LigandDiff: 3D Transition Metal Complex Generation with Diffusion Models

Hongni Jin[a] and Kenneth M. Merz, Jr.[a,b]*

[a]Department of Chemistry, Michigan State University,

East Lansing, Michigan 48824, United States

[b]Department of Biochemistry and Molecular Biology, Michigan State University,

East Lansing, Michigan 48824, United States

*Email: merz@chemistry.msu.edu

Abstract

Transition metal complexes are a class of compounds with varied and versatile properties making them of great technological importance. Their applications cover a wide range of fields, either as metallodrugs in medicine or as materials, catalysts, batteries, solar cells, *etc*. The demand for the novel design of transition metal complexes with new properties remains of great interest . However, the traditional high-throughput screening approach is inherently expensive and laborious since it depends on human expertise. Here, we present LigandDiff, a generative model to design novel transition metal complexes. Unlike the existing methods which simply extracts and combine ligands to the metal to get new complexes, LigandDiff aims at designing novel ligands from scratch, which opens new pathways for the discovery of organometallic complexes. Moreover, it overcomes the limitations of current methods where the diversity of new complexes highly relies on the diversity of available ligands while LigandDiff can enumerate novel ligands without human intervention. Our results indicate that LigandDiff designs unique and novel ligands under different

contexts that are synthetically accessible. Moreover, LigandDiff shows good transferability by generating successful ligands for any transition metal complex.

Introduction

Molecular generation is an important tool for new materials discovery and drug design. It aims to create new structures with desirable properties. However, traditional methods can take a long time and are expensive. For example, the estimated expenditure for a new drug from design to production ranges from $314 million to $ 2.8 billion and keeps rising[1] and it usually takes over 12 years to develop a new drug with suitable bioavailability.[2] It has been estimated that nearly 10^{23}~10^{60} potential drug-like molecules are synthesizable in chemical space,[3] wherein only 10^{8} ~10^{10} molecules have already been synthesized.[4] It is extremely time-consuming to identify novel drugs via brute-force high-throughput screening and human intuition can bias small molecule searches thereby missing novel molecules with optimal properties.

Recently, generative models are opening new pathways for molecular generation. Examples include generative models based on SMILES strings, like the variational autoencoder (VAE)-based[5] and Sequence to Sequence Autoencoder (seq2seq AE)-based[6]; 3D full-molecule generative models, such as molGAN,[7] GraphRNN[8] as well as scaffold-based generative models, like DeepScaffold[9] and EMPIRE[10]. To the best of our knowledge, however, all previous work was aimed at generating small organic molecules. The introduction of metal ions into a biological system has promising applications in clinical therapy and diagnostics.[11,12] In 1965 Barnett Rosenberg *et al.* [13] at Michigan State University serendipitously discovered the anticancer properties of cisplatin which kickstarted modern research on metallodrugs. Unlike organic drugs, metal-based agents have versatile electronic and structural properties. The flexible oxidation state of the metal enables it to coordinate with different types of ligands in various geometries. Such flexibility also offers novel reaction mechanisms, such as ligand exchange, metal/ligand-based

redox activity and photoactivation.[14] With these unique reaction pathways, metallodrugs can easily bind with DNA or proteins at target sites to cause structural lesions, ultimately resulting in cellular apoptosis.[15,16] Metallodrugs can also modulate the proliferation of tumor cells via catalyzing chemical transformations *in vivo*.[17] In both cases, the metal is the foci of the metallodrug and tunes the 3D shape of small organic ligands attached to the metal center. Metallodrugs open different and unique pathways for disease treatment, which traditional organic drugs cannot achieve due to drug resistance.[18] In addition, organometallic complexes also have a wide range of applications in materials, like solar cells,[19] electrocatalysts,[20] batteries,[21] *etc*.

Given the great importance of organometallic complexes in both medicine and industrial applications, much attention has been paid to the design of new organometallic complexes with desirable properties. However, current methods simply extract the already available ligands from the Cambridge Structural Database (CSD) and then combine the ligands with the metal to generate new complexes.[22,23] This limits the investigation of novel ligand domains which further restricts the discovery of novel organometallic complexes since once the ensemble of ligands is determined, the corresponding number of complexes is also determined. In addition, in this workflow, much work, like ligand curation and combination, still needs human involvement and intuition, which also slows down the entire process.

