Stereochemistry-aware string-based molecular generation

PNAS NEXUS

PNAS Nexus, 2025, 4, pgaf329
https://doi.org/10.1093/pnasnexus/pgaf329
Advance access publication 14 October 2025
Research Report

Stereochemistry-aware string-based molecular generation

Gary Tom [a,b,*], Edwin Yu [a], Naruki Yoshikawa [b,c], Kjell Jorner [a,c,d,e,*] and Alán Aspuru-Guzik [a,b,c,f,g,h,*]

[a]Department of Chemistry, University of Toronto, 80 St. George Street, Toronto, ON M5S 3H6, Canada
[b]Vector Institute for Artificial Intelligence, 108 College Street, Toronto, ON M5G 0C6, Canada
[c]Department of Computer Science, University of Toronto, 40 St George Street, Toronto, ON M5S 2E4, Canada
[d]Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Kemigården 4, Gothenburg SE-412 96, Sweden
[e]Department of Chemistry and Applied Biosciences, Institute of Chemical and Bioengineering, ETH Zurich, Vladimir-Prelog-Weg 1, Zurich CH-8093, Switzerland
[f]Department of Chemical Engineering & Applied Chemistry, University of Toronto, 200 College Street, Toronto, ON M5S 3E5, Canada
[g]Department of Materials Science & Engineering, University of Toronto, 184 College Street, Toronto, ON M5S 3E4, Canada
[h]Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), 661 University Avenue, Toronto, ON M5G 1M1, Canada
*To whom correspondence should be addressed: Email: gary.tom@mail.utoronto.ca (G.T.); Email: kjell.jorner@chem.ethz.ch (K.J.); Email: alan@aspuru.com (A.A.-G.)
Edited By Christopher Dupont

Abstract

This study investigates the impact of incorporating stereochemical information, a crucial aspect of computational drug discovery and materials design, in molecular generative modeling. We present a detailed comparison of stereochemistry-aware and conventionally stereochemistry-unaware string-based generative approaches, utilizing both genetic algorithms and reinforcement learning-based techniques. To evaluate these models, we introduce novel benchmarks specifically designed to assess the importance of stereochemistry-aware generative modeling. Our results demonstrate that stereochemistry-aware models generally perform on par with or surpass conventional algorithms across various stereochemistry-sensitive tasks. However, we also observe that in scenarios where stereochemistry plays a less critical role, stereochemistry-aware models may face challenges due to the increased complexity of the chemical space they must navigate. This work provides insights into the trade-offs involved in incorporating stereochemical information in molecular generative models and offers guidance for selecting appropriate approaches based on specific application requirements.

Keywords: molecular generation, stereochemistry, generative modeling, drug design, machine learning

Significance Statement

This research explores the incorporation of stereochemistry—the relative 3D arrangement of atoms of the molecule—into machine learning algorithms to better design molecules. We compare different modeling approaches, and their performance in optimizing stereochemistry-sensitive properties, such as structure similarity, drug activity, and optical activity. We find that explicitly considering stereochemistry can improve optimization in certain tasks. This work provides a benchmark for evaluating and developing stereochemistry-aware generative models for molecular design, and will help create more effective ways to find new medicines and materials.

Introduction

Generative models have become increasingly prominent in the fields of inverse design and molecular discovery, offering a computational approach to explore vast chemical spaces efficiently (1–10). These models employ machine learning techniques to generate molecular structures with targeted properties, potentially expediting the traditionally lengthy and resource-intensive process of molecular design (11–14). Generative models can propose new and potentially viable compounds, adhering to specified criteria. Methods such as genetic algorithms define heuristics for exploring the space of chemicals, while deep-learning methods learn the chemical space distribution from molecular databases. The literature presents a diverse array of approaches in this domain, including but not limited to variational autoencoders (VAEs) (2, 15–19), generative adversarial networks (GANs) (4, 20, 21), reinforcement learning (RL) (22–27), genetic algorithms (GAs) (28–31), and transformer-based architectures (32–35). These methodologies have demonstrated utility across various

Competing Interest: The authors declare no competing interests.
Received: July 21, 2025. **Accepted:** September 28, 2025

applications in drug discovery and materials science, facilitating rapid in silico screening and optimization of molecular structures (36).

The evaluation and benchmarking of generative models for molecular discovery initially focused on determining the goodness of the reproduction of the structures in the dataset chemical space—generation not conditioned on the functional properties of the molecules. These metrics typically emphasize distribution learning, examining the model's ability to capture and reproduce the underlying distribution of the training data (3, 37–39). Other evaluation criteria include the (i) novelty of generated molecules, which measures the proportion of unique structures not present in the training set; (ii) diversity, which assesses the structural variation among the generated molecules; and validity, which ensures that the proposed structures adhere to chemical feasibility constraints (e.g. valid Lewis structures and valency constraints) (18, 40).

While these metrics provide insights into a model's generative capabilities, there is a growing recognition of the need for more realistic and task-specific benchmarks (41–44). The emphasis on general distribution learning, while important, may not fully capture the model's performance in addressing specific chemical challenges. Additionally, performances on task-oriented benchmarks based on simple heuristic fitness functions, such as penalized log water-octanol partition coefficient (2, 45), similarity/rediscovery tasks (18, 40), or quantitative estimate of drug-likeliness (QED) (46), are handily maximized by modern generative models (47–49), and even trivially satisfied by randomly inserting carbon atoms into the molecules (50). These simplistic fitness functions often fail to capture chemical constraints, allowing models to exploit failure modes by reward hacking, and generate molecules with high scores but undesirable properties, such as chemical instability or synthetic infeasibility (51). As the field advances, there is an increasing demand for benchmarks that are more closely aligned with real-world applications in drug discovery, materials design, and other domains of chemistry (52). This shift towards more targeted evaluation methods would provide a more nuanced and practically relevant assessment of generative models, potentially accelerating their adoption and impact in real-world molecular discovery scenarios.

Despite the advances in generative models for molecular design, the incorporation of stereochemical information and the resulting effects on molecular optimization are often overlooked. Molecular stereochemistry, the relative 3D arrangement of atoms within a molecule, significantly influences its chemical properties and biological activity (53). Many current methods either ignore stereochemistry (2, 3, 40) or consider it as a postprocessing step after molecule generation (16, 42, 44). This approach is suboptimal, as stereochemistry plays a crucial role in determining a molecule's properties and biological activity. While certain graph-based generative methods are able to handle stereochemical information (29, 54–56), the lack of realistic benchmarking tasks that incorporate stereochemistry means that models are not evaluated on their ability to produce stereochemically optimal molecules. Furthermore, there is no direct comparison of methods with and without stereochemistry-awareness on these stereochemistry-sensitive tasks.

The importance of stereochemistry is particularly evident in drug discovery, where the spatial arrangement of atoms can significantly influence a compound's pharmacological properties (57, 58). Properties such as binding affinity to target proteins, metabolic stability, and toxicity can be profoundly affected by stereochemistry. For example, the synthesis of methadone produces racemic mixtures of enantiomers—molecules that are mirror images of each other—(R)-methadone and (S)-methadone. While (R)-methadone acts as an opioid for pain relief, (S)-methadone has been identified to bind to the hERG protein and can lead to severe side-effects, such as heart attacks or cardiac arrest (59). In materials science, stereochemistry can impact crystal packing, optical properties, synthesis, and reactivity (60–63). By not explicitly accounting for stereochemistry during the generative process, models may overlook critical aspects of molecular behavior, potentially leading to inefficiencies in the discovery pipeline and missed opportunities for identifying optimal candidates for a given application.

