# Supplementary material: Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules

## Architecture

Here we summarize the exact settings used in all layers of our neural network architecture. The structure of the interaction blocks can be found in Schütt et al. [14]. The number of atom features was set to 128 and used in all atom-wise dense layers of the interaction block and filter-generating layers. Distances are expanded in the filter-generating layers using 25 Gaussians with equally spaced centers 0 Å$\leq \mu \leq$ 10 Å. Overall, we use nine interaction blocks for feature extraction. We re-use one embedding layer with 128 features at all steps depicted in Fig. 1 of the paper. The output network for type predictions consists of five atom-wise dense layers with shifted-softplus non-linearity and 128, 96, 64, 32, and 1 atom features, respectively. The output network for distance predictions also consists of five atom-wise dense layers with shifted-softplus non-linearity and 128, 171, 214, 257, and 300 atom features. Both output networks contain a final softmax layer.

## Training details

The neural networks were trained with stochastic gradient descent using the ADAM optimizer [48]. We start with a learning rate of $10^{-4}$ which is reduced using a decay factor of $0.5$ after $10$ epochs without improvement of the validation loss. The training is stopped at lr $\leq 10^{-6}$. Afterwards, the model with lowest validation error is selected for generation.

While the atom type labels $\mathbf{q}_i^{\text{type}}$ can be directly obtained from the training data, the labels for distance distributions $\mathbf{q}_{ij}^{\text{dist}}$ are obtained using:

$$\left[\mathbf{q}_{ij}^{\text{dist}}\right]_l = \frac{\exp(-\frac{1}{\gamma}(d_{(t+i)j} - \mu_l)^2)}{\sum_{l=1}^{300}\exp(-\frac{1}{\gamma}(d_{(t+i)j} - \mu_l)^2)} \qquad \forall\, j < t+i.$$

The width of the Gaussians can be tuned with the $\gamma$ parameter, which we set to 10% of the bin size in our experiments, resulting in very peaky, uni-modal label vectors.

## Controlling randomness with the temperature parameter

In order to control the randomness during generation, we do not directly implement Eq. 3 but include a temperature parameter $T$:

$$p(\mathbf{r}_{t+i}|\mathbf{R}^t_{\leq i-1}, \mathbf{Z}^t_{\leq i}) = \frac{1}{\alpha}\exp\left(\frac{\sum_{j=1}^{t+i-1}\log p(d_{(t+i)j}|\mathbf{x}_j)}{T}\right). \tag{5}$$

Increasing $T$ will smoothen the grid distribution, effectively increasing randomness, whereas small values lead to a peaky distribution and thus less randomness. For all experiments, we chose a fixed temperature of $T = 0.1$ according to the following procedure.

We used the G-SchNet model trained on 50k equilibrium structures from QM9 [26–28] and generated 20k molecules for each $T \in \{2, 1, 0.1, 0.01, 0.001\}$. From each set, 1k valid and unique molecules were randomly chosen, where 800 resembled test structures, 100 resembled training structures, and the remaining 100 were novel structures with more than 9 heavy atoms. We relaxed the five sets of 1k molecules at the PBE/def2-SVP level of theory[59, 60] using the Orca program, employing the resolution of identity (RI) approximation[61, 62] and Grimme D3 dispersion correction with Becke-Johnson damping.[63] The root-mean-square deviation (RMSD) between atomic positions before and after relaxation was measured. A smaller RMSD means that the generated structures are closer to a true equilibrium configuration.

In Fig. 6 we show boxplots of the RMSD for the five different temperatures $T$. We see that the smallest median values and interquartile ranges are observed for values of $T$ smaller than 1. Since increasing $T$ corresponds to increasing randomness during sampling, the increase in the RMSD for $T = 1$ and $T = 2$ is expected. However, decreasing $T$ to smaller values than $0.1$ does not lead to smaller RMSDs. Generally, we expect the number of unique sampless to decrease as $T$ gets too small. Therefore, for the other experiments, we chose the highest value for $T$ that still produces structures close to equilibrium, namely $T = 0.1$.

