

AGDIFF: Attention-Enhanced Diffusion for Molecular Geometry Prediction

André Wyzykowski, Fatemeh Fathi Niazi, Alex Dickson

Michigan State University - Department Of Biochemistry and Molecular Biology

Abstract

Accurate prediction of molecular geometries is crucial for drug discovery and materials science. Existing fast conformer prediction algorithms often rely on approximate empirical energy functions, resulting in low accuracy. More accurate methods like ab initio molecular dynamics and Markov chain Monte Carlo can be computationally expensive due to the need for evaluating quantum mechanical energy functions. To address this, we introduce AGDIFF, a novel machine learning framework that utilizes diffusion models for efficient and accurate molecular structure prediction. AGDIFF extends GeoDiff by enhancing the global, local, and edge encoders with attention mechanisms, an improved SchNet architecture, batch normalization, and feature expansion techniques. AGDIFF outperforms GeoDiff on both the GEOM-QM9 and GEOM-Drugs datasets. For GEOM-QM9, with a threshold (δ) of 0.5 Å, AGDIFF achieves a mean COV-R of 93.08% and a mean MAT-R of 0.1965 Å. On the more complex GEOM-Drugs dataset, using $\delta = 1.25\text{Å}$, AGDIFF attains a median COV-R of 100.00% and a mean MAT-R of 0.8237 Å. These findings demonstrate AGDIFF's potential to advance molecular modeling techniques, enabling more efficient and accurate prediction of molecular geometries, thus contributing to computational chemistry, drug discovery, and materials design. <https://github.com/ADicksonLab/AGDIFF>

1. Introduction

3D structures of molecules, characterized by their atomic Cartesian coordinates, dictate their biological and physical properties, making them crucial for advancements in healthcare and materials science. Traditional methodologies, such as molecular dynamics (MD) and Markov chain Monte Carlo (MCMC), although theoretically robust, are computationally demanding, especially when used with quantum mechanical energy functions, which are costly to evaluate [1]. In addition, widely used cheminformatics toolkits like RDKit [2] often struggle to generate accurate molecular geometries. A comparison of RDKit and GeoDiff [3] in conformer generation on the GEOM-Drugs dataset reveals that RDKit has limited coverage of the reference conformational

space and generates conformers of suboptimal quality. New methods for predicting chemical structures are needed that can balance the needs for efficiency and accuracy.

The prediction of molecular geometries has evolved significantly over the decades, with each new approach addressing the limitations of its predecessors. Force field methods like MM2 and MM3 in the 1980s and 1990s were followed by semi-empirical methods such as AM1 and PM3 in the 1990s and 2000s, which combined empirical data with quantum mechanical calculations for improved accuracy [4, 5, 6]. Simultaneously, Monte Carlo simulations and simulated annealing were used for low-energy conformer searches [7, 8]. The 2000s saw the rise of quantum mechanical methods, including Hartree-Fock and DFT, which provided accurate 3D structures by solving the Schrödinger equation [9, 10]. In the 2010s, cheminformatics toolkits like RDKit and Open Babel emerged, facilitating 3D structure generation from SMILES strings [2]. Early machine learning approaches, such as autoencoders [11], began exploring molecular structure prediction, laying the foundation for modern deep learning techniques in this field.

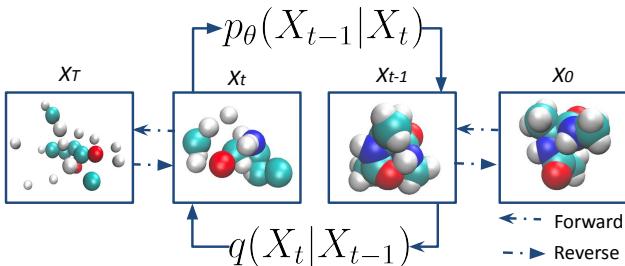
The general goal for structure prediction algorithms is to take the graph structure of a molecule as input ("2D" structure) and return a prediction of the 3D coordinates for each atom. The graph structure of a molecule is comprised of a set of nodes, one for each atom, and a set of edges, which show interactions between the atoms. The edges typically represent covalent bonds between atoms, but can also represent other things, such as spatial proximity. Each node can be annotated with a set of features, encoding information such as the element of the atom and the presence of an explicit charge. Similarly, edges are also given features, typically encoding the covalent bond type (single, double, aromatic, non-bonded, etc.). Here we denote a molecular graph with n nodes as G_n , and the set of all molecular graphs as \mathcal{G} . The task of molecular structure prediction is then to develop a function (f) that maps $\mathcal{G} \rightarrow \mathbb{R}^{3n}$.

These functions can be informed by earlier work in approaches developed for the prediction of molecular properties ($\mathcal{G} \rightarrow \mathbb{R}$).¹ Early Quantitative Structure-Activity

¹We note that algorithms incorporating the 3D structure for property prediction can be seen as a subclass of $\mathcal{G} \rightarrow \mathbb{R}$ algorithms that include the atomic positions as node features.

108	Relationship (QSAR) models relied on statistical methods	162																																																																						
109	and descriptors to correlate molecular structures with ac-	163																																																																						
110	tivities [12]. Later, topological descriptors provided al-	164																																																																						
111	gebraic characterizations of molecular structures [13], and	165																																																																						
112	physicochemical descriptors improved predictive accuracy	166																																																																						
113	by considering properties like lipophilicity and solubility	167																																																																						
114	[14]. The integration of machine learning algorithms, such	168																																																																						
115	as Support Vector Machines (SVM), Random Forest (RF),	169																																																																						
116	and Gradient Boosting Machines (GBM), further enhanced	170																																																																						
117	the predictive models by leveraging computational power to	171																																																																						
118	analyze complex datasets [15]. These approaches have been	172																																																																						
119	particularly useful for predicting scalar molecular proper-	173																																																																						
120	ties, including internal energy, solubility constant, and par-	174																																																																						
121	tion coefficient ($\log P$) [16, 17]. All of these approaches	175																																																																						
122	use projections of the graph onto an intermediate subspace	176																																																																						
123	of relevant features. However, the emergence of GNNs	177																																																																						
124	revolutionized the field by providing a powerful frame-	178																																																																						
125	work for learning directly from molecular graph represen-	179																																																																						
126	tations, eliminating the need for manual feature engineering	180																																																																						
127	and outperforming traditional methods in various predictive	181																																																																						
128	tasks [18].	182																																																																						
129	Recent advancements in graph representation learning	183																																																																						
130	have shown promising results in various tasks within the	184																																																																						
131	domain of computational chemistry and drug discovery, as	185																																																																						
132	extensively reviewed elsewhere [19]. We note that this field	186																																																																						
133	is rapidly evolving and we cite only a few recent exam-	187																																																																						
134	ples. Xiong et al. [20] introduced Attentive FP, a graph	188																																																																						
135	neural network architecture that leverages a graph atten-	189																																																																						
136	tion mechanism for enhanced molecular representation in	190																																																																						
137	drug discovery. Attentive FP achieves state-of-the-art pre-	191																																																																						
138	dictive performance for properties such as solubility, bioac-	192																																																																						
139	tivity, and lipophilicity. Additionally, it offers interpretable	193																																																																						
140	insights by highlighting important features like intramolecu-	194																																																																						
141	lar hydrogen bonding and aromatic systems, aiding in the	195																																																																						
142	understanding of complex molecular interactions.	196																																																																						
143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215
	Similar approaches have also been used for the $\mathcal{G} \rightarrow \mathbb{R}^{3n}$																																																																							
	problem with the help of diffusion models. Diffusion mod-																																																																							
	els are a class of generative models that gradually transform																																																																							
	simple random noise into complex data distributions. These																																																																							
	models have achieved superior quality compared to state-of-																																																																							
	the-art Generative Adversarial Networks (GANs) [21]. Lat-																																																																							
	ent Diffusion Models (LDMs), which operate in a learned																																																																							
	latent space, can capture more complex patterns and gen-																																																																							
	erate high-quality samples. For example, Luo and Hu (2021)																																																																							
	[22] treated point clouds as particles in a thermodynamic																																																																							
	system, utilizing a heat bath to facilitate diffusion from the																																																																							
	original distribution to a noise distribution. These advan-																																																																							
	cements showcase the versatility of diffusion models in han-																																																																							
	dling complex data structures (see Figure 1).																																																																							
	A key requirement for GNN-based diffusion models in																																																																							
	molecular structure prediction is equivariance. That is, a																																																																							
	transformation of the input should be equivalent to per-																																																																							
	forming the same transformation on the output. Mathemat-																																																																							
	ically, for a given rotation matrix, R : $f(R(\vec{x})) = R(f(\vec{x}))$.																																																																							
	In a diffusion model, a GNN must transform input struc-																																																																							
	tures to a slightly “de-noised” version of the input. It is																																																																							
	thus natural for these GNNs to be equivariant to displace-																																																																							
	ments and rotations, in order to ensure adherence to un-																																																																							
	derlying physical laws. A number of equivariant graph																																																																							
	models have been developed that carry explicit geom-																																																																							
	etric information with each node, and where all intermediate																																																																							
	calculations satisfy the equivariance property [24, 25, 26].																																																																							
	There has recently been a growing interest in the applica-																																																																							
	tion of these models to molecular structure prediction in a																																																																							
	diffusion model framework. Hoogeboom et al. [27] intro-																																																																							
	duced the E(3) Equivariant Diffusion Model (EDM) for 3D																																																																							
	molecule generation, offering a model equivariant to Eu-																																																																							
	clidean transformations that markedly improves the quality																																																																							
	and efficiency of generated molecular samples. Similarly,																																																																							
	Liao and Smidt [28] presented Equiformer, incorporating																																																																							
	SE(3)/E(3)-equivariant features in a Transformer network																																																																							
	for 3D atomistic graphs, demonstrating significant perfor-																																																																							
	mance across datasets in the domain of 3D atomistic graphs.																																																																							
	However, while increasing model complexity can improve																																																																							
	performance, it must be balanced with efficiency in training																																																																							
	and inference.																																																																							
	Inspired by these innovations, we introduce AGDIFF																																																																							
	(Figure 2), a computational framework that leverages the																																																																							
	strengths of diffusion models for efficient prediction of																																																																							
	molecular structures. This direct $\mathcal{G} \rightarrow \mathbb{R}^{3n}$ modeling ap-																																																																							
	proach circumvents the limitations inherent in indirect mod-																																																																							
	eling and aims to capture the complex distribution of molecu-																																																																							
	lar geometries with high accuracy and efficiency. AGDIFF																																																																							
	builds upon the work of Xu et al. [3], who introduced																																																																							
	GeoDiff, a diffusion-based model for molecular conforma-																																																																							
	tion generation. We extend GeoDiff by incorporating sev-																																																																							
	eral enhancements to the global and local encoders, as well																																																																							
	as the edge encoder, to improve the model’s expressiveness																																																																							

