FROM BOND ANGLES TO MOLECULAR CURVATURE: A GEOMETRY-AWARE GRAPH TRANSFORMER FOR GROUND-STATE CONFORMATION PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate prediction of molecular ground-state conformations is essential for understanding structure-property relationships and guiding rational molecular design. Despite progress in deep learning, most current methods lack explicit modeling of key geometric priors, particularly atomic bond angles and their derived higher-order curvature features. This often results in conformations with excessive local strain and suboptimal non-bonded interactions that deviate from the true energy minimum. To address this limitation, we propose GeoGT, a Graph Transformer-based model that integrates two geometric constraints into its core design. The first constraint quantifies local geometric non-ideality using an atomic bond angle variance metric, guiding the model away from high-strain structures. The second is a molecular curvature measure derived from angular defect, which extends Gaussian curvature to discrete molecular graphs to capture global spatial bending and packing, thereby improving long-range non-bonded interactions. GeoGT achieves the best distance-based performance across all evaluated datasets and splits, with particularly pronounced gains on QM9. These results highlight the critical role of explicit geometric regularization in improving high-fidelity ground-state conformation prediction. The code and data are available on https://anonymous.4open.science/r/GeoGT-D826.

1 INTRODUCTION

The ground-state conformation of a molecule (Tinkham & Strandberg, 1955; Ospelkaus et al., 2010; Danzl et al., 2010) refers to its three-dimensional structure at the global minimum of the potential energy surface. This conformation fundamentally determines the molecule's physicochemical properties and biological activities (Wang et al., 2024; Pei et al., 2023), rendering it central to diverse applications including drug discovery, biomolecular modeling, and functional material design (Satorras et al., 2021; Schütt et al., 2021; Liu et al., 2022). Consequently, the development of accurate and computationally efficient methods for predicting molecular ground-state conformations represents an enduring challenge in computational chemistry and structural biology (Thölke & De Fabritiis, 2022; Zhou et al., 2023; Zaidi et al., 2023).

Conventional approaches to this task primarily rely on force field–based energy minimization (Landrum et al., 2013; De Vivo et al., 2016; Rappé et al., 1992), quantum chemical calculations (Møller & Plesset, 1934; Parr et al., 1979), or systematic conformational search strategies (Mohamadi et al., 1990). Each of these methods faces substantial challenges. Force fields depend on empirically derived parameters, limiting their transferability and accuracy for complex or unseen molecular systems (Halgren, 1996). Quantum mechanical methods, while highly accurate, incur prohibitive computational costs that preclude large-scale application (Purvis III & Bartlett, 1982). Systematic search algorithms, despite their conceptual simplicity, often fail to escape local energy minima and scale poorly with molecular size (Kolossváry & Guida, 1996). These limitations have motivated growing interest in data-driven alternatives, particularly deep learning frameworks that promise high accuracy at reduced computational cost (Xu et al., 2021a;b).

Nevertheless, current deep learning approaches (Mansimov et al., 2019; Li et al., 2024; Noé et al., 2019) largely neglect explicit geometric constraints (Ho et al., 2020; Zhang et al., 2023). Fundamen-

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.

tal geometric quantities such as bond angles and their higher-order derivatives remain underutilized in conformation modeling, leading to generated structures that lack physical realism. To address this limitation, we introduce a geometric regularization framework that jointly incorporates local and global structural priors into conformation generation. The foundation of our approach is the bond angle, the fundamental geometric primitive that governs local atomic coordination. As shown in Figure 1, all higher-order shape descriptors in our framework are derived from this basic quantity. And in Section F.3, we systematically analyze the bond angle distributions based on the QM9 dataset.

For local geometry, we employ bond angle variance as a physically grounded measure of angular strain. Specifically, the variance centered on atom i, denoted $Var(\Theta_i)$, quantifies distortion in its coordination environment arising from electronic and steric perturbations. Figure 1(b) illustrates how this descriptor captures subtle deviations from ideal hybridization. This is exemplified by a series of neopentane derivatives: in neopentane itself, the central carbon exhibits near-ideal sp^3 geometry with minimal variance; substitution with an electron-withdrawing chlorine in neopentyl chloride perturbs the local electronic structure, increasing angular deviation; and attachment of a bulky phenyl group in neopentylbenzene introduces steric repulsion that further elevates variance. These examples demonstrate that bond angle variance sensitively encodes both electronic and steric contributions to local strain.

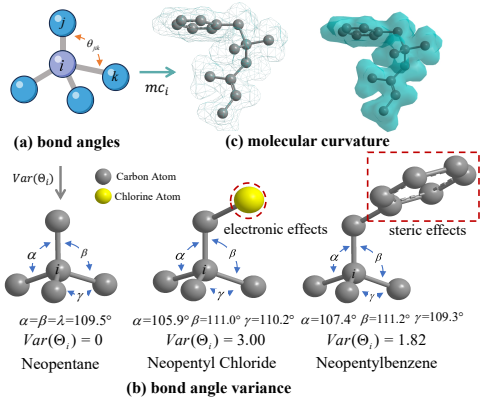


Figure 1: From bond angles to angle variance and molecular curvature in conformation modeling. (a) Bond angles define the basic local geometry around an atom. (b) Bond angle variance quantifies local non-ideality arising from electronic and steric effects. (c) Molecular curvature characterizes global surface morphology.

Complementing this, the framework includes a global shape descriptor based on molecular curvature. Even when local geometries are near-ideal, an unfavorable overall topology can disrupt van der Waals packing or electrostatic complementarity, preventing convergence to the global energy minimum. To capture such large-scale effects, we define molecular curvature using angular defect, a discrete analogue of Gaussian curvature adapted to molecular graphs, as shown in Figure 1(c). This curvature reflects the local convexity or concavity of the molecular surface and directly influences atomic packing density, solvent accessibility, and long-range spatial coordination. These structural attributes govern the strength and complementarity of non-bonded interactions. By embedding this global geometric signal into the conformation generation process, our model enforces spatial regularization that steers the molecular scaffold toward topologies with more favorable non-bonded interaction energies.

In summary, bond angle variance and molecular curvature provide learnable, embeddable geometric constraints for conformation generation from two core dimensions—local strain energy and global interaction energy, respectively. Through the synergistic utilization of these two types of constraints, our method can systematically guide the model to generate three-dimensional structures with reduced local distortion and optimized overall arrangement, thereby more accurately approximating the true ground-state conformation of the molecule.

We implement this framework within a graph transformer architecture, yielding a method named GeoGT that predicts three-dimensional ground-state conformations directly from two-dimensional molecular graphs. By embedding bond angle variance, discrete curvature, adjacency relation, and other geometric cues into the attention mechanism, GeoGT respects both quantum-mechanically informed local geometries and physically plausible global topologies. This physically grounded design achieves accurate conformation prediction without post-hoc refinement or expensive energy minimization, offering a scalable and principled approach to molecular structure modeling.

The main contributions of this work are as follows:

- We propose bond angle variance as a quantitative metric to effectively characterize local geometric distortion, thereby introducing explicit geometric constraints on local atomic arrangements into the generation process.
- Following recent works GTMGC(Xu et al., 2023) and ReBIND(Kim et al., 2025), we adopt QM9 (Ramakrishnan et al., 2014) and Molecule3D (Xu et al., 2021c) as benchmark datasets to evaluate the effectiveness of these geometric constraints. Our experiments demonstrate strong overall performance, with our method achieving the best results on most evaluation metrics across the two datasets.

2 Related Work

2.1 Molecular Generative Model

In recent years, generative models (Xu et al., 2021a; Ho et al., 2020; Mansimov et al., 2019; Luo et al., 2021) have made remarkable progress in the task of molecular conformation generation, becoming important tools for predicting multi-conformational states. Among them, one mainstream approach relies on generative architectures such as diffusion models (Zhang et al., 2023; Fan et al., 2024) or variational autoencoders (VAEs) (Xu et al., 2021b), which directly or indirectly generate atomic coordinates or their derived representations. For instance, methods like GeoDiff (Xu et al., 2022) model the distribution of atomic coordinates through a diffusion process, gradually denoising to produce reasonable conformations; whereas techniques such as Torsional Diffusion (Jing et al., 2022) focus on modeling torsional angles within molecules to reduce the degrees of freedom in the generation process and improve efficiency. Another category of methods adopts autoregressive generation strategies (Li et al., 2024), sequentially determining atomic positions. Additionally, some approaches integrate energy constraints into the generation process, employing conditional generative models to incorporate physicochemical constraints in the output, thereby enhancing the physical plausibility of conformations. Examples include energy-constrained diffusion methods (Yu et al., 2023) or Boltzmann generator-based conformation sampling techniques (Noé et al., 2019). Although these methods (Shi et al., 2021; Ganea et al., 2021) can efficiently generate diverse and low-energy conformational ensembles, they generally place greater emphasis on covering conformational space rather than precisely identifying the global energy-minimum ground-state structure. As a result, they still exhibit significant limitations in tasks aimed at predicting the single most stable ground-state conformation.

2.2 Molecular Graph Transformer

In recent years, Graph Neural Networks (GNNs) (Wu et al., 2020) have made significant progress in the field of molecular modeling, particularly with graph-structured representation learning methods emerging as essential tools for handling molecular conformation and property prediction. Against this backdrop, Graph Transformer—an architecture that integrates graph neural networks with self-attention mechanisms (Dwivedi & Bresson, 2020; Ying et al., 2021; Luo et al., 2022) has gradually demonstrated unique advantages in molecular representation learning (Maziarka et al., 2020; 2024; Choukroun & Wolf, 2021; Wu et al., 2021). Compared to traditional graph neural networks, Graph Transformer (Kim et al., 2022; Min et al., 2022) can effectively capture long-range dependencies between atoms through the self-attention mechanism, overcoming the potential over-smoothing issue that may arise from multi-layer stacking in conventional message-passing mechanisms (Dwivedi & Bresson, 2020). Thereby, it enables a more accurate description of complex non-local interactions within molecules (Dwivedi et al., 2023). Moreover, Graph Transformer has exhibited superior performance over traditional methods in multiple molecular property prediction tasks (Fang et al., 2022). Some studies have further integrated it with equivariant neural networks (Fuchs et al., 2020) to explicitly model roto-translational invariance, thereby enhancing the physical consistency of molecular ground-state conformation generation. However, existing Graph Transformer-based methods for ground-state conformation generation (e.g., GTMGC (Xu et al., 2023)), despite achieving significant performance improvements, tend to focus primarily on modeling lower-order geometric features such as distances, without explicitly incorporating higher-order geometric constraints such as bond angles and curvature. This, to some extent, limits the model's ability to accurately capture both the local geometric rationality and the global morphological coordination of molecules.

3 METHODOLOGY

3.1 NOTATION AND PROBLEM DEFINITION

For ease of reference, the main notation used in this work is summarized in Appendix A, and the preliminaries are presented in Appendix B. We represent the two-dimensional graph structure of a molecule as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ denotes the set of atomic nodes containing N atoms, and $\mathcal{E} = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M\}$ represents the set of chemical bond edges containing M bonds. Its adjacency matrix is denoted as $A \in \mathbb{R}^{N\times N}$, where $A_{ij} = 1$ if there exists a bond between atom i and atom j, and 0 otherwise. And for each central atom i, its neighbor set is defined as $\mathcal{N}(i) = \{j \mid A_{ij} = 1\}$. The ground-state conformation of the molecule is represented by a matrix $\mathcal{C} \in \mathbb{R}^{N\times 3}$, where each row c_i corresponds to the three-dimensional spatial coordinates of atom i. The Euclidean distance matrix between atoms is denoted as $D \in \mathbb{R}^{N\times N}$, where D_{ij} represents the spatial distance between atom i and atom j. Given only the two-dimensional molecular graph structure $\mathcal{G}$, the objective is to predict the corresponding ground-state three-dimensional conformation $\mathcal{C}$. This task requires inferring the most stable three-dimensional structure with the lowest energy from topological information, which is a key challenge in the field of molecular computation.

Overview. The overall framework of GeoGT is shown in Figure 2. First, the input molecule is preprocessed using a tokenizer and then fed into the encoder to obtain latent molecular representations, which are used to generate an initial three-dimensional intermediate conformation. Spatial geometric information, such as atomic bond angles and molecular curvature, is then extracted from this conformation and integrated into the decoder's positional encoding. Subsequently, through the use of MHSA layers based on geometric attention scores, along with the complete decoder module, the final three-dimensional ground-state conformation of the molecule is predicted after multiple rounds of iterative optimization.

3.2 BOND ANGLE COMPUTATION AND VARIANCE ENCODING

Input Preprocessing. We first process the raw molecular data using the MoleBERT tokenizer, converting it into a representation suitable for model input. To further incorporate structural information, in the positional encoding stage, we adopt the Laplacian positional encoding method and integrate the molecular adjacency matrix information into the computation process. Finally, through the above steps, the input representation of the model is obtained as $E \in \mathbb{R}^{N\times d}$.

Encoder Outputs. The embeddings E is fed into an encoder for feature enhancement, where each layer of the encoder consists of MHSA and FFN. Training stability is ensured through residual connections and layer normalization. The final output of the encoder, $X \in \mathbb{R}^{N\times d}$, serves as a deep representation of the entire molecule and is used for subsequent three-dimensional conformation generation.

The bond angle θ_{jik} describes the local spatial geometric configuration formed by three atoms $j \to i \to k$ (with i as the central atom), serving as a key parameter for understanding local chemical environments (e.g., hybridization modes, steric effects) and overall molecular conformations.

In molecular graph representation learning, while the adjacency matrix $A \in \mathbb{R}^{n\times n}$ defines atomic connectivity, it cannot directly provide three-dimensional spatial structural information, particularly critical geometric features like bond angles. To address the limitation of two-dimensional graphs in representing spatial information and to capture geometric distribution characteristics of atomic local environments, the encoder embeddings $X \in \mathbb{R}^{n\times d}$ are fed into a two-layer FFN $f_w(\cdot)$ during processing. This $f_w(\cdot)$ predicts **intermediate conformation** $C_{im} \in \mathbb{R}^{n\times 3}$ containing three-dimensional coordinates of all atoms.

The bond angle, in its essence, is the spatial angle between two chemical bond vectors. Calculating these displacement vectors is the geometric basis for defining this bond angle. Let the three-dimensional coordinates of atom i be $\mathbf{c}_i \in \mathbb{R}^3$.