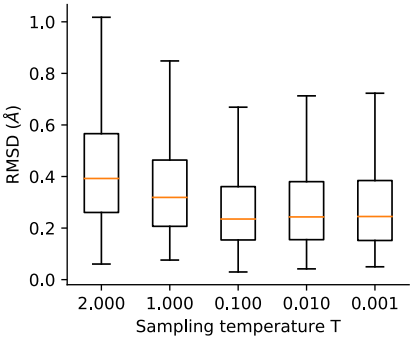


*Figure 6: The boxplots show the RMSD of atomic positions between generated and relaxed structures for different values of $T$, where the boxes extend from the lower to the upper quartile values, the red lines indicate the medians, and the whiskers reach up to 1.5 times the interquartile ranges. Outliers are not shown.*

### Sampling generation traces for training

The generation traces start with the focus and origin tokens set to the same position, which is the center of mass of the respective training molecule. The first new type and position are taken from the atom closest to the center of mass. At each subsequent step, we randomly select one of the already placed non-token atoms as focus point (which is only a single choice for the second step). The new position and type are then taken from the neighbor which is closest to the origin token, where neighbors are all unplaced atoms of the training molecule that are connected to the focus by a bond. If no unplaced neighbors are left, the new type is set to the stop token and the focused atom is marked as finished, i.e. it cannot be chosen as focus anymore. For the next step, another already placed (unfinished) atom is randomly chosen as focus and the procedure is repeated until all atoms have been placed and marked as finished.

### Generating molecules

For the first step of molecule generation, the focus and origin token are set to the origin of a 3d grid. The grid extends up to 1.7 Å into all dimensions with a step-size of 0.05 Å. The token are processed by G-SchNet to sample the type of the first non-token atom and to obtain the two predicted distance distributions. The probabilities of the grid cell positions are calculated according to Eq. 5 in order to sample the position of the first atom. Then, at each subsequent step, a random unfinished non-token atom is selected as focus token (which is only a single choice for the second step). The grid is centered on the focus (but the origin token stays at the former origin of the grid) and G-SchNet predictions are obtained to sample the type and position of the next atom (see Fig. 1 in the paper for an exemplary generation step). If the stop token is predicted as next type, the currently focused atom is marked as finished and no position is sampled. Instead the next generation step is initialized by randomly selecting one of the remaining unfinished atoms as focus point. The generation process stops if no unfinished atoms are left. For our experiments, we also stopped the process if molecules were not finished after placing 35 atoms and discarded these structures as invalid (usually ~1% of generated molecules).

### Matching molecules

To remove non-unique structures and identify moleclues that resemble training or test data, we calculate the Tanimoto similarity of path-based fingerprints (FP2) [52] for pairs of molecules with Open Babel. If the similarity is one, we compare the canonical SMILES representations in a second step. If they match, the two molecules are treated as equal. Note that this is a conservative approach as it filters out some spatial isomers which cannot be distinguished with SMILES.

a

| valid | unique | in train | in test | novel | ≤9 heavy | with origin token | without origin token |
|---|---|---|---|---|---|---|---|
| ✓ | | | | | | 77.0% | 57.9% |
| ✓ | ✓ | | | | | 70.6% | 56.7% |
| ✓ | ✓ | ✓ | | | | 9.4% | 1.9% |
| ✓ | ✓ | | ✓ | | | 12.0% | 2.6% |
| ✓ | ✓ | | | ✓ | | 49.2% | 52.2% |
| ✓ | | | | | ✓ | 60.4% | 13.7% |

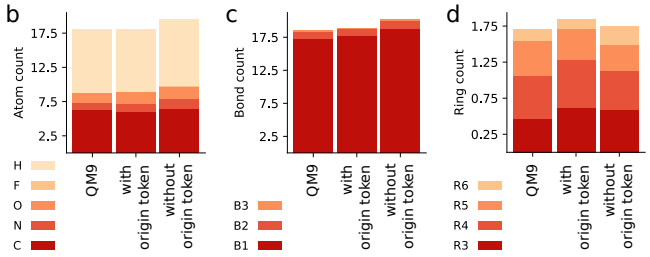


*Figure 7: Ablation study on the effect of the origin token. Statistics are compared for 20k molecules generated by a G-SchNet model with origin token and 20k molecules generated by a G-SchNet model without origin token. Table (a) shows the percentage of generated molecules for which the properties indicated by the check marks in each row hold. The average numbers of atoms, bonds, and rings per generated molecule and per QM9 molecule are compared in (b), (c), and (d), respectively. B1, B2, and B3 correspond to single, double, and triple bonds. R3, R4, R5, and R6 are rings of size 3 to 6.*

