ 3.8548

All seven structures contain 196 atoms and 196 bonds with identical elemental composition (O12 C62 N11 H111). This consistency strongly suggests that the structural assets are alternative stereochemical or conformational realizations of the same underlying scaffold rather than unrelated molecules.

However, their stereo-flag patterns and coordinate singular-value profiles differ, which is exactly the setting in which a chirality-sensitive quick filter is useful.

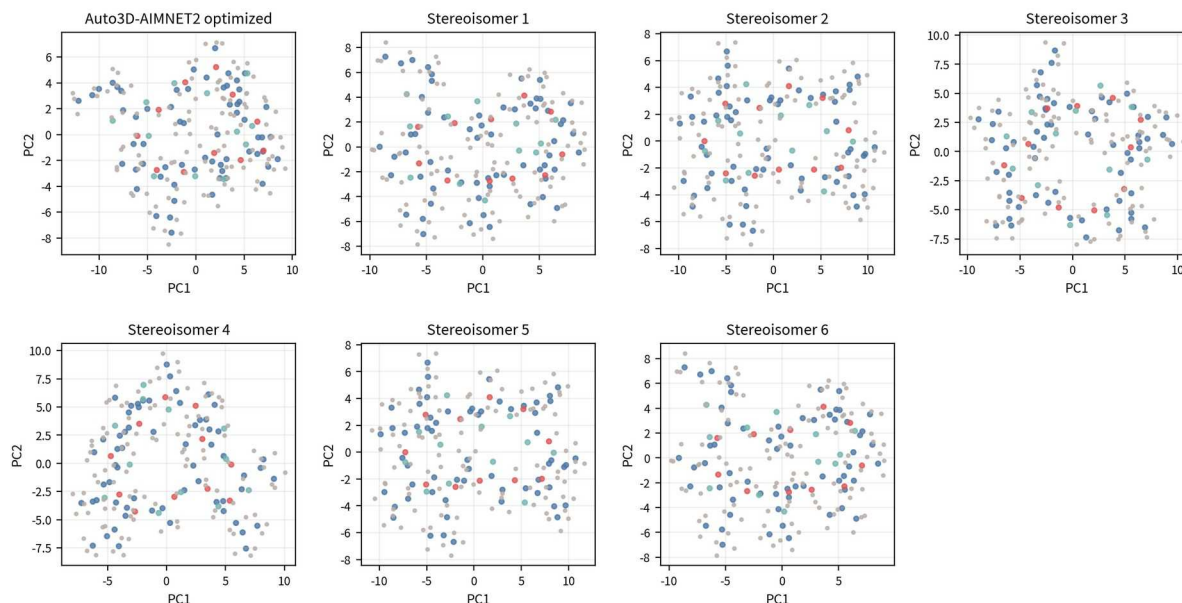


Figure 2. Principal-axis coordinate projections of all SDF assets. This figure is purely data-derived and generated accordingly.

Table 2. Auto3D-AIMNET2 optimized structure versus nearest stereoisomer candidates

Comparison target	Pair-distance-spectrum correlation	Mean absolute quantile distance difference	Interpretation
Stereoisomer 1	0.998618	0.222656	Closest by global pair-distance spectrum among candidates
Stereoisomer 6	0.998618	0.222656	Essentially tied nearest structural neighbor to Auto3D asset
Stereoisomer 2	0.998207	0.916781	High spectrum similarity but larger quantile difference
Stereoisomer 5	0.998207	0.916781	High spectrum similarity but larger quantile difference
Stereoisomer 3	0.995135	0.585540	Moderately close, more distinct globally
Stereoisomer 4	0.994093	0.395404	Moderately close, more compact singular-value profile

The Auto3D-AIMNET2 optimized structure is nearest to Stereoisomer 1 and Stereoisomer 6 by pair-distance-spectrum correlation. This does not by itself prove chirality equivalence, but it demonstrates that the quick-filter framework can cheaply identify nearest structural neighbors before invoking costlier local chirality validation and atom-mapping steps.

