Hybrid Diffusions for Stable Molecular Structure Generation via Explicit Energy-based Model

Youngwoo Cho [1] **Seunghoon Yi** [2] **Soo Kyung Kim** [3 4] **Hongkee Yoon** [1] **Joonseok Lee** [2 5]

Abstract

Generation of 3D molecules utilizing diffusion models often encounters difficulties in producing stable structures, primarily due to the emergence of unstable intermediate structures during diffusion steps. To account for this issue, we introduce a diffusion-based molecule generation model that incorporates an energy-based model (EBM), pretrained on density functional theory (DFT) data. Specifically, we propose three strategic use of EBM: 1) guided exploration using the EBM, 2) stability evaluation to accept the structure or to reject and restart the generation at the end of diffusion steps, and 3) performing post-relaxation refinement. With these three strategies, we demonstrate that the energy estimator significantly enhances the generated molecule's stability.

1. Introduction

Machine learning has been widely applied to molecule and crystal structure generation, which is essential for discovering innovative materials endowed with desired properties. Many strategies have been employed, including generative adversarial networks (GANs) (Goodfellow et al., 2014), variational autoencoders (VAEs) (Kingma & Welling, 2014), and diffusion models (Sohl-Dickstein et al., 2015).

Within this dynamic landscape, we propose a novel approach that marries the diffusion model with EBM, thereby presenting a promising path for generating molecular structures with enhanced stability. Our approach leverages the power of diffusion models and further enhances it with an EBM, accounting for physical energy in structure generation.

Specifically, we introduce three strategies for utilizing the EBM as follows:

- **EBM-guided exploration**: The EBM directly guides the structure exploration during the diffusion process, leveraging the energy information it provides. This approach enhances the stability of the molecules.
- **Iterative energy-based assessment of generated structures**: During the generation, we periodically evaluate the stability of generated structures using their energy profiles and discard unstable structures.
- **Post-relaxation refinement**: By performing additional optimization steps without time constraints, we enhance the structural stability of the generated molecules.

These strategies focus on distinct application areas for the EBM to help the diffusion model. Their effectiveness is gauged by evaluating the stability of optimized structures, as determined by DFT calculation. By incorporating DFT, we validate and demonstrate the effectiveness of our methods in generating stable and reliable molecular configurations.

2. Preliminary

We first introduce the diffusion probabilistic model and its application in molecular generation tasks. Subsequently, we overview energy-based structure relaxation methods.

2.1. Diffusion model

Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021a; Song & Ermon, 2019; Song et al., 2021b) are designed to learn a denoising process, a kind of Markov chain capable of producing a data sample $\mathbf{x}_0$ from a random Gaussian noise $\mathbf{x}_T$. It consists of two chains, forward and reverse processes. The forward process $q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$, *i.e.*, *diffusion process*, is a Markov chain that transforms a given sample into the standard normal distribution. Conversely, the reverse process $p(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p(\mathbf{x}_{t-1}|\mathbf{x}_t)$ where $p(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \mu(\mathbf{x}_t, t), \mathbf{\Sigma}(\mathbf{x}_t, t))$, referred to as

[1]Korea Advanced Institute of Science and Technology, Daejeon, Korea [2]Seoul National University, Seoul, Korea [3]Palo Alto Research Center, Palo Alto, CA, USA [4]Stanford Research Institute, Menlo Park, CA, USA [5]Google Research, Mountain View, CA, USA. Correspondence to: Hongkee Yoon <hongkeeyoon7@gmail.com>, Joonseok Lee <joonseok@snu.ac.kr>.

Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023.

the *denoising process*, reconstructs a sample by gradually removing the small amount of noise from an input random Gaussian noise. However, since this reverse process is complicated and nontrivial, a neural network $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ is employed to estimate it. The approximation of the denoising process is achieved by minimizing the KL divergence between the original input and the recovered sample.