### Ablation study

In order to assess the effect of the origin token, we conduct an ablation study where we remove the origin token from the input to a G-SchNet model during training and generation. All other hyperparameters are identical to the ones used when training the standard G-SchNet with origin token. After training, we generate 20k molecules with each architecture and compare their statistics. Fig. 7a shows that the validity of generated molecules drops by almost 20 percent without the origin token. Furthermore, the amount of generated molecules that match QM9 training or test structures significantly decreases. All QM9 structures consist of at most 9 heavy atoms but only 13.7% of the molecules generated without origin token have 9 or less heavy atoms (compared to 60.4% with origin token). The diverging atom count can also be seen in Fig. 7b. Moreover, the bond count in Fig. 7c also diverges from the training data distribution. The ring count in Fig. 7d, on the other hand, is not noticeably better or worse without origin token. There is an increase in six-membered rings and a decrease in all smaller ring structures compared to the model with origin token. Both models slightly diverge from the QM9 training data ring count. Overall, we conclude that the origin token has a significant, positive effect on the approximated probability distribution. It enables G-SchNet to better capture the characteristics of the training data, leading to a model that generates more valid molecules which are more faithful to the training structures but still equally unique and unseen (in test or novel) compared to a model without origin token.

### Targeted discovery with respect to further electronic properties

In a similar fashion to the experiments for molecules with a small HOMO-LUMO gap, we bias G-SchNet models towards large values of three more electronic properties available in QM9, namely the isotropic polarizability, the dipole moment, and the electronic spatial extent. The results depicted in Fig. 8 show clear shifts in the distribution of targeted properties for molecules generated with biased G-SchNets. We again use small subsets of molecules from QM9 exhibiting the respective property for fine-tuning of the G-SchNet model previously trained on 50k structures. These subsets consist of $2100/500$ molecules with an isotropic polarizability $\geq 91$ Bohr$^3$, $3000/500$ molecules with a dipole moment $\geq 5.75$ Debye, and $4400/500$ molecules with an electronic spatial extent $\geq 1785$ Bohr$^2$ for training/validation, respectively. In contrast to the HOMO-LUMO gap experiments, where we relaxed generated structures and calculated the gap numerically with time-consuming DFT simulations, we train three separate SchNet models to predict the three electronic properties. We use 100k molecules from QM9 for training, 10k for validation, and the remaining structures as a test set. The mean absolute test error is $0.070$ Bohr$^3$ for the electronic polarizability, $0.016$ Debye for the dipole moment, and $0.126$ Bohr$^2$ for the electronic spatial extent.

### Detailed statistics

We provide two tables which report more detailed statistics on relevant properties of generated molecules. In Table 1 we compare 20k molecules generated by our standard G-SchNet model, by the G-SchNet model trained on QM9 without structures with 3- or 4-membered rings, by the

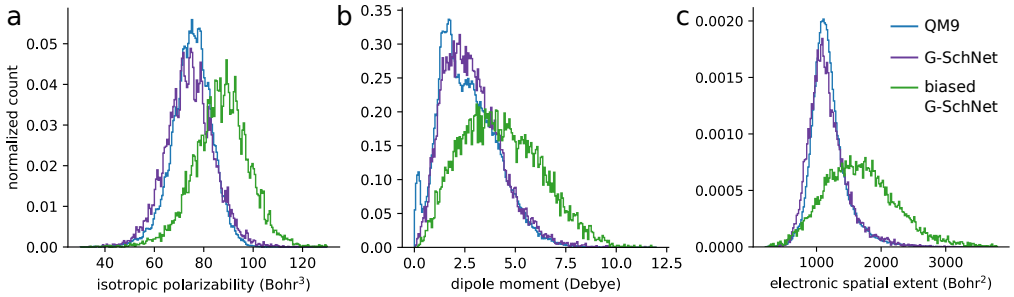


*Figure 8: Distribution of three quantum-chemical properties for molecules from QM9 (blue), generated by an unbiased G-SchNet (purple), and generated by G-SchNets biased towards larger values of the respective property (green).*