In this work, we introduce LigandDiff, a scaffold-based diffusion model to generate 3D transition complexes from scratch. Diffusion models[24] are a class of probabilistic generative models which destroy the initial clean input data by progressively introducing random noise, then reverse the whole process for new sample generation. This method has been widely used in inpainting,[25] video

generation,[26] Natural Language Generation,[27] 3D small organic molecule generation,[28,29] medical image reconstruction,[30] *etc*. Denoising Diffusion Probabilistic Models (DDPMs)[31] are a type of diffusion model and are inspired by non-equilibrium thermodynamics. A DDPM includes two Markovian chains, namely the forward chain and the reverse chain. The forward chain keeps adding random noise to a clean data point x_0 with predefined steps T to transform the input data to a simplified predefined distribution, *e.g*. Gaussian distribution. At a given step $t = 0, \ldots, T$, the noised data x_t is derived by

$$q(x_t|x_0) = \mathcal{N}(x_t|\alpha_t x, \sigma_t^2 I) \tag{1}$$

where $\alpha_t \in \mathbb{R}^+$ determines how much information is kept while $\sigma_t \in \mathbb{R}^+$ determines how much noise is added. As proposed by Sohl-Dickstein *et al.*[32], $\sigma_t^2 = 1 - \alpha_t$. Usually, this noise schedule is predefined, and it smoothly transitions from $\alpha_0 \approx 1$ towards $\alpha_T \approx 0$. By sampling a Gaussian distribution $\epsilon \in \mathcal{N}(0, \mathrm{I})$, a sample of x_t is obtained as

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon \tag{2}$$

where $\bar{\alpha}_t = \prod_{k=0}^{t} \alpha_k$. Intuitively, x_T is pure noise without any structural information included. The denoising process or the reverse step is derived as

$$q(x_{t-1}|x_0, x_t) = \mathcal{N}(x_{t-1}|\mu_t(x_0, x_t), \sigma_{t\to t-1}^2 I) \tag{3}$$

and the mean and variance are defined as

$$\mu_t(x_0, x_t) = \frac{\alpha_{t|t-1}\sigma_{t-1}^2}{\sigma_t^2} x_t + \frac{\alpha_{t-1}\sigma_{t|t-1}^2}{\sigma_t^2} x_0 \tag{4}$$

$$\sigma_{t\to t-1} = \frac{\sigma_{t|t-1}\sigma_{t-1}}{\sigma_t} \tag{5}$$

where $\alpha_{t|t-1} = \frac{\alpha_t}{\alpha_{t-1}}$ and $\sigma_{t|t-1}^2 = \sigma_t^2 - \alpha_{t|t-1}^2 \sigma_{t-1}^2$. Eq. 3 indicates that any intermediate state x_t in this diffusion trajectory can be derived from the initial state x_0 and the final state x_T. With this property in mind, the generative denoising process starts from a prior distribution $p(x_T)$,

$$p(x_T) = \mathcal{N}(x_T; 0, I) \tag{6}$$

and it aims to invert the diffusion trajectory while x_0 is unknown. To achieve this, a neural network ϕ is introduced and the generative transition is then defined as

$$p(x_{t-1}|x_t) = q(x_{t-1}|\hat{x}, x_t) \tag{7}$$

where $\hat{x} = \phi(x_t, t)$, an estimate of x_0 predicted by ϕ. Inspired by Ho et al[31], the neural network is further adapted to predict the added noise, i.e. $\hat{\epsilon}_t = \phi(x_t, t)$. Then the estimated $\hat{x}$ is derived as

$$\hat{x} = (1/\alpha_t)x_t - (\sigma_t/\alpha_t)\hat{\epsilon}_t \tag{8}$$

The object of this model is to minimize $\mathcal{L}(t) = \|\epsilon - \hat{\epsilon}_t\|^2$ via mini-batch gradient descent optimization. Once this model is well trained, new sample points can be generated. Any sample point $x_T \in \mathcal{N}(0, \mathrm{I})$ is iteratively denoised via eq. 7 for $t = T, \ldots, 0$ to obtain a new data point x_0.

Another feature of LigandDiff is that it is scaffold-based, *i.e.*, it only diffuses or denoises one ligand while other ligands as well as the central metal are fixed at each step. In drug discovery such "scaffold modeling" is widely used where a large portion of the molecule is kept fixed while the remaining parts of the molecule are modified.[33] Keeping the main scaffold structure, while modifying small functional groups, allows for accurate and quick design of new drugs with desirable properties.[34] Generative models can further speed up such targeted exploration of the chemical space due to its powerful flexibility with little or no human intervention. In addition, the generation of one new ligand is similar to ligand substitution which is a useful tool for new material discovery in organometallic complexes.[35,36] Overall, LigandDiff can be used to investigate the structure-activity relationship (SAR) in the area of metal-ligand interactions. For example, Ferrocene (Fc), a 'sandwich' organometallic complex, has two stable cyclopentadienyl groups

which can be easily redesigned by either replacing the whole group with different organic groups or simply attaching extra functional groups to the five-membered rings, leading to a large variety of derivatives. Indeed, Fc analogues have shown potential as drug candidates against malaria as well as cancer and each modified cyclopentadienyl moiety has specific mechanisms by which they interact with biomolecules, which improves the overall therapeutic efficacy.[37] We anticipate that LigandDiff has the potential to accelerate the process of lead optimization for this and other classes of organometallic compounds.