In the context of the molecular domain, the diffusion models operate on both atomic species and coordinates. This can be formulated using a graph $\mathbf{G}(\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ and $\mathbf{E}$ correspond to node (*i.e.*, atomic information) and edge (*i.e.*, bonding information) features, respectively. Diffusion models predict the noise that should be removed from each node and edge feature at a given time step t. Our baseline model, $E(3)$ equivariant diffusion model (EDM) (Hoogeboom et al., 2022), aligns with this approach. EDM applies diffusion and denoising processes on node features containing atomic coordinates and embedding. In addition, EDMs employ $E(n)$ equivariant graph neural networks (EGNNs) (Satorras et al., 2021) to predict the noise at time step t. Adaptation of EGNN guarantees equivariance with respect to the atomic coordinates.

2.2. Energy-based model and structure relaxation

EBMs have been utilized to predict the total energy of molecules and materials, trained on datasets labeled by DFT calculations. The study aims to accelerate the exploration and discovery of stable structures using the diffusion model by combining the accuracy of DFT calculations and the speed of machine learning potentials. The stability of generated molecules is evaluated through DFT-based structure relaxation, assessing their structural and energy differences from DFT-optimized structures. Here, we characterize the quality of generated molecules by the structural (ΔP) and energy (ΔE) differences from DFT-optimized structures, with smaller values indicating the better stability of generated molecules following our previous study (Yi et al., 2023). The results provide valuable insights into the stability and quality of the generated molecules.

3. Related work

This section introduces the flow of molecular generation research in three folds.

3.1. Classical method: DFT as a scoring function

Notable codes in this field, such as XtalOpt (Lonie & Zurek, 2011), USPEX (Glass et al., 2006), CALYPSO (Wang et al., 2012), and AMADEUS (Lee et al., 2016), employ evolutionary algorithms, cluster optimization algorithms, and conformal space annealing global optimization methods, respectively. All of these methods rely on massive DFT calculations, which are computationally intensive as usual.

3.2. GANs and VAEs

To address the inherent limitation of DFT in terms of computational speed, machine learning techniques have garnered significant interest in molecular structure generation. Starting from a random noise input with chemical properties, VAEs (Kingma & Welling, 2014) and GANs (Goodfellow et al., 2014) are employed to produce a molecular structure, which has the desired properties. This scheme is commonly referred to as inverse design (Sanchez-Lengeling & Aspuru-Guzik, 2018).

Initially, inverse design has focused on generating 1D or 2D representations of molecules, such as SMILES strings (Weininger, 1988; Kusner et al., 2017; Gómez-Bombarelli et al., 2018; Li et al., 2022) or graphs (Jin et al., 2018; De Cao & Kipf, 2018; Mitton et al., 2020), and subsequently expanded to produce 3D molecular structures (Eguchi et al., 2022; Huang et al., 2022; Wang et al., 2022). VAEs are known to suffer from generating low-fidelity samples (Xiao et al., 2022), while GANs suffer from mode collapse (Thanh-Tung & Tran, 2018).

3.3. Diffusion models

EDM (Hoogeboom et al., 2022) employs the diffusion model to generate molecular structures exhibiting desired properties. Corso et al. (2023) and Igashov et al. (2022) address protein docking and binding, to identify molecules capable of effectively binding to a given protein structure. Wu et al. (2022) defines simple equation-based potential and couples them with diffusion models.

4. Method

Our framework consists of two distinct models, an explicit EBM (*i.e.*, machine learning potential) and the diffusion model to generate stable molecular structures as shown in Fig. 1. These two models are trained independently, collaborating only during the actual generation of molecular structures using diffusion.

The EBM serves as a surrogate function that models the potential energy surface landscapes of molecular structures based on the positions and species of atoms. On the other hand, the diffusion model is trained to explore molecules' conformational space and generate diverse molecular structures. From the perspective of a score-based approach (Song & Ermon, 2019; Song et al., 2021b), the diffusion model assigns an internal score to each molecular conformation. Here, the internal scores that the diffusion model learns to predict, can be interpreted as virtual energies rather than the total energy of a given state. While training the models, molecular structures along with their corresponding energy

information are utilized.

Our objective is to investigate the advantages of integrating the explicit EBM into the diffusion model over solely focusing on enhancing each model independently. Thus, we explore three approaches to improve the diffusion model by leveraging the EBM, described in the following subsections.