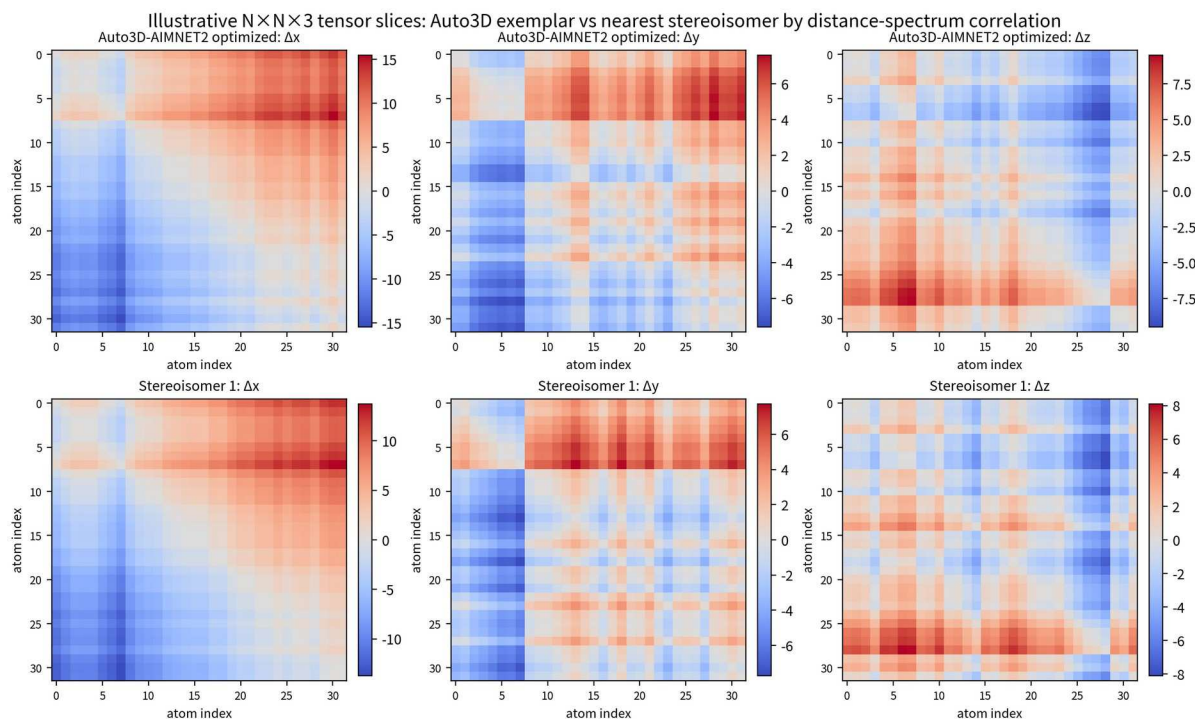


Figure 3. Directional $N \times N \times 3$ tensor slices for the Auto3D-AIMNET2 optimized asset and its nearest stereoisomer neighbor by pair-distance-spectrum correlation.