Methods

Dataset

Naveen and co-workers reported a set of ~86k mononuclear octahedral transition metal complexes,[38] from which we curated M complexes, M= Cr, Mn, Fe, Co, Ni, Cu, Zn with 100 atoms or less. We further constrained the nonmetal elements to {H, C, N, O, F, P, S, Cl, Br}. Next, complexes with missing hydrogens or disorder were excluded, leading to a set of 23308 complexes, each of which has at least two ligands. We then used molSimplify[39,40] to break the complexes apart to obtain ligand information. Each ligand is masked for diffusion/generation: For example, for a complex with six ligands, we can obtain six variations of this complex, each of which has a unique ligand to diffuse/generate. With such an implementation, we finally obtained 87531 samples. All hydrogen atoms were removed to reduce the computational cost. Two subsets of 400 samples are used for validation and testing, while the remaining data was used for training.

Molecule representation

All complexes are regarded as 3D point clouds in space. A point cloud x is denoted as $x = [r, h, h_L]$, where r is the atom coordinates $r = (r_1, \ldots, r_N) \in \mathbb{R}^{N\times 3}$ and h is the one-hot representations of atom type $h = (h_1, \ldots, h_N) \in \mathbb{R}^{N\times m}$, N is the total number of atoms, m is the number of atom types. h_L is a one-hot embedding to decode the ligand group information, *i.e.*, for a given atom, it belongs to which ligand; $h_L = \left(h_{L_1}, \ldots, h_{L_N}\right) \in \mathbb{R}^{N\times l}$, where l is the number of ligands. In LigandDiff, before passing through the neural network ϕ, the noise is added to only the coordinates and the atom types which belong to the diffused ligand, while the ligand-group embeddings are unchanged. Although the whole atomic embeddings $x = [r, h, h_L]$ are updated through the neural network, we only consider the predicted coordinates and the predicted atom type features.

3D-conditional diffusion models

In LigandDiff, each assigned ligand x^L is diffused or denoised under a fixed context u, *i.e.*, the undestroyed ligands as well as the central metal. u has the same embedding constituents as x. Under this context, the generative process in eqs. 7 and 8 are adapted as

$$p(x_{t-1}^L | x_t^L, u) = q(x_{t-1}^L | \hat{x}^L, x_t^L) \tag{9}$$

$$\hat{x}^L = (1/\alpha_t) x_t^L - (\sigma_t/\alpha_t)\, \phi(x_t^L, u, t) \tag{10}$$

The schematic process of LigandDiff is given in Figure 1.

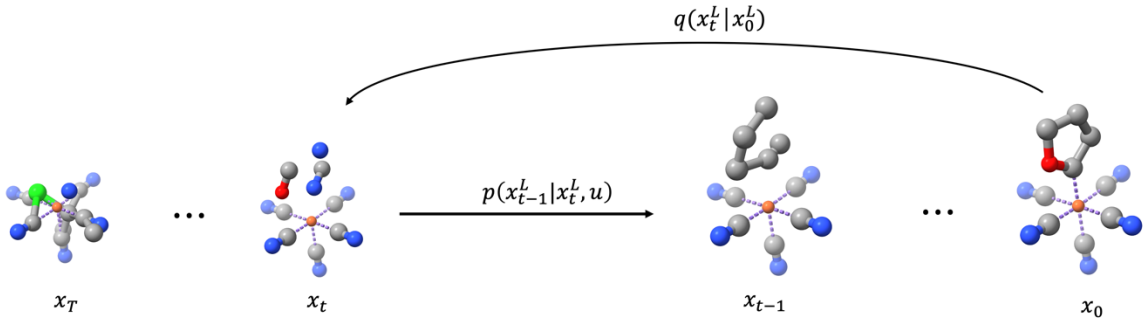

Figure 1. Overview of LigandDiff. It starts from the forward diffusion process q from x_0^L to x_t^L to sample noised data for a given ligand x^L. Once the model is well trained, any new ligand can be generated from $x_T^L \in \mathcal{N}(0, I)$ by iteratively denoising x_t^L through the conditional distributions p.

The framework of Geometric Vector Perceptrons (GVPs)[41] is used to model the dynamics of the diffusion model, *i.e*., the learnable function ϕ. GVP is based on graph neural networks (GNNs)[42] where each molecule is regarded as a graph and each atom is a node, while each bond is an edge. GVPs define scalar and vector features to embed the nodes and edges. The edges $e = (s, V)$ consist of a normalized direction vector $V \in \mathbb{R}^{N\times 1\times 3}$ as well as the distance between two nodes $s \in \mathbb{R}^{N\times 1}$, where N is the number of edges. Under the linear transformations and nonlinear activation, the edges are embedded as $e' = (s', V')$, where $s' \in \mathbb{R}^{N\times F}$ and $V' \in \mathbb{R}^{N\times 1\times 3}$, F is the size of the hidden layer. The nodes follow a similar transformation but start with only scalar features, $h = (s)$, where $s \in \mathbb{R}^{N\times m}$ and m is the features of each node. The nodes are transformed as $h' = (s', V')$, where $s' \in \mathbb{R}^{N\times F}$ and $V' \in \mathbb{R}^{N\times (F/2)\times 3}$. In the following section, unless explicitly stated otherwise, the embedded edges are denoted as e and the embedded nodes are denoted as h for clarity. In LigandDiff, each graph is fully connected and interactions between all atoms are counted during the message passing process which is defined as