Using the definitions of the vector dot product and magnitude, the cosine of the included angle is computed directly:

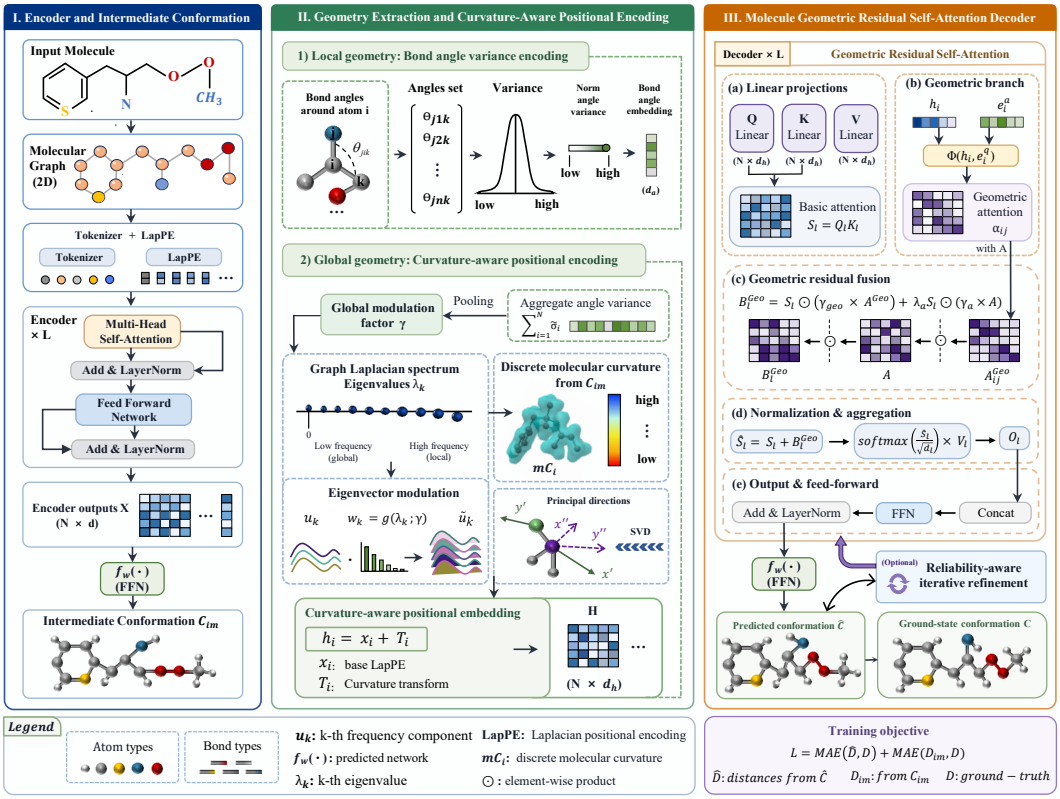


Figure 2: Overview of the proposed GeoGT framework. **(I).** The input two-dimensional molecular graph is tokenized and encoded with Laplacian positional information to generate an intermediate conformation. **(II).** Local bond-angle variance and global molecular curvature are extracted from the intermediate conformation and embedded as geometry-aware positional priors for the decoder. **(III).** The decoder incorporates these geometric priors through geometric residual self-attention and iteratively refines the structure to obtain the final molecular ground-state conformation.

$$\mathbf{v}_{ji} = \mathbf{c}_j - \mathbf{c}_i, \mathbf{v}_{ki} = \mathbf{c}_k - \mathbf{c}_i, \cos\theta_{jik} = \frac{\mathbf{v}_{ji} \cdot \mathbf{v}_{ki}}{\|\mathbf{v}_{ji}\| \cdot \|\mathbf{v}_{ki}\| + \epsilon}. \tag{1}$$

where ϵ is set to $1e{-}6$ to prevent division by zero. And the bond angle θ_{jik} is accurately computed using the arccosine function. For the central atom i, its set of all bond angles Θ_i is defined as $\Theta_i = \{\theta_{jik} \mid j, k \in \mathcal{N}(i), j < k\}$. The variance of atomic bond angles represents the extent to which the atomic environment deviates from the ideal hybridization state. The bond angle variance of atom i is given by:

$$\mathrm{Var}(\Theta_i) = \frac{1}{|\Theta_i|} \sum_{\theta \in \Theta_i} \left(\theta^2 - \left(\frac{1}{|\Theta_i|} \sum_{\theta \in \Theta_i} \theta \right)^2 \right), \tilde{\sigma}_i = \frac{\mathrm{Var}(\Theta_i) - \min_{1 \le l \le N} \mathrm{Var}(\Theta_l)}{\max_{1 \le l \le N} \mathrm{Var}(\Theta_l) - \min_{1 \le l \le N} \mathrm{Var}(\Theta_l) + \epsilon}, \tag{2}$$

where ϵ = $1e{-}6$. Map $\tilde{\sigma}_i$ to the bond angle variance embedding of the corresponding atom via a feed-forward network:

$$e_i^a = W_2 \cdot ReLU\left(W_1 \cdot \tilde{\sigma}_i + b_1\right) + b_2, \tag{3}$$

where W_1, W_2 are weight matrices, and b_1, b_2 are bias vectors.

3.3 CURVATURE-AWARE POSITIONAL ENCODING

The global bond angle variance of a molecule measures its deviation from ideal hybridization geometry. A larger variance typically reflects the accumulation of local distortions caused by steric hindrance, strain in small or bridged rings, and strongly electron-withdrawing or electron-donating substituents. In such cases, the initial conformation often deviates from a reasonable structure, requiring enhanced high-frequency components to capture and correct local deviations. A smaller variance indicates that local structures more closely approximate ideal hybridization geometries, as in the conjugated planes of aromatic systems or extended conformations of long-chain alkanes. Further optimization toward the ground-state conformation then relies primarily on improving long-range interactions and the continuity of conjugation. Low-frequency components should therefore be enhanced to extend the propagation range of global information, improving the overall consistency and convergence of structural predictions.

Theorem 1. *Let $\mathcal{G}$ be a connected undirected graph, and let $\mathbf{L} = \mathbf{D} - \mathbf{A}$ be its graph Laplacian. Suppose the eigenvalues of $\mathbf{L}$ satisfy $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$, with corresponding orthogonal eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. For each eigenvector $\mathbf{u}_k$, its smoothness over the graph is measured by the Rayleigh quotient:*

$$\mathcal{R}_{\mathbf{L}}(\mathbf{u}_k) = \frac{\mathbf{u}_k^\top \mathbf{L} \mathbf{u}_k}{\|\mathbf{u}_k\|_2^2} = \lambda_k.$$

When $\lambda_k \approx 0$, $\mathbf{u}_k$ has small edge-wise variation and therefore varies smoothly over the graph, encoding global structural information. Conversely, when λ_k is large, $\mathbf{u}_k$ has large graph-gradient energy and therefore represents high-frequency edge-wise variations, which are useful for capturing local structural details.

In Appendix C, we provide a theoretical derivation and proof about Theorem 1. Based on the above analysis, we design an adaptive weighting mechanism that modulates molecular feature vectors in the frequency domain, enabling them to dynamically balance the contributions of high-frequency and low-frequency components according to the molecule's global bond angle variance. Specifically, the molecular eigenvectors are adjusted using the following formula:

$$w_k = \gamma \cdot \frac{1}{\log(e + \lambda_k/\epsilon)} + (1-\gamma) \cdot \frac{\lambda_k}{\lambda_{\max}}, \; \tilde{u}_k = w_k \cdot u_k, \tag{4}$$

where $\gamma = \exp\left(-\frac{1}{N}\sum_{i=1}^{N} \tilde{\sigma}_i\right)$ is a modulation factor computed based on atomic bond angle variance, λ_k denotes the eigenvalue corresponding to the k-th frequency component u_k, $\lambda_{\max}$ is the maximum eigenvalue, and ϵ is a small constant introduced to ensure numerical stability.

In differential geometry, the continuous Gaussian curvature quantifies the intrinsic bending of a surface at a given point:

$$K = \lim_{r \to 0} \frac{2\pi - \sum \theta_i'}{A(r)}, \tag{5}$$

where the $2\pi - \sum \theta_i'$ represents the angle deficit, $A(r)$ is the area of a small disk centered at the target point with radius r.

Molecular structures can be modeled as discrete surfaces (polyhedra), where atoms correspond to vertices, chemical bonds correspond to edges, and planar or polygonal regions formed by atomic rings correspond to faces. Based on this discrete geometric model, the definition of continuous Gaussian curvature particularly its angle deficit characterization can be naturally extended to molecules. This enables computation of discrete curvature for the local surface around atomic centers:

$$R_i = \max(A_i \odot D_i), mc_i = \frac{4\pi - \sum_{\theta \in \Theta_i} 2\pi \left(1 - \cos\left(\frac{\theta}{2}\right)\right)}{4\pi R_i^2 + \epsilon}. \tag{6}$$

where A_i is the adjacency vector of atom i, D_i is the distance vector from atom i to all other atoms, $\odot$ denotes hadamard product, and R_i represents the distance between atom i and its farthest bonded atom. The constant ϵ is introduced to ensure numerical stability.

Regions exhibiting extreme curvature values, such as flat ligand-binding pockets or specific catalytic sites, often encode critical chemical functional information. Therefore, curvature should be projected along the most geometrically relevant principal directions to enhance feature representation in these key regions. Integrating discrete geometric curvature with modulated feature vectors, followed by singular value decomposition (SVD) to extract principal directional features.

$$P_i = \mathrm{Diag}\left(\exp\left(-\beta|mc_i - mc_{\min}|\right)\right) + A_i + D_i, b = \sum_{i=1}^{N} mc_i \tilde{u}_i,\ [U, \Sigma, V] = \mathrm{SVD}(b),\ T_i = P_i + U. \tag{7}$$

where $mc_{\min} = \min_{1 \le j \le N} mc_j$, β is a learnable parameter. Ultimately, the curvature-aware positional representation $H \in \mathbb{R}^{N \times d}$ is obtained:

$$h_i = x_i + T_i, \qquad H = [h_1, h_2, \ldots, h_N]^\top, \tag{8}$$

where $\mathbf{x}_i$ denotes the i-th representation from the encoder output X.

3.4 Molecular Geometric Residual Self-Attention

After obtaining the curvature-aware positional embeddings H, they are fed into MHSA layer. For each attention head $l \in [1, \mathrm{P}]$, the input features H are linearly projected into a query matrix Q_l and a key matrix K_l. The dot product similarity between them is then computed as the basic attention score matrix $S_l = HW_l^Q (HW_l^K)^T = Q_l K_l^T$.

Combine the curvature-aware positional encoding embeddings $\mathbf{h}_i$ and bond angle variance embeddings $\mathbf{e}_i^a$ to calculate geometric attention score for the adjacent atom pair (v_i, v_j):

$$\alpha_{ij} = \frac{\exp\left(h_i^T h_j + w_a^T (e_i^a + e_j^a)\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(h_i^T h_k + w_a^T (e_i^a + e_k^a)\right)}, A_{ij}^{Geo} = A_{ij} + \lambda_\alpha \alpha_{ij}. \tag{9}$$

where A_{ij} equals 1 if atoms i and j are bonded, otherwise 0, and λ_α is a learnable parameter. Building upon the original self-attention mechanism that captures global chemical contexts, the framework further extends its modeling capability to geometric constraints through B_l^{Geo}:

$$B_l^{Geo} = S_l \odot (\gamma_{geo} \times A^{Geo}) + \lambda_a S_l \odot (\gamma_a \times A). \tag{10}$$

where γ_{geo} and γ_a are learnable weight vectors, and λ_a is a scalar parameter. The integrated self-attention score $\widehat{S}_l$ and MHSA's output O is generated by:

$$\widehat{S}_l = S_l + B_l^{Geo}, O_l = \mathrm{softmax}\left(\frac{\widehat{S}_l}{\sqrt{d_l}}\right) V_l, O = \mathrm{Concat}(O_1, \ldots, O_P) W^O. \tag{11}$$

where $\sqrt{d_l}$ is the scaling factor and W^O denotes the output projection weight matrix.

The resulting representation O is then passed through a feed-forward network, followed by residual connections and layer normalization, to obtain the final output representation $G \in \mathbb{R}^{N \times d}$. G is further fed into the same feed-forward network $f_w(\cdot)$ previously used for encoder conformation generation, yielding the predicted molecular ground-state conformation $\widehat{C} \in \mathbb{R}^{N \times 3}$.

Optional iterative refinement. The intermediate conformation C_{im} generated by the encoder provides an initial geometric reference for the decoder, which uses the extracted geometric information to predict the molecular ground-state conformation $\widehat{C}$. In this process, the intermediate conformation provides informative geometric cues and does not itself need to attain the accuracy of the final

predicted conformation. To further improve the geometric reference, the model supports optional iterative refinement: the conformation predicted in the previous round is reused as the geometric reference to update the corresponding geometric information and perform another decoding pass. We denote the number of decoding passes by T, where $T = 1$ indicates a single decoding pass without additional iterative refinement. As shown in Appendix F.7, a moderate number of refinement steps can further reduce prediction errors but incurs substantial inference overhead, and additional iterations do not guarantee continued improvements. Therefore, we adopt $T = 1$ by default and directly use $\widehat{C}$ as the final predicted conformation $C \in \mathbb{R}^{N\times 3}$, balancing prediction accuracy and computational efficiency. For applications that can accommodate additional computational costs, iterative refinement can be enabled, with the number of decoding passes selected based on validation performance. Additionally, Appendix D provides the complexity analysis and pseudocode for GeoGT.

3.5 Loss Function

Ultimately, we employ the Mean Absolute Error (MAE) of the inter-atomic distance matrix as the core component of the loss function. Specifically, the MAE between the predicted distance matrix $\widehat{D}$ of the molecular ground-state conformation C and the true distance matrix D is calculated to ensure the model accurately learns the spatial structure of the target conformation. To further enhance the representational capability of the encoder, we also compute the MAE between the distance matrix D_{im} of the intermediate conformation C_{im} and the true matrix D. This constraint enables the intermediate representations generated by the encoder to progressively approximate the true geometric relationships. The overall loss function is thus constituted by the sum of the aforementioned two loss terms:

$$\mathcal{L} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\widehat{\boldsymbol{D}}_{ij} - \boldsymbol{D}_{ij}\right| + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\boldsymbol{D}_{im,ij} - \boldsymbol{D}_{ij}\right|. \tag{12}$$

4 Experiments

4.1 Experimental Setup

Following recent works GTMGC (Xu et al., 2023) and ReBIND (Kim et al., 2025), we adopt QM9 and Molecule3D as benchmark datasets to evaluate GeoGT. QM9 comprises small organic molecules with up to nine heavy atoms, while Molecule3D contains approximately 3.9 million molecules. For Molecule3D, we consider both random and scaffold splits (Appendix E.1). Baselines include RDKit DG, RDKit ETKDG, GINE, GATv2, GraphGPS (RW/LP), GTMGC, and ReBIND, with DeeperGCN-DAGNN additionally included for Molecule3D (Appendix E.4). We measure conformation accuracy using D-MAE and D-RMSE for interatomic distances, and K-RMSD for heavy-atom coordinates after Kabsch alignment (Appendix E.2). Dataset partitions and training configurations are detailed in Appendix E.3. Additional analyses, including ablation studies, chemical validity, and inference efficiency, are provided in Appendix F.