216 and adaptability.
217



228 Figure 2. Illustration of the diffusion process for molecular conformation
229 generation. The forward process involves adding noise
230 to the molecular structure, transitioning from an initial structure
231 X_0 to a noisy version X_t . The reverse process, modeled by
232 $p_\theta(X_{t-1}|X_t)$, reconstructs the original structure by gradually
233 removing the noise.

234 In this work, we propose an enhanced version of the
235 SchNet architecture [29] for the global encoder, introducing
236 modifications such as learnable activation functions,
237 attention mechanisms, adaptive scaling modules, dual path-
238 way processing, and enhanced CFConv layers. For the
239 local encoder, we build upon the GINEncoder from GeoDiff
240 [3], incorporating batch normalization layers to stabilize
241 training and improve convergence. Furthermore, we extend
242 the MLP Edge Encoder proposed by GeoDiff [3] by intro-
243 ducing feature expansion and processing, combination of
244 processed features and bond embeddings, and an attention
245 mechanism to generate more expressive and adaptable edge
246 representations.

247 As we delve into AGDIFF, we will detail our methods,
248 present our findings, and discuss their implications for drug
249 discovery and computational chemistry at large.

251 2. Methods

252 Our computational model for molecular geometry pre-
253 diction, AGDIFF, leverages a dual encoder architecture to
254 capture both global and local structural information (Fig-
255 ure 3). The global encoder learns a holistic representation
256 of the entire molecular graph, focusing on the overall geo-
257 metric arrangement and interactions between atoms. In
258 contrast, the local encoder captures fine-grained details and
259 local chemical environments of individual atoms and their
260 immediate neighborhoods.

261 This dual encoder design is motivated by the hierar-
262 chical nature of molecular structures. Molecules exhibit global
263 conformational preferences, which refer to the overall spa-
264 tial arrangement of atoms that minimizes the molecule’s poten-
265 tial energy, and local chemical constraints, such as bond
266 lengths, angles, and torsions, dictate the local geometry
267 around individual atoms. AGDIFF can effectively capture
268 these multi-scale patterns and dependencies by employing

270 separate encoders for global and local information.
271

272 The global encoder, operating on the complete molecu-
273 lar graph, processes node features, edge indices, and edge
274 attributes to learn a comprehensive representation of the
275 molecule’s overall structure. This enables the model to
276 reason about long-range interactions, symmetry, and global
277 conformational trends. On the other hand, the local encoder,
278 with its focus on individual atoms and their local environ-
279 ments, captures information such as bond lengths, angles,
280 and local chemical properties. This local perspective allows
281 the model to accurately predict fine-grained structural de-
282 tails and maintain chemically valid geometries, providing a
283 comprehensive understanding of the molecular structure.

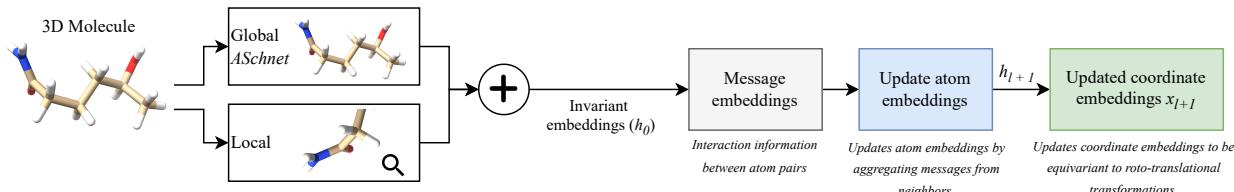
284 AGDIFF can generate molecularly consistent and ener-
285 getically favorable 3D conformations by combining the
286 complementary strengths of global and local encoders. The
287 global encoder ensures that the generated structures adhere
288 to overall geometric constraints, while the local encoder re-
289 fines the atomic positions based on local chemical rules.
290 This synergistic approach enables AGDIFF to predict accu-
291 rate and realistic molecular geometries across diverse chem-
292 ical spaces.

293 In the following subsections, we will delve into the de-
294 tails of each encoder and its respective architectures, high-
295 lighting the key components and enhancements introduced
296 in AGDIFF.

297 2.1. Global Encoder

298 In this work, we propose an enhanced version of the
299 SchNet architecture, a deep learning framework introduced
300 by Schütt et al. (2017) for learning molecular representa-
301 tions. SchNet utilizes continuous-filter convolutions (CF-
302 Conv) and interaction blocks to capture the complex inter-
303 actions between atoms in a molecule. The architecture
304 consists of an embedding layer followed by multiple inter-
305 action blocks, each comprising an atom-wise dense layer,
306 a CFConv layer, and a residual connection. The CFConv
307 layer applies a filter-generating network to the atom embed-
308 dings and aggregates the results based on the interatomic
309 distances. By stacking multiple interaction blocks, SchNet
310 enables multiple rounds of message passing and feature ex-
311 traction, allowing the model to learn intricate patterns and
312 relationships within the molecular structure. Finally, the
313 atom-wise features, enriched through these interactions, are
314 pooled to obtain a comprehensive global molecular rep-
315 resentation. This pooling step aggregates the detailed atomic
316 information into a singular representation that captures the
317 overall characteristics of the molecule.