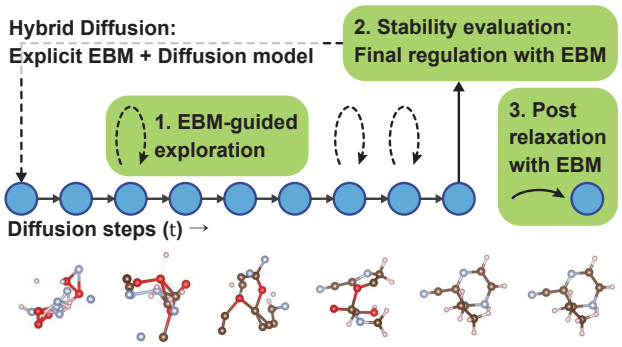

Figure 1. Schematic of the diffusion model combined with an energy-based model. EBM intervenes during and after the diffusion process in three aspects. These three approaches using EBM contribute to generating and validating more stable molecular structures.

4.1. Guided exploration: Incorporating energy-based model into diffusion process

First, we may direct the diffusion process by conducting relaxation using the EBM. Specifically, energy-based structure relaxation steps are intermittently inserted between the diffusion steps to guide the model to output a stable molecular structure. We selectively use the EBM intervention to minimize disruptions during the update of atomic elements within the structure, reducing the risk of the process falling into local energy minima with fixed species. This approach leverages the EBM's ability to explore the conformational space, enhancing the stability of the molecular structures generated through diffusion.

4.2. Stability evaluation: Energy-based assessment of generated structures

At the end of each diffusion process, a decision-making process is employed to determine whether to accept the generated molecule or to restart the generation process again. A few previous models (Hoogeboom et al., 2022) have incorporated chemical intuition, leading to reasonable performance for organic molecules composed of majorly C, H, O, and N atoms. However, such approaches are hardly applicable to arbitrary molecules, bulk crystals, transition metals, or other complex systems.

We evaluate the stability of EBM-guided optimized structures based on criteria such as; 1) minimal structural changes during relaxation, 2) monotonic energy decrease, and 3) reasonably short optimization time. This allows us to determine whether to retry the diffusion process, preventing premature termination of the generation process and encouraging further attempts to refine unstable structures.

Exp. #	Heuristics reg.	Guided explore.	Final reg.	Post relax.	$(\sigma^1_{\Delta E}, \sigma^2_{\Delta E})$	$(\sigma^1_{\Delta P}, \sigma^2_{\Delta P})$
Base	✓				(9.7, 1.7)	(8.3, 0.8)
1					(11.3, 1.9)	(8.1, 0.7)
2		✓			(12.6, 2.2)	(7.0, 0.6)
3		✓	✓		(3.8, 1.6)	(4.3, 0.6)
4		✓	✓	✓	(2.5, 0.4)	(4.5, 0.5)

Table 1. Comparison with baseline models for energy deviations ($\sigma^1_{\Delta E}$, $\sigma^2_{\Delta E}$) and average distortion ($\sigma^1_{\Delta P}$, $\sigma^2_{\Delta P}$) after structure optimization by DFT. (σ^1, σ^2) indicates the two width components of the double Gaussian fit shown in Fig. 3(f, g). Lower is better for all metrics. We report $\sigma_{\Delta E}$ in eV and $\sigma_{\Delta P}$ in Å. Note that our baseline utilizes *heuristics regularization*, which naively checks bond types (*e.g.*, C-H bond should be a single bond) based on the bond length.

4.3. Post-relaxation refinement

The diffusion model utilizes an implicit score to generate structurally plausible molecules, which may approximate the energetic minima but does not guarantee it. An EBM may be employed to perform relaxation tasks starting from the generated structure to further enhance the stability of the generated molecular structures. Similar to the earlier approach, we utilize the force, which is the derivative of the energy with respect to position, to apply the BFGS method (Liu & Nocedal, 1989) allowing a large number of optimization steps. This approach provides a straightforward improvement, depending upon the accuracy of the EBM.

5. Experiments and Results

We conduct experiments to answer the following questions: **Q1.** How stable are the molecules generated by the baseline model using DFT? **Q2.** How much do the three proposed adaptations of the EBM contribute to generating more stable structures?

5.1. Experimental setups

Datasets. We use a publicly available QM9 dataset (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014), which consists of optimal structures of 130,000 molecules containing species with up to 9 heavy atoms of (C, O, N, and F) out of the GDB-17 (Ruddigkeit et al., 2012) database.

