



Fine-tuning of a generative neural network for designing multi-target compounds

Thomas Blaschke¹ · Jürgen Bajorath¹

Received: 4 March 2021 / Accepted: 23 May 2021 / Published online: 28 May 2021
© The Author(s) 2021

Abstract

Exploring the origin of multi-target activity of small molecules and designing new multi-target compounds are highly topical issues in pharmaceutical research. We have investigated the ability of a generative neural network to create multi-target compounds. Data sets of experimentally confirmed multi-target, single-target, and consistently inactive compounds were extracted from public screening data considering positive and negative assay results. These data sets were used to fine-tune the REINVENT generative model via transfer learning to systematically recognize multi-target compounds, distinguish them from single-target or inactive compounds, and construct new multi-target compounds. During fine-tuning, the model showed a clear tendency to increasingly generate multi-target compounds and structural analogs. Our findings indicate that generative models can be adopted for de novo multi-target compound design.

Keywords Multi-target activity · Deep learning · Generative modeling · Structure-promiscuity relationships · Multi-target ligand design

Introduction

Computational de novo and multi-target ligand design are important topics in the pharmaceutical research community. During the early stages of drug discovery, computational compound design is often applied to complement experimental or virtual screening and identify new molecules with desired properties in a time-efficient manner [1]. Recently, deep generative models have become popular for de novo compound design [2, 3]. De novo design using generative models typically involves a two-step process. First, a generative model is trained on a large data set of known compounds using their SMILES [4] representations; then, the model is fine-tuned to generate only compounds with desired properties. The first training step enables the generative model to learn the syntax of molecular string representations and generate new syntactically correct strings without restrictions.

For example, Arús-Pous et al. [5] have shown that generative models trained with one million SMILES were capable of covering the chemical space of a fully enumerated set of all possible molecules with up to 13 atoms. Fine-tuning of the generative model is either carried out using reinforcement or transfer learning [6, 7]. In reinforcement learning, the generative model first constructs molecules and then receives property-based feedback for the compounds, for example, by applying a bioactivity classifier. Depending on the feedback, the generative model updates its output to increase or decrease the number of structurally related compounds. Through iterative feedback, the model generates compounds that increasingly meet desired properties. In transfer learning, the model does not rely on feedback when generating compounds, but enters a second learning phase on a smaller subset of compounds with desired properties. By repeatedly exposing the generative model to preferred molecules, it learns common features and then generates new compounds with such features. Both fine-tuning approaches have been applied in different studies to optimize compounds with activity against individual targets [6–11].

Multi-target activity of small molecules, often also referred to as promiscuity, has gained much attention in the medicinal chemistry community over the last two decades. Of note, multi-target activity is often viewed controversially.

✉ Jürgen Bajorath
bajorath@bit.uni-bonn.de

¹ Department of Life Science Informatics and Data Science, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Friedrich-Hirzebruch-Allee 6, 53115 Bonn, Germany

On the one hand, promiscuity is associated with non-specific ligand-target interactions and assay artifacts caused by aggregators or other assay interference compounds [12–15]. On the other hand, true multi-target activity provides the fundamental basis for polypharmacology of drugs, which is caused by concomitant *in vivo* interactions with multiple targets [16–19]. Polypharmacology is often essential for therapeutic efficacy in the treatment of multifactorial diseases [16–21], but may also cause undesired side effects. Accordingly, the study of multi-target activity is important not only to better understand the fundamental basis of polypharmacology, but also to control potential side effects of new drugs. To rationalize multi-target activity and predict multi-target compounds, different computational approaches have been adopted [17, 22, 23]. Most of these predictions have focused on the identification of additional targets for known active compounds [24–31], while only few have attempted to predict different types of promiscuous compounds directly [32–35]. These latter studies have shown that promiscuous and non-promiscuous compounds could be differentiated with reasonable accuracy on the basis of chemical structure, indicating the presence of structural patterns that distinguish compounds with single- and multi-target activity. These studies have also revealed that nearest neighbor relationships between multi-target or single-target compounds strongly contributed to the predictions. Further exploring structure-promiscuity relationships is expected to aid in the design of compounds with pre-defined multi-target activities, which is currently mostly attempted by combining pharmacophore information for different targets [36–39]. However, another potential route to designing multi-target compounds would be adapting deep generative models for this task, which to our knowledge has not been attempted so far.