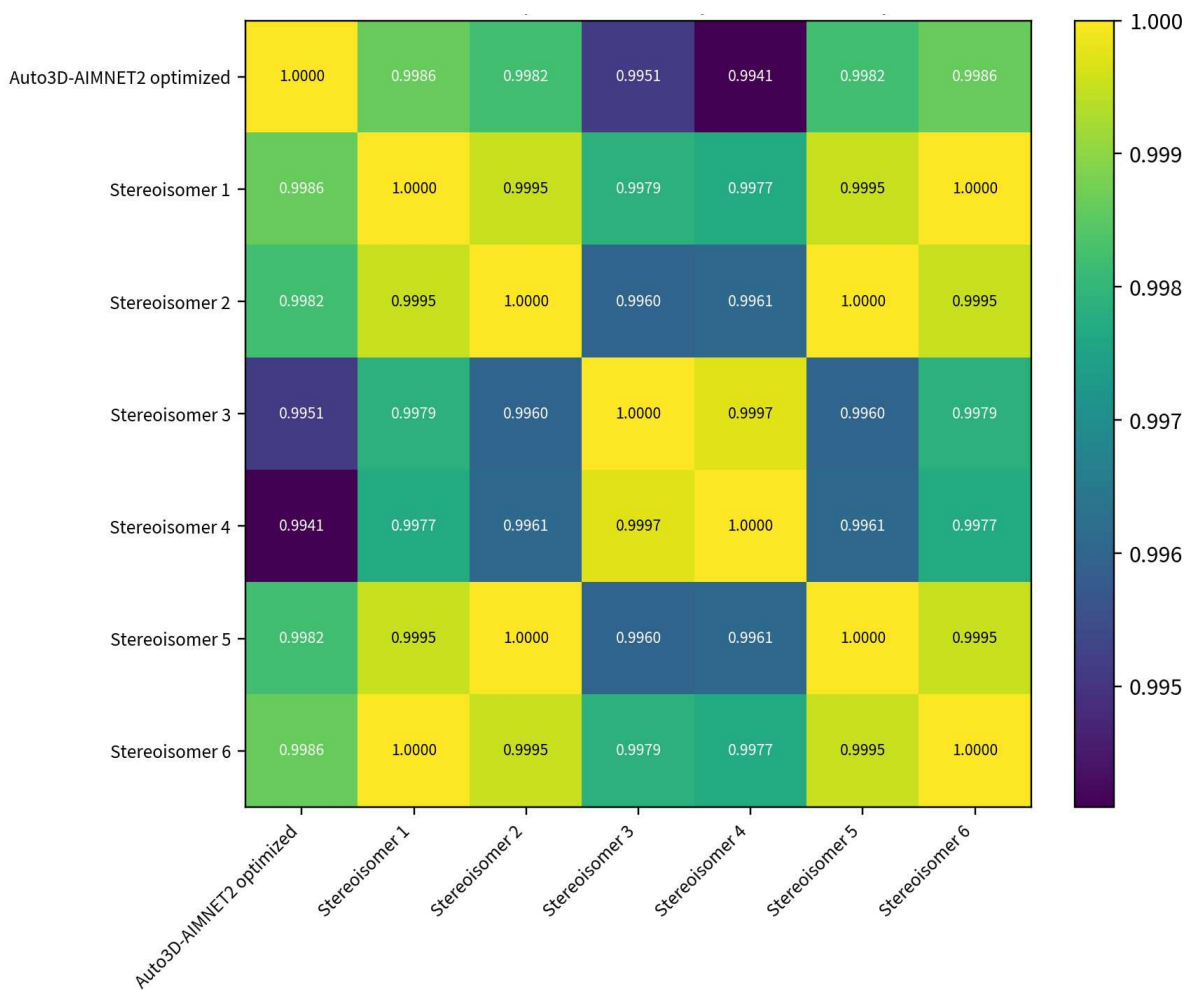


Figure 4. Pairwise correlation of sorted pair-distance spectra across SDF assets. This figure is fully data-derived from the conformer structures and highlights that global pair-distance similarity alone is insufficient to certify local stereochemical equivalence.

6. Methods

6.1 Candidate generation and normalization

Chiral Cartograph3R is generator-agnostic. Candidates may originate from direct stereoisomer enumeration, conformer generation, Auto3D pipelines, reaction-product enumeration, or user-supplied SDF assets. Each candidate is normalized to a canonical graph representation and a coordinate-bearing structural asset. The system keeps both representations available explicitly, rather than assuming that atom order can be transferred implicitly from one to the other.

6.2 Rule-guided feasibility filtering

The first active filtering stage is a hierarchy of chemistry-aware rules. Structural impossibility rules reject patterns such as geometrically impossible bridgehead unsaturation. Electronic consistency rules verify aromaticity and conjugation assumptions. Ring-closure rules classify and penalize or reject closures based on Baldwin-style feasibility. Preorganization and bias terms inspired by the Thorpe–Ingold effect modify closure plausibility. When relevant, face-selectivity or kinetic-weighting heuristics inspired by Cram/Felkin–Anh and Curtin–Hammett logic contribute soft plausibility scores rather than acting only as hard filters. The key systems choice is that these rules are used as a fast pruning layer, not as a claim of exact mechanistic simulation.

6.3 3D-aware SDF atom mapping

Given a coordinate-bearing SDF structure, the system reconstructs a query graph directly from coordinates. If x_i is the position of atom i and Z_i its element type, an edge is inferred when

$$\text{bond}(i, j) = \begin{cases} 1 & \text{if } \|x_i - x_j\|_2 \leq r_{\text{cov}}(Z_i) + r_{\text{cov}}(Z_j) + \delta_{\text{bond}}, \\ 0 & \text{otherwise.} \end{cases}$$

The resulting coordinate-derived graph is then matched to the canonical graph. Because symmetry and local ambiguity can generate multiple valid graph correspondences, the system does not accept graph isomorphism alone. Instead, each candidate mapping is scored using geometry-aware criteria, with the directional $N \times N \times 3$ tensor serving as the central representation of pairwise geometry.