Baselines. We compare our model to EDM (Hoogeboom et al., 2022). For our experiments, we adopt the EDM (Hoogeboom et al., 2022) as our backbone for the diffusion model and utilize SpookyNet (Unke et al., 2021) for the energy-based model. In the case of EDM, the model is trained from scratch with the default hyperparameters,

and we used pretrained parameters for SpookyNet[1]. Since our method is orthogonal to the training procedure, there was no finetuning on both models before and during the generation.

Stability comparison through DFT-based structural optimization. We have generated over 500 molecules with lengths ranging from 7 to 27 atoms and evaluated them through structural optimization using DFT. This allows us to determine whether a visually plausible structure actually corresponds to stable conformations. Additionally, we investigate the structural distortion of unstable structures by measuring the deviations in total energy (ΔE_{DFT}) and positional (ΔP_{DFT}) before and after relaxation, following the concept introduced by Yi et al. (2023). In detail, ΔP_{DFT} is defined as the average distance of each atom's distortion from the DFT-relaxed structure in the initially generated structure.

5.2. Evaluation of the baseline

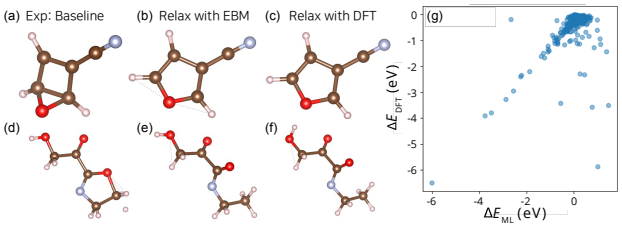

Figure 2. (a,d) Examples of molecules generated by the baseline model. (b,e) Relaxed structure by the EBM. (c,f) Relaxed structures by DFT. (g) Distribution of ΔE_{DFT} and ΔE_{EBM} obtained by relaxing multiple samples from the baseline model. Smaller values of ΔE_{DFT} indicate stabler molecules are generated by the baseline model. Linearity between ΔE_{DFT} and ΔE_{EBM} reflects the accuracy of EBM. The discrepancy in EBM from DFT leads to generating unstable structures, but this can be alleviated by adopting a more accurate model in the future.

We utilize DFT and EBM to evaluate the stability of generated molecules from the baseline model and estimate the EBM quality. Fig. 2(a) shows that, while most positions are stable, the C-O-C atoms segment forms a separate C-O-C bond in addition to the C-C bond. By optimizing it using the EBM, Fig. 2(a) transforms into Fig. 2(b). The C-C distance increases, resulting in a more plausible molecule without direct bonding. Additionally, it is evident that Fig. 2(c), obtained through DFT relaxation, resembles Fig. 2(b).

Fig. 2(d-f) shows the relaxation results for a slightly larger molecule. The EBM relaxation successfully recovers improper C-O and C-H bonding lengths in Fig. 2(e). We generally observe reasonable concordance when comparing the EBM and DFT optimization results. However, slight deviations are observed in the EBM model, such as the O-H angle at the top of Fig. 2(f).

In Fig. 2(g), we illustrate a more extensive set of molecules, each one subjected to relaxation through both EBM and DFT. A lower ΔE_{DFT} signifies that the diffusion model has generated more stable molecules. In general, the linear relationship between ΔE_{DFT} and ΔE_{EBM} implies the accuracy of EBM. However, it also reveals the presence of substantial outliers, indicating both the resemblance and divergence between our employed EBM and the actual DFT results. Despite the observed deviations, the results suggest that EBM may adequately detect and rectify unstable structures.