318 The global encoder plays a crucial role in the overall ar-
319 chitecture, as it is responsible for learning a compact and
320 informative representation of the entire molecule. This
321 learned representation serves as the foundation for various
322 downstream tasks, such as property prediction or molecule

324
325
326
327
328
329
330331 Figure 3. Schematic overview of the AGDiff workflow. The input to the model is the 3D structure of the molecule, which includes
332 information about atom types, bond indices, bond types, and bond lengths. Following the implementation in Ref. [3], local and global
333 node embeddings are later combined to form the invariant embeddings, h_0 .334
335 generation. Therefore, enhancing the expressiveness and
336 adaptability of the global encoder can significantly improve
337 the model’s performance and generalization capabilities for
338 molecular conformer prediction.
339340

2.2. Enhancements to Global Encoder

341 We introduce several modifications to the SchNet ar-
342 chitecture to enhance its performance and adaptability in
343 molecular representation learning tasks. Details of the en-
344 hanced SchNet architecture are shown in Figure 4.345 These enhancements are designed to be computationally
346 efficient, ensuring that the model remains fast and scalable,
347 which is particularly important when considering the time-
348 consuming nature of diffusion model steps. The proposed
349 modifications include the incorporation of learnable activa-
350 tion functions, attention mechanisms, and adaptive scaling
351 modules, enabling the model to dynamically focus on rel-
352 evant features and interactions. Additionally, we introduce
353 dual pathway processing and enhanced CFConv layers to
354 increase the model’s expressiveness and learning capacity.
355 Each enhancement is discussed in detail below:356
357 **1. Enhanced CFConv Layers:** We make several mod-
358 ifications to the CFConv layer to improve its learning
359 capacity and robustness. Batch normalization is
360 applied to the feature maps to stabilize training and
361 speed up convergence. We replace the ReLU activa-
362 tion with LeakyReLU, which maintains a small gra-
363 dient for inactive units, preventing them from becom-
364 ing completely inactive during training. Additionally,
365 we integrate a learnable distance weighting into the
366 CFConv layer to enable dynamic feature priorita-
367 tion and distance-based weighting. This network takes
368 the interatomic distances as input and learns to as-
369 sign appropriate weights to the interactions based on
370 these distances. By capturing more complex distance-
371 dependent relationships, the model can better repre-
372 sent the molecular data and adapt to different molec-
373 ular structures.374 To refine the radial basis function that defines inter-
375 acting atoms in the CFConv layer, we implement a
376 Gaussian smoothing function that provides a smooth
377378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431and differentiable transition for interaction strengths
as a function of distance. This approach mitigates the
abrupt cutoff that traditionally limits the interaction be-
tween atoms, better reflecting the physical decay of
molecular forces. The function is mathematically ex-
pressed as:

$$C(d) = \exp\left(-\frac{(d - d_{\text{cutoff}})^2}{2\sigma^2}\right)$$

where d represents the interatomic distance, d_{cutoff} is
the predefined maximum distance for interactions, and
 σ is set to d_{cutoff} to control the spread of the Gaussian
envelope. In order to ensure that physically irrelevant
interactions are not considered we set $C(d)$ to zero for
 $d > d_{\text{cutoff}}$.2. **Enhanced interaction block:** To increase the expres-
siveness of the interaction blocks, we introduce a dual
pathway processing scheme. Each interaction block
now consists of two parallel CFConv layers with sep-
arate filter-generating networks (Pathway 1 and 2 in
Figure 4). The motivation behind this design is to al-
low the model to learn different aspects of the molec-
ular data simultaneously. One pathway uses a higher
number of hidden dimensions for the graph convolu-
tion operation, and one uses a lower number. The ra-
tionale for incorporating both pathways is grounded in
their complementary strengths:

- (a) **Pathway 1 (Standard CFConv):** This pathway
employs a higher number of hidden dimensions
in the graph convolution operation. The primary
advantage of this approach is its capacity to cap-
ture more complex and detailed patterns within
the molecular structure. By leveraging a higher-
dimensional feature space, the model can learn
intricate relationships and subtle interactions be-
tween atoms. This pathway is particularly effec-
tive for identifying long-range dependencies and
high-order interactions.
- (b) **Pathway 2 (Reduced CFConv):** In contrast,
the reduced CFConv pathway operates with a

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

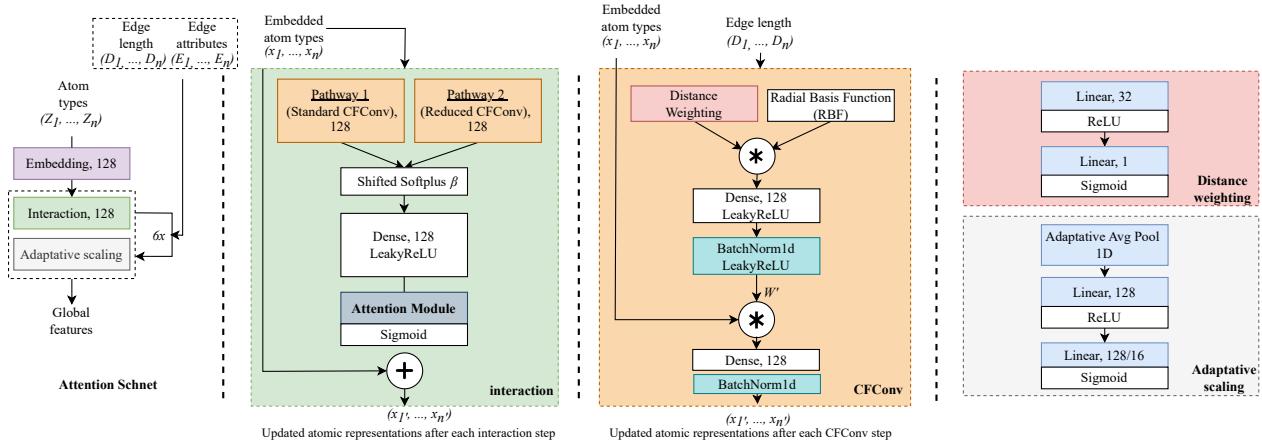


Figure 4. Diagram of the enhanced SchNet architecture. This figure illustrates the modifications introduced to improve performance and adaptability in molecular representation learning tasks, including learnable activation functions, attention mechanisms, adaptive scaling modules, dual pathway processing, and enhanced CFCConv layers.

lower number of hidden dimensions. This design choice enhances the model's efficiency by reducing computational complexity and mitigating the risk of overfitting, especially in scenarios where the dataset may not be large enough to support high-dimensional feature learning. Additionally, the reduced pathway can capture more general patterns and interactions that might be missed by the more complex pathway. It provides a form of regularization by focusing on broader trends and ensuring that the model remains robust and generalizable.

By integrating both pathways, the enhanced interaction block achieves a balance between detail and generality. This dual pathway approach allows the model to simultaneously leverage the strengths of both high-dimensional and low-dimensional feature spaces, resulting in a more robust and versatile molecular representation.

The outputs of the CFCConv layers are each passed through a *ShiftedSoftplus* activation layer. This introduces a set of learnable parameters, β that enable the model to dynamically adapt the activation function's response curve for each element during training. These outputs are then concatenated and processed using an attention mechanism to effectively integrate the information from both pathways by focusing on the most relevant features. The attention mechanism learns a set of attention weights using a neural network comprising two linear layers. The first linear layer reduces the dimensionality by half and applies a ReLU activation function. The second linear layer further processes these features, followed by a Sigmoid activation

function to produce the final attention weights. This allows the model to highlight the most relevant features from both pathways, enhancing the integration of information.

3. Adaptive Feature Scaling: We propose an adaptive scaling module that dynamically scales the features based on global information. This module applies average pooling followed by fully connected layers to compute scaling factors for each feature channel. The scaled features are then combined with the original features using element-wise multiplication. This adaptive scaling mechanism enables the model to emphasize or suppress certain features based on their relevance to the global molecular representation. By dynamically adjusting the feature scales, the model can adapt to different molecular properties and prioritize the most informative features for the given task.

4. Integration of Adaptive Scaling: We integrate adaptive scaling modules into the main encoder loop of the SchNet architecture. Each interaction block's output is passed through an adaptive scaling module before being added to the residual connection. This approach enables dynamic feature scaling at each stage, enhancing the model's ability to adapt to various data distributions and feature importances. The adaptive scaling module operates on a global scale, using average pooling followed by neural network transformations to adjust feature magnitudes, ensuring they are appropriately normalized and impactful for subsequent layers.