$$m_{ij} = \phi_e\left(h_i, h_j, e_{ij}\right) \tag{11}$$

$$\tilde{e}_{ij} = \phi_{att} m_{ij} \tag{12}$$

$$m_i = \sum_j^{N-1} \tilde{e}_{ij} m_{ij} \tag{13}$$

$$h_i = \phi_h(h_i, m_i) \tag{14}$$

where h_i is the center node, h_j is its neighbor, e_{ij} is the edge attributes, ϕ_e includes three GVPs to collect messages from its neighbors. ϕ_{att} is an attention layer defined by one GVP for edges. And ϕ_h consists of two GVPs to update the center node. The whole message passing process is iterated several times to fully extract geometric information from molecules. Finally, another GVP is used to transform the scalar and vector features back to the point cloud format x, from which the

predicted noise $\hat{\epsilon} = [\hat{\epsilon}^r, \hat{\epsilon}^h]$ is extracted, where $\hat{\epsilon}^r$ is the noise from the coordinates and $\hat{\epsilon}^h$ is the noise from the atom types.

Training

LigandDiff was trained with T = 500 diffusion steps. The model has 5 layers with 192 hidden features with a batch size of 64. It uses the Adam optimizer with the learning rate of 1.0×10^{-4}. The model was trained on a single NVIDIA A100 GPU. And it took about 7 mins for one epoch.

Assessment metrics

Various metrics are used to fully assess the performance of LigandDiff. We first use OpenBabel[43] to add bonds to the generated data points x^L. Then the validity of the generated ligands is evaluated by RDKit[44]. The metric of the model's ability to generate valid ligands is given as

$$p_l^{val} = \frac{N_l^{valid}}{N_{total}} \tag{15}$$

where N_l^{valid} is the number of valid ligands, N_{total} is the number of total ligands generated. The connectivity of the ligands is used to check whether all atoms in the valid ligands are fully connected, which is calculated as

$$p_l^{con} = \frac{N_l^{valid\&connected}}{N_l^{valid}} \tag{16}$$

where $N_l^{valid\&connected}$ is the number of valid and connected ligands. The uniqueness and novelty are also evaluated as

$$p_l^{uniq} = \frac{N_l^{unique}}{N_l^{valid\&connected}} \tag{17}$$

$$p_l^{nov} = \frac{N_l^{nov}}{N_l^{valid\&connected}} \tag{18}$$

where N_l^{unique} is the number of unique ligands among outputs and N_l^{nov} is the number of the ligands outside the training dataset. Finally, we check the validity of the whole complex using molSimplify, calculated as

$$p_c^{val} = \frac{N_c^{valid}}{N_{total}} \tag{19}$$

where N_c^{valid} is the number of valid complexes.

Results and discussion

Random sample

In the test set, 105 out of 400 samples have only one heavy atom as a ligand to generate and half of the complexes have less than 5 heavy atoms to generate. We believe it is easy for LigandDiff to sample valid complexes under such conditions. To strictly evaluate the performance of LigandDiff, we randomly sample the size of the generated ligands in the range of 6 to 20. This range was chosen because it only covers 41.7% of the size distribution in the training dataset while 51.5% of the diffused ligands in the training set have 5 heavy atoms or less. The results are reported in Table 1. Although this size sample is challenging for LigandDiff, it still shows remarkable performance. Unlike other generative models[45,46] which explicitly employ valency rules to improve validity, LigandDiff is able to learn these rules implicitly and thus generates what are perceived as valid ligands. And these valid ligands are highly connected and unique, leading to 90 % valid complexes.

Table 1. Performance of LigandDiff [a]

	N_{atom}	p_l^{val}	p_l^{con}	p_l^{uniq}	p_l^{nov}	p_c^{val}
Random sample	6 ~ 20	0.94 ± 0.012	0.96 ± 0.008	0.97 ± 0.009	0.96 ± 0.009	0.90 ± 0.016
	6	0.97 ± 0.006	0.94 ± 0.012	0.56 ± 0.013	0.81 ± 0.016	0.91 ± 0.015

Fix the ligand size	7	0.97 ± 0.009	0.95 ± 0.015	0.70 ± 0.094	0.83 ± 0.018	0.92 ± 0.014
	8	0.97 ± 0.007	0.95 ± 0.005	0.89 ± 0.018	0.94 ± 0.011	0.91 ± 0.011
	9	0.96 ± 0.011	0.95 ± 0.007	0.90 ± 0.012	0.98 ± 0.007	0.90 ± 0.014
	10	0.96 ± 0.012	0.95 ± 0.008	0.92 ± 0.030	0.98 ± 0.004	0.91 ± 0.010
	11	0.96 ± 0.008	0.96 ± 0.010	0.95 ± 0.009	0.99 ± 0.007	0.91 ± 0.014
PPR_100	11 ~ 40	0.94 ± 0.017	0.94 ± 0.015	0.92 ± 0.019	1.0 ± 0	0.87 ± 0.026

[a]The results are reported as 'mean ± std' over 10 independent runs.