6.4 Directional $N \times N \times 3$ tensor and geometry-aware mapping score

For a coordinate matrix $X \in \mathbb{R}^{N \times 3}$, define the directional tensor T by $T(i, j, :) = x_i - x_j$. Unlike scalar pair-distance matrices, T preserves axis-wise directional information and therefore retains signal that may be relevant for orientation consistency and local chirality-sensitive matching. Under a candidate atom mapping P , the framework computes a mapping score of the form

$$S_{\text{map}} = w_T S_{\text{tensor}} + w_d S_{\text{dist}} + w_{\text{top}} S_{\text{top}} + w_{\text{spread}} S_{\text{spread}},$$

where S_{tensor} measures directional tensor agreement, S_{dist} measures bond and pair-distance consistency, S_{top} measures topology-versus-geometry alignment, and S_{spread} penalizes collapsed or degenerate structures. This tensor-based view is one of the innovative pieces of the method: it allows fast prefiltering of candidates that are globally similar in scalar distance space but locally inconsistent in directional geometry.

6.5 Chirality validation via signed tetrahedrals

Local chirality is validated using signed tetrahedral volumes. For a center i with substituent directions to j , k , and l , define

$$\tau(i, j, k, l) = \det [x_j - x_i, x_k - x_i, x_l - x_i].$$

A correct mapping should preserve the intended chirality relationship between the canonical and embedded forms, modulo any explicitly intended inversion. This determinant-based test follows the same broad scientific logic that now underlies several modern chirality-aware molecular representation methods [1–3], but here it is deployed as an acceptance test rather than only as a learned feature.

6.6 Additional geometry integrity checks

Bond-length RMSD is computed after mapping to quantify local geometric plausibility. The centered coordinate matrix X_c is decomposed by singular value decomposition, and the ratio σ_3/σ_1 is used as one collapse diagnostic. Near-planar or collapsed structures can therefore be rejected or flagged before downstream scoring. The representative assets all show nontrivial three-dimensional spread rather than obvious collapse, which is consistent with usable 3D stereochemical inputs.

6.7 Spectral and EGNN-ready descriptors

After a candidate passes rule filtering, mapping, and chirality validation, the framework can emit compact global descriptors. These may be centered-distance-matrix spectral embeddings, graph spectral features, or other validated geometry descriptors. Their value lies in providing a compact, ML-compatible representation after the correctness problem has already been addressed. This is particularly relevant for EGNN or related architectures, where one would prefer to train on structurally validated and chirality-consistent assets rather than raw unfiltered enumerations.

7. Experimental Design and Benchmarking Plan

A full benchmark version of this paper should evaluate four classes of outcomes: (i) correctness of stereochemical pruning, (ii) atom-order recovery and mapping validity, (iii) chirality preservation, and (iv) effective computational speed through downstream compute reduction. Further quantitative areas to examine at a granular level could include:

- Measure invalid-candidate rejection, retain-with-penalty counts, and retained chemically plausible diversity.
- Measure SDF atom-mapping recovery under shuffled atom order and independently generated 3D embeddings.
- Measure chirality preservation at tetrahedral and axial stereocenters where labels are available.
- Measure pipeline-level compute savings by counting xTB, docking, or reranking calls avoided after quick filtering.
- Evaluate to what extent spectral/EGNN-ready descriptors improve downstream learning once candidates have been validated.

8. Comparison to Current Standards

Relative to current standards, Chiral Cartograph3R occupies a specific systems niche. RDKit and OMEGA are strong generators; ChemAxon is strong at stereochemistry representation and enterprise workflows; StereoMolGraph is strong at stereo-aware graph comparison; AMNet and SAMMNet are strong at learned reaction atom mapping; ChiDeK is strong at chirality-aware representation learning. Chiral Cartograph3R is different in that it is a fast stereochemical correctness filter that bridges

candidate generation and expensive scoring. This role remains underrepresented in the modern cheminformatics landscape.