To clarify the distinction between this and the attention mechanism used within the interaction blocks: the attention mechanism dynamically prioritizes and integrates features from dual pathways based on immedi-

540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593

ate contextual relevance. In contrast, the adaptive scaling module adjusts the magnitude of features globally, based on a broader view of the data context at each stage of processing. It uses global average pooling to reduce feature dimensions to a single value per channel, which is then transformed by a neural network to produce scaling factors. These factors adjust the magnitude of features before further processing.

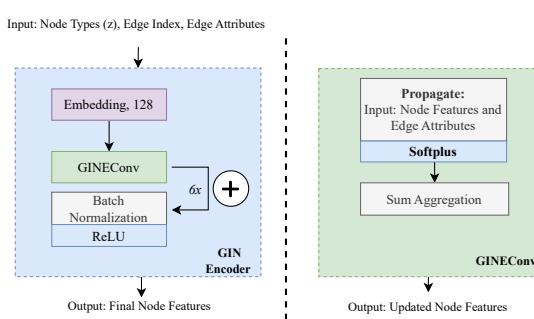
By incorporating both mechanisms, the model achieves different levels of feature adjustment. The attention mechanism refines feature integration within each block, while adaptive scaling optimizes feature scales for the entire network, ensuring well-tuned feature amplitudes for predictive accuracy.

The proposed enhancements to the SchNet architecture are designed to improve its ability to learn rich and expressive molecular representations while maintaining computational efficiency. By incorporating dynamic feature attention, learnable distance weighting, and dual pathway processing, the model can capture complex interactions and adaptively focus on the most informative aspects of the molecular structure. The adaptive feature scaling further enhances the model’s adaptability, enabling it to prioritize relevant features based on the global molecular context. These modifications aim to increase the model’s capacity to learn intricate patterns and relationships within the molecular data, ultimately leading to improved performance in conformer prediction. The combination of these enhancements represents a significant step towards more flexible and powerful models for molecular representation learning. By leveraging the strengths of attention mechanisms, learnable parameters, and adaptive scaling, the proposed architecture can effectively capture the complex nature of molecular interactions and adapt to various molecular properties.

2.3. Local Encoder

The local encoder, based on the GINEncoder from GeoDiff [3], utilizes layers from GINEConv [30] to process node features, edge indices, and edge attributes. We introduce batch normalization layers after each convolution in the GINEncoder to stabilize training and improve convergence.

As illustrated in Figure 5, the GINEncoder begins with node type embeddings and passes them through a series of GINEConv layers. Each GINEConv layer performs a propagation step that aggregates and updates node features, leveraging the residual connections present in the architecture to enhance expressiveness. Batch normalization and ReLU activation are applied after each GINEConv layer, except for the last layer. The final output represents the updated node features.



594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

Figure 5. The architecture of the local encoder used in this work. This is similar to that used in Ref. [3], except that a batch normalization is used to stabilize the training.

2.4. Edge Encoder

The edge encoder is responsible for capturing the geometric and chemical information of the molecular graph by processing edge attributes such as edge types and edge lengths. The edge types include different covalent bond types (single, double, triple, aromatic) as well as a non-covalent edge type. It takes both edge lengths and edge types as input and outputs a feature vector representing the encoded edge information (Figure 6). Our work expands upon the edge encoder proposed by GeoDiff [3], which employed a multi-layer perceptron (MLP) to process edge lengths and combined them with bond embeddings using element-wise multiplication.

Our enhanced MLP Edge Encoder incorporates the following components:

1. Feature Expansion and Processing: The MLP Edge Encoder includes a feature expansion layer (FE), implemented as a linear layer, to transform the 1D edge length features to a higher-dimensional space. The expanded features are concatenated with the edge type embeddings, which are also learned using an embedding layer. The concatenated features are then processed through a multi-layer perceptron (MLP_1). MLP_1 introduces nonlinearity in the processing of edge features, allowing the model to learn more abstract representations. The Gaussian Error Linear Unit (GELU) activation function is used in MLP_1 for smooth and efficient activation [31].

2. Combination of Processed Features and Bond Embeddings: To integrate the processed edge features and bond embeddings, a combination multi-layer perceptron (MLP_2) is introduced. MLP_2 learns an adaptive combination of the processed edge features and bond embeddings, enabling the model to balance the importance of different feature types and capture their interactions. The GELU activation function is em-

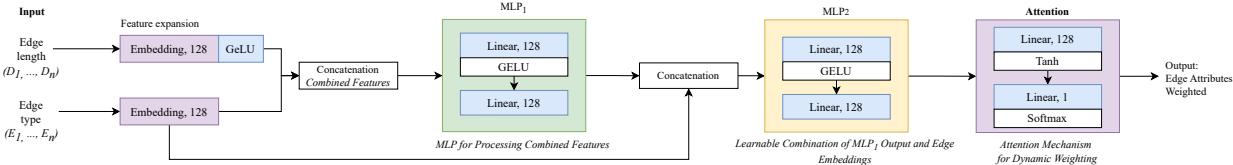


Figure 6. Architecture of the MLP Edge Encoder

ployed in MLP_2 to introduce nonlinearity and improve the expressiveness of the combined representations.

3. Attention Mechanism: An attention mechanism is incorporated to enhance the representational power of the MLP Edge Encoder. The attention mechanism consists of a sequence of layers, including a linear layer, followed by a hyperbolic tangent (\tanh) activation, another linear layer with a single output unit, and a softmax function applied along the feature dimension. The attention mechanism dynamically weights the edge features based on their relevance and importance. The attention weights are expanded to match the dimensions of the edge attributes, and then element-wise multiplication is performed to obtain the weighted edge attributes.

In the forward pass of the model, the MLP Edge Encoder is applied to all edges in the molecular graph, regardless of their context. The encoded edge attributes are then utilized by the subsequent components of the model for further processing and molecular representation learning.

3. Experimental Results

We present the experimental results of our AGDIFF model for molecular geometry prediction on the GEOM QM9 and Drugs datasets [32]. We compare the generated conformations with the reference conformations using various metrics, including quantitative evaluation of conformation quality and diversity (Section 3.1), RMSD average distributions (Section 3.2), scatter plot comparisons (Section 3.3), and correlation matrix (Section 3.4). Additionally, we perform a case study on alanine dipeptide molecular conformer analysis (Section 3.5) and provide qualitative analysis of generated conformations for alanine dipeptide (Figure 10) and drug-like molecules (Figure 11).

3.1. Evaluation of Generated Conformation Quality and Diversity

Evaluating the quality and diversity of generated molecular conformations is crucial in computational chemistry for accurate molecular simulations and drug design. High-quality conformations ensure reliable simulations, while diversity captures molecular flexibility, enabling the exploration of various conformational states.

To assess conformation generation models, we use root-mean-square deviation (RMSD) metrics, which quantify the average distance between corresponding atoms in two structures after optimal alignment using the Kabsch algorithm [33].

We utilize two primary metrics, Coverage (COV) and Matching (MAT), as introduced by [34], to compare generated conformations (S_g) with a reference set (S_r). These metrics are defined as follows:

$$\text{COV-R}(S_g, S_r) = \frac{1}{|S_r|} \left| \left\{ \mathcal{X} \in S_r \mid \text{RMSD}(\mathcal{X}, \hat{\mathcal{X}}) \leq \delta, \hat{\mathcal{X}} \in S_g \right\} \right|, \quad (1)$$

$$\text{MAT-R}(S_g, S_r) = \frac{1}{|S_r|} \sum_{\mathcal{X} \in S_r} \min_{\hat{\mathcal{X}} \in S_g} \text{RMSD}(\mathcal{X}, \hat{\mathcal{X}}), \quad (2)$$

where δ is a predefined RMSD threshold. COV-R measures the proportion of reference conformations covered by the generated set, indicating diversity. A higher COV-R score suggests better coverage of the reference conformational space.

MAT-R calculates the average RMSD between each reference conformation and its nearest neighbor in the generated set, indicating quality. A lower MAT-R score implies higher quality generated conformations.