In our work, we study the effects of stereochemistry on string-based generative models. We evaluate the models, both with and without stereochemistry-awareness, on a variety of molecular design tasks that are sensitive to the stereochemistry of molecules. Additionally, we explore different string representations of molecular graphs, and create a workflow for benchmarking the models, which includes a novel fitness function based on the circular dichroism spectra of molecules. We find that stereo-aware models perform as well as, or better than nonstereo models, but the performance increase of the stereo models are dependent on the sensitivity of the task to stereochemistry, and the additional complexity in the chemical search space due to the inclusion of stereoisomers. The models and the fitness functions are all available at https://github.com/aspuru-guzik-group/stereogeneration.

Results

To study the effects of stereochemistry on molecular generative models, we implement RL and GA methods, which have been shown to be strong baselines for molecular generation tasks (25, 31, 41). We modify the REINVENT (22) and JANUS (30) methods to permit the representation of stereochemical information. In these models, the molecular graphs are represented as strings, where REINVENT uses Simplified Molecular-Input Line-Entry System (SMILES) (64), and JANUS uses SELF-Referencing Embedded Strings (SELFIES) (65), or GroupSELFIES (66), which we dub GroupJANUS. We choose to use string-based generative models due to their expressiveness and flexibility in exploring chemical space when compared to graph-based methods (38), and their native support of stereochemical string tokens. By directly comparing the models with and without the stereochemistry-awareness across the various tasks, we can elucidate the effect of stereochemistry in the molecular generation process. Details of the models are found in the Methods section.

Stereochemistry

We focus on two primary forms of stereoisomerism: E/Z geometric diastereomers, arising from restricted rotation around double bonds, and R/S diastereomers and enantiomers, determined by the arrangement of substituents around chiral centers. Enantiomers are nonsuperimposable mirror images of each other and often have different optical activity and physical properties. Diastereomers, stereoisomers that are not mirror images of each other, also often exhibit different physical and chemical properties.

While we incorporate E/Z and R/S isomerism, we do not explicitly account for axial chirality, a type of chirality arising from hindered rotation around single bonds (67), or ring isomers. This omission limits our model's ability to generate and differentiate atropisomers, a specific class of axially chiral molecules. We also do not account for nontetrahedral forms of isomerisms,

which are mostly relevant for transition metals and are not within the scope of our benchmark.

String representations

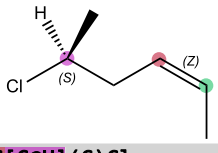
SMILES were initially created as a compact representation of molecular graphs for purposes of database retrieval, and substructure searching. When used in generative models, SMILES of generated molecules can sometimes violate the grammar of the representation, resulting in invalid SMILES. To address this, SELFIES made use of overloaded tokens, and local definitions of rings and branches to create a robust representation that will always translate to a valid molecular graph. GroupSELFIES further extended SELFIES by allowing for custom tokens which can encode groups with specified attachment points. For more details on string representations of molecules, we direct the readers to Krenn et al. (68). We also note that there are other string representations that incorporate stereochemistry which are not explored in this work (69).

All three representations natively encode stereochemical information (Fig. 1). SMILES encode counter-clockwise and clockwise chirality with "@" and "@@" tokens, respectively. *E*–*Z* stereoisomers are denoted with "\" and "/" before the characters to indicate the position of a bond relative to an adjacent double bond. The same characters are used in the SELFIES stereochemical tokens, while also maintaining the robustness of the representation. GroupSELFIES defines *E*–*Z* stereoisomers in the same way as SMILES and SELFIES but defines chirality through unique tokens for each chiral center and for all possible attachment points. The attachment points directly encode the chirality of the chiral center, with different attachment indices in the tokens specifying the order of substituents around the chiral center.

For all experiments, we use a subset of the ZINC15 database that was randomly sub-sampled by Gomez-Bombarelli et al. (2, 70). This dataset is composed of about 250,000 commercially available drug-like molecules. Stereoinformation is defined for most molecules in the dataset. Any molecules with ambiguous stereochemistry are assigned stereochemistry by randomly selecting from a list enumerating all unspecified stereocenters using RDKit cheminformatics software (71). For the nonstereo experiments, the stereoinformation is discarded, and duplicates resulting from the loss of stereoinformation are removed. Subsequently, the unique string tokens are collected to create an alphabet, with stereo and nonstereo alphabets for each representation. The GroupSELFIES representation has an additional essential set of chiral tokens, which are appended to the alphabet generated from the dataset.

Experiments

We perform three stereochemistry-sensitive generative experiments to benchmark the models. We study REINVENT, JANUS, and GroupJANUS with SMILES, SELFIES, and GroupSELFIES representations, respectively. Both the stereo-aware and nonstereo-aware models can generate molecules with unspecified stereocenters. For stereo-aware models, all unspecified stereocenters are randomly assigned, while maintaining any stereocenters that were specified by the model. The modified string is then used as the input for the model in subsequent generations. For nonstereo models, molecules with undefined stereocenters are only assigned stereochemical information for fitness evaluation, which simulates the postgeneration treatment of stereochemistry. However, the stereochemistry is not fed back into the nonstereo-aware generative model. While stereo models can output molecules with stereochemical tokens, nonstereo models cannot generate such tokens or distinguish between different stereoisomers with the same atomic connectivity. All models are not permitted to generate repeated compounds observed in current or previous generations, hence, for the nonstereo model, the assigned stereochemistry for a compound will not be resampled.



SMILES:	C/C=C\C[C@H](C)Cl
SELFIES:	[C][/C][=C][\C][C@H1][Branch1][C][C][Cl]
Group SELFIES:	[C][/C][=C][\C][:0chiral][Ring1][C][pop] [Ring1][Cl][pop]

Fig. 1. Example of isomeric molecule encoded with SMILES, SELFIES, and GroupSELFIES.

We evaluate optimization performance by looking at the optimization trace, which plots the cumulative top-1 score achieved as a function of the generation of the campaign. We include cumulative top-10 and top-100 traces as well in Figs. S10–S12. We do this across the models, stereo- and nonstereo-aware, on the aforementioned tasks. Further experimentation details provided in Methods.

Additionally, we use the area-under-curve (AUC) of the optimization traces as a quantitative measure of the optimization performance. For the AUC calculation, the number of generations is normalized from 0 to 1. For the rediscovery tasks, the similarity score and the AUC are both bounded by 0 and 1. For the docking and CD tasks, there is no maximum achievable score. Therefore, we normalize the AUC scores by the best score in the initial ZINC dataset. Higher AUC indicates the generation of higher scoring molecules, and also earlier discovery of such molecules. The AUC scores are found in Table 1. Alternatively, we report the maximum top-1 score and fraction of top-100 identified in Tables S1 and S2. We also report the AUC for the top-10 and top-100 optimization traces in Tables S3 and S4, respectively.

Stereoisomer rediscovery task

Rediscovery tasks in molecular generative modeling benchmarking aim to evaluate a model's ability to recreate the structure of known molecules. The structural similarity is measured by the Tanimoto similarity of molecular fingerprints—typically, extended circular fingerprints (ECFPs), bit vectors based on the topological features of a certain radius in the molecular graph (72). The model successfully rediscovers a target when the similarity is 1.0. While rediscovery tasks are not useful in practice, since the target molecules are known a priori, they serve as useful baselines to study the generative capabilities of the models in directly optimizing molecular structures, rather than chemical function. Previous benchmarking rediscovery tasks ignore the stereochemistry of the molecular structures (40). We include the stereoinformation as part of the target through the use of isomeric ECFPs. For this, we chose to perform rediscovery of (R)-albuterol (used in asthma treatment) and mestranol (used as estrogen medication for hormone therapy), with one and five chiral centers, respectively. The chemical structure of the compounds are shown in Fig. S1.

In the rediscovery scoring, nonstereo-aware methods first generate a molecule without explicit stereochemical information. To assign a score, stereochemistry is then considered postgeneration

Table 1. AUC of optimization traces for all tasks, for stereo and nonstereo-aware models.