4.2 Molecular Ground-State Conformation Prediction

As shown in Table 1, GeoGT achieves the best results on both the validation and test sets of QM9 and the Molecule3D random split. Compared with the best baseline for each metric, GeoGT reduces D-MAE by 50.8% and 49.2% on the QM9 validation and test sets, respectively, and D-RMSE and K-RMSD by 33.4%~38.4%. On the Molecule3D random split, D-MAE decreases by 16.0% on both sets. Under the more challenging scaffold split, GeoGT maintains the best D-MAE and D-RMSE results, while ranking second to ReBIND in K-RMSD. Furthermore, GeoGT achieves the highest inference efficiency (Appendix F.4) and the highest overall pass rate in the chemical validity evaluation (Appendix F.5). These results demonstrate that GeoGT improves conformation prediction accuracy through explicit geometric priors while achieving high inference efficiency and chemical validity, and maintaining good generalization across different data splits.

Table 1: Ground-state conformation prediction results on QM9 and Molecule3D.

Method	Validation			Test		
	D-MAE↓	D-RMSE↓	K-RMSD↓	D-MAE↓	D-RMSE↓	K-RMSD↓
(a) QM9						
RDKit DG	0.358	0.616	0.722	0.358	0.615	0.722
RDKit ETKDG	0.355	0.621	0.691	0.355	0.621	0.689
GINE	0.357	0.673	0.685	0.357	0.669	0.693
GATv2	0.339	0.663	0.661	0.339	0.659	0.666
GraphGPS(RW)	0.328	0.629	0.628	0.327	0.624	0.628
GraphGPS(LP)	0.283	0.500	0.544	0.283	0.499	0.546
GTMGC	0.262	0.468	0.362	0.264	0.470	0.367
ReBIND	0.252	0.442	0.320	0.254	0.446	0.321
GeoGT(Ours)	**0.124**	**0.288**	**0.197**	**0.129**	**0.297**	**0.209**
Reduction (%)↑	50.8	34.8	38.4	49.2	33.4	34.9
(b) Molecule3D Scaffold Split						
RDKit DG	0.542	0.872	1.001	0.524	0.857	0.973
RDKit ETKDG	0.531	0.874	0.928	0.511	0.859	0.898
DeeperGCN-DAGNN	0.617	0.930	–	0.763	1.176	–
GINE	0.883	1.517	1.407	1.400	2.224	1.960
GATv2	0.778	1.385	1.254	1.238	2.069	1.752
GraphGPS(RW)	0.503	0.853	0.978	0.595	1.024	1.065
GraphGPS(LP)	0.417	0.690	0.827	0.411	0.690	0.827
GTMGC	0.406	0.675	0.678	0.400	0.679	0.693
ReBIND	0.391	0.661	**0.640**	0.386	0.663	**0.667**
GeoGT(Ours)	**0.344**	**0.618**	0.660	**0.359**	**0.641**	0.686
Reduction (%)↑	12.0	6.5	-3.1	7.0	3.3	-2.8
(c) Molecule3D Random Split						
RDKit DG	0.581	0.930	1.054	0.582	0.932	1.055
RDKit ETKDG	0.575	0.941	0.998	0.576	0.942	0.999
DeeperGCN-DAGNN	0.509	0.849	–	0.571	0.961	–
GINE	0.590	1.014	1.116	0.592	1.018	1.116
GATv2	0.563	0.983	1.082	0.564	0.986	1.083
GraphGPS(RW)	0.512	0.900	1.006	0.513	0.903	1.008
GraphGPS(LP)	0.440	0.730	0.854	0.441	0.732	0.854
GTMGC	0.432	0.719	0.712	0.433	0.721	0.713
ReBIND	0.418	0.706	0.698	0.419	0.708	0.699
GeoGT(Ours)	**0.351**	**0.656**	**0.679**	**0.352**	**0.659**	**0.681**
Reduction (%)↑	16.0	7.1	2.7	16.0	6.9	2.6

5 Conclusion

This paper presents GeoGT, a graph transformer framework for predicting molecular ground-state conformations. The method explicitly incorporates bond angle variance and molecular curvature as geometric priors to jointly regularize local atomic geometry and global molecular topology. By embedding these geometric constraints into the attention mechanism and positional encoding, GeoGT preserves local geometric rationality while maintaining the physical plausibility of the overall molecular shape during end-to-end learning. These results demonstrate that explicit modeling of geometric structure can significantly improve the quality and physical realism of predicted conformations.

AI use statement

We used generative AI tools to assist with language polishing and translation of the manuscript.

Ethics statement

This work investigates molecular ground-state conformation prediction using the publicly available QM9 and Molecule3D datasets. It does not involve experiments on human participants or animals. To the best of our knowledge, this study raises no specific ethical concerns.

Reproducibility statement

To facilitate reproduction of our results, we provide a link to the source code in the abstract.

Acknowledgments

We sincerely thank everyone who provided assistance and support throughout this research and the preparation of the manuscript.

References

Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific data*, 9(1):185, 2022.

Sarp Aykent and Tian Xia. Gotennet: Rethinking efficient 3d equivariant graph neural networks. In *International Conference on Learning Representations*, volume 2025, pp. 40622–40646, 2025.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *ICLR*, 2021.

Yoni Choukroun and Lior Wolf. Geometric transformer for end-to-end molecule properties prediction. *arXiv preprint arXiv:2110.13721*, 2021.

Johann G Danzl, Manfred J Mark, Elmar Haller, Mattias Gustavsson, Russell Hart, Jesus Aldegunde, Jeremy M Hutson, and Hanns-Christoph Nägerl. An ultracold high-density sample of rovibronic ground-state molecules in an optical lattice. *Nature Physics*, 6(4):265–270, 2010.

Marco De Vivo, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli. Role of molecular dynamics and related methods in drug discovery. *Journal of medicinal chemistry*, 59(9):4035–4061, 2016.

Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*, 2020.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *Journal of Machine Learning Research*, 24(43):1–48, 2023.

Zhiguang Fan, Yuedong Yang, Mingyuan Xu, and Hongming Chen. Ec-conf: A ultra-fast diffusion model for molecular conformation generation with equivariant consistency. *Journal of Cheminformatics*, 16(1):107, 2024.

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

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33:1970–1981, 2020.

Octavian Ganea, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *Advances in Neural Information Processing Systems*, 34:13757–13769, 2021.

Thomas A Halgren. Merck molecular force field. v. extension of mmff94 using experimental data, additional computational data, and empirical rules. *Journal of Computational Chemistry*, 17(5-6): 616–641, 1996.

Timothy F Havel. Distance geometry: Theory, algorithms, and chemical applications. *Encyclopedia of Computational Chemistry*, 120(723-742):5–4, 1998.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265*, 2019.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in neural information processing systems*, 35:24240–24253, 2022.

Jinwoo Kim, Dat Nguyen, Seonwoo Min, Sungjun Cho, Moontae Lee, Honglak Lee, and Seunghoon Hong. Pure transformers are powerful graph learners. *Advances in Neural Information Processing Systems*, 35:14582–14595, 2022.

Taewon Kim, Hyunjin Seo, Sungsoo Ahn, and Eunho Yang. Rebind: Enhancing ground-state molecular conformation prediction via force-based graph rewiring. In *ICLR 2025*. International Conference on Learning Representations (ICLR), 2025.

István Kolossváry and Wayne C Guida. Low mode search. an efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides. *Journal of the American Chemical Society*, 118(21):5011–5019, 1996.

Greg Landrum et al. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 8(31.10):5281, 2013.

Xinze Li, Penglei Wang, Tianfan Fu, Wenhao Gao, Chengtao Li, Leilei Shi, and Junhong Liu. Autodiff: autoregressive diffusion modeling for structure-based drug design. *arXiv preprint arXiv:2404.02003*, 2024.

Meng Liu, Cong Fu, Xuan Zhang, Limei Wang, Yaochen Xie, Hao Yuan, Youzhi Luo, Zhao Xu, Shenglong Xu, and Shuiwang Ji. Fast quantum property prediction via deeper 2d and 3d graph networks. *arXiv preprint arXiv:2106.08551*, 2021.

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 (ICLR)*, 2022.

Shengjie Luo, Tianlang Chen, Yixian Xu, Shuxin Zheng, Tie-Yan Liu, Liwei Wang, and Di He. One transformer can understand both 2d & 3d molecular data. *ICLR*, 2022.

Shitong Luo, Chence Shi, Minkai Xu, and Jian Tang. Predicting molecular conformation via dynamic graph score matching. *Advances in neural information processing systems*, 34:19784–19795, 2021.

Elman Mansimov, Omar Mahmood, Seokho Kang, and Kyunghyun Cho. Molecular geometry prediction using a deep generative graph neural network. *Scientific reports*, 9(1):20381, 2019.

Łukasz Maziarka, Tomasz Danel, Sławomir Mucha, Krzysztof Rataj, Jacek Tabor, and Stanisław Jastrzębski. Molecule attention transformer. *arXiv preprint arXiv:2002.08264*, 2020.

Łukasz Maziarka, Dawid Majchrowski, Tomasz Danel, Piotr Gaiński, Jacek Tabor, Igor Podolak, Paweł Morkisz, and Stanisław Jastrzębski. Relative molecule self-attention transformer. *Journal of Cheminformatics*, 16(1):3, 2024.

Erxue Min, Runfa Chen, Yatao Bian, Tingyang Xu, Kangfei Zhao, Wenbing Huang, Peilin Zhao, Junzhou Huang, Sophia Ananiadou, and Yu Rong. Transformer for graphs: An overview from architecture perspective. *arXiv preprint arXiv:2202.08455*, 2022.

Fariborz Mohamadi, Nigel GJ Richards, Wayne C Guida, Rob Liskamp, Mark Lipton, Craig Caufield, George Chang, Thomas Hendrickson, and W Clark Still. Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of Computational Chemistry*, 11(4):440–467, 1990.

Chr Møller and Milton S Plesset. Note on an approximation treatment for many-electron systems. *Physical review*, 46(7):618, 1934.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

S Ospelkaus, K-K Ni, G Quéméner, B Neyenhuis, D Wang, MHG De Miranda, JL Bohn, J Ye, and DS Jin. Controlling the hyperfine state of rovibronic ground-state polar molecules. *Physical review letters*, 104(3):030402, 2010.

Robert G Parr, Shridhar R Gadre, and Libero J Bartolotti. Local density functional theory of atoms and molecules. *Proceedings of the National Academy of Sciences*, 76(6):2522–2526, 1979.

Qizhi Pei, Kaiyuan Gao, Lijun Wu, Jinhua Zhu, Yingce Xia, Shufang Xie, Tao Qin, Kun He, Tie-Yan Liu, and Rui Yan. Fabind: Fast and accurate protein-ligand binding. *Advances in Neural Information Processing Systems*, 36:55963–55980, 2023.

George D Purvis III and Rodney J Bartlett. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of chemical physics*, 76(4):1910–1918, 1982.

Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems*, 35:14501–14515, 2022.

Anthony K Rappé, Carla J Casewit, KS Colwell, William A Goddard III, and W Mason Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society*, 114(25):10024–10035, 1992.

Sereina Riniker and Gregory A Landrum. Better informed distance geometry: using what we know to improve conformation generation. *Journal of chemical information and modeling*, 55(12): 2562–2574, 2015.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International conference on machine learning*, pp. 9377–9388. PMLR, 2021.

Chence Shi, Shitong Luo, Minkai Xu, and Jian Tang. Learning gradient fields for molecular conformation generation. In *International conference on machine learning*, pp. 9558–9568. PMLR, 2021.

Philipp Thölke and Gianni De Fabritiis. Torchmd-net: equivariant transformers for neural network based molecular potentials. *arXiv preprint arXiv:2202.02541*, 2022.

Michael Tinkham and Malcolm Woodrow Pershing Strandberg. Theory of the fine structure of the molecular oxygen ground state. *Physical Review*, 97(4):937, 1955.

Debby D Wang, Wenhui Wu, and Ran Wang. Structure-based, deep-learning models for protein-ligand binding affinity prediction. *Journal of Cheminformatics*, 16(1):2, 2024.

Fang Wu, Qiang Zhang, Dragomir Radev, Jiyu Cui, Wen Zhang, Huabin Xing, Ningyu Zhang, and Huajun Chen. 3d-transformer: Molecular representation with transformer in 3d space. 2021.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

Guikun Xu, Yongquan Jiang, PengChuan Lei, Yan Yang, and Jim Chen. Gtmgc: Using graph transformer to predict molecule's ground-state conformation. In *The Twelfth International Conference on Learning Representations*, 2023.

Minkai Xu, Shitong Luo, Yoshua Bengio, Jian Peng, and Jian Tang. Learning neural generative dynamics for molecular conformation generation. *ICLR*, 2021a.

Minkai Xu, Wujie Wang, Shitong Luo, Chence Shi, Yoshua Bengio, Rafael Gomez-Bombarelli, and Jian Tang. An end-to-end framework for molecular conformation generation via bilevel programming. In *International conference on machine learning*, pp. 11537–11547. PMLR, 2021b.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *ICLR*, 2022.

Zhao Xu, Youzhi Luo, Xuan Zhang, Xinyi Xu, Yaochen Xie, Meng Liu, Kaleb Dickerson, Cheng Deng, Maho Nakata, and Shuiwang Ji. Molecule3d: A benchmark for predicting 3d geometries from molecular graphs. *arXiv preprint arXiv:2110.01717*, 2021c.

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.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23174–23184, 2023.

Sheheryar Zaidi, Michael Schaarschmidt, James Martens, Hyunjik Kim, Yee Whye Teh, Alvaro Sanchez-Gonzalez, Peter Battaglia, Razvan Pascanu, and Jonathan Godwin. Pre-training via denoising for molecular property prediction. *ICLR*, 2023.