5.3. Collaborative stable diffusion with energy-based model

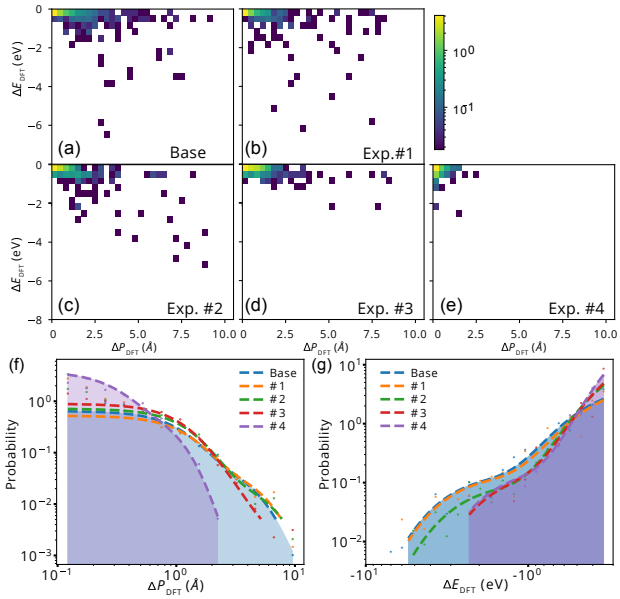

Figure 3. (a-e) 2D normalized histograms of ΔP_{DFT} and ΔE_{DFT} for each experiments in Tab. 1. Concentrated distribution on the top-left ΔP_{DFT} and ΔE_{DFT} indicates better structural stability. (f-g) 1D normalized histograms for ΔP_{DFT} and ΔE_{DFT} for experiments in (a-e), respectively. The dashed lines are fitted using a double Gaussian model on a logarithmic scale. Note that the range of typical carbon bonding is 1.2 to 1.5 Å, and the energy scale of room temperature is about 25 meV.

We use the diffusion model and EBM to generate molecules using three distinct approaches. Since the diffusion model relies on probabilistic generation, we sample multiple molecules of varying sizes, enabling us to examine the overall impact through diverse samples. We incrementally add each idea in the order of the expected impact, namely, EBM-guided exploration, stability evaluation after diffusion, and post-relaxation refinement.

[1]Source code and pretrained parameter we used in our experiments are available at https://github.com/OUnke/SpookyNet and https://github.com/ehoogeboom/e3_diffusion_for_molecules.

EBM-guided exploration. The energy calculated through the EBM and the derived forces assists in the position updates of atoms during the diffusion process. Here, one may consider the direct use of DFT in this update procedure. However, because of the substantial computational cost of DFT, leveraging EBM accelerates the process by approximately 100 to 1,000 times when generating a number of molecules. In practice, guiding based on DFT instead of EBM requires computation time ranging from a few minutes to a day depending on the size of the molecules, whereas EBM typically necessitates only a few seconds.

In our setup, we apply an intervention after 80% of the total diffusion steps have been completed, performing a maximum of 30 iterations of BFGS optimization at each EBM-assisting stage. This approach is designed to balance speed with the need to avoid the risk of settling into local minima, a potential issue if the BFGS optimization is applied too frequently.

With respect to the number of BFGS iterations, our experiments demonstrated little advantage in exceeding 20 steps. Extended iteration sequences consume more time and could obstruct the diffusion process, particularly the updating of element types, possibly leading to undesirable local minima. Given this, we observed a comparable performance between 20 and 50 iterations, and chose 30 iterations as a compromise between optimization efficacy and computational efficiency. It's worth noting that the diffusion process simultaneously updates the elements' types and positions. Therefore, early interference in this process with the aim of optimizing positions for specific elements based on energy could inadvertently prompt the system to settle in local minima. Our experiments show that introducing interventions between 70-90% of the diffusion steps has no significant impact, but initiating interventions prior to the 50% stage can, surprisingly, generate unnatural molecules.

Fig. 3(c) shows the results with a more confined distribution of ΔE_{DFT} and ΔP_{DFT} within a smaller area than the baseline model of Fig. 3(a). Nevertheless, there are instances where both ΔE_{DFT} and ΔP_{DFT} remain significant. This can be attributed to the fact that we do not always proceed with complete relaxation using EBM. The rationale behind not fully relaxing is two-folds: firstly, we aim to update only the unstable parts within the given species and positions using EBM, and secondly, we strive to prevent the system from falling into local minima, which can occur if we overemphasize position optimization at the expense of updating the species.