Fixing the ligand size

To assess whether the model learns chemical principles, or just memorized the ligands in the training set, we fixed the size of the generated ligand, *i.e.*, each complex in the test set now has a ligand with the same size to generate. We start from the ligand size n = 6 and increase it to 11. As shown in Table 1, the metrics of validity are at a consistently high level, where the validity of ligands and the whole complexes are around 0.96 and 0.91, respectively. And the connectivity is also noteworthy, yielding more than 94% connected ligands. The rapid increase of uniqueness and novelty strongly indicates that LigandDiff "learns chemistry" to some extent and can use this knowledge to generate new and valid ligands. For n = 6, uniqueness is only 0.56 which means nearly half of the successfully generated ligands are duplicates, while this metric reaches 0.95 when n increases to 11. The observed improvement is consistent with our understanding of chemistry, *i.e.*, the diversity of structures increases as the size of the system increases. And LigandDiff appears to understand this and keeps generating different and diverse ligands. This also applies to novelty. Instead of simply generating the ligands which already exist in the training set, LigandDiff tends to design completely new ligands. Even with n = 6, LigandDiff still can

generate 81% novel ligands and when n reaches 11, the generated ligands are almost exclusively outside the training set. All these results show that in this extreme situation where all the generated ligands have to have the same size, LigandDiff still can generate different ligands under the given context. Some examples are given in Figure 2.

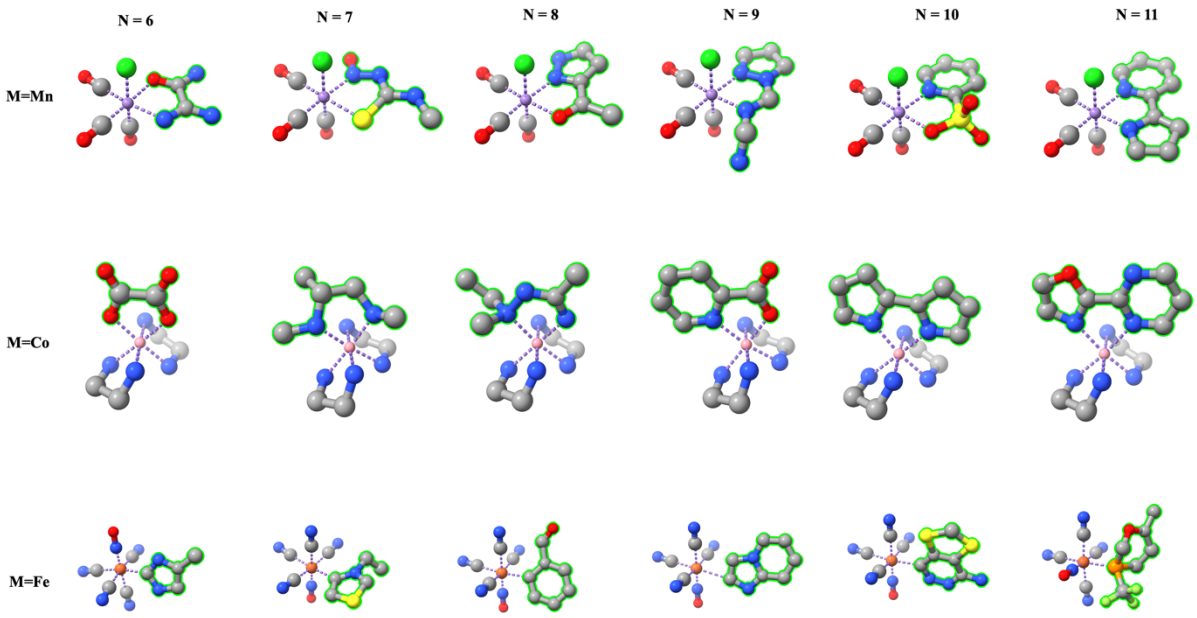

Figure 2. The generated complexes under a fixed ligand size. Each column has the ligand with the identical size to generate but under a different context, while each row has the ligand with an increasing size to sample under the same context. The generated ligand is highlighted in green outline. Atoms include carbon: gray, nitrogen: blue, oxygen: red, fluorine: greenyellow, phosphorous: orange, sulfur: yellow and chlorine: green.