Haotian Zhang, Shengming Li, Jintu Zhang, Zhe Wang, Jike Wang, Dejun Jiang, Zhiwen Bian, Yixue Zhang, Yafeng Deng, Jianfei Song, et al. Sdegen: learning to evolve molecular conformations from thermodynamic noise for conformation generation. *Chemical Science*, 14(6): 1557–1568, 2023.

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 Notation

For clarity, Table 2 summarizes the principal notation used in GeoGT. The symbols are grouped according to the main components of the proposed framework.

B Preliminary

B.1 Laplace Positional Encoding

Consider a graph structure with n nodes. Its adjacency matrix is denoted as $A \in \mathbb{R}^{n\times n}$, where $A_{ij} = 1$ indicates an edge exists between node i and node j, and $A_{ij} = 0$ otherwise. The degree matrix $Deg \in \mathbb{R}^{n\times n}$ is a diagonal matrix defined as:

$$Deg_{ii} = \sum_{j=1}^{n} A_{ij}. \tag{13}$$

Table 2: Summary of principal notation used in GeoGT.

Symbol	Description	Symbol	Description
Molecular graph and conformations			
$\mathcal{G}$	2D molecular graph.	$\mathcal{V}, \mathcal{E}$	Atom set and bond set.
N, M	Numbers of atoms and bonds.	$\mathbf{A}$	Adjacency matrix.
$\mathcal{N}(i)$	Neighbor set of atom i.	$\mathbf{C}$	Ground-truth conformation.
$\mathbf{C}_{\text{im}}$	Intermediate conformation.	$\hat{\mathbf{C}}$	Predicted conformation.
$\mathbf{D}$	Ground-truth distance matrix.	$\hat{\mathbf{D}}$	Predicted distance matrix.
$\mathbf{D}_{\text{im}}$	Intermediate distance matrix.	$\mathbf{c}_i$	Coordinate vector of atom i.
Representations and spectral encoding			
$\mathbf{E}$	Input molecular embedding.	$\mathbf{X}$	Encoder output representation.
$\mathbf{H}$	Curvature-aware decoder input.	d	Hidden feature dimension.
$\mathbf{L}$	Graph Laplacian.	$\mathbf{L}_{\text{sym}}$	Normalized graph Laplacian.
λ_k	k-th Laplacian eigenvalue.	$\mathbf{u}_k$	k-th Laplacian eigenvector.
$\tilde{\mathbf{u}}_k$	Modulated eigenvector.	w_k	Adaptive spectral weight.
γ	Frequency modulation factor.	$f_w(\cdot)$	Conformation prediction network.
Geometric priors			
θ_{jik}	Bond angle centered at atom i.	Θ_i	Bond angle set of atom i.
$\text{Var}(\Theta_i)$	Bond angle variance.	$\tilde{\sigma}_i$	Normalized bond angle variance.
$\mathbf{e}_i^a$	Bond angle variance embedding.	mc_i	Discrete molecular curvature.
$\mathbf{T}_i$	Curvature transform.	$\mathbf{h}_i$	Curvature-aware atom representation.
Geometric residual self-attention			
$\mathbf{S}_l$	Basic attention score.	$\mathbf{A}^{\text{Geo}}$	Geometry-enhanced adjacency.
$\mathbf{B}_l^{\text{Geo}}$	Geometric residual bias.	$\hat{\mathbf{S}}_l$	Geometry-enhanced attention score.
$\mathbf{O}$	Geometric attention output.	$\mathbf{G}$	Decoder output representation.
Training objective			
$\mathcal{L}$	Training loss.	ϵ	Numerical stability constant.

The symmetric normalized Laplacian matrix is computed as:

$$L_{\text{sym}} = I - Deg^{-1/2} A Deg^{-1/2}, \tag{14}$$

where $I \in \mathbb{R}^{n \times n}$ is the identity matrix, $Deg^{-1/2}$ is the inverse square root of the degree matrix (defined only for nodes with non-zero degree, with diagonal elements $Deg_{ii}^{-1/2}$). $L_{\text{sym}} \in \mathbb{R}^{n \times n}$ is the Laplacian matrix.

Laplacian positional encoding utilizes the eigenvectors of the Laplacian matrix to encode positional relationships between nodes. The eigenvalue decomposition of L_{sym} is performed as:

$$L_{\text{sym}} = U \Lambda U^{\top}, \tag{15}$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^{n \times n}$ is the diagonal matrix of eigenvalues (with eigenvalues sorted in ascending order: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$), $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ is the matrix of eigenvectors.

The input H is obtained by element-wise addition of the eigenvectors matrix U and the feature matrix E:

$$H = E + U. \tag{16}$$

B.2 GRAPH TRANSFORMER

Graph Transformer extends the standard Transformer architecture to graph-structured data by incorporating graph topology into the self-attention mechanism. Compared with conventional Graph Neural Networks (GNNs), which mainly aggregate information from local neighborhoods, Graph Transformer can model both local connectivity and long-range dependencies among nodes. This property makes it particularly suitable for molecular representation learning, where both local chemical bonds and global molecular interactions are important.

Given a molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $N = |\mathcal{V}|$ atoms, let $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N]^\top \in \mathbb{R}^{N \times d}$ denote the node representation matrix, where $\mathbf{h}_i \in \mathbb{R}^d$ is the representation of the i-th atom. Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ denote the adjacency matrix encoding chemical bonds. Since the vanilla self-attention mechanism is permutation-invariant and does not explicitly encode graph topology, Graph Transformer usually introduces structural information, such as adjacency relations, edge types, shortest-path distances, or spatial distances, into the attention computation.

For the l-th attention head, the query, key, and value matrices are computed as

$$\mathbf{Q}_l = \mathbf{H}\mathbf{W}_l^Q, \quad \mathbf{K}_l = \mathbf{H}\mathbf{W}_l^K, \quad \mathbf{V}_l = \mathbf{H}\mathbf{W}_l^V, \tag{17}$$

where $\mathbf{W}_l^Q, \mathbf{W}_l^K, \mathbf{W}_l^V \in \mathbb{R}^{d \times d_k}$ are learnable projection matrices. The i-th rows of $\mathbf{Q}_l$, $\mathbf{K}_l$, and $\mathbf{V}_l$ correspond to the node-level query, key, and value vectors $\mathbf{q}_i^{(l)}$, $\mathbf{k}_i^{(l)}$, and $\mathbf{v}_i^{(l)}$, respectively.

To make the attention mechanism aware of graph structure, a structural bias matrix $\mathbf{B} = [b_{ij}] \in \mathbb{R}^{N \times N}$ is added to the scaled dot-product attention. The attention score from node i to node j in the l-th head is therefore given by

$$e_{ij}^{(l)} = \frac{\mathbf{q}_i^{(l)} (\mathbf{k}_j^{(l)})^\top}{\sqrt{d_k}} + b_{ij}, \qquad \alpha_{ij}^{(l)} = \frac{\exp(e_{ij}^{(l)})}{\sum_{m=1}^{N} \exp(e_{im}^{(l)})}. \tag{18}$$

Here, b_{ij} encodes the structural relation between nodes i and j. For example, it can be determined by whether two atoms are adjacent, by the type of chemical bond between them, or by a more general distance-based structural encoding.

Equivalently, the above node-level attention can be written in matrix form as

$$\mathrm{Head}_l = \mathrm{Softmax}\left(\frac{\mathbf{Q}_l \mathbf{K}_l^\top}{\sqrt{d_k}} + \mathbf{B}\right)\mathbf{V}_l, \tag{19}$$

where the softmax operation is applied row-wise. The outputs of all attention heads are then concatenated and projected to obtain the multi-head self-attention output:

$$\mathrm{MHSA}(\mathbf{H}) = \mathrm{Concat}\left(\mathrm{Head}_1, \ldots, \mathrm{Head}_h\right)\mathbf{W}^O, \tag{20}$$

where $\mathbf{W}^O \in \mathbb{R}^{hd_k \times d}$ is the output projection matrix.

A Graph Transformer layer further applies residual connections, layer normalization, and a feed-forward network (FFN) to update the node representations:

$$\widetilde{\mathbf{H}} = \mathrm{LayerNorm}\left(\mathbf{H} + \mathrm{MHSA}(\mathbf{H})\right), \mathrm{GraphTransformerLayer}(\mathbf{H}) = \mathrm{LayerNorm}\left(\widetilde{\mathbf{H}} + \mathrm{FFN}(\widetilde{\mathbf{H}})\right). \tag{21}$$

By stacking multiple Graph Transformer layers, the model can progressively refine atomic representations by integrating both local topological information and long-range graph dependencies. This provides a flexible foundation for molecular representation learning and downstream molecular conformation prediction.

C PROOF OF THEOREM 1

We provide a rigorous proof showing that the eigenvalues of the graph Laplacian characterize the frequency of the corresponding eigenvectors over the graph. Consider an undirected, connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n$. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be the symmetric adjacency matrix,

$\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ be the degree matrix with $d_i = \sum_{j=1}^{n} A_{ij}$, and let $\mathbf{L} = \mathbf{D} - \mathbf{A}$ be the combinatorial graph Laplacian.

For any graph signal $\mathbf{f} \in \mathbb{R}^n$, the Laplacian quadratic form satisfies

$$\begin{aligned}
\mathbf{f}^\top \mathbf{L}\mathbf{f} &= \mathbf{f}^\top \mathbf{D}\mathbf{f} - \mathbf{f}^\top \mathbf{A}\mathbf{f} \\
&= \sum_{i=1}^{n} d_i f_i^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij} f_i f_j \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\left(f_i^2 + f_j^2 - 2 f_i f_j\right) \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(f_i - f_j)^2 \\
&\geq 0.
\end{aligned} \tag{22}$$

Therefore, $\mathbf{L}$ is symmetric positive semidefinite. Since the graph is connected, the null space of $\mathbf{L}$ is $\ker(\mathbf{L}) = \mathrm{span}\{\mathbf{1}\}$. Thus, the eigenvalues of $\mathbf{L}$ can be ordered as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$, and there exists an orthonormal eigenbasis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, such that $\mathbf{L}\mathbf{u}_k = \lambda_k \mathbf{u}_k, \mathbf{u}_i^\top \mathbf{u}_j = \delta_{ij}$. Equivalently, the spectral decomposition of the graph Laplacian is

$$\mathbf{L} = \sum_{k=1}^{n} \lambda_k \mathbf{u}_k \mathbf{u}_k^\top. \tag{23}$$

The first eigenvector can be chosen as the normalized constant vector $\mathbf{u}_1 = \frac{1}{\sqrt{n}}\mathbf{1}, \lambda_1 = 0$.

We now show that λ_k measures the oscillation frequency of $\mathbf{u}_k$ over the graph. The Rayleigh quotient of a nonzero graph signal $\mathbf{f}$ is defined as

$$\mathcal{R}_{\mathbf{L}}(\mathbf{f}) = \frac{\mathbf{f}^\top \mathbf{L}\mathbf{f}}{\mathbf{f}^\top \mathbf{f}}. \tag{24}$$

Using the quadratic form above, this can be rewritten as

$$\mathcal{R}_{\mathbf{L}}(\mathbf{f}) = \frac{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(f_i - f_j)^2}{\sum_{i=1}^{n} f_i^2}. \tag{25}$$

Hence, the Rayleigh quotient is precisely the normalized Dirichlet energy of the graph signal. It measures how strongly the signal varies across adjacent vertices.

For an eigenvector $\mathbf{u}_k$, we have

$$\mathcal{R}_{\mathbf{L}}(\mathbf{u}_k) = \frac{\mathbf{u}_k^\top \mathbf{L}\mathbf{u}_k}{\mathbf{u}_k^\top \mathbf{u}_k} = \frac{\mathbf{u}_k^\top \lambda_k \mathbf{u}_k}{\mathbf{u}_k^\top \mathbf{u}_k} = \lambda_k. \tag{26}$$

Since $\|\mathbf{u}_k\|_2 = 1$, this further gives

$$\lambda_k = \mathbf{u}_k^\top \mathbf{L}\mathbf{u}_k = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(u_{k,i} - u_{k,j})^2. \tag{27}$$

Therefore, λ_k is exactly the total squared variation of $\mathbf{u}_k$ along graph edges.

To express this result in operator form, fix an arbitrary orientation of the edge set and define the incidence matrix $\mathbf{B} \in \mathbb{R}^{|\mathcal{E}| \times n}$ by $B_{e,i} = 1, B_{e,j} = -1$, for each oriented edge $e = (i, j)$, with all other entries equal to zero. Then $\mathbf{L} = \mathbf{B}^\top \mathbf{B}$. Consequently,

$$\|\mathbf{B}\mathbf{f}\|_2^2 = \mathbf{f}^\top \mathbf{B}^\top \mathbf{B}\mathbf{f} = \mathbf{f}^\top \mathbf{L}\mathbf{f}. \tag{28}$$

For the Laplacian eigenvector $\mathbf{u}_k$, this yields

$$\|\mathbf{B}\mathbf{u}_k\|_2^2 = \mathbf{u}_k^\top \mathbf{L}\mathbf{u}_k = \lambda_k. \tag{29}$$

Thus, $\|\mathbf{B}\mathbf{u}_k\|_2 = \sqrt{\lambda_k}$. This shows that $\sqrt{\lambda_k}$ is the graph-gradient magnitude of $\mathbf{u}_k$. Therefore, λ_k provides a rigorous measure of the graph frequency of the eigenvector.

We first consider the low-frequency case. If $\lambda_k \approx 0$, then

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(u_{k,i} - u_{k,j})^2 \approx 0. \tag{30}$$

Since each term in the summation is nonnegative, the values $u_{k,i}$ and $u_{k,j}$ must be close for most adjacent vertices $(i,j) \in \mathcal{E}$. More formally, for any threshold $\tau > 0$, define $\mathcal{E}_\tau = \{(i,j) \in \mathcal{E} : |u_{k,i} - u_{k,j}| \geq \tau\}$.

Then

$$\begin{aligned}\lambda_k &= \sum_{(i,j)\in\mathcal{E}} (u_{k,i} - u_{k,j})^2 \\ &\geq \sum_{(i,j)\in\mathcal{E}_\tau} (u_{k,i} - u_{k,j})^2 \\ &\geq \tau^2 |\mathcal{E}_\tau|.\end{aligned}$$

Hence, $|\mathcal{E}_\tau| \leq \frac{\lambda_k}{\tau^2}$.