Evaluating stability after diffusion. While we generally expect stable molecular structures to emerge from a complete diffusion process, it's worth noting that this might not always hold true. In such instances, we employ two criteria to decide whether to accept the molecule as the final outcome or restart the diffusion process from the beginning. Our decision-making hinges on two main criteria. The first assesses whether the energy monotonically decreases when we optimize a generated molecule using EBM. The non-monotonic decrease might indicate the presence of more stable structures beyond an activation barrier, suggesting the molecule has not reached sufficient stability. The second criterion pertains to the energy difference before and after optimization. We accept only these molecules where the energy difference is not significantly large. We specifically set the threshold at 25 meV/atom (corresponding to room temperature, 300 K), in anticipation that the molecule would likely maintain stability at room temperature. Fig. 3(d) illustrates the results of this decision-making process, showing that the molecules generated through this iterative approach exhibit a significantly lower number and degree of outliers in ΔE_{DFT} and ΔP_{DFT}. This demonstrates the effectiveness of the decision-making mechanism in guiding the generation of more stable molecular structures. Given our approach's reliance on EBM over chemical intuition, this approach may be extended to all types of molecules and crystal structures.

Post-relaxation refinement. We optimize the final structure using the EBM with the results obtained from the two techniques above. Unlike previous steps, no time constraint is imposed on the optimization time step, allowing the system to relax fully. As depicted in Fig. 3(e), this process even further reduces the distribution of ΔP and ΔE. Further improvements in the EBM's accuracy and extrapolability could yield even better outcomes. Nonetheless, even in its current state, significant improvements have been observed.

6. Conclusion

By integrating the diffusion model with EBM, we have successfully improved the stability of the structures generated from the diffusion model. Exploiting the EBM in guided exploration and post-relaxation enhances the stability of the emerging structures. Looking ahead, future improvements to the phase space coverage and accuracy of the EBM are expected to generate even more stable molecules. Consequently, this advancement would enhance the screening process, paving the way for more robust and reliable molecular and crystal structure generation.

Acknowledgement

This work was supported by National Research Foundation (2021H1D3A2A03038607, 2022R1C1C1010627), Institute of Information & communications Technology Planning & Evaluation (IITP; 2022-0-00264, 2019-0-00075) grants funded by the government of Korea (MSIT).

Broader impact

Integrating an explicit energy-based model (EBM) with a diffusion model presents a promising direction for generating more stable molecular structures. This hybrid methodology is expected to increase the stability of machine learning-predicted molecules, thereby enhancing their likelihood of being realized through actual synthesis. The discovery of novel molecular structures with unique properties and functions, catalyzing advancements in various scientific and technological domains such as drug design, materials science, catalysis, and electronics, will be facilitated. In addition, this convergence of EBM and diffusion models not only contributes to the overall advancement of computational materials research but also offers a viable strategy for solving natural science problems where *energy* is well-defined. Consequently, our proposed method holds the potential to revolutionize molecule design and discovery, paving the way for scientific breakthroughs and innovations.

References

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. S. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *Proc. of the International Conference on Learning Representations (ICLR)*, 2023.

De Cao, N. and Kipf, T. MolGAN: An implicit generative model for small molecular graphs. In *Proc. of the International Conference on Machine Learning (ICML) Workshop on Theoretical Foundations and Applications of Deep Generative Models*, 2018.

Eguchi, R. R., Choe, C. A., and Huang, P.-S. Ig-VAE: Generative modeling of protein structure by direct 3D coordinate generation. *PLOS Computational Biology*, 18 (6):e1010271, 2022. doi: 10.1371/journal.pcbi.1010271.

Glass, C. W., Oganov, A. R., and Hansen, N. USPEX—Evolutionary crystal structure prediction. *Computer Physics Communications*, 175(11):713–720, 2006.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In *Proc. of the Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276, 2018. doi: 10.1021/acscentsci.7b00572. PMID: 29532027.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *Proc. of the Advances in Neural Information Processing Systems (NeurIPS)*, pp. 6840–6851, 2020.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3D. In *Proc. of the International Conference on Machine Learning (ICML)*, pp. 8867–8887, 2022.

Huang, Y., Peng, X., Ma, J., and Zhang, M. 3DLinker: An E(3) Equivariant Variational Autoencoder for Molecular Linker Design. In *Proc. of the International Conference on Machine Learning (ICML)*, 2022.