	Tasks	REINVENT	JANUS	GroupJANUS
Nonstereo	(R)-albuterol rediscovery	0.487 ± 0.058	0.790 ± 0.105	0.840 ± 0.109
	Mestranol rediscovery	0.292 ± 0.034	0.633 ± 0.031	0.672 ± 0.032
	1SYH docking	0.900 ± 0.020	1.033 ± 0.031	1.084 ± 0.053
	1OYT docking	0.954 ± 0.013	1.064 ± 0.028	1.068 ± 0.028
	6Y2F docking	0.987 ± 0.015	1.068 ± 0.052	1.067 ± 0.029
	CD spectral peak score	0.413 ± 0.117	2.007 ± 0.352	2.066 ± 0.761
Stereo	(R)-albuterol rediscovery	0.403 ± 0.053	0.931 ± 0.044	0.923 ± 0.035
	Mestranol rediscovery	0.280 ± 0.032	0.843 ± 0.087	0.918 ± 0.074
	1SYH docking	0.887 ± 0.011	1.065 ± 0.031	1.106 ± 0.070
	1OYT docking	0.940 ± 0.021	1.099 ± 0.027	1.059 ± 0.035
	6Y2F docking	0.979 ± 0.023	1.088 ± 0.043	1.065 ± 0.042
	CD task	0.385 ± 0.111	2.884 ± 1.009	2.198 ± 0.563

The mean and SD are reported. Statistically significantly better (higher) AUC scores between the nonstereo and stereo variants are bolded.

by randomly sampling one stereoisomer from all possible stereoisomers of the generated molecular graph. This approach means that nonstereo-aware methods have the potential to achieve the maximum rediscovery score if the randomly sampled stereoisomer matches the target molecule.

The optimization traces for the rediscovery tasks are shown in Fig. 2. The higher number of chiral centers in mestranol make it a more difficult target for rediscovery. This is clearly shown in the optimization traces of the REINVENT models—mestranol rediscovery does not achieve similarity higher than the initial dataset. When compared to rediscovery in other studies (27, 40, 41), REINVENT optimization performance is greatly reduced when stereochemistry is introduced. There are no statistically significant differences in the performance of REINVENT when comparing stereo and nonstereo models. Samples of the top generated molecules of each method for (R)-albuterol and mestranol rediscovery tasks are shown in Figs. S4 and S5, respectively.

The stereo-aware JANUS and GroupJANUS models significantly outperform the nonstereo-aware models, indicating the ability of the stereo model in learning specific stereochemistries in molecular structures. For (R)-albuterol, both stereo GAs successfully rediscover the structure for all runs, with no significant differences between JANUS and GroupJANUS. For mestranol, the use of GroupSELFIES slightly improves the optimization for the stereo-aware model, when compared to JANUS with SELFIES.

Protein–ligand docking task

Protein–ligand interactions are associated with the bioactivity of drug molecules. Ligands are molecules that bind inside the protein binding pockets, forming intermolecular interactions with the amino acids of the protein, activating or inhibiting biological functions of the protein. For the benchmark, we use the high-throughput docking score implemented in the Tartarus benchmark (44), which uses the smina software to simulate the protein–ligand binding affinity (74).

Because the scoring function takes in a 3D conformer of the molecules, a conformer search is performed using RDKit to find the lowest energy conformer, respecting all specified stereoinformation, followed by energy relaxation with the Merck Molecule Force Field 94 (MMFF94) (75). The molecule is placed inside the binding pocket to sample binding poses; the resulting docking score is maximized. The binding pocket is defined as the bounding box encompassing the volume occupied by the protein's native ligand with 3 Å padding.

We perform the protein–ligand docking task for three different targets, visualized in Fig. 3 with their respective bounding boxes. Both 1SYH and 6Y2F are targets from Tartarus: 1SYH is associated with neurological diseases, and 6Y2F is responsible for the translation of the SARS-CoV-2 virus RNA. We also include the 1OYT protein, which is associated with blood coagulation (76), and has a binding pocket with a volume between those of 1SYH and 6Y2F. The structures for the native ligands of the proteins are shown in Fig. S2.

Moving beyond simple structural reproduction, the protein–ligand docking task assesses the practical utility of generative models in a drug discovery context. The optimization traces for the docking tasks are in Fig. 4. We again observe that REINVENT struggles to improve upon the results of the ZINC dataset, with the exception of the 6Y2F protein. There are no differences between the stereo and nonstereo variants of REINVENT.

Meanwhile, both GAs optimize better than REINVENT. For JANUS, we observe consistent improvements in optimization performance with stereo-aware models for generating ligands for 1SYH and 1OYT. The faster optimization of the stereo GAs are also reflected in the AUC score (Table 1). In the case of the 6Y2F target, possessing a comparatively larger and more flexible binding pocket, the difference in performance between stereo and nonstereo models was less pronounced. This observation implies that for certain targets, the impact of stereochemistry on binding affinity might be less critical, with other molecular features playing a dominant role. The increasing size of the generated ligands are seen in the chemical structures of the top scoring molecules, shown for each method and protein target in Figs. S6–S8.

Circular dichroism task

We finally developed a task based on circular dichroism (CD), which directly probes the chirality of structures, making it the ideal task for studying the effects of stereoinformation in molecular generation. CD produces spectra of the absorption of left- and right-handed polarized light in chemical species, and can be used to study folding structures in proteins (77), or chiral optical properties of materials, which have light manipulation and photonics applications (60, 78–80). Previous works have studied the ML prediction of CD spectra of molecules (81) and proteins (82), and the development of chiral molecular representations to predict the interaction between circularly polarized light and chiral compounds (83).

In this task, like before, the molecules are 3D embedded with RDKit. The conformer search and geometry optimization is performed using crest (84–86) and semiempirical extended tight-binding (xTB) (87) at the GFN2 level of theory (88). Representative spectra for example chiral molecules are shown