This inequality implies that when λ_k is small, the number of edges across which $\mathbf{u}_k$ changes sharply must also be small. Therefore, low-frequency eigenvectors vary smoothly over the graph.

The path-wise behavior of low-frequency eigenvectors further supports their global nature. For any two vertices $p, q \in \mathcal{V}$, let $P_{p\to q} = (i_0 = p, i_1, \ldots, i_\ell = q)$ be a path connecting p and q. Then

$$u_{k,p} - u_{k,q} = \sum_{r=0}^{\ell-1} \left(u_{k,i_r} - u_{k,i_{r+1}}\right). \tag{31}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}|u_{k,p} - u_{k,q}|^2 &\leq \ell \sum_{r=0}^{\ell-1} \left(u_{k,i_r} - u_{k,i_{r+1}}\right)^2 \\ &\leq \ell \sum_{(i,j)\in\mathcal{E}} (u_{k,i} - u_{k,j})^2 \\ &= \ell\lambda_k.\end{aligned}$$

Thus, if λ_k is small, the variation of $\mathbf{u}_k$ is also small along graph paths. This means that low-frequency eigenvectors encode slowly varying, large-scale structural patterns of the graph.

The variational characterization of Laplacian eigenvectors gives a second rigorous interpretation. By the Courant-Fischer theorem,

$$\lambda_k = \min_{\substack{\mathcal{S}\subset\mathbb{R}^n \\ \dim(\mathcal{S})=k}} \max_{\substack{\mathbf{f}\in\mathcal{S} \\ \mathbf{f}\neq\mathbf{0}}} \frac{\mathbf{f}^\top\mathbf{L}\mathbf{f}}{\mathbf{f}^\top\mathbf{f}}. \tag{32}$$

Equivalently, the k-th eigenvector satisfies

$$\mathbf{u}_k \in \operatorname*{arg\,min}_{\substack{\mathbf{f}\perp\mathrm{span}\{\mathbf{u}_1,\ldots,\mathbf{u}_{k-1}\} \\ \|\mathbf{f}\|_2=1}} \mathbf{f}^\top\mathbf{L}\mathbf{f}. \tag{33}$$

Substituting the Laplacian quadratic form gives

$$\mathbf{u}_k \in \operatorname*{arg\,min}_{\substack{\mathbf{f}\perp\mathrm{span}\{\mathbf{u}_1,\ldots,\mathbf{u}_{k-1}\} \\ \|\mathbf{f}\|_2=1}} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(f_i - f_j)^2. \tag{34}$$

Thus, $\mathbf{u}_1$ is the constant mode, $\mathbf{u}_2$ is the smoothest nonconstant mode, and higher-indexed eigenvectors are successively less smooth directions under orthogonality constraints. This is exactly the graph analogue of increasing Fourier frequencies in the Euclidean domain.

We next consider the high-frequency case. Since

$$\lambda_k = \sum_{(i,j)\in\mathcal{E}} (u_{k,i} - u_{k,j})^2, \tag{35}$$

a large λ_k implies large graph-gradient energy. In particular,

$$\max_{(i,j)\in\mathcal{E}} |u_{k,i} - u_{k,j}|^2 \geq \frac{1}{|\mathcal{E}|} \sum_{(i,j)\in\mathcal{E}} (u_{k,i} - u_{k,j})^2 = \frac{\lambda_k}{|\mathcal{E}|}. \tag{36}$$

Therefore,

$$\max_{(i,j)\in\mathcal{E}} |u_{k,i} - u_{k,j}| \geq \sqrt{\frac{\lambda_k}{|\mathcal{E}|}}. \tag{37}$$

This inequality shows that high-frequency eigenvectors must exhibit pronounced edge-wise variations. In particular, for the largest eigenvalue,

$$\lambda_n = \max_{\mathbf{f}\neq\mathbf{0}} \frac{\mathbf{f}^\top \mathbf{L}\mathbf{f}}{\mathbf{f}^\top \mathbf{f}} = \max_{\|\mathbf{f}\|_2=1} \frac{1}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(f_i - f_j)^2. \tag{38}$$

Hence, $\mathbf{u}_n$ is the unit-norm graph signal with maximal edge-wise oscillation. This establishes the sense in which large eigenvalues correspond to high-frequency graph components.

The same conclusion can also be obtained from graph Fourier analysis. Since $\{\mathbf{u}_k\}_{k=1}^n$ forms an orthonormal basis of $\mathbb{R}^n$, any graph signal $\mathbf{f}$ can be decomposed as

$$\mathbf{f} = \sum_{k=1}^{n} \alpha_k \mathbf{u}_k, \qquad \alpha_k = \mathbf{u}_k^\top \mathbf{f}. \tag{39}$$

By Parseval's identity,

$$\|\mathbf{f}\|_2^2 = \sum_{k=1}^{n} |\alpha_k|^2. \tag{40}$$

Moreover,

$$\begin{aligned}
\mathbf{f}^\top \mathbf{L}\mathbf{f} &= \left(\sum_{k=1}^{n} \alpha_k \mathbf{u}_k\right)^\top \mathbf{L} \left(\sum_{\ell=1}^{n} \alpha_\ell \mathbf{u}_\ell\right) \\
&= \sum_{k=1}^{n}\sum_{\ell=1}^{n} \alpha_k \alpha_\ell \mathbf{u}_k^\top \mathbf{L}\mathbf{u}_\ell \\
&= \sum_{k=1}^{n}\sum_{\ell=1}^{n} \alpha_k \alpha_\ell \lambda_\ell \mathbf{u}_k^\top \mathbf{u}_\ell \\
&= \sum_{k=1}^{n} \lambda_k |\alpha_k|^2.
\end{aligned}$$

Therefore, components associated with small eigenvalues contribute little to the graph Dirichlet energy, while components associated with large eigenvalues contribute strongly to it.

Let $\mathcal{U}_{\leq r} = \mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, $\mathcal{U}_{>r} = \mathrm{span}\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$. For the low-frequency projection

$$\mathbf{f}_{\leq r} = \sum_{k=1}^{r} \alpha_k \mathbf{u}_k, \tag{41}$$

we have

$$\frac{\mathbf{f}_{\leq r}^\top \mathbf{L}\mathbf{f}_{\leq r}}{\|\mathbf{f}_{\leq r}\|_2^2} = \frac{\sum_{k=1}^{r} \lambda_k |\alpha_k|^2}{\sum_{k=1}^{r} |\alpha_k|^2} \leq \lambda_r. \tag{42}$$

For the high-frequency projection

$$\mathbf{f}_{>r} = \sum_{k=r+1}^{n} \alpha_k \mathbf{u}_k, \tag{43}$$

we have

$$\frac{\mathbf{f}_{>r}^\top \mathbf{L} \mathbf{f}_{>r}}{\|\mathbf{f}_{>r}\|_2^2} = \frac{\sum_{k=r+1}^{n} \lambda_k |\alpha_k|^2}{\sum_{k=r+1}^{n} |\alpha_k|^2} \geq \lambda_{r+1}. \tag{44}$$

Thus, low-frequency subspaces consist of smooth graph signals, whereas high-frequency subspaces consist of signals with strong local variations.

Finally, we give a diffusion-based interpretation. Consider the graph heat equation

$$\frac{\partial \mathbf{f}(t)}{\partial t} = -\mathbf{L}\mathbf{f}(t), \qquad \mathbf{f}(0) = \mathbf{f}_0. \tag{45}$$

Its solution is

$$\mathbf{f}(t) = e^{-t\mathbf{L}} \mathbf{f}_0. \tag{46}$$

Using the spectral decomposition of $\mathbf{L}$, we obtain

$$e^{-t\mathbf{L}} = \sum_{k=1}^{n} e^{-t\lambda_k} \mathbf{u}_k \mathbf{u}_k^\top. \tag{47}$$

If

$$\mathbf{f}_0 = \sum_{k=1}^{n} \alpha_k(0) \mathbf{u}_k, \tag{48}$$

then

$$\mathbf{f}(t) = \sum_{k=1}^{n} e^{-t\lambda_k} \alpha_k(0) \mathbf{u}_k. \tag{49}$$

Since $e^{-t\lambda_k}$ decays faster when λ_k is larger, high-frequency components vanish rapidly under diffusion, while low-frequency components persist for a longer time. More precisely, for $r < n$, let $P_{\leq r}$ and $P_{>r}$ denote the orthogonal projections onto $\mathcal{U}_{\leq r}$ and $\mathcal{U}_{>r}$, respectively. Then

$$\|P_{>r}\mathbf{f}(t)\|_2^2 = \sum_{k=r+1}^{n} e^{-2t\lambda_k} |\alpha_k(0)|^2 \leq e^{-2t\lambda_{r+1}} \|P_{>r}\mathbf{f}_0\|_2^2, \tag{50}$$

whereas

$$\|P_{\leq r}\mathbf{f}(t)\|_2^2 = \sum_{k=1}^{r} e^{-2t\lambda_k} |\alpha_k(0)|^2 \geq e^{-2t\lambda_r} \|P_{\leq r}\mathbf{f}_0\|_2^2. \tag{51}$$

Therefore,

$$\frac{\|P_{>r}\mathbf{f}(t)\|_2^2}{\|P_{\leq r}\mathbf{f}(t)\|_2^2} \leq e^{-2t(\lambda_{r+1}-\lambda_r)} \frac{\|P_{>r}\mathbf{f}_0\|_2^2}{\|P_{\leq r}\mathbf{f}_0\|_2^2}. \tag{52}$$

This inequality shows that graph diffusion suppresses high-frequency components exponentially faster than low-frequency components. Thus, low-frequency eigenvectors correspond to persistent, smooth, global patterns, while high-frequency eigenvectors correspond to transient, rapidly varying, local patterns.

Combining the Dirichlet-energy identity, the incidence-matrix formulation, the Courant-Fischer variational characterization, the graph Fourier decomposition, and the heat-kernel interpretation, we conclude that the eigenvalue λ_k is a rigorous measure of the frequency of the eigenvector $\mathbf{u}_k$ over the graph. When $\lambda_k \approx 0$, $\mathbf{u}_k$ has small edge-wise variation and therefore varies smoothly over the graph, encoding global structural information. Conversely, when λ_k is large, $\mathbf{u}_k$ has large graph-gradient energy and therefore exhibits strong local oscillations, capturing local structural details. This proves Theorem 1. □

D Complexity Analysis And Pseudocode Algorithm

The time complexity and space complexity of GeoGT primarily depend on the number of atoms N and the feature dimension d in the molecule.

Time complexity. In terms of time complexity, both the multi-head self-attention mechanism in the encoder and the sub-geometry multi-head attention mechanism in the decoder exhibit a typical complexity of $O(N^2d)$, due to the need to compute attention scores for all atom pairs. The FFN has a time complexity of $O(Nd^2)$. During the conformation generation and geometric information extraction stages, the computation of bond angles and curvature relies on the local coordination number of each atom; in real molecules the number of bonds per atom is limited and can be treated as a constant, hence this step contributes $O(N)$ time complexity. However, computing the variance of all molecular bond angles and performing singular value decomposition involve operations on an $N \times d$ matrix, with complexity $O(Nd^2)$. Overall, the model's total time complexity is dominated by the attention mechanisms and matrix decomposition, yielding $O(N^2d + Nd^2)$.

Space complexity. Regarding space complexity, the model needs to store the attention score matrices $S_i \in \mathbb{R}^{N\times N}$, which require $O(N^2)$ memory. At the same time, it maintains atom embeddings $E, X, H, G \in \mathbb{R}^{N\times d}$, 3D coordinates $C_{im}, \hat{C} \in \mathbb{R}^{N\times 3}$, and geometric information (bond angle variance embeddings and curvature vectors), collectively occupying $O(Nd + N)$ space. Therefore, the overall space complexity is $O(N^2 + Nd)$, where the $O(N^2)$ term arises from storing the attention matrices—a characteristic feature of Transformer-based models. The pseudocode of the GeoGT algorithm flow is shown in Algorithm 1.

Algorithm 1: GeoGT

Input: Molecular representation E, Adjacency matrix A.
Output: Predicted molecular ground-state conformation C.

begin
- The input E is passed through encoder to obtain the embedding representation X.
- Using $f_w(\cdot)$, generate the initial intermediate conformation C_{im} from X.
- Based on C_{im}, compute bond angle variance $\tilde{\sigma}_i$ and its embedding representation e_i^a via Equations (10)-(13).
- In the Laplacian positional encoding of the decoder, modulate the eigenvectors via Equation (14) and Theorem 1 to obtain $\tilde{u}$.
- Compute the molecular curvature mc_i via Equation (16).
- Integrate mc_i with $\tilde{u}$ through Equations (17)-(18) to complete the positional encoding of X in the decoder, yielding H.
- In decoder's MHSA, incorporate the geometric constraint B_l^{Geo} via Equations (19)-(21).
- After passing through the entire decoder, obtain the output embedding G.
- Using $f_w(\cdot)$, obtain the predicted molecular ground-state conformation $\widehat{C}$ from G and set $C = \widehat{C}$ by default.

end

return C.

E Experimental Setup

E.1 Dataset Descriptions

To evaluate the performance of the proposed method, experiments were conducted on two widely recognized benchmark datasets for molecular systems, QM9 (Ramakrishnan et al., 2014) and Molecule3D (Xu et al., 2021c).

QM9. A classic benchmark dataset in quantum chemistry. Curated from the GDB-17 chemical space, it contains precisely 130,831 stable small organic molecules, each possessing a maximum of 9 heavy atoms. The dataset provides density functional theory (DFT)-computed three-dimensional ground-state conformations alongside diverse quantum chemical properties. Owing to its moderate size, well-balanced molecular structural diversity, and reliable computational accuracy, QM9

remains the gold standard dataset for evaluating the performance of molecular models. Detailed statistics of the QM9 dataset are reported in Table 3 (Axelrod & Gomez-Bombarelli, 2022).

Molecule3D. A large-scale, multi-task, multi-modal molecular dataset. It comprises approximately 3.9 million molecules along with their corresponding ground-state conformations. The dataset further integrates two-dimensional molecular graphs and multiple quantum chemical properties, providing exceptionally rich and critical resources for training highly generalizable and performant molecular representation learning models. In this study, we adopt two dataset partitioning strategies: random split and scaffold split. The detailed dataset statistics and split information are summarized in Table 4 (Xu et al., 2021c).

Table 3: Statistics of the QM9 dataset.