Igashov, I., Stärk, H., Vignac, C., Satorras, V. G., Frossard, P., Welling, M., Bronstein, M., and Correia, B. Equivariant 3d-conditional diffusion models for molecular linker design. *arXiv preprint arXiv:2210.05274*, 2022.

Jin, W., Barzilay, R., and Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. In *Proc. of the International Conference on Machine Learning (ICML)*, pp. 2323–2332, 2018.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. In *Proc. of the International Conference on Learning Representations (ICLR)*, 2014.

Kusner, M. J., Paige, B., and Hernández-Lobato, J. M. Grammar variational autoencoder. In *Proc. of the International Conference on Machine Learning (ICML)*, pp. 1945–1954, 2017.

Lee, I.-H., Oh, Y. J., Kim, S., Lee, J., and Chang, K. J. Ab initio materials design using conformational space annealing and its application to searching for direct band gap silicon crystals. *Computer Physics Communications*, 203:110–121, 2016.

Li, C., Yamanaka, C., Kaitoh, K., and Yamanishi, Y. Transformer-based objective-reinforced generative adversarial network to generate desired molecules. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3884–3890, 2022. doi: 10.24963/ijcai.2022/539.

Liu, D. C. and Nocedal, J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1):503–528, 1989. doi: 10.1007/BF01589116.

Lonie, D. C. and Zurek, E. XtalOpt: An open-source evolutionary algorithm for crystal structure prediction. *Computer Physics Communications*, 182(2):372–387, 2011.

Mitton, J., Senn, H. M., Wynne, K., and Murray-Smith, R. A graph vae and graph transformer approach to generating molecular graphs. In *Proc. of the International Conference on Machine Learning (ICML) Workshop on Graph Representation Learning and Beyond*, 2020.

Ramakrishnan, R., Dral, P. O., Rupp, M., and von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1, 2014.

Ruddigkeit, L., Van Deursen, R., Blum, L. C., and Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *Journal of chemical information and modeling*, 52(11):2864–2875, 2012.

Sanchez-Lengeling, B. and Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400):360–365, 2018. doi: 10.1126/science.aat2663.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks. In *Proc. of the International Conference on Machine Learning (ICML)*, pp. 9323–9332, 2021.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proc. of the International Conference on Machine Learning (ICML)*, pp. 2256–2265, 2015.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *Proc. of the International Conference on Learning Representations (ICLR)*, 2021a.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In *Proc. of the Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *Proc. of the International Conference on Learning Representations (ICLR)*, 2021b.

Thanh-Tung, H. and Tran, T. On catastrophic forgetting and mode collapse in generative adversarial networks. *arXiv preprint arXiv:1807.04015*, 2018.

Unke, O. T., Chmiela, S., Gastegger, M., Schütt, K. T., Sauceda, H. E., and Müller, K.-R. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications*, 12(1):7273, 2021. doi: 10.1038/s41467-021-27504-0.

Wang, L., Bai, R., Shi, X., Zhang, W., Cui, Y., Wang, X., Wang, C., Chang, H., Zhang, Y., Zhou, J., Peng, W., Zhou, W., and Huang, B. A pocket-based 3d molecule generative model fueled by experimental electron density. *Scientific Reports*, 12(1):15100, 2022. doi: 10.1038/s41598-022-19363-6.

Wang, Y., Lv, J., Zhu, L., and Ma, Y. CALYPSO: A method for crystal structure prediction. *Computer Physics Communications*, 183(10):2063–2070, 2012.

Weininger, D. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988. doi: 10.1021/ci00057a005.

Wu, L., Gong, C., Liu, X., Ye, M., and Liu, Q. Diffusion-based molecule generation with informative prior bridges. In *Proc. of the Advances in Neural Information Processing Systems (NeurIPS)*, pp. 36533–36545, 2022.

Xiao, Z., Kreis, K., and Vahdat, A. Tackling the generative learning trilemma with denoising diffusion gans. In *Proc. of the International Conference on Learning Representations (ICLR)*, 2022.

Yi, S., Cho, Y., Sul, J., Ko, S. W., Kim, S. K., Choo, J., Yoon, H., and Lee, J. Towards physically reliable molecular representation learning. In *Proc. of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2023.