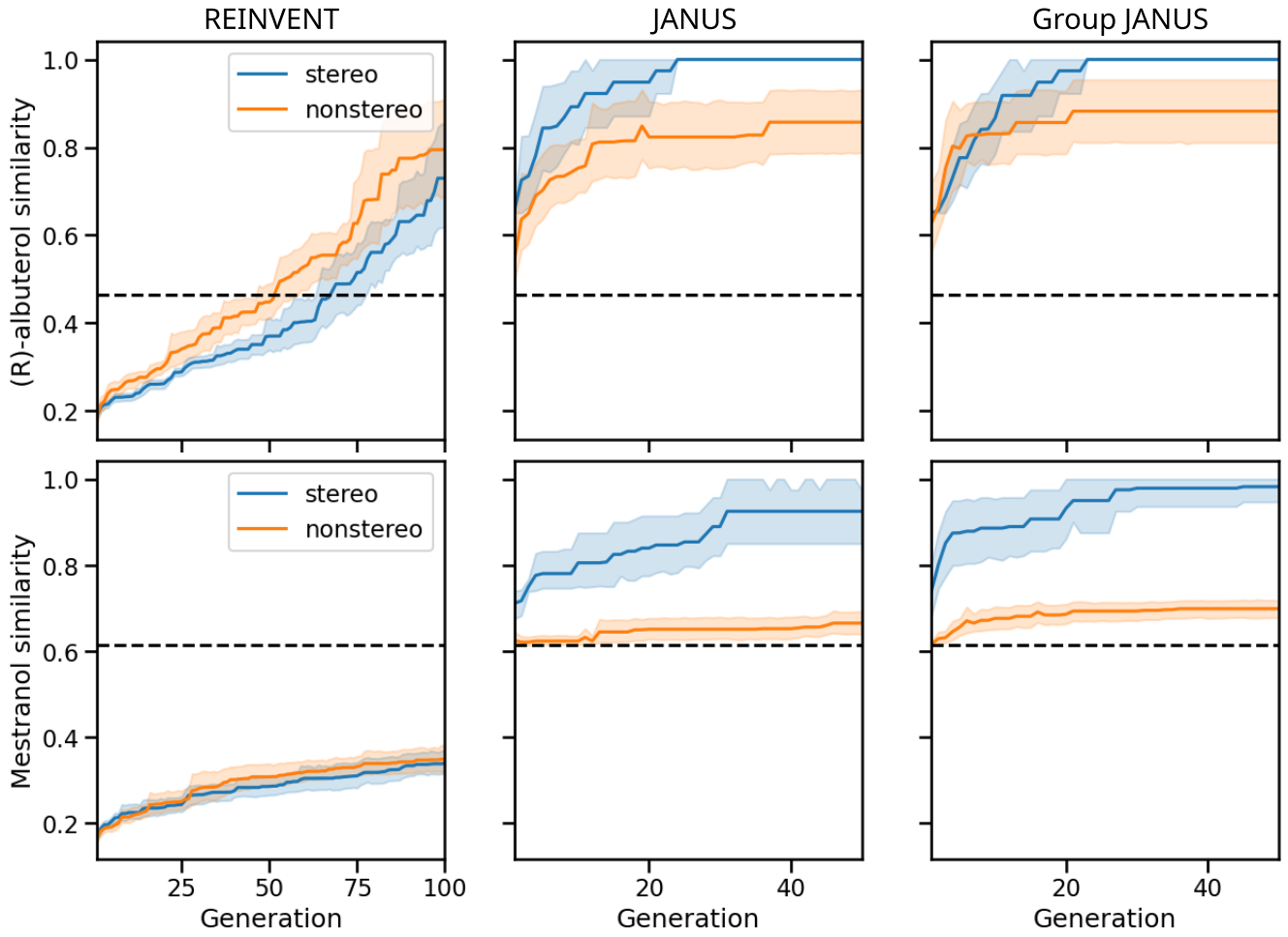


Fig. 2. Optimization traces for rediscovery tasks. The cumulative top-1 similarity score to the target molecule as a function of generation of optimization. Shaded regions indicate the 95% CI. The dashed line is the best score found in the starting dataset.

in Fig. S3. For a benchmark involving the generation and evaluation of thousands of molecules, computational efficiency is important. sTDA-xTB offers a balance between realistic simulations of chiroptical properties and computational cost, making it feasible for high-throughput screening within a generative model loop. While higher-level calculations would provide more accurate spectra, they would be computationally prohibitive for this scale of study. This computational approach is therefore well-suited, as the benchmark is explicitly stereochemistry-dependent and serves as a task for comparing the effect of including stereochemistry in the generation process.

The xTB calculation produces orbitals and orbital energies, which can be treated using simplified Tamm-Dancoff approximated (sTDA) time-dependent density functional theory (TD-DFT). This workflow, `sTDA-xTB`, produces CD spectra of the lowest energy conformers relatively quickly, even for molecular systems with hundreds of atoms (89, 90). A peak score is defined as the signed area under the spectrum for wavelengths 450–550 nm, a region where small organic molecules can have CD signals, and is also within the visible range for possible materials applications. Maximizing the peak score produces chiral optically active materials within the blue region of visible light.

The results of CD peak score optimization task are shown in Fig. 5. The structures of the top scoring CD peak score molecules are shown in Fig. S9. There are no differences between the nonstereo and stereo `REINVENT`, which are unable to improve upon the scores of the initial dataset. The `GroupJANUS` optimization also shows no difference between the stereo and nonstereo models. However, the stereo-aware `JANUS` is capable of generating molecules with stronger CD signals than the nonstereo-aware counterpart. The results indicate that the CD task is a suitably stereochemistry-sensitive optimization task for molecular generative modeling.

Discussion

While it might be hypothesized that incorporating more detailed physical and chemical information, such as stereochemistry, would lead to improved model performance on realistic chemical benchmarks, our findings show the practical impact is moderated by the task's intrinsic sensitivity to this information and the model's ability to effectively navigate the increased complexity—including an expanded search space and, for string-based methods, a larger action space—that accompanies such representations. For explicit optimization of molecular structures, stereo-aware GAs perform better than nonstereo counterparts. In the docking task, we observe that stereo GAs boost the optimization performance for 1SYH and 1OYT. In the case of 6Y2F, the generated ligand molecules are larger, in order to fit in the bigger protein binding pocket. Larger structural changes such as additional of fragments and functional groups allow the models to more quickly traverse the permitted molecule space, while slight changes in stereochemistry only result in small changes in the docking score. In these tasks, stereo models

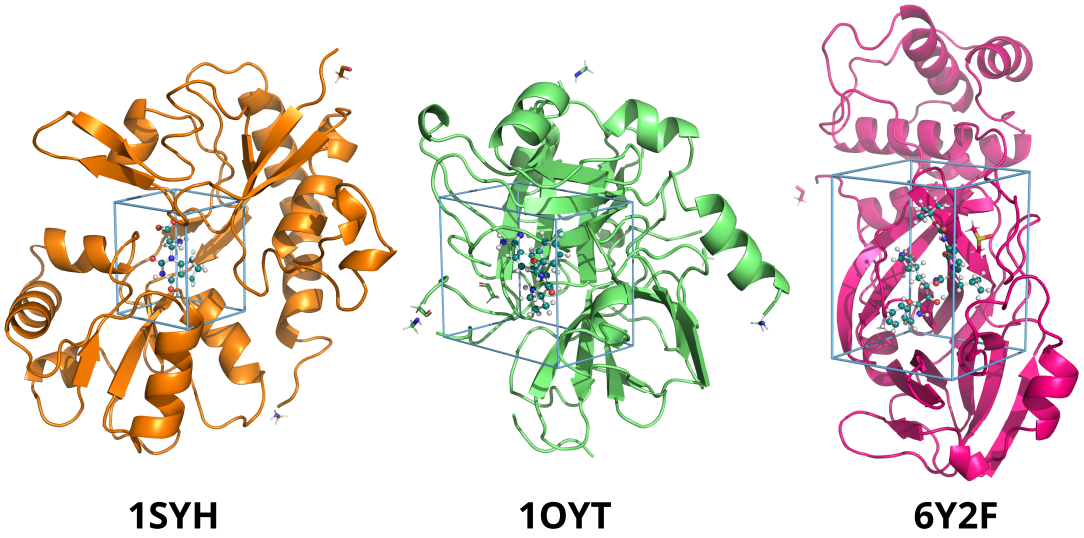


Fig. 3. Structures of proteins with native ligands. The structures are from the Protein Data Bank (73). The native ligand in the binding pocket is shown inside a bounding box.

still perform as well as nonstereo models. The CD spectra task directly probes the effects of chirality, and the spectra is less related to specific molecular size or functional groups. In this task, stereo JANUS outperforms nonstereo JANUS.

Except for the rediscovery tasks, unlike JANUS, the stereo and nonstereo variants of GroupJANUS perform similarly. This may be due to inefficiencies of the GroupSELFIES representation of stereochemistry. The addition of stereochemistry tokens increase the alphabet size by almost 10 times. Also, all additional tokens and group tokens are overloaded to ensure robustness. We hypothesize that the increased number of tokens interferes with the decoding of stereoisomeric GroupSELFIES, truncating molecules at rings and branches.