Statistic	**Mean**	**Std. dev.**	**Max.**
Number of atoms	18.0	3.0	29
Number of heavy atoms	8.8	0.51	9
Molecular weight (amu)	122.7	7.6	152.0
Number of rotatable bonds	2.2	1.6	8
Stereochemistry	95,734	–	–
Total molecules	130,831	–	–

Table 4: Statistics of the Molecule3D dataset.

Statistic	**Value**
Total molecules	3,899,647
Molecules in training set	2,339,788
Molecules in validation set	779,929
Molecules in test set	779,930
Split ratio	6:2:2
Split type	Random / Scaffold
Average atoms per molecule	29.11
Average heavy atoms per molecule	14.08
Average bonds per molecule	29.53

E.2 Evaluation Metrics

To quantify the discrepancies between the predicted molecular ground-state conformation and the true ground-state conformation, this study employs the following three evaluation metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Root Mean Square Deviation (RMSD).

The accuracy of the model in predicting relative interatomic positioning is assessed by calculating the differences between the predicted distance matrix $\hat{D}$ and the true distance matrix D. The specific metrics are defined as follows:

$$\text{D-MAE}(\hat{D}, D) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\hat{D}_{ij} - D_{ij}\right|, \text{D-RMSE}(\hat{D}, D) = \sqrt{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{D}_{ij} - D_{ij})^2} \tag{53}$$

Since the predicted conformation $\hat{G}$ and the true conformation G may reside in different coordinate systems, direct comparison of atomic coordinates would be misleading. Therefore, this study first utilizes the Kabsch algorithm to perform a rigid alignment of the predicted conformation $\hat{G}$ to the true conformation G, thereby eliminating the effects of rotational and translational degrees of

freedom. Subsequently, RMSD of the corresponding heavy atom coordinates in the aligned conformations is calculated. This metric evaluates the overall structural discrepancy between the two conformations in three-dimensional space:

$$\text{K-RMSD}(\hat{G}, G) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{G}_i - G_i\right\|_2^2} \tag{54}$$

E.3 Implementation Details

Regarding model implementation specifics, we rigorously followed the settings below. For QM9, we employed the standard split scheme, with the training, validation, and test sets containing 110000, 10000, and 11000 samples respectively. For Molecule3D, to evaluate model robustness, experiments were conducted under both random and scaffold split strategies. The training, validation, and test sets were divided in a ratio of 6:2:2 under both strategies. The model was built in an environment with PyTorch 2.1.0, Python 3.10, and CUDA 12.1. Experiments were performed on machines equipped with NVIDIA RTX 4090 GPUs. The model architecture had a hidden layer dimension size of 256 with 8 attention heads. The AdamW optimizer was selected for gradient optimization. On QM9, the batch size was set to 1200 for 50 training epochs, whereas on Molecule3D, the batch size was set to 600 for 20 training epochs. A fixed random seed of 42 was used across all experiments to ensure reproducibility of results.

E.4 Baselines

To rigorously evaluate the effectiveness of GeoGT for molecular ground-state conformation prediction, we compare it with representative baselines covering classical distance-geometry methods, conventional GNNs, and recent Graph Transformer-based models.

- **RDKit DG** (Havel, 1998; Landrum et al., 2013): RDKit DG is a classical distance-geometry-based conformer generation method implemented in RDKit. It embeds atoms into 3D space using geometric constraints such as interatomic distance bounds, serving as a non-learning baseline.
- **RDKit ETKDG** (Riniker & Landrum, 2015; Landrum et al., 2013): RDKit ETKDG extends distance geometry by incorporating experimental torsion-angle preferences and knowledge-based constraints, producing more chemically reasonable conformations.
- **DeeperGCN-DAGNN** (Liu et al., 2021): DeeperGCN-DAGNN combines DeeperGCN and DAGNN to enhance deep message passing and adaptive receptive-field aggregation. It is used as a strong deep GNN baseline for molecular graph representation learning.
- **GINE** (Hu et al., 2019): GINE is an edge-aware variant of GIN that incorporates bond features into message passing. It serves as a representative conventional GNN baseline for molecular graphs.
- **GATv2** (Brody et al., 2021): GATv2 improves the expressiveness of graph attention by addressing the static attention limitation of GAT. It is included as an attention-based GNN baseline.
- **GraphGPS (RW)** (Rampášek et al., 2022): GraphGPS combines local message passing with global attention. The RW variant uses random-walk-based encodings to introduce structural information.
- **GraphGPS (LP)** (Rampášek et al., 2022): GraphGPS (LP) uses Laplacian positional encoding to inject spectral graph information, allowing comparison with the random-walk encoding variant.
- **GTMGC** (Xu et al., 2023): GTMGC is a Graph Transformer model for end-to-end molecular ground-state conformation prediction from 2D molecular graphs. It introduces Molecule Structural Residual Self-Attention for molecular structure modeling.
- **ReBIND** (Kim et al., 2025): ReBIND is a force-aware graph rewiring framework that adds edges between non-bonded atomic pairs based on Lennard-Jones forces, improving the modeling of non-bonded interactions.

These baselines provide a comprehensive benchmark from classical cheminformatics algorithms to advanced Graph Transformer models, enabling a reliable evaluation of GeoGT.

F Additional Experimental Results

F.1 Molecular Ground-State Conformation Visualization

To rigorously validate the effectiveness of GeoGT in generating molecular ground-state conformations, this study designed a systematic comparative experiment, with a focus on conformational visualization analysis. As illustrated in Figure 3 and 4, benchmark conformations were obtained using high-accuracy quantum chemical calculations (the DFT method) as ground-truth references. Subsequently, the conformations generated by RDKit ETKDG and GTMGC were comprehensively compared against the predictions from GeoGT. By integrating visualizations of both three-dimensional spatial conformations and two-dimensional molecular structure diagrams, the analysis intuitively demonstrates the notable accuracy advantage of GeoGT in conformation prediction.

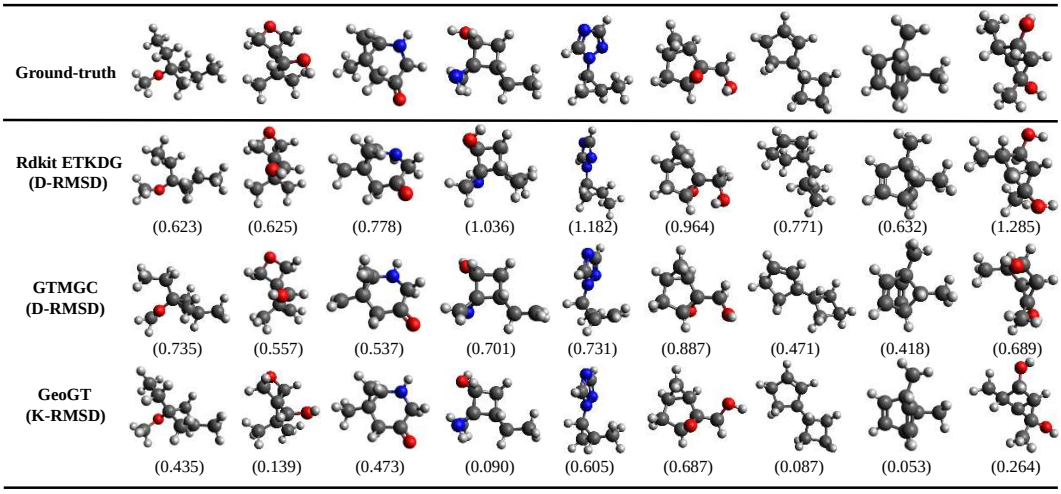


Figure 3: Molecular ground-state conformation visualization.

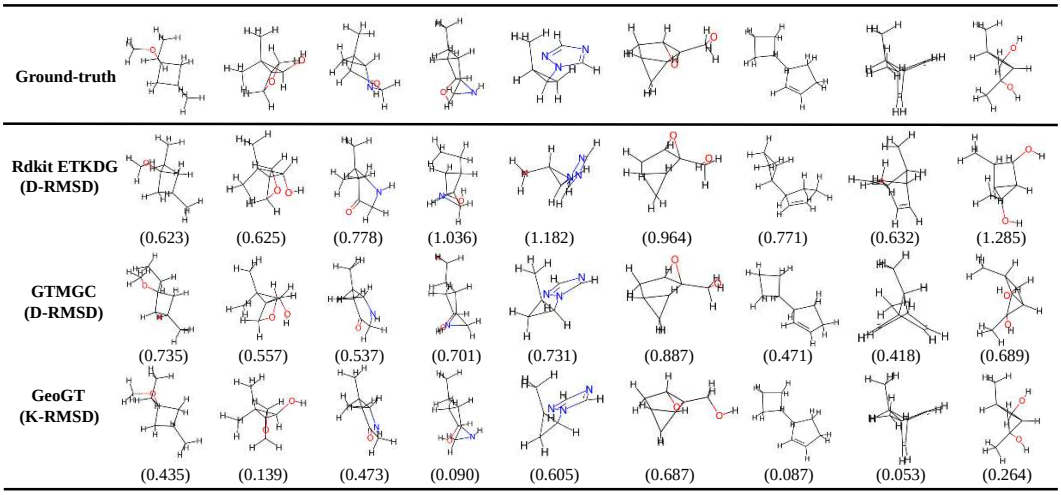


Figure 4: Two-dimensional structural diagrams of the corresponding molecules in the ground-state.

F.2 Hyperparameter Sensitivity Analysis

To evaluate the sensitivity of GeoGT to hyperparameters, we conducted a systematic experimental analysis on the QM9 dataset focusing on the number of epochs, the number of heads in the multi-head self-attention mechanism, and learning rate. As shown in Figure 5, despite substantial variations in these hyperparameters, the model still demonstrated consistently strong performance across multiple evaluation metrics, including D-MAE, D-RMSE, and K-RMSD. Specifically, we adjusted the number of epochs (30, 40, 50, 60, 70), number of attention heads (2, 4, 8, 16, 32), and learning rate (6e-4, 7e-4, 8e-4, 9e-4, 1e-3), and recorded the model's performance on both validation and test sets. The experimental results indicate that GeoGT exhibits considerable robustness to changes in these hyperparameters: even under significantly different parameter configurations, the model maintains low prediction errors and high consistency without notable performance degradation.

F.3 Statistical Analysis of Bond Angles in QM9

As shown in Table 5, the statistical information of the QM9 dataset highlights the central role of bond angles in characterizing molecular three-dimensional structures. The dataset comprises over 130,000 molecules, with each molecule containing approximately 35 bond angles on average (median: 36), demonstrating the significant structural complexity of these small organic molecules. The number of bond angles varies widely, ranging from 1 to 60 (standard deviation: 8.82), reflecting considerable diversity in molecular size and topology. Molecules with 44 bond angles occur most frequently (7,504 instances). Figure 6 illustrates the distribution of bond angle counts in QM9. Figure 6(a) shows the molecule counts together with the cumulative percentage and the mean value, indicating that most molecules contain approximately 20–45 bond angles, with the mean close to the median reported in Table 5. Figure 6(b) presents the same distribution on a logarithmic scale with the kernel density estimate, making the sparse regions more visible and showing that only a small number of molecules have extremely low or high bond angle counts. These bond angle data not only quantify molecular geometry but also directly determine electronic distribution, energy states, and physicochemical properties. Thus, they provide an essential structural basis for GeoGT to accurately understand and predict molecular ground-state.

Table 5: Statistical summary in the QM9 dataset.

Statistic	**Value**
Total Number of Molecules	130,831
Mean Number of Bond Angles	34.99
Median Number of Bond Angles	36.00
Standard Deviation	8.82
Minimum Number of Bond Angles	1.0
Maximum Number of Bond Angles	60.0
Most Frequent Number of Bond Angles	44.0

F.4 Molecular Inference Efficiency Analysis

As shown in the Table 6, GeoGT demonstrates highly competitive inference efficiency across different datasets, with particularly clear advantages on the larger-scale and structurally more complex Molecule3D datasets. On the QM9 dataset, GeoGT achieves throughputs of 2225.28 and 2225.05 molecules/s on the validation and test sets, respectively, ranking second in both cases, only behind GTMGC. This indicates that on datasets such as QM9, where molecules are relatively small and the computational cost is comparatively low, efficiency differences among different methods may be more strongly affected by model implementation and fixed computational overhead. Nevertheless, GeoGT is still significantly faster than baseline methods such as GINE, GATv2, GPS, and ReBIND, demonstrating the high computational efficiency of its overall inference process.

On the Molecule3D Random dataset, GeoGT achieves the best inference speed, reaching 1147.62 and 1066.05 molecules/s on the validation and test sets, respectively, corresponding to improvements

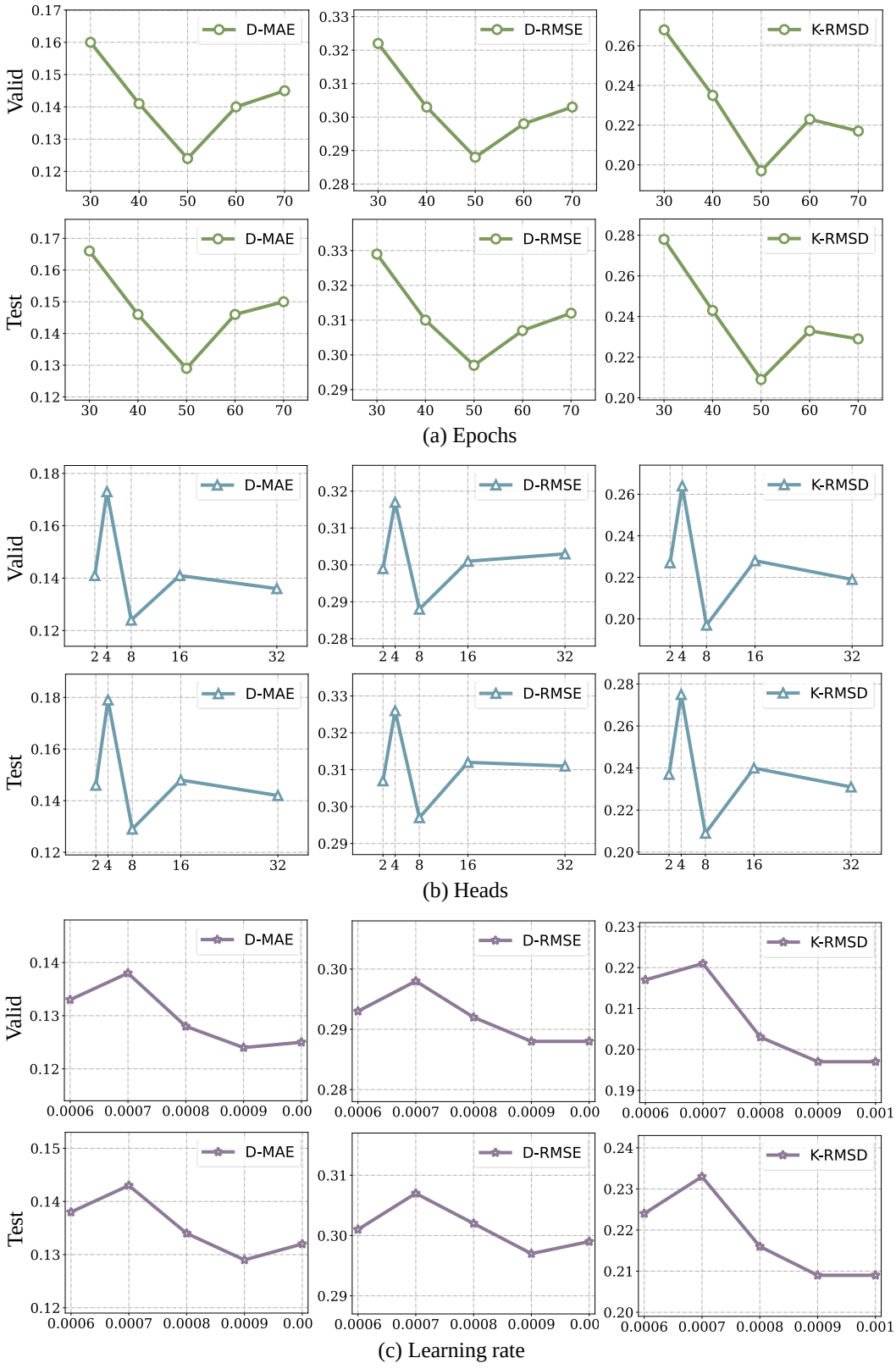


Figure 5: Hyperparameter sensitivity analysis on the QM9.