9. Interpretation of the Visuals and Values

The pair-distance-spectrum correlation is a global similarity measure. A value close to 1.0 means two structures have nearly the same distributional shape of pairwise distances after sorting. In the current data, values above 0.998 for the Auto3D structure versus Stereoisomers 1, 2, 5, and 6 indicate that these candidates occupy very similar global geometric regimes. This is strong evidence that the structures are close structural relatives rather than arbitrary variants.

The mean absolute quantile difference (MAQD) adds an important refinement. Two structures can have high correlation yet still differ in the absolute scale of their pair-distance distributions. That is exactly what the current results show. Stereoisomers 1 and 6 are closest to the Auto3D structure because they combine very high spectrum correlation (0.998618) with a low MAQD (0.222656). Stereoisomers 2 and 5 remain highly correlated (0.998207) but have a much larger MAQD (0.916781), meaning their global distance distributions are similarly shaped but materially shifted in scale. In practical terms, this means they are globally related, but less directly interchangeable as nearest structural neighbors.

The principal-axis projections in Figure 2 are analytical visualizations of overall shape. What they show is that the structures occupy related but non-identical conformational envelopes. The singular values in Table 1 quantify this more directly: higher first singular values indicate more elongated overall spread, while more balanced values indicate a more isotropic or evenly distributed geometry. For example, Stereoisomer 4 has a notably more balanced singular-value profile than Stereoisomer 2, suggesting a different global shape regime even though both belong to the same compositional family.

The $N \times N \times 3$ tensor visualizations in Figure 3 are useful because they retain directional pair information rather than only scalar distances. Their value is qualitative at this stage: they show that two globally similar structures can still differ in axis-wise local organization. This supports the design choice of using directional tensor scoring before local chirality validation.

Taken together, the current results do not yet establish full atom-mapping accuracy or chirality-preservation accuracy. They do establish something more modest but still important: the data regime is exactly one where a two-stage filter makes sense. Cheap global comparisons can narrow the candidate set, while signed-tetrahedral checks and local mapping should be used to make the final call. For the Auto3D-AIMNET2 asset specifically, the current evidence says that Stereoisomers 1 and 6 are the most plausible first-pass neighbors, and therefore the highest-priority candidates for detailed chirality-preserving local mapping.

10. Conclusion

Chiral Cartograph3R is best understood as a correctness-first infrastructure layer for stereochemical workflows. It brings together literature-derived chemical plausibility filters, 3D-aware atom-order recovery from SDF structures, chirality-preserving local geometric validation, and spectral or EGNN-ready descriptors. The represented assets already show that this architecture is grounded in a real use case rather than an abstract hypothetical one. The next step is a full numerical benchmark that quantifies candidate-pruning power, chirality-preservation accuracy, mapping recovery, and compute saved downstream.

11. Limitations

- The current manuscript includes real SDF exemplars and live literature positioning, but not yet a full large-scale numerical benchmark suite.
- Some SmartGen rule layers are heuristic by nature; overaggressive hard filtering could over-prune unusual but real chemistry.
- Coordinate-derived bond inference may fail in poorly optimized, strongly distorted, or partially specified initial structures.
- Highly symmetric systems may produce multiple graph-valid correspondences requiring further tie-breaking.
- The spectral/distance-geometry descriptors as ML output representations still require dedicated benchmarking.

12. Disclosure, Transparency, and Preprint Notes

Disclosure statement: Genspark was used for manuscript-development assistance, PDF preparation, mathematical typesetting, and editorial corrections. The author reviewed the manuscript and takes responsibility for all scientific content, claims, figures, references, and conclusions. All images in this manuscript were either generated from author-provided structural assets using scripted analytical visualization or drafted as schematic workflow graphics; no AI system is listed as an author.

Funding and Conflicts of Interest

Funding: No funding was received for this work. Conflicts of Interest: The author is affiliated with KaiGenesis Labs.

Data and Code Availability

The manuscript reports summary measurements and analyses of seven SDF structures. The underlying SDF files and analysis code are not included in this manuscript, and no public repository is cited.