Additionally, GAs perform better than the REINVENT model for the same number of oracle calls, results which are consistent with previous studies (31). The evolutionary approach of GAs will always select the members of the population that maximize the fitness, meaning the GA cannot perform worse than the previous (or initial) generations. GAs are also not encumbered by the prior chemical space distribution of the training set, unlike deep learning methods like REINVENT, allowing the generation of more diverse molecules. Due to the prior model, the REINVENT agent requires more oracle calls and more frequent retraining to condition for higher rewards. No improvement was observed for stereo REINVENT. The expanded alphabet increases the number of possible actions the RL, requiring the model to learn a more complex policy which may have impeded stereo REINVENT.

Conclusion

This study presents a detailed investigation into the incorporation of stereochemical information within molecular generative models, focusing on established techniques such as string-based RL and GAs. We aim to provide a nuanced understanding of the impact of stereochemistry awareness by employing a suite of evaluation metrics, including both conventional benchmarks and newly designed tasks specifically tailored to assess the role of stereochemistry. A key contribution of this work is the introduction of a novel CD-based task, which proved to be suitable for probing the effects of chirality in the generated molecules.

Our findings highlight the importance of considering task-specific requirements when deciding whether to include stereochemical information within the generative process. In cases where different stereoisomers can significantly influence the desired molecular properties, the inclusion of stereochemistry led to improved performance. Specifically, we observed that stereochemistry-aware GA JANUS consistently outperformed their nonstereo counterparts in generating molecules for stereoisomer rediscovery, docking to proteins 1SYH and 1OYT, and CD spectra peak optimization.

However, our results also suggest that the benefits of incorporating stereochemistry are less pronounced in tasks where other molecular features, such as size or functional group presence, may play a more dominant role. This was seen in the protein–ligand docking task for target 6Y2F, where the impact of stereochemistry was less substantial relative to larger structural molecular changes, due to the bigger and less constraining binding pocket. The increased number of atoms and bonds of the molecules generated for the 6Y2F pocket naturally have more possible stereoisomers. Consequently, larger structural modifications, such as the addition of functional groups, can be more efficient in exploring the chemical space than smaller stereochemical adjustments. Specific to our string-based approaches, we observe that the addition of stereochemical tokens considerably increases the action space of the models, which may frustrate the optimization. For example, in GroupJANUS, the inclusion of stereochemical groups and chiral tokens increases the alphabet size by a factor of 10, greatly increasing the number of possible mutations in the GA. This observation underscores the need for a considered approach when deciding on the necessity of stereochemical information in generative models.

Extending the methodologies to encompass additional stereochemical complexities represents a valuable direction for future work. It is important to note that our evaluation primarily focused on E/Z and R/S isomerism. This scope was found to be sufficient for the drug-like small organic molecules of the ZINC dataset and the benchmarking tasks that we examined here. However, the explicit handling of other isomerisms, such as axial chirality or nontetrahedral isomerisms, was not within the scope of this study.

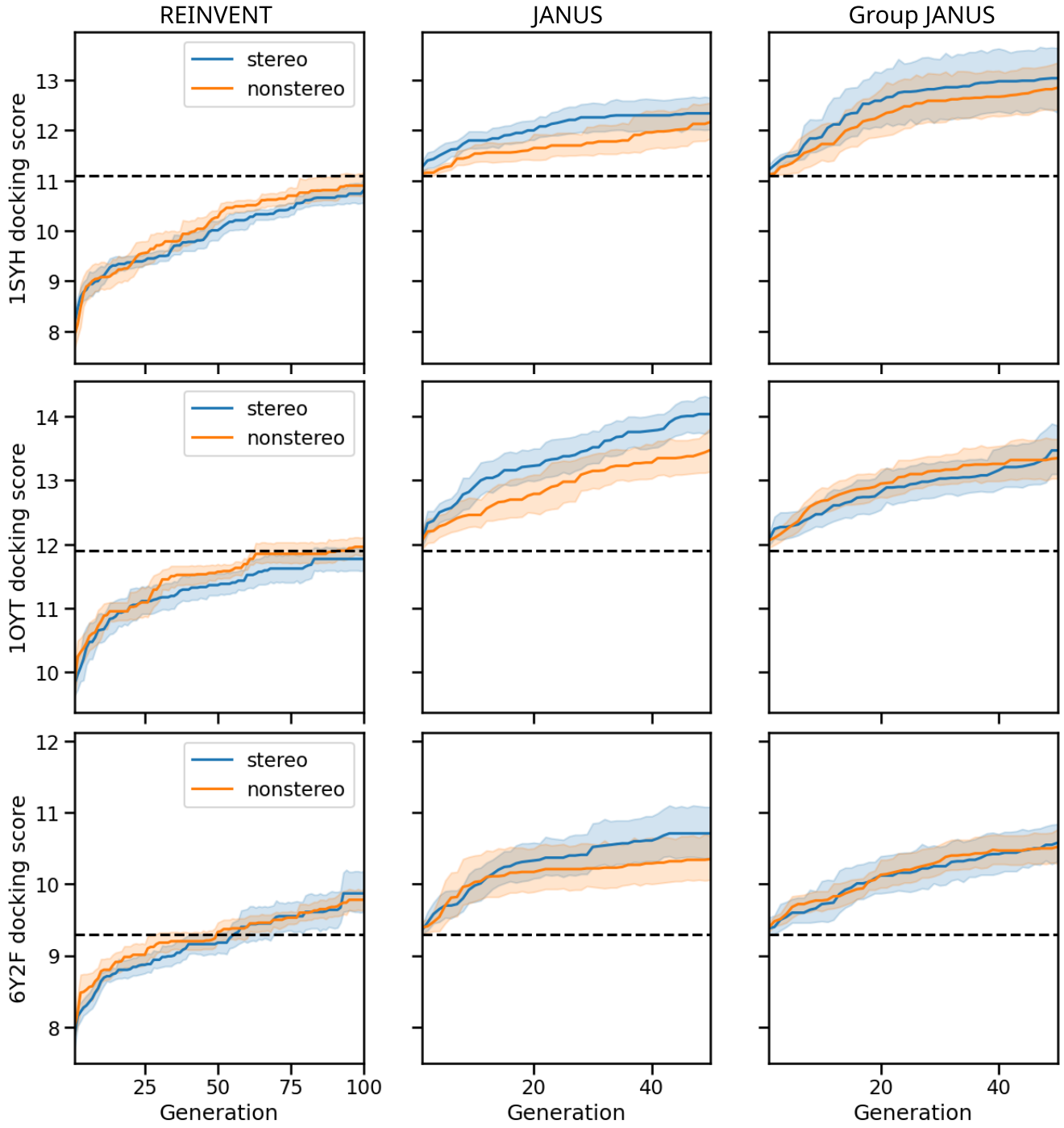


Fig. 4. Optimization traces for docking tasks. The cumulative top-1 docking score for protein targets as a function of generation of optimization. Shaded regions indicate the 95% CI. The dashed line is the best score found in the starting dataset.

Further investigation into more efficient and robust representations and handling of stereochemistry is warranted. For instance, while this study focused on string-based methods, models using other molecular representations, such as graphs (16, 29), voxel grids (91), or 3D representations (92, 93), offer alterative approaches. Our benchmark provides realistic stereochemistry-sensitive tasks to benchmark such models for further development. Regardless of the specific generative method or molecular representation employed, the fundamental trade-off observed—balancing the advantages in optimizing stereochemistry-sensitive tasks against the challenges posed by an expanded chemical space—will likely persist. Concurrently, for string-based deep learning models such as REINVENT, the development of architectures or training strategies better equipped to handle the additional stereochemical tokens within molecular strings may be necessary to fully harness their potential in stereoselective design.