*Table 1: Statistics for all of our models and CGVAE [22] (molecules provided by the authors). CGVAE was trained on all molecules in QM9. Our models were trained on 50k randomly selected molecules from QM9. For the second model molecules with three- and four-membered rings were excluded. The third model was fine-tuned on 3k molecules with a HOMO-LUMO gap $\leq$ 4.5 eV. The numbers are percent of all validly generated molecules (which are 77%, 78%, 69% and 100% of 20k generated for the models from left to right). The molecules are categorized according to the checkmarks in the first five columns. Unseen refers to molecules not found in the training data and novel stands for molecules not found in QM9. $\leq$9 heavy marks molecules with nine or less heavy atoms and gap refers to the HOMO-LUMO gap and was calculated for relaxed structures.*

| unique | unseen | novel | $\leq$9 heavy | gap $\leq$5eV | G-SchNet 50k | G-SchNet 50k no R3/R4 | G-SchNet 50k + 3k gap$\leq$4.5eV | CGVAE ~134k |
|---|---|---|---|---|---|---|---|---|
| ✓ | | | | | 91.6% | 89.4% | 73.8% | 98.4% |
| ✓ | ✓ | | | | 79.4% | 68.3% | 66.8% | 87.4% |
| ✓ | | ✓ | | | 63.8% | 68.3% | 57.5% | 87.4% |
| ✓ | | | ✓ | | 70.2% | 58.2% | 57.8% | 30.0% |
| ✓ | ✓ | | ✓ | | 58.0% | 37.1% | 50.8% | 18.9% |
| ✓ | | ✓ | ✓ | | 42.4% | 37.1% | 41.5% | 18.9% |
| ✓ | ✓ | | | ✓ | 6.8% | 21.2% | 43.4% | – |

G-SchNet model biased towards small HOMO-LUMO gaps, and by the constrained graph variational autoencoder (CGVAE). The provided numbers are percent of all validly generated structures.

In Table 2 we compare the number of valid, novel, and unique molecules generated by the standard G-SchNet as well as six related generative models that rely either on graphs or SMILES strings as molecule representation. The shown numbers are the respective percentage of 20k generated molecules for each model. Note that this means that a high novelty score can also come from a low number of valid molecules, as each invalid structure is not included in the training data and therefore novel. Similarly, a high uniqueness score is only interesting if the majority of generated molecules is valid as otherwise the model generates many unique but invalid structures. In general, the methods cannot directly be compared since generating 3d molecular structures is a different task than generating graphs or strings. For example, all of the valid G-SchNet molecules correspond to one proper 3d structure whereas valid graphs and SMILES strings may have no, one, or many different corresponding 3d conformers which cannot easily be found without expensive quantum-chemical simulations.

*Table 2: Percent of valid, novel (not in training data), and unique molecules among 20k structures generated after training on QM9 for G-SchNet and related models working with SMILES or graph representations. Molecules are considered valid if the valency-constraints of all its atoms are met. In order to identify duplicate and novel molecules, we use molecular fingerprints and canonical SMILES strings as explained in section "matching molecules" above. Statistics for graph- and SMILES-based models are taken from Fig. 3 in Liu et al. [22].*

| | **G-SchNet**<br>3d structure | **CGVAE***[22]<br>graph | **GraphVAE***[64]<br>graph | **NeVAE***[65]<br>graph | **LSTM***[22]<br>SMILES | **CVAE***[19]<br>SMILES | **GVAE***[32]<br>SMILES |
|---|---|---|---|---|---|---|---|
| **valid** | 77.07% | 100.00% | 61.00% | 98.00% | 94.78% | 10.00% | 30.00% |
| **novel** | 87.47% | 94.30% | 85.00% | 100.00% | 82.98% | 90.00% | 95.44% |
| **unique** | 91.91% | 98.57% | 40.90% | 99.86% | 96.94% | 64.50% | 9.30% |

*statistics are taken from Fig. 3 in Liu et al. [22]

## References


[1] J. Behler and M. Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98(14):146401, 2007.