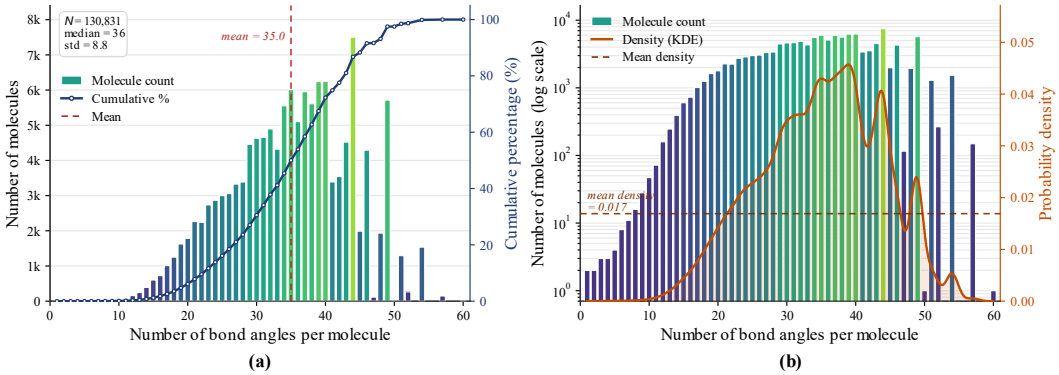


Figure 6: Distribution of the number of bond angles per molecule in QM9. (a) Molecule counts (bars) with the cumulative percentage (line) and the mean; (b) the same distribution on a logarithmic scale, overlaid with the kernel-density estimate and its mean density.

of approximately 67.0% and 59.3% over the second-best method, GPS. Meanwhile, the average per-molecule inference times of GeoGT are 0.000871 s and 0.000938 s, respectively, which are clearly lower than those of the other baseline methods. Its standard deviations are also relatively small, indicating that GeoGT not only achieves higher throughput but also exhibits more stable inference times across different molecular samples. Similarly, on the Molecule3D Scaffold dataset, GeoGT maintains the highest throughput, achieving 1055.47 and 1048.89 molecules/s on the validation and test sets, respectively, with improvements of approximately 15.3% and 17.2% over the second-best method, GTMGC. In addition, GeoGT achieves the best or near-best average and median inference times on this dataset, further validating its efficiency in complex molecular conformation prediction tasks.

GeoGT maintains near-optimal inference efficiency on QM9 and achieves highest throughput and low per-molecule inference latency on the more challenging Molecule3D Random and Molecule3D Scaffold datasets. These results indicate that GeoGT does not introduce excessive computational overhead while incorporating geometric information modeling. Instead, it enables more efficient and stable inference in complex three-dimensional molecular scenarios, demonstrating strong scalability and practical potential for molecular ground-state conformation prediction tasks.

F.5 Chemical Validity Analysis of Generated Conformations

As shown in Table 7, GeoGT achieves the highest Overall pass rate across all datasets and splits, indicating that its generated conformations better satisfy chemical validity constraints. On QM9, GeoGT obtains Overall pass rates of 0.9923 and 0.9935 on the validation and test sets, respectively. Although the performance gap is small due to the relatively simple molecular structures in QM9, GeoGT still consistently outperforms all baselines.

On Molecule3D random, GeoGT shows a clearer advantage, achieving Overall pass rates of 0.9432 and 0.9426 on the validation and test sets. It also obtains the best Bond angles scores of 0.9822 and 0.9821, demonstrating its stronger ability to preserve reasonable local geometric structures compared with both general graph models and stronger conformation-generation baselines. On Molecule3D scaffold, GeoGT maintains the best Overall pass rates of 0.9260 and 0.9019 on the validation and test sets, showing better robustness under scaffold distribution shift. Its strong performance on Bond angles, Internal steric clash, and Aromatic ring flatness further indicates that GeoGT can generate chemically more reliable conformations with reasonable local geometry, fewer steric conflicts, and better ring planarity.

Overall, the PoseBusters results demonstrate that GeoGT achieves the best overall chemical validity and stable performance across key chemical constraints, further confirming its effectiveness for molecular ground-state conformation prediction.

Table 6: Molecular inference efficiency analysis on the QM9 and Molecule3D datasets. * indicates the most important metric, bold indicates the best result, and underline indicates the second-best result.

	Validation						Test					
Methods	Molecules/s *	Mean(s)	Std(s)	Median(s)	Min(s)	Max(s)	Molecules/s *	Mean(s)	Std(s)	Median(s)	Min(s)	Max(s)
						(a) QM9						
GINE	1344.27	0.000744	0.003123	0.000433	0.000277	0.046437	1522.28	0.000657	0.002448	0.000438	0.000263	0.045484
GATv2	1013.03	0.000987	0.002273	0.000711	0.000459	0.045519	444.04	0.002252	0.003364	0.001255	0.000906	0.047366
GPS	1451.58	0.000689	0.002288	0.000461	0.000290	0.044588	1509.16	0.000663	0.002269	0.000454	0.000309	0.045496
GTMGC	<u>2408.40</u>	0.000415	0.000963	0.000369	0.000259	0.042961	<u>2633.47</u>	0.000380	0.000506	0.000357	0.000242	0.039877
ReBIND	744.43	0.001343	0.004050	0.000908	0.000587	0.049882	713.50	0.001402	0.004250	0.000915	0.000518	0.047359
GeoGT	**3084.62**	0.000324	0.000763	0.000274	0.000199	0.029594	**3330.62**	0.000300	0.000412	0.000274	0.000201	0.028848
						(b) Molecule3D Random						
GINE	503.36	0.001987	0.007359	0.001061	0.000252	0.202265	479.78	0.002084	0.007682	0.001135	0.000243	0.263593
GATv2	419.56	0.002383	0.006975	0.001501	0.000611	0.183745	413.05	0.002421	0.007339	0.001466	0.000606	0.185619
GPS	<u>687.23</u>	0.001455	0.006013	0.000803	0.000285	0.204569	<u>669.37</u>	0.001494	0.006060	0.000823	0.000280	0.158423
GTMGC	621.45	0.001609	0.006813	0.000746	0.000255	0.394657	618.49	0.001617	0.006844	0.000745	0.000247	0.176748
ReBIND	321.84	0.003107	0.008316	0.001855	0.000456	0.196299	322.22	0.003103	0.008352	0.001840	0.000447	0.107199
GeoGT	**1147.62**	0.000871	0.003202	0.000638	0.000252	0.180682	**1066.05**	0.000938	0.003676	0.000640	0.000254	0.089053
						(c) Molecule3D Scaffold						
GINE	545.13	0.001834	0.006371	0.000915	0.000252	0.174914	550.18	0.001818	0.006397	0.000925	0.00028	0.189549
GATv2	346.09	0.002889	0.007709	0.001744	0.000477	0.179641	340.86	0.002934	0.007741	0.001821	0.000443	0.282951
GPS	488.91	0.002045	0.007347	0.001072	0.00032	0.18459	474.01	0.00211	0.007438	0.001162	0.000323	0.177427
GTMGC	<u>915.66</u>	0.001092	0.004412	0.000661	0.000269	0.088949	<u>895.19</u>	0.001117	0.00442	0.000682	0.000266	0.125626
ReBIND	356.24	0.002807	0.007459	0.001728	0.000488	0.082622	343.02	0.002915	0.007546	0.001798	0.000436	0.096536
GeoGT	**1055.47**	0.000947	0.003628	0.000655	0.000291	0.367169	**1048.89**	0.000953	0.003641	0.000655	0.000279	0.074181

F.6 Ablation Study

To evaluate the actual contributions of the core components in the GeoGT, we systematically conducted ablation experiments on the QM9 dataset. As shown in Table 8, "$w/o\ Angle$" denotes removing the component related to the bond angle variance in GeoGT, meaning that the feature vectors u in the decoder are no longer modulated, and the geometric constraint B_l^{Geo} is not added in the decoder's MHSA. "$w/o\ T$" indicates that in the positional encoding of the decoder, the curvature-based positional encoding T is omitted, retaining only the original laplacian positional encoding. The experimental results demonstrate that removing any component leads to a significant decline in model performance, fully underscoring the critical role of geometric constraints in the attention mechanism, as well as the necessity of geometric encodings like curvature and interatomic distances for enhancing the expressive capacity of positional information. These findings consistently validate the indispensable contribution of each component to improving the accuracy of conformational prediction.

Furthermore, we conducted ablation experiments on samples from the Molecule3D scaffold split for which GeoGT exhibits higher K-RMSD than ReBIND. As shown in Table 9, the full model achieves the lowest errors across all three metrics on the corresponding validation and test subsets, while removing both components results in the worst performance. These results indicate that both types of geometric information still contribute to improving GeoGT's conformation prediction accuracy, even on samples where its K-RMSD is worse than that of ReBIND.

Table 7: Chemical validity analysis on the QM9 and Molecule3D datasets. * indicates the most important metric, and bold indicates the best performance.

Metrics	Validation						Test					
	GINE	GATV2	GPS	GTMGC	ReBIND	GeoGT	GINE	GATV2	GPS	GTMGC	ReBIND	GeoGT
(a) QM9												
Overall pass rate*	0.4521	0.6383	0.7128	0.9676	0.9891	**0.9916**	0.4422	0.6323	0.7167	0.9680	0.9892	**0.9926**
Bond lengths	0.6154	0.7919	0.8520	0.9797	0.9957	0.9955	0.6042	0.7851	0.8550	0.9776	0.9957	0.9958
Bond angles	0.6073	0.7463	0.7955	0.9892	0.9982	0.9986	0.6021	0.7436	0.7973	0.9901	0.9973	0.9980
Internal steric clash	0.8968	0.9433	0.9570	0.9999	0.9998	0.9999	0.8982	0.9446	0.9597	0.9993	0.9999	0.9997
Aromatic ring flatness	0.9822	0.9997	0.9987	1.0000	1.0000	1.0000	0.9814	0.9996	0.9990	1.0000	1.0000	1.0000
Non-aromatic ring non-flatness	0.9981	0.9963	0.9979	0.9943	0.9944	0.9976	0.9986	0.9977	0.9982	0.9955	0.9954	0.9982
Double bond flatness	0.9968	0.9990	0.9988	0.9994	0.9997	0.9991	0.9982	0.9988	0.9988	0.9996	0.9997	0.9997
(b) Molecule3D random												
Overall pass rate*	0.3383	0.4039	0.4499	0.9322	0.9347	**0.9432**	0.3392	0.4044	0.4503	0.9320	0.9346	**0.9426**
Bond lengths	0.6500	0.7617	0.8225	0.9817	0.9855	0.9742	0.6507	0.7611	0.8228	0.9817	0.9858	0.9739
Bond angles	0.4368	0.4700	0.5043	0.9720	0.9670	0.9822	0.4374	0.4710	0.5049	0.9721	0.9670	0.9821
Internal steric clash	0.7221	0.7335	0.7429	0.9972	0.9969	0.9965	0.7220	0.7326	0.7426	0.9972	0.9969	0.9965
Aromatic ring flatness	0.9886	0.9954	0.9949	1.0000	1.0000	0.9999	0.9884	0.9952	0.9948	1.0000	1.0000	0.9999
Non-aromatic ring non-flatness	0.9936	0.9905	0.9877	0.9802	0.9832	0.9847	0.9936	0.9908	0.9877	0.9801	0.9833	0.9846
Double bond flatness	0.9901	0.9927	0.9946	0.9946	0.9964	0.9966	0.9903	0.9928	0.9946	0.9944	0.9963	0.9965
(c) Molecule3D scaffold												
Overall pass rate*	0.1611	0.2080	0.3906	0.9129	0.9163	**0.9260**	0.0296	0.0418	0.2461	0.9003	0.8963	**0.9019**
Bond lengths	0.2852	0.3966	0.6772	0.9491	0.9518	0.9566	0.0575	0.1208	0.4306	0.9415	0.9336	0.9429
Bond angles	0.2368	0.2694	0.4811	0.9610	0.9670	0.9813	0.0550	0.0687	0.3540	0.9662	0.9728	0.9783
Internal steric clash	0.4705	0.4763	0.7131	0.9780	0.9874	0.9971	0.1486	0.1601	0.6131	0.9863	0.9853	0.9949
Aromatic ring flatness	0.9373	0.9863	0.9811	0.9893	0.9987	0.9992	0.9174	0.9816	0.9554	0.9924	0.9894	0.9979
Non-aromatic ring non-flatness	0.9847	0.9805	0.9826	0.9537	0.9627	0.9836	0.9834	0.9819	0.9798	0.9560	0.9516	0.9765
Double bond flatness	0.9843	0.9894	0.9909	0.9844	0.9936	0.9955	0.9767	0.9779	0.9787	0.9914	0.9828	0.9928

F.7 Effect of Iterative Refinement on Prediction Accuracy and Inference Efficiency

As shown in Table 10, two decoding passes achieve the lowest errors across all three metrics, reducing D-MAE, D-RMSE, and K-RMSD by approximately 8.6%, 3.6%, and 6.5%, respectively, compared with a single decoding pass, while increasing inference time by approximately 62.7%. Further increasing the number of decoding passes yields no additional improvement, with prediction errors generally rising; at six decoding passes, both D-RMSE and K-RMSD exceed their single-pass counterparts. These results indicate that moderate iterative refinement can improve prediction accuracy, whereas excessive iterations may degrade performance and substantially increase computational overhead. Therefore, to balance accuracy and efficiency, we adopt a single decoding pass by default and retain additional iterations as an optional refinement strategy when the computational budget permits.