In conclusion, this work offers insights into how string-based generative models handle stereochemical information for molecular design, outlining their capabilities and limitations. Our findings indicate that while incorporating such information can be beneficial, particularly when 3D molecular structure is a key

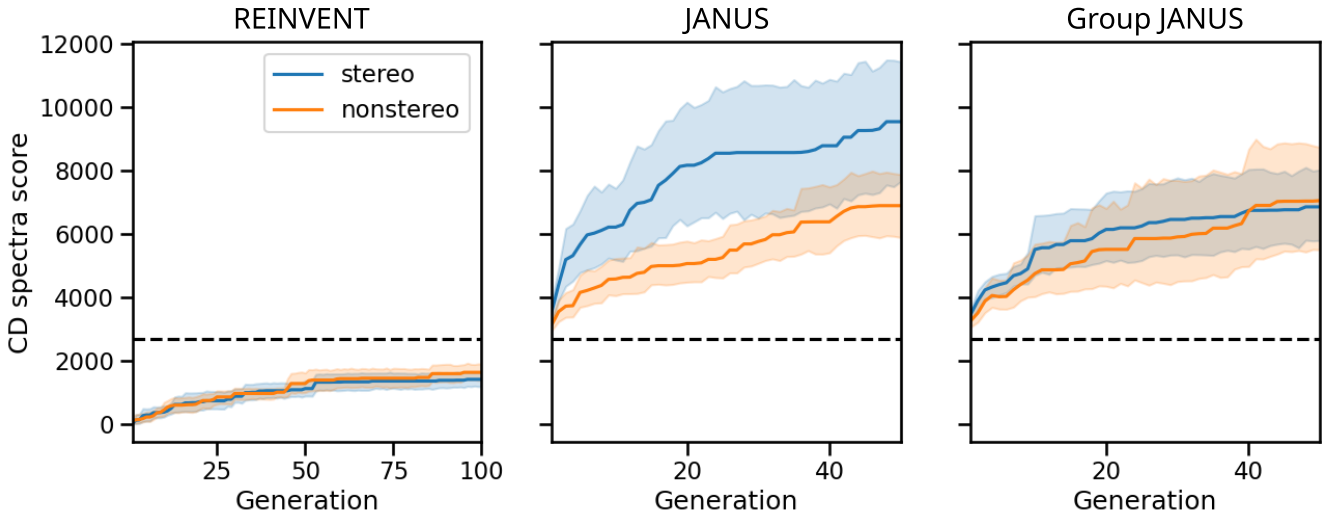


Fig. 5. Optimization traces for CD task. The cumulative top-1 CD peak score as a function of generation of optimization. Shaded regions indicate the 95% CI. The dashed line is the best score found in the starting dataset.

consideration, the decision should be guided by a careful assessment of task-specific requirements and the trade-offs associated with increased model complexity. Further investigation into more efficient and robust representations of stereochemistry, such as graph-based approaches, is a worthwhile direction for future research to potentially improve these generative models.

Methods

Reinvent

REINVENT is an RL algorithm that uses a recurrent neural network (RNN) pretrained on a dataset of SMILES as a chemical language model agent (22, 23, 25). For this work, we follow the procedures described in Olivecrona et al. (22). While later versions of REINVENT have additional features, such as scaffold and linker design using transformer models, the underlying RL method for de novo design is unchanged. When provided a token from a SMILES string, the RNN is trained to generate a conditional distribution of the subsequent tokens in the sequence. A memory state is passed into the model as well, retaining information about previous tokens of the sequence observed by the model. The RNN is first pretrained on the initial ZINC dataset, allowing it to learn the grammar of the stereo and nonstereo SMILES in the dataset, producing 94 and 91% average validity of generated SMILES, respectively. During the RL optimization, the prior RNN is fine-tuned after each generation by a loss function augmented by the fitness score achieved by the molecule $S \in [0, 1]$, with good candidates scoring $S = 1$, and poor candidates and invalid SMILES scoring $S = 0$. Fitness functions that are outside of this range are scaled using a sigmoid function. With each iteration, the RL algorithm will aim to optimize the molecules to maximize the fitness function. Note that SELFIES can also be used with REINVENT, but previous studies have demonstrated that the RNN model is sufficiently capable of generating valid SMILES, and no significant performance gain is observed for SELFIES-REINVENT (41).

JANUS and GroupJANUS

JANUS admits only SELFIES-based representations. Leveraging the robustness of SELFIES representation, JANUS can perform mutation and crossover operations, as defined in the STONED algorithm (94). JANUS maintains two separate populations for exploration and exploitation of chemical space. The exploration set is generated by mutation and crossover operations within the entire population, while the exploitation set is generated through a series of mutations on the fittest molecules. The best candidates found in the exploitation set are then exchanged with the worst candidates in the exploration set. At each iteration, selection pressure from the fitness function allows the model to converge toward the optimum.

In our workflow, we implement the GroupSELFIES version of JANUS, dubbed GroupJANUS, which operates in the same fashion as JANUS. In order to isolate the effect of the stereochemical tokens, only the chiral group tokens are used in GroupJANUS; no other groups are encoded in the GroupSELFIES grammar. For both JANUS and GroupJANUS, the mutation operations depend on the random sampling of tokens in the alphabet. For both models, the inclusion of stereochemical tokens greatly increases the size of the alphabet, and structural tokens which are responsible for encoding molecular rings and branches are less likely to be sampled. To account for this imbalance, structural tokens—such as [RingX], [BranchX], and the GroupSELFIES specific [pop] tokens—are weighted such that they are sampled with the same probability as in the nonstereo alphabet.

Experimental setup

For all tasks, the models are allotted 10,000 fitness oracle calls. The GAs run 50 generations with a population size of 200 molecules; the initial populations for the GA are fixed, starting with the top 5,000 scoring molecules from the initial dataset. REINVENT runs 100 generations of 100 molecules to allow for more policy updates throughout the optimization. REINVENT is fine-tuned on the same top 5,000 scoring molecules before optimization.

Each experiment was repeated 10 times with different random seeds, which affect the sampling operations within REINVENT and the GAs, and in any random assignment of stereochemical information. The reported mean and SD of the optimization trace AUC are calculated over these 10 runs. Statistical significance was determined using a two-sample Student's t-test ($P < 0.05$), comparing the stereo-aware and nonstereo-aware variants for each

model and task based on the results from these independent runs. The 95% CI shown in the figures are also derived from the independent runs.

Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions. The authors thank Austin Cheng, Cher Tian Ser, Leon Schlosser, and AkshatKumar Nigam for helpful discussions. Computations were made on the supercomputers Béluga and Narval from École de technologie supérieure, managed by Calcul Québec and the Digital Research Alliance of Canada. The operation of the supercomputers is funded by the Canada Foundation for Innovation (CFI), Ministère de l'Économie, des Sciences et de l'Innovation du Québec (MESI) and le Fonds de recherche du Québec–Nature et technologies (FRQ-NT).

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

G.T. acknowledges the support of the Vector Institute for Artificial Intelligence, and the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Canada Graduate Research Scholarship (PSGD3-559078-2021). K.J. acknowledges funding through an International Postdoc grant from the Swedish Research Council (no. 2020-00314). A.A.-G. acknowledges support from the Canada 150 Research Chairs program and CIFAR, as well as the generous support of Anders G. Frøseth.

Author Contributions

G.T., K.J., and A.A.-G. designed research; G.T., E.Y., and N.Y. performed research; G.T. and E.Y. analyzed data; and G.T. and K.J. wrote the article.

Preprints

A preprint of this article is published at DOI: 10.26434/chemrxiv-2024-tkjr1.

Data Availability

The code, including the implemented models and benchmarking tasks, are available at https://github.com/aspuru-guzik-group/stereogeneration under the MIT license. The data and the result files are provided at DOI: 10.5281/zenodo.14545514.