We also use precision counterparts, COV-P and MAT-P, which swap the roles of generated and reference sets. COV-P measures the proportion of generated conformations close to any reference conformation, while MAT-P computes the average RMSD between each generated conformation and its nearest reference conformation, focusing on realism and plausibility.

The RMSD threshold δ is dataset-specific, based on previous studies [34, 35]. For the QM9 dataset, δ is set to 0.5\AA , and for the Drugs dataset, δ is 1.25\AA . These thresholds balance precise matching with tolerance for conformational variations.

In our experiments, we generate a set of conformations S_g that is twice the size of the reference set S_r for each molecule. This allows us to evaluate the model's ability to generate diverse conformations covering the reference space while maintaining high quality.

We evaluate our models on two datasets: QM9 and Drugs. The QM9 dataset consists of small organic molecules with up to 9 heavy atoms (C, O, N, and F), while the Drugs dataset contains larger and more complex drug-like molecules. These datasets provide a diverse range of

756 molecular structures to assess the generalization capabili-
757 ties of our AGDIFF framework.
758
759

760 Table 1 presents the results on the GEOM-QM9 dataset.
761 Our AGDIFF model outperforms the GEODIFF-A and
762 GEODIFF-C baselines across all metrics. In terms of COV-
763 R, AGDIFF achieves a mean of 93.08% and a median of
764 96.25%, surpassing the baselines. This indicates that
765 AGDIFF generates conformations that better cover the ref-
766 erence conformational space. Similarly, AGDIFF achieves
767 lower MAT-R values (mean: 0.1965 Å, median: 0.1919
768 Å) compared to the baselines, demonstrating higher qual-
769 ity generated conformations. The precision metrics, COV-
770 P and MAT-P, further confirm the superior performance of
771 AGDIFF, with higher COV-P values and lower MAT-P val-
772 ues compared to the baselines.
773
774
775
776
777

Table 1. Results on the GEOM-QM9 dataset (T = 5000 steps)

Models	COV-R (%) ↑		MAT-R (Å) ↓	
	Mean	Median	Mean	Median
GEODIFF-A	90.54	94.61	0.2104	0.2021
GEODIFF-C	90.07	93.39	0.2090	0.1988
AGDIFF	93.08	96.25	0.1965	0.1919
COV-P (%) ↑				
	Mean	Median	Mean	Median
	52.35	50.10	0.4539	0.4399
GEODIFF-C	52.79	50.29	0.4448	0.4267
AGDIFF	56.62	54.69	0.4156	0.3987

797 Table 2 shows the results on the GEOM-Drugs dataset,
798 which contains larger and more complex molecules com-
799 pared to QM9. AGDIFF consistently outperforms the
800 GEODIFF-A and GEODIFF-C baselines across all metrics.
801 Notably, AGDIFF achieves a median COV-R of 100.00%,
802 indicating that it generates conformations that fully cover
803 the reference conformational space for the majority of
804 molecules in the dataset. The lower MAT-R values (mean:
805 0.8237 Å, median: 0.8058 Å) demonstrate the high qual-
806 ity of the generated conformations. The precision metrics,
807 COV-P and MAT-P, also showcase the superior performance
808 of AGDIFF, with higher COV-P values and lower MAT-P
809 values compared to the baselines.

Table 2. Results on the GEOM-Drugs dataset (T = 5000 steps)

Models	COV-R (%) ↑		MAT-R (Å) ↓	
	Mean	Median	Mean	Median
GEODIFF-A	88.36	96.09	0.8704	0.8628
GEODIFF-C	89.13	97.88	0.8629	0.8529
AGDIFF	91.31	100.00	0.8237	0.8058
COV-P (%) ↑				
	Mean	Median	Mean	Median
	60.14	61.25	1.1864	1.1391
GEODIFF-C	61.47	64.55	1.1712	1.1232
AGDIFF	64.34	69.36	1.1316	1.0805

3.2. Analysis of the RMSD Average Distribution

Figure 7 shows the average distribution of the best 5 RMSD (Root Mean Square Deviation) values for molecular conformations generated by AGDIFF compared to the reference conformations in the QM9 and Drugs datasets, each consisting of 200 examples. RMSD quantifies the deviation between the generated and reference geometries, with lower values indicating higher accuracy of the generated conformations.

For QM9, the data points are clustered at the lower end of the RMSD scale, with a significant peak near 0.0 Å - 0.1 Å, indicating that the predicted conformations are very close to the reference geometries for most molecules. The distribution is heavily skewed towards the left, suggesting consistent high-quality conformer generation with minimal errors. A few data points extend towards higher RMSD values (up to 0.8 Å), representing more challenging cases with less accurate predictions.

In contrast, the Drugs dataset shows a wider spread of RMSD values, with the majority falling between 0.5 Å and 1.0 Å, reflecting the increased complexity and variability of the molecular structures compared to QM9. The distribution exhibits a peak around 0.75 Å, suggesting that the most common RMSD values fall within this range. Higher RMSD values, extending up to 2.0 Å, indicate cases where the model’s predictions deviate significantly from the reference structures.

The differences in RMSD distributions between QM9 and Drugs datasets highlight the challenges associated with predicting conformations for more complex molecules. While the model performs exceptionally well on the simpler QM9 dataset, it faces increased difficulty in accurately predicting geometries for the more diverse and intricate structures in the Drugs dataset.

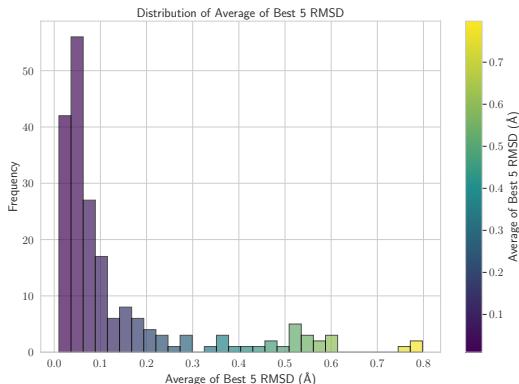
3.3. Scatter Plot Comparison

Figure 8 presents scatter plots comparing the RMSD values of the average of the best five generated conformers versus reference conformations from the GEOM QM9

810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

864

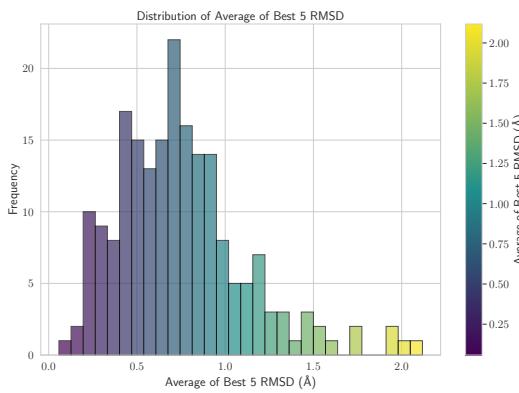
865



(a) RMSD Distribution for QM9 Dataset

877

878



(b) RMSD Distribution for Drugs Dataset

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

and Drugs test sets, each containing 200 examples. These plots visually illustrate how closely the generated conformers match the reference conformers.

For the QM9 dataset, a significant concentration of data points lies in the region where generated RMSD values are lower than reference RMSD values. Most generated RMSD values fall within a constrained range of 0.0 Å to 0.8 Å, while reference RMSD values range between 0.4 Å and 1.2 Å. This clustering demonstrates the AGDIFF model's high accuracy in predicting molecular geometries for the QM9 dataset, consistently generating conformers closely aligned with the reference structures.