F.8 Downstream Molecular Property Prediction

To evaluate the downstream utility of generated conformations, we use the officially released QM9-base model of GotenNet(Aykent & Xia, 2025) to predict eight QM9 properties from conformations generated by GTMGC, ReBIND, and GeoGT. All methods are evaluated on the same molecules with consistent atom ordering, and reference property labels are matched using the original QM9

Table 8: Ablation results on QM9.

Validation				
Angle	T	**D-MAE↓**	**D-RMSE↓**	**K-RMSD↓**
✗	✓	0.197	0.372	0.323
✓	✗	0.235	0.445	0.219
✗	✗	0.287	0.496	0.382
✓	✓	**0.133**	**0.299**	**0.214**
Test				
Angle	T	**D-MAE↓**	**D-RMSE↓**	**K-RMSD↓**
✗	✓	0.199	0.377	0.321
✓	✗	0.243	0.459	0.228
✗	✗	0.291	0.494	0.390
✓	✓	**0.139**	**0.309**	**0.224**

Table 9: Ablation study on challenging Molecule3D scaffold samples.

Validation				
Angle	T	**D-MAE↓**	**D-RMSE↓**	**K-RMSD↓**
✗	✓	0.504	0.772	0.997
✓	✗	0.613	0.917	0.686
✗	✗	0.753	1.015	1.179
✓	✓	**0.357**	**0.631**	**0.672**
Test				
Angle	T	**D-MAE↓**	**D-RMSE↓**	**K-RMSD↓**
✗	✓	0.515	0.781	0.983
✓	✗	0.631	0.936	0.699
✗	✗	0.759	1.125	1.268
✓	✓	**0.364**	**0.659**	**0.691**

identifiers. The results show that GeoGT achieves the lowest MAE and RMSE across all eight properties, outperforming both GTMGC and ReBIND. These findings support the effectiveness of GeoGT-generated conformations as geometric inputs for downstream molecular property prediction.

F.9 Performance Analysis across Molecular Sizes

To analyze the predictive performance and stability of different methods across molecular sizes, we group evaluation samples by molecular size and report MAE, RMSE, and RMSD. As shown in Figure 7 and Figure 8, the prediction errors of all methods generally increase with molecular size on both QM9 and Molecule3D, indicating that ground-state conformation prediction is more challenging for larger molecules.

On QM9, as shown in Figure 7, GeoGT consistently achieves the lowest MAE, RMSE, and RMSD across all molecular size groups on both the validation and test splits, with errors increasing relatively gradually as molecular size grows. In contrast, GINE, GATv2, and GPS exhibit more pronounced error increases, particularly for larger molecules. Although ReBIND and GTMGC outperform these graph neural network baselines, their prediction errors remain higher than those of GeoGT. These results indicate that GeoGT handles variations in molecular size effectively and maintains high accuracy in predicting molecular geometry across different size groups.

On Molecule3D, similar trends are observed under the random and scaffold splits, as illustrated in Figure 8. Under both splits, GeoGT generally maintains low prediction errors, with particularly pro-

Table 10: Effect of iterative refinement on the QM9 test.

Decoding passes T	D-MAE ↓	D-RMSE ↓	K-RMSD ↓	Throughput ↑ (molecules/s)	Relative time increase ↓
1 (default)	0.1393	0.3092	0.2235	**3330**	**0%**
2	**0.1273**	**0.2980**	**0.2089**	2047	62.67%
3	0.1298	0.3036	0.2156	1490	123.48%
4	0.1291	0.3027	0.2148	1155	188.31%
5	0.1343	0.3065	0.2224	944	252.75%
6	0.1384	0.3110	0.2326	797	317.82%

Table 11: Downstream molecular property prediction on QM9 using conformations generated by different methods. Lower is better; best results are highlighted in bold.

Methods	μ (D)	α (bohr^3)	HOMO (eV)	LUMO (eV)	ZPVE (eV)	U (eV)	G (eV)	C_v (kcal mol^{-1} K^{-1})
MAE ↓								
GTMGC	0.6596	1.5682	0.1815	0.2737	0.1087	0.2640	0.2519	0.2124
ReBIND	0.6954	1.4346	0.1927	0.2517	0.1392	0.2600	0.2451	0.1740
GeoGT (Ours)	**0.4522**	**0.6150**	**0.1128**	**0.1230**	**0.0625**	**0.1410**	**0.1319**	**0.0653**
RMSE ↓								
GTMGC	0.9922	2.4275	0.2572	0.4720	0.1446	0.8026	0.3714	5.0393
ReBIND	0.9743	2.0990	0.2632	0.4283	0.1747	0.3718	0.3386	5.4841
GeoGT (Ours)	**0.8528**	**1.4030**	**0.1796**	**0.2454**	**0.1012**	**0.2204**	**0.2020**	**0.6213**

nounced advantages over GINE, GATv2, and GPS for medium-sized and larger molecules. GeoGT also remains competitive overall with stronger baselines such as ReBIND and GTMGC, indicating its effectiveness under both random splits and the distribution shifts associated with scaffold splits.

Limitation. GeoGT still exhibits limitations in predicting the conformations of larger molecules. In particular, all methods show substantial error increases in the group of Molecule3D molecules containing 50–70 atoms, and GeoGT follows this trend. This phenomenon may be related to the relatively limited representation of larger molecules in the training data, as well as their typically greater conformational degrees of freedom and more complex long-range geometric relationships, which make it more difficult for models to accurately recover the overall conformation. Thus, despite its generally low prediction errors, GeoGT still has room for improvement on larger molecules.

The molecular size analysis demonstrates that GeoGT performs well across different size ranges and data splits, while also highlighting the remaining challenges in modeling larger molecules. To address these limitations, future work could explore increasing the coverage of larger molecules in the training data and strengthening the modeling of torsion angles and geometric relationships between molecular fragments. These directions may help further improve prediction accuracy and generalization for larger molecules.

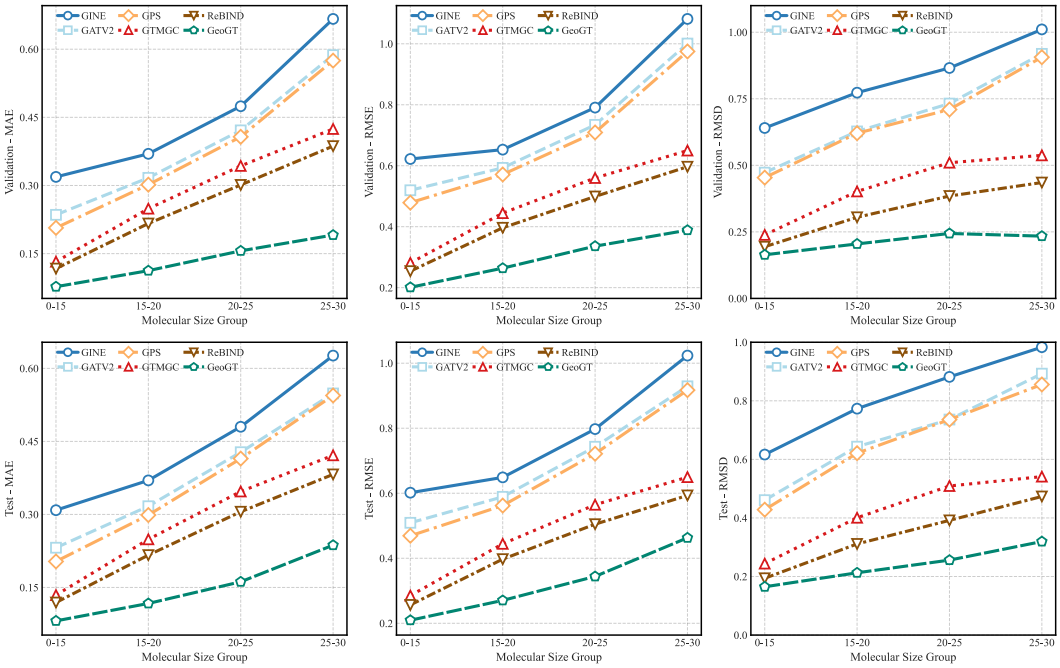
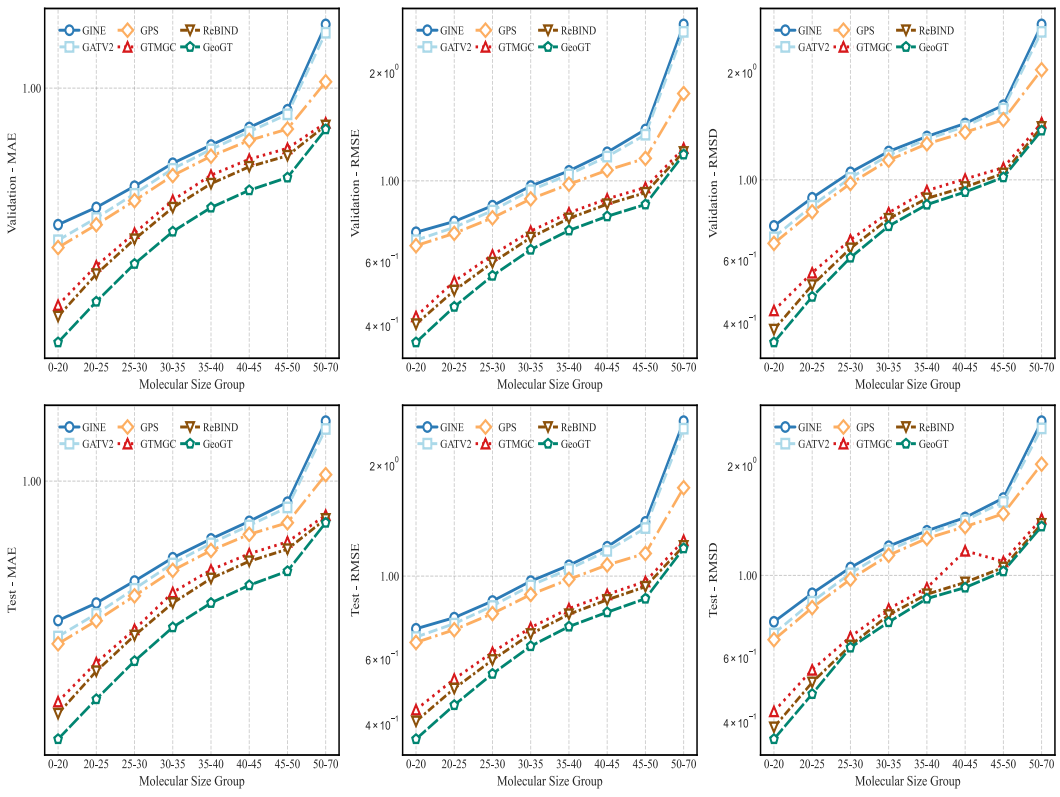
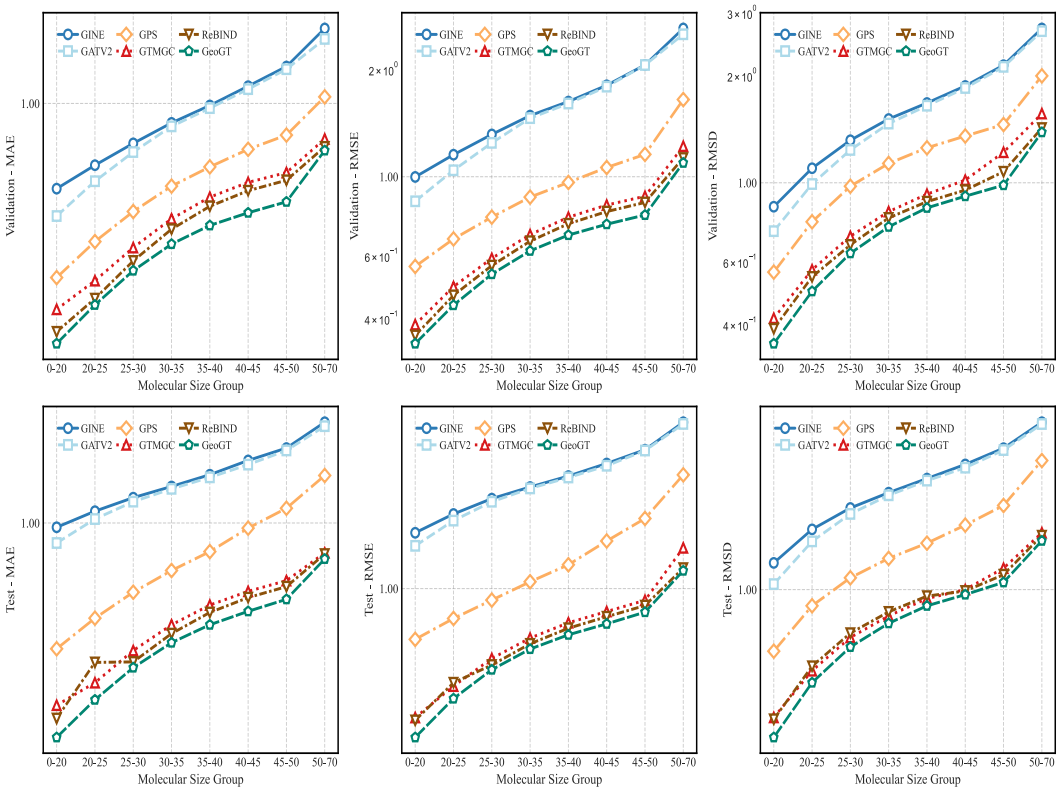


Figure 7: Molecular size analysis on the QM9.



(a) Molecule3D Random



(b) Molecule3D Scaffold

Figure 8: Molecular size analysis on the Molecule3D.