References

1 Sanchez-Lengeling B, Aspuru-Guzik A. 2018. Inverse molecular design using machine learning: generative models for matter engineering. *Science*. 361(6400):360–365.
2 Gómez-Bombarelli R, *et al.* 2018. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci*. 4(2):268–276.
3 Segler MHS, Kogej T, Tyrchan C, Waller MP. 2018. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent Sci*. 4(1):120–131.
4 De Cao N, Kipf T. 2018. Molgan: an implicit generative model for small molecular graphs [preprint], arXiv, arXiv:1805.11973. https://doi.org/10.48550/arXiv.1805.11973
5 Meyers J, Fabian B, Brown N. 2021. De novo molecular design and generative models. *Drug Discov Today*. 26(11):2707–2715.
6 Bengio E, Jain M, Korablyov M, Precup D, Bengio Y. 2021. Flow network based generative models for non-iterative diverse candidate generation. *Adv Neural Inf Process Syst*. 34:27381–27394.
7 Yao Z, *et al.* 2021. Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nat Mach Intell*. 3(1):76–86.
8 Grisoni F, *et al.* 2021. Combining generative artificial intelligence and on-chip synthesis for de novo drug design. *Sci Adv*. 7(24):eabg3338.
9 Zhavoronkov A, *et al.* 2019. Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Nat Biotechnol*. 37(9):1038–1040.
10 Ren F, *et al.* 2023. Alphafold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel cdk20 small molecule inhibitor. *Chem Sci*. 14(6):1443–1452.
11 Schneider G, Fechner U. 2005. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov*. 4(8):649–663.
12 Schneider P, *et al.* 2020. Rethinking drug design in the artificial intelligence era. *Nat Rev Drug Discov*. 19(5):353–364.
13 Pollice R, *et al.* 2021. Data-driven strategies for accelerated materials design. *Acc Chem Res*. 54(4):849–860.
14 Tom G, *et al.* 2024. Self-driving laboratories for chemistry and materials science. *Chem Rev*. 124(16):9633–9732.
15 Kusner MJ, Paige B, Hernández-Lobato JM. Grammar variational autoencoder. In: *International Conference on Machine Learning*. PMLR; 2017. p. 1945–1954.
16 Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation. In: *International Conference on Machine Learning*. PMLR; 2018. p. 2323–2332.
17 Winter R, Montanari F, Noé F, Clevert D-A. 2019. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem Sci*. 10(6):1692–1701.
18 Polykovskiy D, *et al.* 2020. Molecular sets (moses): a benchmarking platform for molecular generation models. *Front Pharmacol*. 11:565644.
19 Liu Q, Allamanis M, Brockschmidt M, Gaunt A. 2018. Constrained graph variational autoencoders for molecule design. *Adv Neural Inf Process Syst*. 31:7806–7815.
20 Guimaraes GL, Sanchez-Lengeling B, Outeiral C, Farias PLC, Aspuru-Guzik A. 2017. Objective-reinforced generative adversarial networks (organ) for sequence generation models [preprint], arXiv, arXiv:1705.10843. https://doi.org/10.48550/arXiv.1705.10843
21 Sanchez-Lengeling B, Outeiral C, Guimaraes GL, Aspuru-Guzik A. 2017. Optimizing distributions over molecular space. an objective-reinforced generative adversarial network for inverse-design chemistry (organic) [preprint], *ChemRxiv*. https://doi.org/10.26434/chemrxiv.5309668.v3
22 Olivecrona M, Blaschke T, Engkvist O, Chen H. 2017. Molecular de-novo design through deep reinforcement learning. *J Cheminform*. 9(1):1–14.
23 Blaschke T, *et al.* 2020. Reinvent 2.0: an AI tool for de novo drug design. *J Chem Inf Model*. 60(12):5918–5922.
24 Nica AC, *et al.* Evaluating generalization in gflownets for molecule design. In: *ICLR2022 Machine Learning for Drug Discovery*. 2022.
25 Loeffler HH, *et al.* 2024. Reinvent 4: modern AI-driven generative molecule design. *J Cheminform*. 16(1):20.