In contrast, the scatter plot for the Drugs dataset reveals a broader distribution of RMSD values, with a significant number of points indicating comparable generated and reference RMSD values. However, the wider spread of points and higher RMSD values, extending to approximately 1.5 Å for generated RMSD and 2.5 Å for reference RMSD, highlight the increased complexity of drug-like molecules and the model's challenges in accurately predicting their conformations. The deviations from the ideal case reflect the

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

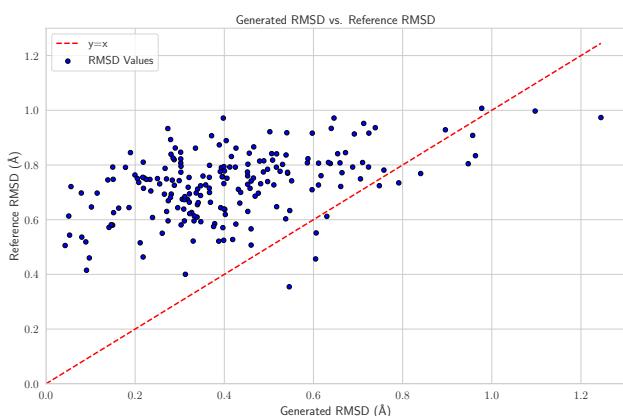
953

954

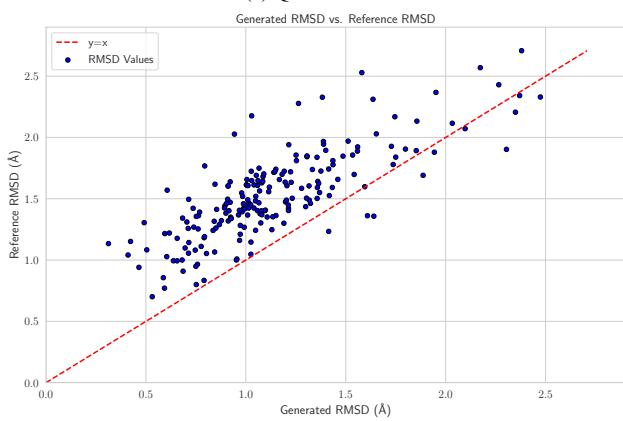
955

956

957



(a) QM9 dataset.



(b) Drugs dataset.

Figure 8. Scatter plots comparing the RMSD values of the average of the best 5 generated conformers versus reference conformations for the QM9 and Drugs datasets. Each dataset contains 200 examples. The red dashed line represents the ideal case where generated and reference RMSD values are equal.

variability in the model's performance, suggesting room for improvement in handling more complex and diverse molecular structures.

3.4. Correlation Matrix

Figure 9 shows the correlation coefficients between molecular properties and the average of the best 5 RMSD values for each molecule.

The correlation matrix for the Drugs dataset reveals significant relationships between specific molecular properties and the average of the best 5 RMSD values, highlighting the challenges in accurately predicting conformations for drug-like molecules. A high positive correlation of 0.57 between the number of atoms and RMSD suggests that larger molecules with more atoms tend to have higher RMSD values, indicating increased difficulty in predicting their conformations. Similarly, the number of rotatable bonds shows a strong positive correlation of 0.50 with RMSD, emphasizing the challenge posed by molecular flexibility. The

topological polar surface area (TPSA) also has a moderate positive correlation of 0.36, suggesting that molecules with larger polar surface areas are more complex to model accurately.

Furthermore, our observations indicate that the global encoder encounters more difficulties with the Drugs dataset than the QM9 dataset, likely due to the molecules' larger size and increased complexity. In contrast, LogP and the number of hydrogen donors display weaker correlations with RMSD, at 0.23 and 0.10, respectively, suggesting a minimal impact on conformer prediction accuracy. The number of hydrogen acceptors exhibits a moderate correlation of 0.34, indicating some influence on RMSD. These findings underscore the need for enhanced modeling techniques that can effectively manage the complexity introduced by larger, more flexible, and more polar molecules, which are prevalent in drug-like compounds.

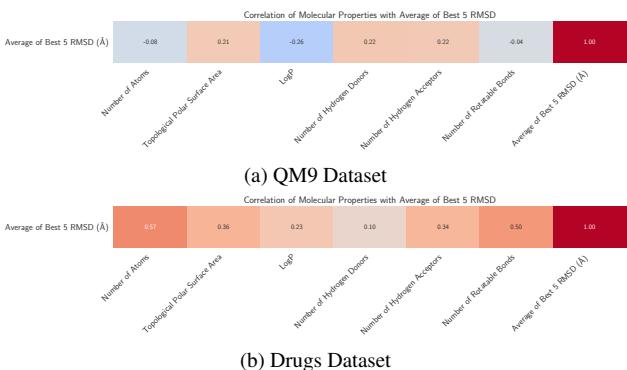


Figure 9. Correlation matrices displaying the Pearson correlation coefficients between molecular properties and the average of the best 5 RMSD values for the QM9 and Drugs datasets. The Drugs dataset (b) shows stronger correlations, particularly for the number of atoms, rotatable bonds, and topological polar surface area (TPSA), indicating challenges in predicting conformations for larger, more flexible, and more polar drug-like molecules.

3.5. Molecular Conformer Analysis

Alanine dipeptide is a small molecule commonly used as a benchmark system for studying the conformational preferences of peptides. To assess the quality and diversity of the conformations generated by AGDIFF, we analyze the Ramachandran plot of alanine dipeptide, which depicts the free energy landscape as a function of the dihedral angles Φ and Ψ .

Figure 12 presents the contour plot of the free energy landscape for alanine dipeptide, overlaid with 200 conformers generated by AGDIFF (represented by orange dots). The color gradient indicates the free energy, with blue regions corresponding to low free energy (more stable conformations) and red regions corresponding to high free energy (less stable conformations).

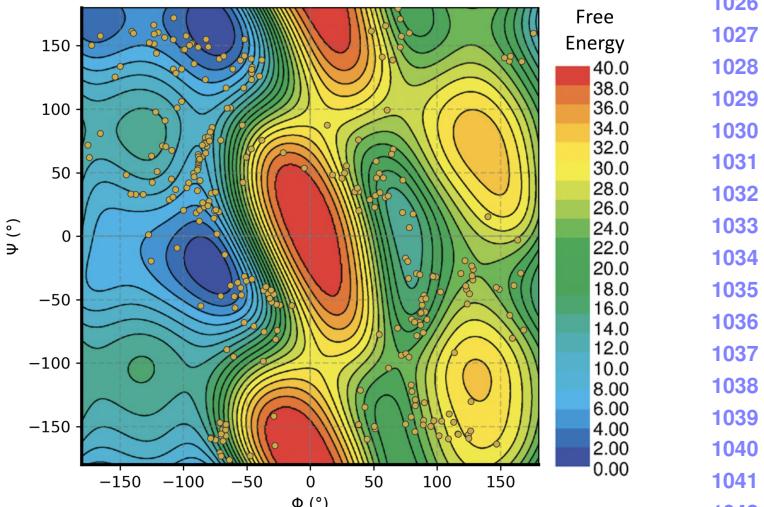


Figure 12. Ramachandran plot for alanine dipeptide, showing the free energy landscape as a function of dihedral angles Φ and Ψ . The plot is overlaid with 200 generated conformers (orange dots) from the AGDIFF model. The color gradient represents the free energy, with blue indicating low energy (stable conformations) and red indicating high energy (less stable conformations). The free energy landscape was obtained using the ff03 force field with TIP3P water, as reported by [36].

The generated conformers are predominantly clustered in the plot's blue and green regions, representing the energetically favorable alanine dipeptide conformations. This adequate coverage of the low free energy areas demonstrates the accuracy of AGDIFF in identifying stable molecular conformations. Moreover, the conformers are well-distributed across the entire free energy landscape, including regions of moderate free energy (yellow and green), indicating that AGDIFF explores a diverse set of conformations, capturing the most stable states and potentially relevant transient states. The dihedral angles Φ and Ψ are crucial for understanding the conformational flexibility of peptides. The generated conformers cover the entire range of Φ and Ψ angles, highlighting AGDIFF's ability to explore the entire conformational space of alanine dipeptide. This comprehensive sampling is particularly valuable for chemists studying the conformational behavior of molecules, as it provides insights into the molecule's stability, reactivity, and dynamic properties.