[2] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104(13):136403, 2010.

[3] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.*, 3(5):e1603015, 2017.

[4] M. Gastegger, J. Behler, and P. Marquetand. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.*, 8(10):6924–6935, 2017.

[5] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148 (24):241722, 2018.

[6] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature communications*, 9(1):3887, 2018.

[7] K. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature communications*, 10(1):5024, 2019.

[8] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. Von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, 108(5):058301, 2012.

[9] M. Eickenberg, G. Exarchakis, M. Hirn, and S. Mallat. Solid harmonic wavelet scattering: Predicting quantum molecular energy from invariant descriptors of 3d electronic densities. In *Advances in Neural Information Processing Systems 30*, pages 6543–6552. Curran Associates, Inc., 2017.

[10] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature Communications*, 8:13890, 2017.

[11] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1263–1272, 2017.

[12] D. Jha, L. Ward, A. Paul, W.-k. Liao, A. Choudhary, C. Wolverton, and A. Agrawal. Elemnet: Deep learning the chemistry of materials from only elemental composition. *Scientific reports*, 8 (1):17593, 2018.

[13] J. S. Smith, O. Isayev, and A. E. Roitberg. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.*, 8(4):3192–3203, 2017.

[14] K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, pages 992–1002, 2017.

[15] J. S. Smith, O. Isayev, and A. E. Roitberg. Ani-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific data*, 4:170193, 2017.

[16] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Physical Review B*, 99(6):064114, 2019.

[17] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *NIPS*, pages 2224–2232, 2015.

[18] B. Ramsundar, S. Kearnes, P. Riley, D. Webster, D. Konerding, and V. Pande. Massively multitask networks for drug discovery. *arXiv preprint arXiv:1502.02072*, 2015.

[19] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, and A. Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.*, 2016.

[20] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley. Molecular graph convolutions: moving beyond fingerprints. *J. Comput. Aided Mol. Des.*, 30(8):595–608, 2016.

[21] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing, and V. Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2): 513–530, 2018.

[22] Q. Liu, M. Allamanis, M. Brockschmidt, and A. Gaunt. Constrained graph variational autoencoders for molecule design. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 7795–7804. Curran Associates, Inc., 2018.

[23] W. Jin, R. Barzilay, and T. Jaakkola. Junction tree variational autoencoder for molecular graph generation. *arXiv preprint arXiv:1802.04364*, 2018.

[24] J. You, B. Liu, Z. Ying, V. Pande, and J. Leskovec. Graph convolutional policy network for goal-directed molecular graph generation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 6410–6421. Curran Associates, Inc., 2018.

[25] Y. Li, L. Zhang, and Z. Liu. Multi-objective de novo drug design with conditional graph generative model. *Journal of cheminformatics*, 10(1):33, 2018.

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

[27] J.-L. Reymond. The chemical space project. *Acc. Chem. Res.*, 48(3):722–730, 2015.

[28] L. Ruddigkeit, R. Van Deursen, L. C. Blum, and J.-L. Reymond. 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.

[29] J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 82–90. Curran Associates, Inc., 2016.

[30] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas. Learning representations and generative models for 3d point clouds. In *International Conference on Machine Learning*, pages 40–49, 2018.

[31] D. Weininger. 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.

[32] M. J. Kusner, B. Paige, and J. M. Hernández-Lobato. Grammar variational autoencoder. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1945–1954. JMLR. org, 2017.

[33] G. L. Guimaraes, B. Sanchez-Lengeling, P. L. C. Farias, and A. Aspuru-Guzik. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. *arXiv preprint arXiv:1705.10843*, 2017.

[34] H. Dai, Y. Tian, B. Dai, S. Skiena, and L. Song. Syntax-directed variational autoencoder for structured data. In *International Conference on Learning Representations*, 2018.