Beyond Cr, Mn, Fe, Co, Ni, Cu, Zn

To further assess the capability of LigandDiff, we curated a challenging dataset termed PPR_100 from the original database[38]. The PPR_100 set includes 100 Pt, Pd, and Ru complexes with more than 50 atoms. These three types of transition metal complexes were chosen since they are the top 3 complexes in the database excepting the transition metal complexes already covered in our dataset. For each complex, we mask the ligand with more than 10 heavy atoms and 148 samples are obtained because some complexes have one more suitable ligand to mask. Since 67.8% of the diffused ligands in the training set have 10 or less heavy atoms, this becomes a challenge for LigandDiff to generate valid complexes. The results are given in Table 1. Although Pt, Pd and Ru are not included in the training set, LigandDiff is still able to generate 94% valid and connected ligands with 100% novelty. Again, LigandDiff generates these ligands with high diversity since only 8% ligands are duplicates. And 87% complexes are valid as indicated by molSimplify. Some examples of the generated complexes as well as the reference complexes are shown in Figure 3. In LigandDiff the metal is constrained in the context, and it is only used to predict the noise of the generated ligand but the metal itself is never involved in the diffusion process. Such flexible design enables LigandDiff to generate novel ligands for any transition metal.

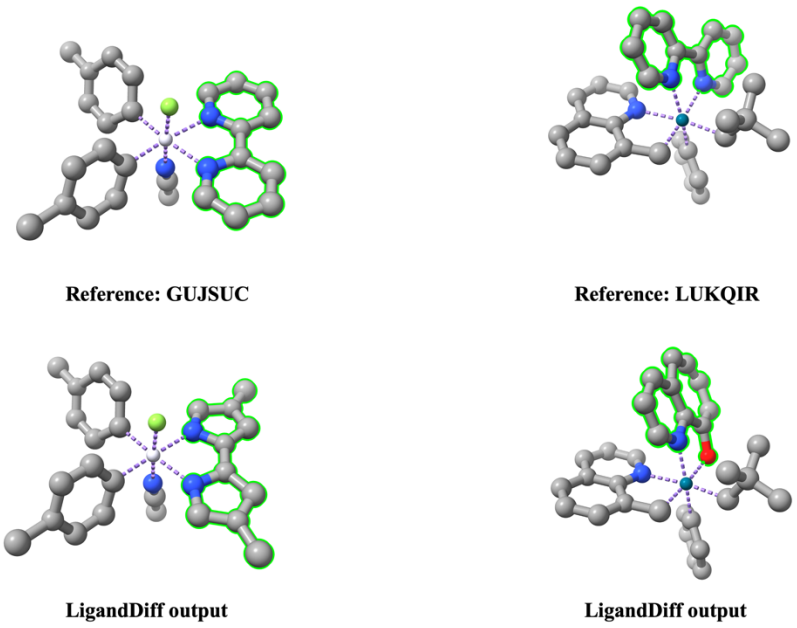

Figure 3. Examples of generated complexes(bottom) and the corresponding reference complexes (up) in the PPR_100 set. The CSD code is given.

Synthetic accessibility

To assess whether the generated ligands are synthesizable, we calculate the average synthetic accessibility (SA) score[47]. As shown in Table 2, LigandDiff generates realistic ligands with high SA. And this capability remains at a high level as the ligand size increases.

Table 2. The SA scores of LigandDiff.

	N_{atom}	SA[a]
Random sample	6 ~ 20	0.69 ± 0.008
Fix the ligand size	6	0.80 ± 0.005
	7	0.77 ± 0.020
	8	0.74 ± 0.005
	9	0.74 ± 0.003
	10	0.73 ± 0.008
	11	0.72 ± 0.008
PPR_100	11 ~ 40	0.68 ± 0.008

[a] 0 = hard, 1 = easy.

Conclusion

The design of novel organometallic complexes is highly demanding but worth the effort given their various applications. In this study, we described LigandDiff, a 3D-conditional diffusion model for transition metal complexes generation. LigandDiff designs realistic ligands under a set of given ligands and their associated metals and it is capable of generating novel and unique ligands which is relevant for molecular design. Moreover, we found LigandDiff to be transferable and can design ligands for transition metals that are not included in the training dataset. Overall, we believe this tool has potential to facilitate the design of novel organometallics for applications ranging from metallodrugs to materials.

ASSOCIATED CONTENT

Data Availability Statement

All data and code are available at https://github.com/Neon8988/LigandDiff.

Supporting Information

The size distributions of the diffused ligands in training dataset and the synthetic accessibility calculation. (PDF)

Author information

Corresponding Author

Kenneth M. Merz, Jr.

Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States; Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States.

Email: merz@chemistry.msu.edu

Author

Hongni Jin

Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States

Notes

The authors declare no competing financial interest.

Acknowledgments

The authors gratefully acknowledge financial support from the NIH (GM130641). The authors also thank the high-performance computing center (HPCC) at Michigan State University for providing all computational resources.

References

(1) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **2020**, *323*, 844-853.

(2) Mohs, R. C.; Greig, N. H. Drug Discovery and Development: Role of Basic Biological Research. *Alzheimers Dement (N Y)* **2017**, *3*, 651–657.

(3) Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the Size of Drug-like Chemical Space Based on GDB-17 Data. *J. Comput. Aided. Mol. Des.* **2013**, *27*, 675–679.

(4) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44,* D1202–D1213.