Figure 10 showcases four representative conformations of alanine dipeptide generated by AGDIFF, each representing a distinct conformational state. These examples demonstrate the ability of AGDIFF to generate diverse and physically realistic conformations for alanine dipeptide, covering various regions of the conformational space. Similarly, Figure 11 presents generated conformations for drug-like molecules from the GEOM-Drugs dataset, showcasing the diversity and complexity of the molecules and the ability of

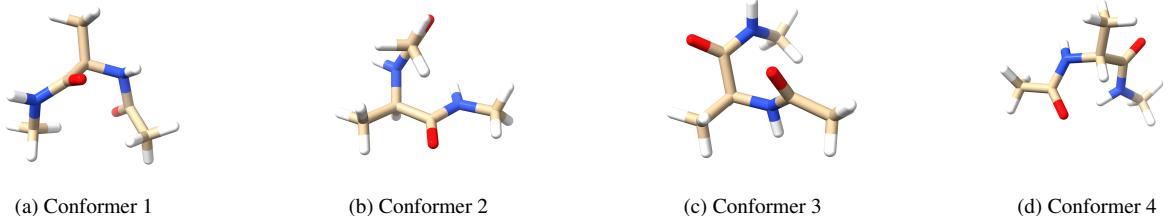


Figure 10. Generated conformations for alanine dipeptide using the AGDIFF model. Each subfigure displays a predicted 3D structure of alanine dipeptide, representing a distinct conformational state.

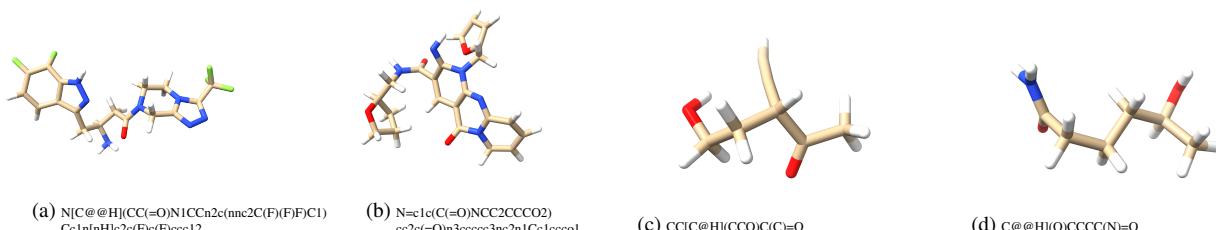


Figure 11. Generated conformations for drug-like molecules from the GEOM-Drugs dataset. Each subfigure displays a 3D structure of a predicted molecular conformation, along with its corresponding SMILES string representation. These examples demonstrate the diversity and complexity of the molecules in the GEOM-Drugs dataset and showcase the ability of the AGDIFF model to generate realistic and chemically valid conformations for drug-like compounds.

AGDIFF to generate realistic and chemically valid conformations for these compounds.

4. Conclusions

We introduced AGDIFF, a novel computational framework that leverages diffusion models for efficient and accurate molecular structure prediction. By extending the GeoDiff model with enhancements to the global, local, and edge encoders, AGDIFF demonstrates superior performance on both the GEOM-QM9 and GEOM-Drugs datasets compared to existing baselines. The incorporation of attention mechanisms, improved SchNet architecture, batch normalization, and feature expansion techniques enables AGDIFF to learn rich and expressive molecular representations while maintaining computational efficiency.

The experimental results highlight the effectiveness of AGDIFF in generating high-quality and diverse molecular conformations. On the GEOM-QM9 dataset, AGDIFF achieves a mean COV-R of 93.08% and a mean MAT-R of 0.1965 Å, surpassing the GeoDiff baselines. Similarly, on the more challenging GEOM-Drugs dataset, AGDIFF attains a median COV-R of 100.00% and a mean MAT-R of 0.8237 Å, demonstrating its ability to capture the conformational space of complex drug-like molecules.

The qualitative analysis of generated conformations, such as the Ramachandran plot for alanine dipeptide and the visualization of drug molecule predictions, further validates the quality and plausibility of the structures generated by AGDIFF. These results underscore the potential of AGDIFF

to advance molecular modeling techniques and contribute to fields such as computational chemistry, drug discovery, and materials design.

Future work could explore the integration of AGDIFF with other molecular property prediction tasks, such as binding affinity, to provide a more comprehensive tool for drug discovery pipelines. Additionally, investigating the interpretability of the learned molecular representations and the incorporation of domain-specific knowledge could further enhance the model’s performance and applicability.

AGDIFF represents a step forward in molecular structure prediction, offering a powerful and efficient framework for generating accurate and diverse molecular conformations. By leveraging the strengths of diffusion models and introducing novel enhancements to the model architecture, AGDIFF has the potential to accelerate drug discovery efforts and advance our understanding of molecular systems.

References

- [1] P. Li, J. Wang, Y. Qiao, H. Chen, Y. Yu, X. Yao, P. Gao, G. Xie, and S. Song, “Learn molecular representations from large-scale unlabeled molecules for drug discovery,” *ArXiv*, vol. abs/2012.11175, 2020.
- [2] G. Landrum, “Rdkit documentation,” *Release*, vol. 1, no. 1-79, p. 4, 2013.
- [3] M. Xu, L. Yu, Y. Song, C. Shi, S. Ermon, and J. Tang, “Geodiff: A geometric diffusion model for molecular conformation generation,” in *International*