[35] D. Janz, J. van der Westhuizen, B. Paige, M. Kusner, and J. M. H. Lobato. Learning a generative model for validity in complex discrete structures. In *International Conference on Learning Representations*, 2018.

[36] M. H. S. Segler, T. Kogej, C. Tyrchan, and M. P. Waller. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science*, 4(1):120–131, 2018.

[37] M. Popova, O. Isayev, and A. Tropsha. Deep reinforcement learning for de novo drug design. *Science Advances*, 4(7), 2018.

[38] J. Lim, S. Ryu, J. W. Kim, and W. Y. Kim. Molecular generative model based on conditional variational autoencoder for de novo molecular design. In *J. Cheminformatics*, 2018.

[39] T. Blaschke, M. Olivecrona, O. Engkvist, J. Bajorath, and H. Chen. Application of generative autoencoder in de novo molecular design. *Molecular Informatics*, 37(1-2):1700123, 2017.

[40] A. Gupta, A. T. Müller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, and G. Schneider. Generative recurrent networks for de novo drug design. *Molecular Informatics*, 37(1-2):1700111, 2017.

[41] P. B. Jørgensen, M. Mesta, S. Shil, J. M. García Lastra, K. W. Jacobsen, K. S. Thygesen, and M. N. Schmidt. Machine learning-based screening of complex molecules for polymer solar cells. *The Journal of Chemical Physics*, 148(24):241735, 2018.

[42] Y. Li, O. Vinyals, C. Dyer, R. Pascanu, and P. Battaglia. Learning deep generative models of graphs. *arXiv preprint arXiv:1803.03324*, 2018.

[43] W. Jin, K. Yang, R. Barzilay, and T. Jaakkola. Learning multimodal graph-to-graph translation for molecular optimization. In *International Conference on Learning Representations*, 2019.

[44] E. Mansimov, O. Mahmood, S. Kang, and K. Cho. Molecular geometry prediction using a deep generative graph neural network. *arXiv preprint arXiv:1904.00314*, 2019.

[45] N. W. A. Gebauer, M. Gastegger, and K. T. Schütt. Generating equilibrium molecules with deep neural networks. *arXiv preprint arXiv:1810.11347*, 2018.

[46] A. van den Oord, N. Kalchbrenner, L. Espeholt, K. Kavukcuoglu, O. Vinyals, and A. Graves. Conditional image generation with PixelCNN decoders. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 4790–4798. Curran Associates, Inc., 2016.

[47] P. Ramachandran and G. Varoquaux. Mayavi: 3D Visualization of Scientific Data. *Computing in Science & Engineering*, 13(2):40–51, 2011. ISSN 1521-9615.

[48] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[49] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller. SchNetPack: A deep learning toolbox for atomistic systems. *Journal of Chemical Theory and Computation*, 15(1):448–455, 2019. doi: 10.1021/acs.jctc.8b00908.

[50] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[51] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.

[52] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. Open babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1):33, Oct 2011.

[53] A. D. Becke. Density-functional thermochemistry. III. the role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.

[54] C. Lee, W. Yang, and R. G. Parr. LYP correlation: Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785, 1988.

[55] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211, 1980.

[56] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98(45):11623–11627, 1994.

[57] F. Neese. The ORCA program system. *WIREs Comput. Mol. Sci.*, 2(1):73–78, 2012.

[58] RDKit, online. RDKit: Open-source cheminformatics. `http://www.rdkit.org`. [Online; accessed 23-May-2019].

[59] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996.

[60] F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7(18):3297–3305, 2005.

[61] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 240(4):283–290, 1995.

[62] O. Vahtras, J. Almlöf, and M. W. Feyereisen. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.*, 213(5-6):514–518, 1993.

[63] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132(15):154104, 2010.

[64] M. Simonovsky and N. Komodakis. GraphVAE: Towards generation of small graphs using variational autoencoders. In *International Conference on Artificial Neural Networks*, pages 412–422. Springer, 2018.

[65] B. Samanta, D. Abir, G. Jana, P. K. Chattaraj, N. Ganguly, and M. Gomez-Rodriguez. NeVAE: A deep generative model for molecular graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1110–1117, 2019.