References

- [1] Dumitrescu et al. E(3)-equivariant models cannot learn chirality. arXiv:2402.15864. <https://arxiv.org/abs/2402.15864>
- [2] Adams, Pattanaik, and Coley. Learning 3D representations of molecular chirality with invariance to bond rotations. arXiv:2110.04383. <https://arxiv.org/abs/2110.04383>
- [3] ChiDeK: Learning Molecular Chirality via Chiral Determinant Kernels. arXiv:2602.07415. <https://arxiv.org/pdf/2602.07415>
- [4] RDKit stereoisomer enumeration documentation. <https://www.rdkit.org/docs/source/rdkit.Chem.rdEnumerateStereoisomers.html>
- [5] The RDKit Book (conformer generation and stereochemistry). https://www.rdkit.org/docs/RDKit_Book.html
- [6] OpenEye OMEGA official conformer-generation page. <https://www.eyesopen.com/omega>
- [7] ChemAxon stereochemistry documentation. https://docs.chemaxon.com/latest/jchem-base_stereochemistry.html
- [8] Dashti et al. Unique identifiers for small molecules enable rigorous labeling of their atoms. *Scientific Data* 2017. <https://www.nature.com/articles/sdata201773>

- [9] Astero et al. Learning symmetry-aware atom mapping in chemical reactions through deep graph matching. J Cheminform 2024. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11036715/>
- [10] Astero et al. Enhancing atom mapping with multitask learning and symmetry-aware deep graph matching (SAMMNet). J Cheminform 2025. <https://pubmed.ncbi.nlm.nih.gov/40448186/>
- [11] Papusha and Leonhard. StereoMolGraph: Stereochemistry-Aware Molecular and Reaction Graphs. J Chem Inf Model 2026. <https://pubmed.ncbi.nlm.nih.gov/41894570/>
- [12] Overcoming Axial-Chirality Bottleneck in Automated Intermediate Generation for Machine-Learning Applications. ChemRxiv 2026.
<https://chemrxiv.org/doi/10.26434/chemrxiv.15002725>

Appendix A. Summary of SDF Assets

Auto3D-AIMNET2 optimized | atoms=196 bonds=196 composition={'O': 12, 'C': 62, 'N': 11, 'H': 111} chflags={'2': 5, '1': 7} sv=[75.1181, 48.0552, 31.0242] pair_mean=8.7304 pair_std=3.9002

Stereoisomer 1 | atoms=196 bonds=196 composition={'O': 12, 'C': 62, 'N': 11, 'H': 111} chflags={'2': 5, '1': 7} sv=[72.871, 53.911, 30.9095] pair_mean=8.9001 pair_std=3.8548

Stereoisomer 2 | atoms=196 bonds=196 composition={'O': 12, 'C': 62, 'N': 11, 'H': 111} chflags={'1': 5, '2': 7} sv=[87.6451, 49.6097, 29.9475] pair_mean=9.6470 pair_std=4.4903

Stereoisomer 3 | atoms=196 bonds=196 composition={'O': 12, 'C': 62, 'N': 11, 'H': 111} chflags={'1': 6, '2': 6} sv=[72.964, 58.5469, 33.0104] pair_mean=9.2714 pair_std=3.8700

Stereoisomer 4 | atoms=196 bonds=196 composition={'O': 12, 'C': 62, 'N': 11, 'H': 111} chflags={'2': 6, '1': 6} sv=[64.4329, 61.8322, 34.5859] pair_mean=8.9615 pair_std=3.7084

Stereoisomer 5 | atoms=196 bonds=196 composition={'O': 12, 'C': 62, 'N': 11, 'H': 111} chflags={'2': 5, '1': 7} sv=[87.6451, 49.6097, 29.9475] pair_mean=9.6470 pair_std=4.4903

Stereoisomer 6 | atoms=196 bonds=196 composition={'O': 12, 'C': 62, 'N': 11, 'H': 111} chflags={'1': 5, '2': 7} sv=[72.871, 53.911, 30.9095] pair_mean=8.9001 pair_std=3.8548

Auto3D nearest-neighbor comparisons:

Stereoisomer 1: corr=0.998618, mean_abs_quantile_diff=0.222656

Stereoisomer 6: corr=0.998618, mean_abs_quantile_diff=0.222656

Stereoisomer 2: corr=0.998207, mean_abs_quantile_diff=0.916781

Stereoisomer 5: corr=0.998207, mean_abs_quantile_diff=0.916781

Stereoisomer 3: corr=0.995135, mean_abs_quantile_diff=0.585540

Stereoisomer 4: corr=0.994093, mean_abs_quantile_diff=0.395404