- 1188 Conference on Learning Representations, 2022. 1242
 1189 [Online]. Available: <https://openreview.net/forum?id=PzcvxEMzvQC> 1243
 1190 1244
 1191 1245
- 1192 [4] R. R. Pissurlenkar, M. S. Shaikh, R. P. Iyer, and E. C. 1246
 1193 Coutinho, "Molecular mechanics force fields and their 1247
 1194 applications in drug design," *Anti-Infective Agents* 1248
 1195 in *Medicinal Chemistry (Formerly Current Medicinal* 1249
 1196 *Chemistry-Anti-Infective Agents)*, vol. 8, no. 2, pp. 1250
 1197 128–150, 2009. 1251
 1198 1252
- 1199 [5] R. A Bryce and I. H Hillier, "Quantum chemical app- 1253
 1200 roaches: Semiempirical molecular orbital and hy- 1254
 1201 brid quantum mechanical/molecular mechanical tech- 1255
 1202 niques," *Current Pharmaceutical Design*, vol. 20, 1256
 1203 no. 20, pp. 3293–3302, 2014. 1257
 1204 1258
- 1205 [6] J. P. McNamara and I. H. Hillier, "Semi-empirical 1259
 1206 molecular orbital methods including dispersion cor- 1260
 1207 rections for the accurate prediction of the full range of 1261
 1208 intermolecular interactions in biomolecules," *Physical* 1262
 1209 *chemistry chemical physics*, vol. 9, no. 19, pp. 2362– 1263
 1210 2370, 2007. 1264
 1211 1265
- 1212 [7] S. R. Wilson, W. Cui, J. W. Moskowitz, and K. E. 1266
 1213 Schmidt, "Applications of simulated annealing to the 1267
 1214 conformational analysis of flexible molecules," *Journal* 1268
 1215 *of computational chemistry*, vol. 12, no. 3, pp. 1269
 1216 342–349, 1991. 1270
 1217 1271
- 1218 [8] C. Parish, R. Lombardi, K. Sinclair, E. Smith, 1272
 1219 A. Goldberg, M. Rappleye, and M. Dure, "A compari- 1273
 1220 son of the low mode and monte carlo conformational 1274
 1221 search methods," *Journal of Molecular Graphics and* 1275
 1222 *Modelling*, vol. 21, no. 2, pp. 129–150, 2002. 1276
- 1223 [9] S. M. LaPointe and D. F. Weaver, "A review of 1277
 1224 density functional theory quantum mechanics as ap- 1278
 1225 plied to pharmaceutically relevant systems," *Current* 1279
 1226 *Computer-Aided Drug Design*, vol. 3, no. 4, pp. 290– 1280
 1227 296, 2007. 1281
 1228 1282
- 1229 [10] L. E. Ratcliff, S. Mohr, G. Huhs, T. Deutsch, 1283
 1230 M. Masella, and L. Genovese, "Challenges in large 1284
 1231 scale quantum mechanical calculations," *Wiley Inter-* 1285
 1232 *disciplinary Reviews: Computational Molecular Sci-* 1286
 1233 *ence*, vol. 7, no. 1, p. e1290, 2017. 1287
- 1234 1288
- 1235 [11] Y. Xu, K. Lin, S. Wang, L. Wang, C. Cai, C. Song, 1289
 1236 L. Lai, and J. Pei, "Deep learning for molecular gen- 1290
 1237 eration," *Future medicinal chemistry*, vol. 11, no. 6, 1291
 1238 pp. 567–597, 2019. 1292
 1239 1293
- 1240 [12] A. Z. Dudek, T. Arodz, and J. Gálvez, "Computational 1294
 1241 methods in developing quantitative structure-activity 1295
- relationships (qsar): a review," *Combinatorial chem- 1242
 istry & high throughput screening*, vol. 9, no. 3, pp. 1243
 213–228, 2006. 1244
- [13] M. Randić, "Generalized molecular descriptors," *Journal of Mathematical Chemistry*, vol. 7, no. 1, pp. 1245
 155–168, 1991. 1246
- [14] O. A. Raevsky, "Physicochemical descriptors in prop- 1247
 erty-based drug design," *Mini reviews in medicinal* 1248
chemistry, vol. 4, no. 10, pp. 1041–1052, 2004. 1249
- [15] A. Nayarisseri, R. Khandelwal, P. Tanwar, M. Mad- 1250
 havi, D. Sharma, G. Thakur, A. Speck-Planche, and 1251
 S. K. Singh, "Artificial intelligence, big data and ma- 1252
 chine learning approaches in precision medicine & 1253
 drug discovery," *Current drug targets*, vol. 22, no. 6, 1254
 pp. 631–655, 2021. 1255
- [16] Z.-M. Win, A. M. Cheong, and W. S. Hopkins, "Us- 1256
 ing machine learning to predict partition coefficient 1257
 (log p) and distribution coefficient (log d) with molec- 1258
 ular descriptors and liquid chromatography retention 1259
 time," *Journal of Chemical Information and Model-* 1260
ing, vol. 63, no. 7, pp. 1906–1913, 2023. 1261
- [17] G. Scalia, C. A. Grambow, B. Pernici, Y.-P. Li, and 1262
 W. H. Green, "Evaluating scalable uncertainty estima- 1263
 tion methods for deep learning-based molecular prop- 1264
 erty prediction," *Journal of chemical information and* 1265
modeling, vol. 60, no. 6, pp. 2697–2717, 2020. 1266
- [18] O. Wieder, S. Kohlbacher, M. Kuenemann, A. Garon, 1267
 P. Ducrot, T. Seidel, and T. Langer, "A compact re- 1268
 view of molecular property prediction with graph neu- 1269
 ral networks," *Drug Discovery Today: Technologies*, 1270
 vol. 37, pp. 1–12, 2020. 1271
- [19] H.-C. Yi, Z.-H. You, D.-S. Huang, and C. K. 1272
 Kwoh, "Graph representation learning in bioinformat- 1273
 ics: trends, methods and applications," *Briefings in* 1274
Bioinformatics, vol. 23, no. 1, p. bbab340, 2022. 1275
- [20] Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, 1276
 Z. Li, X. Luo, K. Chen, H. Jiang, and M. Zheng, 1277
 "Pushing the boundaries of molecular representation 1278
 for drug discovery with graph attention mechanism." 1279
Journal of medicinal chemistry, 2020. 1280
- [21] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, 1281
 "Diffusion models in vision: A survey," *IEEE Trans-* 1282
actions on Pattern Analysis and Machine Intelligence, 1283
 2023. 1284
- [22] S. Luo and W. Hu, "Diffusion probabilistic models 1285
 for 3d point cloud generation," in *Proceedings of the* 1286
IEEE/CVF Conference on Computer Vision and Pat- 1287
tern Recognition, 2021, pp. 2837–2845. 1288

- 1296 [23] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan,
 1297 and S. Ganguli, “Deep unsupervised learning
 1298 using nonequilibrium thermodynamics,” *CoRR*,
 1299 vol. abs/1503.03585, 2015. [Online]. Available:
 1300 <http://arxiv.org/abs/1503.03585>
 1301
- 1302 [24] Y.-L. Liao and T. Smidt, “Equiformer: Equivariant
 1303 graph attention transformer for 3d atomistic graphs,”
 1304 *arXiv preprint arXiv:2206.11990*, 2022.
 1305
- 1306 [25] Y.-L. Liao, B. Wood, A. Das, and T. Smidt,
 1307 “EquiformerV2: Improved equivariant transformer
 1308 for scaling to higher-degree representations,” *arXiv
 1309 preprint arXiv:2306.12059*, 2023.
- 1310 [26] V. G. Satorras, E. Hoogeboom, and M. Welling, “E (n)
 1311 equivariant graph neural networks,” in *International
 1312 conference on machine learning*. PMLR, 2021, pp.
 1313 9323–9332.
- 1314 [27] E. Hoogeboom, V. G. Satorras, C. Vignac, and
 1315 M. Welling, “Equivariant diffusion for molecule gen-
 1316 eration in 3d,” *ArXiv*, vol. abs/2203.17003, 2022.
 1317
- 1318 [28] Y. Liao and T. Smidt, “Equiformer: Equivariant graph
 1319 attention transformer for 3d atomistic graphs,” *ArXiv*,
 1320 vol. abs/2206.11990, 2022.
 1321
- 1322 [29] K. Schütt, P.-J. Kindermans, H. E. Sauceda Fe-
 1323 lix, S. Chmiela, A. Tkatchenko, and K.-R. Müller,
 1324 “Schnet: A continuous-filter convolutional neural net-
 1325 work for modeling quantum interactions,” *Advances in
 1326 neural information processing systems*, vol. 30, 2017.
 1327
- 1328 [30] W. Hu, B. Liu, J. Gomes, M. Zitnik, P. Liang,
 1329 V. Pande, and J. Leskovec, “Strategies for pre-
 1330 training graph neural networks,” *arXiv preprint
 1331 arXiv:1905.12265*, 2019.
- 1332 [31] D. Hendrycks and K. Gimpel, “Gaussian error lin-
 1333 ear units (gelus),” *arXiv preprint arXiv:1606.08415*,
 1334 2016.
- 1335 [32] S. Axelrod and R. Gomez-Bombarelli, “Geom,
 1336 energy-annotated molecular conformations for prop-
 1337 erty prediction and molecular generation,” *Scientific
 1338 Data*, vol. 9, no. 1, p. 185, 2022.
- 1339 [33] W. Kabsch, “A solution for the best rotation to
 1340 relate two sets of vectors,” *Acta Crystallographica Sec-
 1341 tion A: Crystal Physics, Diffraction, Theoretical and
 1342 General Crystallography*, vol. 32, no. 5, pp. 922–923,
 1343 1976.
- 1344 [34] M. Xu, S. Luo, Y. Bengio, J. Peng, and J. Tang,
 1345 “Learning neural generative dynamics for molecu-
 1346 lar conformation generation,” in *International Confer-
 1347 ence on Learning Representations*, 2021.
- 1348
- 1349
- 1350 [35] O.-E. Ganea, L. Pattanaik, C. W. Coley, R. Barzi-
 1351 lay, K. F. Jensen, W. H. Green, and T. S.
 1352 Jaakkola, “Geomol: Torsional geometric generation
 1353 of molecular 3d conformer ensembles,” *arXiv preprint
 1354 arXiv:2106.07802*, 2021.
- 1355
- 1356 [36] J. Vymetal and J. Vondrasek, “Metadynamics as a
 1357 tool for mapping the conformational and free-energy
 1358 space of peptides the alanine dipeptide case study,”
 1359 *The Journal of Physical Chemistry B*, vol. 114, no. 16,
 1360 pp. 5632–5642, 2010.
- 1361
- 1362
- 1363
- 1364
- 1365
- 1366
- 1367
- 1368
- 1369
- 1370
- 1371
- 1372
- 1373
- 1374
- 1375
- 1376
- 1377
- 1378
- 1379
- 1380
- 1381
- 1382
- 1383
- 1384
- 1385
- 1386
- 1387
- 1388
- 1389
- 1390
- 1391
- 1392
- 1393
- 1394
- 1395
- 1396
- 1397
- 1398
- 1399
- 1400
- 1401
- 1402
- 1403