(5) 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.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci*. **2018**, *4*, 268–276.

(6) Gao, K.; Duc Duy Nguyen; Tu, M.; Wei, G.-W. Generative Network Complex for the Automated Generation of Drug-like Molecules. *J. Chem. Inf. Model.* **2020**, *60*, 5682–5698.

(7) Cao, N. D.; Kipf, T. MolGAN: An implicit generative model for small molecular graphs. *arXiv (Machine learning*), 1805.11973, September 27, 2022, ver. 2. https://doi.org/10.48550/arXiv.1805.11973 (accessed).

(8) You, J.; Ying, R.; Ren, X.; Hamilton, W. L.; Leskovec, J.GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models. *arXiv (Machine learning*), 1802.08773, June 23, 2018, ver. 2. https://doi.org/10.48550/arXiv.1802.08773 (accessed).

(9) Li, Y.; Hu, J.; Wang, Y.; Zhou, J.; Zhang, L.; Liu, Z. DeepScaffold: A Comprehensive Tool for Scaffold-Based de Novo Drug Discovery Using Deep Learning. *J. Chem. Inf. Model.* **2019**, *60*, 77–91.

(10) Kaitoh, K.; Yamanishi, Y. Scaffold-Retained Structure Generator to Exhaustively Create Molecules in an Arbitrary Chemical Space. *J. Chem. Inf. Model.* **2022**, *62*, 2212-2225.

(11) Gaynor, D.; Griffith, D. M. The Prevalence of Metal-Based Drugs as Therapeutic or Diagnostic Agents: Beyond Platinum. *Dalton Trans.* **2012**, *41*, 13239.

(12) de Paiva, R. E.; Marçal Neto, A.; Santos, I. A.; Jardim, A. C.; Corbi, P. P.; Bergamini, F. R. What Is Holding Back the Development of Antiviral Metallodrugs? A Literature Overview and Implications for SARS-COV-2 Therapeutics and Future Viral Outbreaks. *Dalton Trans.* **2020**, *49*, 16004–16033.

(13) Rosenberg, B.; Van Camp, L.; Krigas, T. Inhibition of Cell Division in Escherichia Coli by Electrolysis Products from a Platinum Electrode. *Nature* **1965**, *205*, 698–699.

(14) Anthony, E. J.; Bolitho, E. M.; Bridgewater, H. E.; Carter, O. W.; Donnelly, J. M.; Imberti, C.; Lant, E. C.; Lermyte, F.; Needham, R. J.; Palau, M.; Sadler, P. J.; Shi, H.; Wang, F.-X.; Zhang, W.-Y.; Zhang, Z. Metallodrugs Are Unique: Opportunities and Challenges of Discovery and Development. *Chem. Sci.* **2020**, *11*, 12888–12917.

(15) Messori, L.; Merlino, A. Cisplatin Binding to Proteins: A Structural Perspective. *Coord. Chem. Rev.* **2016**, *315*, 67–89.

(16) Jamieson, E. R.; Lippard, S. J. Structure, Recognition, and Processing of Cisplatin−DNA Adducts. *Chem. Rev.* **1999**, *99*, 2467–2498.
(17) Sasmal, P. K.; Streu, C. N.; Meggers, E. Metal Complex Catalysis in Living Biological Systems. *Chem. Commun*. **2013**, *49*, 1581–1587.
(18) Ndamse, C. C.; Masamba, P.; Kappo, A. P. Bioorganometallic Compounds as Novel Drug Targets against Schistosomiasis in Sub-Saharan Africa: An Alternative to Praziquantel? *Adv. Pharm. Bull.* **2022**, *12*, 283-297.
(19) Gao, H.; Yu, R.; Ma, Z.; Gong, Y.; Zhao, B.; Lv, Q.; Tan, Z. Recent Advances of Organometallic Complexes in Emerging Photovoltaics. *J. Polym. Sci.* **2021**, *60*, 865–916.
(20) Bellini, M.; Bevilacqua, M.; Marchionni, A.; Miller, H. A.; Filippi, J.; Grützmacher, H.; Vizza, F. Energy Production and Storage Promoted by Organometallic Complexes. *Eur. J. Inorg. Chem*. **2018**, *40*, 4393–4412.
(21) Wang, D.-Y.; Liu, R.; Guo, W.; Li, G.; Fu, Y. Recent Advances of Organometallic Complexes for Rechargeable Batteries. *Coordination Chemistry Reviews* **2021**, *429*, 213650.
(22) Duan, C.; Nandy, A.; Terrones, G.; Kastner, D. W.; Kulik, H. J. Active Learning Exploration of Transition-Metal Complexes to Discover Method-Insensitive and Synthetically Accessible Chromophores. *JACS Au* **2022**, *3*, 391–401.
(23) Nandy, A.; Duan, C.; Goffinet, C.; Kulik, H. J. New Strategies for Direct Methane-To-Methanol Conversion from Active Learning Exploration of 16 Million Catalysts. *JACS Au* **2022**, *2*, 1200–1213.
(24) Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In International Conference on Machine Learning, pp. 2256–2265. PMLR, 2015.
(25) Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. 2022. Repaint: Inpainting using denoising diffusion probabilistic models. In IEEE Conference on Computer Vision and Pattern Recognition. 11461–11471.
(26) William Harvey, Saeid Naderiparizi, Vaden Masrani, Christian Weilbach, and Frank Wood. 2022. Flexible Diffusion Modeling of Long Videos. *arXiv (Machine learning*), 2205.11495, December 15, 2022, ver. 3. https://doi.org/10.48550/arXiv.2205.11495 (accessed).
(27) Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. 2021. Structured denoising diffusion models in discrete state-spaces. In Advances in Neural Information Processing Systems.
(28) Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d. In International Conference on Machine Learning, pp. 8867–8887. PMLR, 2022.
(29) Igashov, I., Stark, 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 (Machine learning*), 2210.05274, October 11, 2022, ver. 1. https://doi.org/10.48550/arXiv.2210.05274 (accessed).
(30) Chentao Cao, Zhuo-Xu Cui, Shaonan Liu, Dong Liang, and Yanjie Zhu. 2022. High-Frequency Space Diffusion Models for Accelerated MRI. *arXiv (Machine learning*), 2208.05481, January 20, 2024, ver. 5. https://doi.org/10.48550/arXiv.2208.05481 (accessed).