Herein, we explore the possibility of fine-tuning a SMILES-based generative neural network to recognize multi-target compounds, distinguish them from single-target or inactive compounds, and construct new multi-target candidates. If structural patterns exist that are characteristic of multi-target compounds, a generative model should be able to detect these patterns via transfer learning and utilize them to create new multi-target compounds. In contrast to other machine learning algorithms, SMILES-based generative models do not rely on the explicit calculation of substructure fingerprints or physicochemical properties. Neither use these models the information that compounds have multi-target activity, nor are they specifically trained to distinguish between multi- and single-target compounds. This added layer of abstraction enables the recognition of non-obvious structure-promiscuity relationships in an unsupervised manner, thereby bridging between transfer learning and multi-target ligand design.

For this study, we used high-confidence compound sets extracted from biological screening data and applied the

publicly available REINVENT model [40] for transfer learning. Through fine-tuning, we determined if the generative model was able to recognize multi-target compounds and distinguish them from single-target or inactive compounds. Furthermore, newly generated SMILES representations were analyzed following each fine-tuning cycle to assess the recovery of known compounds and structural neighbors as a proof-of-concept measure for the principal capacity of the model to generate new multi-target compounds.

Methods and materials

Data extraction

For the analysis, a comprehensive collection of publicly available PubChem screening data [41] was used after applying a number of confidence criteria. Only qualitative compound assay results for human targets with the designation ‘active’ and ‘inactive’ were considered. Assays imported from ChEMBL [42], BindingDB [43], or Tox21 [44] and revoked or ambiguously annotated assays were excluded from the analysis. Because the assessment of multi-target compounds is particularly vulnerable to false positive activity assignments, assays with a hit rate higher than 2% were also excluded. Furthermore, compounds with potential liabilities were omitted including designated pan-assay interference compounds [45] detected with publicly available filters from ChEMBL, ZINC [46], and RDKit [47], molecules violating empirical medicinal chemistry rules [48], and others yielding aggregation alerts [49]. Finally, compounds with inconsistent activity annotations across different assays for the same target were also discarded.

Qualifying compounds were assigned to three different sets depending on the number of human targets they were active against. Screening molecules with activity against five or more different targets were classified as multi-target compounds. In addition, compounds with activity against only one target and confirmed inactivity against at least four other targets were categorized as single-target compounds. Furthermore, compounds with no reported activity but inactivity in assays against at least five different targets were classified as inactive (“no-target”) compounds. Qualifying compounds not meeting any of these selection criteria were not further considered. Multi-target compounds were required to be active against at least five different targets to ensure that they were promiscuous in nature, setting them clearly apart from single-target compounds.

Generative model

For generative modeling, we used REINVENT, a publicly available model that was originally trained on ~1.4 million bioactive compounds from ChEMBL [40].

For our study, REINVENT was fine-tuned using a random selection of 1000 multi-target compounds. Fine-tuning was carried out for 200 epochs using the ADAM optimizer [50]. The loss function used during fine-tuning minimized the negative log-likelihood (NLL) of the SMILES of multi-target training compounds. After each training epoch, the NLL for the canonical SMILES of all detected compounds was calculated. To avoid overfitting, the SMILES representations of multi-target compounds were randomized [51].

Compound design

After each fine-tuning epoch, 1,000,000 SMILES were sampled. The sampling of molecules over different epochs was performed using the same random seed. Accordingly, two identical models would sample the same SMILES. Consequently, any difference in the sampled SMILES directly resulted from fine-tuning of the underlying generative model and was not a result of random sampling. The generated SMILES were canonicalized with RDKit and all unique valid SMILES were considered to represent newly generated compounds.

Molecular similarity

For each generated molecule, the extended-connectivity fingerprint with bond diameter 6 (ECFP6) [52] and constant 2048 bit format was used as a representation and Tanimoto similarity to multi-, single-, and no-target compounds from PubChem was calculated. If the Tanimoto similarity of a generated molecule to a PubChem compound was at least 0.6, it was classified as a structural (fingerprint) neighbor.

Results and discussion

Compound sets

Applying high-confidence data selection criteria taking positive as well as negative assay results into account, 2809 multi-target, 61,928 single-target, and 295,395 no-target compounds were extracted from PubChem screening assays. As expected, the compound data set was imbalanced, containing comparably few multi-target compounds.