26 Zhou Z, Kearnes S, Li L, Zare RN, Riley P. 2019. Optimization of molecules via deep reinforcement learning. *Sci Rep*. 9(1):10752.
27 Guo J, Schwaller P. 2024. Saturn: sample-efficient generative molecular design using memory manipulation [preprint], arXiv, arXiv:2405.17066. https://doi.org/10.48550/arXiv.2405.17066
28 Brown N, McKay B, Gilardoni F, Gasteiger J. 2004. A graph-based genetic algorithm and its application to the multiobjective evolution of median molecules. *J Chem Inf Comput Sci*. 44(3):1079–1087.
29 Jensen JH. 2019. A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space. *Chem Sci*. 10(12):3567–3572.
30 Nigam AK, Pollice R, Aspuru-Guzik A. 2022. Parallel tempered genetic algorithm guided by deep neural networks for inverse molecular design. *Digit Discov*. 1(4):390–404.
31 Tripp A, Hernández-Lobato JM. 2023. Genetic algorithms are strong baselines for molecule generation [preprint], arXiv, arXiv:2310.09267. https://doi.org/10.48550/arXiv.2310.09267
32 Dollar O, Joshi N, Beck DAC, Pfaendtner J. 2021. Attention-based generative models for de novo molecular design. *Chem Sci*. 12(24): 8362–8372.
33 Bagal V, Aggarwal R, Vinod PK, Priyakumar UD. 2021. MolGPT: molecular generation using a transformer-decoder model. *J Chem Inf Model*. 62(9):2064–2076.
34 Born J, Manica M. 2023. Regression transformer enables concurrent sequence regression and generation for molecular language modelling. *Nat Mach Intell*. 5(4):432–444.
35 Jablonka KM, Schwaller P, Ortega-Guerrero A, Smit B. 2024. Leveraging large language models for predictive chemistry. *Nat Mach Intell*. 6(2):161–169.
36 Du Y, *et al*. 2024. Machine learning-aided generative molecular design. *Nat Mach Intell*. 6(6):589–604.
37 Preuer K, Renz P, Unterthiner T, Hochreiter S, Klambauer G. 2018. Fréchet chemnet distance: a metric for generative models for molecules in drug discovery. *J Chem Inf Model*. 58(9):1736–1741.
38 Flam-Shepherd D, Zhu K, Aspuru-Guzik A. 2022. Language models can learn complex molecular distributions. *Nat Commun*. 13(1):3293.
39 Skinnider MA, Stacey RG, Wishart DS, Foster LJ. 2021. Chemical language models enable navigation in sparsely populated chemical space. *Nat Mach Intell*. 3(9):759–770.
40 Brown N, Fiscato M, Segler MHS, Vaucher AC. 2019. Guacamol: benchmarking models for de novo molecular design. *J Chem Inf Model*. 59(3):1096–1108.
41 Gao W, Fu T, Sun J, Coley C. 2022. Sample efficiency matters: a benchmark for practical molecular optimization. *Adv Neural Inf Process Syst*. 35:21342–21357.
42 García-Ortegón M, *et al*. 2022. Dockstring: easy molecular docking yields better benchmarks for ligand design. *J Chem Inf Model*. 62(15):3486–3502.
43 Huang K, *et al*. 2021. Therapeutics data commons: machine learning datasets and tasks for drug discovery and development. *Adv Neural Inf Process Syst*.
44 Nigam AK, *et al*. 2023. Tartarus: a benchmarking platform for realistic and practical inverse molecular design. *Adv Neural Inf Process Syst*. 36:3263–3306.
45 Wildman SA, Crippen GM. 1999. Prediction of physicochemical parameters by atomic contributions. *J Chem Inf Comput Sci*. 39(5):868–873.
46 Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. 2012. Quantifying the chemical beauty of drugs. *Nat Chem*. 4(2):90–98.
47 BenevolentAI. 2020. BenevolentAI/guacamol_results. https://github.com/BenevolentAI/guacamol_results. Accessed on August 19, 2024.
48 Ahn S, Kim J, Lee H, Shin J. 2020. Guiding deep molecular optimization with genetic exploration. *Adv Neural Inf Process Syst*. 33:12008–12021.
49 Tripp A, Simm GNC, Hernández-Lobato JM. A fresh look at de novo molecular design benchmarks. In: *NeurIPS 2021 AI for Science Workshop*. 2021.
50 Renz P, Van Rompaey D, Wegner JK, Hochreiter S, Klambauer G. 2019. On failure modes in molecule generation and optimization. *Drug Discov Today Technol*. 32–33(1):55–63.
51 Gendreau P, *et al*. 2023. Molecular assays simulator to unravel predictors hacking in goal-directed molecular generations. *J Chem Inf Model*. 63(13):3983–3998.
52 Bender A, *et al*. 2022. Evaluation guidelines for machine learning tools in the chemical sciences. *Nat Rev Chem*. 6(6):428–442.
53 Chhabra N, Aseri ML, Padmanabhan D. 2013. A review of drug isomerism and its significance. *Int J Appl Basic Med Res*. 3(1):16–18.
54 Ganea O, *et al*. 2021. Geomol: torsional geometric generation of molecular 3D conformer ensembles. *Adv Neural Inf Process Syst*. 34:13757–13769.
55 Ochiai T, *et al*. 2023. Variational autoencoder-based chemical latent space for large molecular structures with 3D complexity. *Commun Chem*. 6(1):249.
56 Morehead A, Cheng J. 2024. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun Chem*. 7(1):150.
57 McConathy J, Owens MJ. 2003. Stereochemistry in drug action. *Prim Care Companion J Clin Psychiatry*. 5(2):70.
58 McVicker RU, O'Boyle NM. 2024. Chirality of new drug approvals (2013–2022): trends and perspectives. *J Med Chem*. 67(4): 2305–2320.
59 Smith SW. 2009. Chiral toxicology: it's the same thing… only different. *Toxicol Sci*. 110(1):4–30.
60 Guy S, *et al*. 2009. Pure chiral organic thin films with high isotropic optical activity synthesized by UV pulsed laser deposition. *J Mater Chem*. 19(38):7093–7097.
61 Carreira EM, Kvaerno L. *Classics in stereoselective synthesis*. John Wiley & Sons, 2009.
62 Albano G, Pescitelli G, Di Bari L. 2020. Chiroptical properties in thin films of π-conjugated systems. *Chem Rev*. 120(18): 10145–10243.
63 Cuvellier A, *et al*. 2019. The influence of stereochemistry on the reactivity of the Diels–Alder cycloaddition and the implications for reversible network polymerization. *Polym Chem*. 10(4): 473–485.
64 Weininger D. 1988. Smiles, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci*. 28(1):31–36.
65 Krenn M, Häse F, Nigam AK, Friederich P, Aspuru-Guzik A. 2020. Self-referencing embedded strings (selfies): a 100% robust molecular string representation. *Mach Learn Sci Technol*. 1(4):045024.
66 Cheng AH, *et al*. 2023. Group selfies: a robust fragment-based molecular string representation. *Digit Discov*. 2(3):748–758.
67 Basilaia M, Chen MH, Secka J, Gustafson JL. 2022. Atropisomerism in the pharmaceutically relevant realm. *Acc Chem Res*. 55(20):2904–2919.
68 Krenn M, *et al*. 2022. Selfies and the future of molecular string representations. *Patterns*. 3(10):100588.
69 Mastrolorito F, *et al*. 2024. fragSMILES as a chemical string notation for advanced fragment and chirality representation. *Commun Chem*. 8(26):1–9.
70 Sterling T, Irwin JJ. 2015. Zinc 15–ligand discovery for everyone. *J Chem Inf Model*. 55(11):2324–2337.
71 Landrum G. 2024. RDKit: ospen-source cheminformatics (v2024.03.5). https://www.rdkit.org.

72 Rogers D, Hahn M. 2010. Extended-connectivity fingerprints. *J Chem Inf Model*. 50(5):742–754.
73 Berman HM, *et al.* 2000. The protein data bank. *Nucleic Acids Res*. 28(1):235–242.
74 Koes DR, Baumgartner MP, Camacho CJ. 2013. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J Chem Inf Model*. 53(8):1893–1904.
75 Halgren TA. 1999. MMFF VI. MMFF94s option for energy minimization studies. *J Comput Chem*. 20(7):720–729.
76 Olsen JA, *et al.* 2003. A fluorine scan of thrombin inhibitors to map the fluorophilicity/fluorophobicity of an enzyme active site: evidence for C–F···C=O interactions. *Angew Chem Int Ed Engl*. 42(22):2507–2511.
77 Greenfield NJ. 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*. 1(6):2876–2890.
78 Herman WN. 2001. Polarization eccentricity of the transverse field for modes in chiral core planar waveguides. *J Opt Soc Am A Opt Image Sci Vis*. 18(11):2806–2818.
79 Bahar E. 2007. Mueller matrices for waves reflected and transmitted through chiral materials: waveguide modal solutions and applications. *J Opt Soc Am B*. 24(7):1610–1619.
80 Lodahl P, *et al.* 2017. Chiral quantum optics. *Nature*. 541(7638):473–480.
81 Mamede R, de Almeida BS, Chen M, Zhang Q, Aires-de Sousa J. 2020. Machine learning classification of one-chiral-center organic molecules according to optical rotation. *J Chem Inf Model*. 61(1):67–75.
82 Zhao L, *et al.* 2021. Accurate machine learning prediction of protein circular dichroism spectra with embedded density descriptors. *JACS Au*. 1(12):2377–2384.
83 Adams K, Pattanaik L, Coley CW. 2021. Learning 3D representations of molecular chirality with invariance to bond rotations [preprint], arXiv, arXiv:2110.04383. https://doi.org/10.48550/arXiv.2110.04383
84 Grimme S. 2019. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J Chem Theory Comput*. 15(5):2847–2862.
85 Pracht P, Bohle F, Grimme S. 2020. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys Chem Chem Phys*. 22(14):7169–7192.
86 Pracht P, *et al.* 2024. CREST—a program for the exploration of low-energy molecular chemical space. *J Chem Phys*. 160(11):114110.
87 Bannwarth C, *et al.* 2021. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip Rev Comput Mol Sci*. 11(2):e1493.
88 Bannwarth C, Ehlert S, Grimme S. 2019. GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J Chem Theory Comput*. 15(3):1652–1671.
89 Bannwarth C, Grimme S. 2014. A simplified time-dependent density functional theory approach for electronic ultraviolet and circular dichroism spectra of very large molecules. *Comput Theor Chem*. 1040–1041:45–53.
90 Grimme S, Bannwarth C. 2016. Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB). *J Chem Phys*. 145(5):054103.
91 O Pinheiro PO, *et al.* 2023. 3D molecule generation by denoising voxel grids. *Adv Neural Inf Process Syst*. 36:69077–69097.
92 Flam-Shepherd D, Aspuru-Guzik A. 2023. Language models can learn complex molecular distributions. *Nat Commun*. 13(3293):1–10.
93 Cheng AH, Sun C, Aspuru-Guzik A. 2025. Scalable autoregressive 3D molecule generation [preprint], arXiv, arXiv:2505.13791. https://doi.org/10.48550/arXiv.2505.13791
94 Nigam AK, Pollice R, Krenn M, dos Passos Gomes G, Aspuru-Guzik A. 2021. Beyond generative models: superfast traversal, optimization, novelty, exploration and discovery (stoned) algorithm for molecules using selfies. *Chem Sci*. 12(20):7079–7090.