(31) Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 2020, 33:6840–6851.
(32) Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In International Conference on Machine Learning, pp. 2256–2265. PMLR, 2015.
(33) Hu, Y.; Stumpfe, D.; Bajorath, J. Computational Exploration of Molecular Scaffolds in Medicinal Chemistry. *J. Med. Chem*. **2016**, *59*, 4062–4076.
(34) Hu, Y.; Stumpfe, D.; Bajorath, J. Recent Advances in Scaffold Hopping. *J. Med. Chem*. **2016**, *60*, 1238–1246.
(35) Cornia, A.; Fabretti, A. C.; Garrisi, P.; Mortalò, C.; Bonacchi, D.; Gatteschi, D.; Sessoli, R.; Sorace, L.; Wernsdorfer, W.; Barra, A.-L. Energy-Barrier Enhancement by Ligand Substitution in Tetrairon(III) Single-Molecule Magnets. Angewandte Chemie **2004**, *116*, 1156–1159.
(36) Langley, S. K.; Chilton, N. F.; Moubaraki, B.; Murray, K. S. Single-Molecule Magnetism in $\{Co_2^{III}Dy_2^{III}\}$-Amine-Polyalcohol-Acetylacetonate Complexes: Effects of Ligand Replacement at the Dy (III) Sites on the Dynamics of Magnetic Relaxation. *Inorg. Chem. Front.* **2015**, *2*,867–875.
(37) Patra, M.; Gasser, G. The Medicinal Chemistry of Ferrocene and Its Derivatives. *Nat Rev Chem* **2017**, *1*, 1–12.
(38) Arunachalam, N.; Gugler, S.; Taylor, M.; Duan, C.; Nandy, A.; Jon Paul Janet; Meyer, R.; Jonas Albrecht Oldenstaedt; Daniel; Kulik, H. J. Ligand Additivity Relationships Enable Efficient Exploration of Transition Metal Chemical Space. *J. Chem. Phys.* **2022**, *157*, 184112-184127.
(39) Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J. MolSimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J.Comput. Chem.* **2016**, *37*, 2106-2117.
(40) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Ind. Eng. Chem. Res.* **2018**, *57*, 13973–13986.
(41) Jing, B., Eismann, S., Suriana, P., Townshend, R. J., and Dror, R. Learning from protein structure with geometric vector perceptrons. *arXiv (Machine learning*), 2009.01411, May 16, 2021, ver. 3. https://doi.org/10.48550/arXiv.2009.01411 (accessed).
(42) Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv (Machine learning*), 1806.01261, 2018, October 17, 2018, ver. 3. https://doi.org/10.48550/arXiv.1806.01261 (accessed).
(43) Noel M. O'Boyle, Michael Banck, Craig A. James, Chris Morley, Tim Vandermeersch, Geoffrey R. Hutchison. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
(44) Landrum, G. RDKit: Open-Source Cheminformatics. http://www.rdkit.org/ (accessed).
(45) Imrie, F.; Bradley, A. R.; van der Schaar, M.; Deane, C. M. Deep Generative Models for 3D Linker Design. *J. Chem. Inf. Model.* **2020**, *60*, 1983–1995.
(46) Huang, Y.; Peng, X.; Ma, J.; Zhang, M. 3DLinker: An E(3) Equivariant Variational Autoencoder for Molecular Linker Design. *arXiv (Machine learning*), 2205.07309, 2022, May 15, 2022, ver. 1. https://doi.org/10.48550/arXiv.2205.07309 (accessed).

(47) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. *J Cheminform* **2009**, *1*, 8.

TOC