A random selection of 1000 multi-target compounds was used as a training set for fine-tuning the general-purpose REINVENT model. The remaining 1809 multi- and 61,928 single-target compounds, as well as an equally sized random subset of 61,928 no-target compounds, were used as test sets.

Fine-tuning

The REINVENT model was fine-tuned for 200 epochs. After each epoch, the NLL for all PubChem compounds was calculated. NLL values provide a quantitative estimate for the probability that the model will (re-)generate a particular compound at a given stage in the process. The resulting NLL value distribution is shown in Fig. 1.

Prior to fine-tuning (epoch 0), there was a notable difference in the NLL distribution between multi-, single-, and no-target compounds. On the basis of the NLL median values, multi-target compounds were 3–4 times less likely to be generated than single- or no-target compounds. Moreover, comparing the 75% quartile, multi-target compounds were 10 times less likely than single- or no-target compounds.

This difference in the likelihood of generating multi-target compounds compared to others could be rationalized

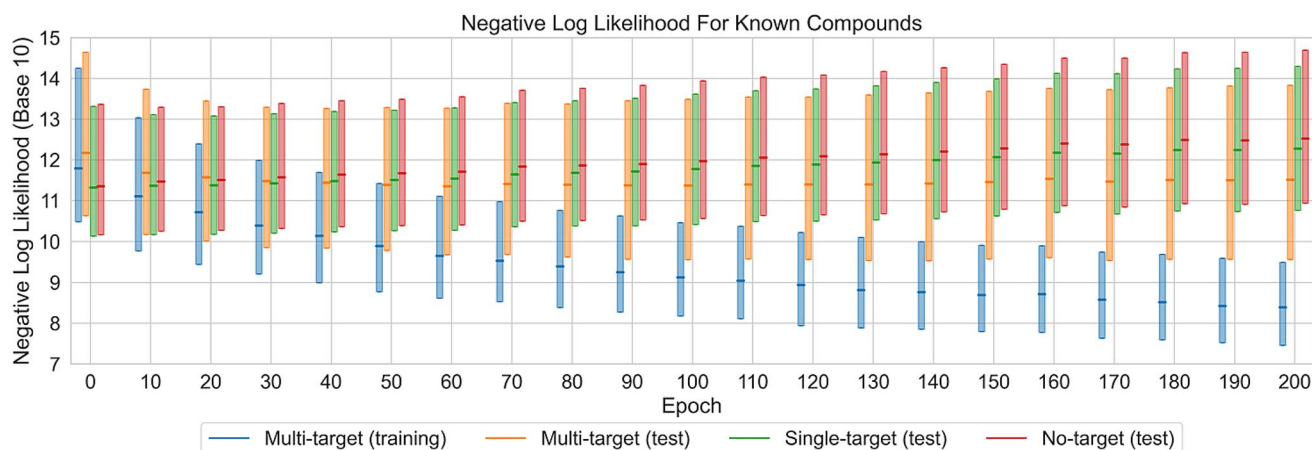


Fig. 1 Distribution of negative log-likelihood values. Boxplots report the NLL distribution for known multi-, single-, and no-target compounds. Shown are the 25% quartile, the median (horizontal line), and the 75% quartile. Whiskers and statistical outliers are omitted for clarity

by considering the derivation of the generative REINVENT model [40]. The REINVENT model was originally trained on a large collection of bioactive compounds from ChEMBL, the majority of which are single-target compounds [53]. ChEMBL only contains a very small proportion of compounds with reported activity against more than five targets (<2%), which has remained essentially constant over time [53]. Accordingly, the REINVENT model was tailored towards single-target compounds. Prior to fine-tuning, the model also preferentially learned structural features from single-target screening compounds (but also inactive compounds), as revealed by the higher likelihood of generating single- and no-target compounds from screening assays; an interesting observation.

However, after only 10 epochs of fine-tuning, the NLL value distributions for multi-, single-, and no-target compounds were very similar including their mean values. At this stage, the 25% and 75% quartile displayed a difference of less than 0.4 NLL units.

After 30 epochs of fine-tuning, the model started to preferentially recognize multi-target compounds from the training set. Compared to the initial state prior to fine-tuning, the median NLL was reduced by 1.5 units. This reduction already corresponded to a 400-fold increase in the likelihood to generate multi-target compounds at this early stage. Moreover, for the 75% quartile, the median NLL was reduced by 2.2. Concomitantly, the NLLs for single- and no-target compounds slightly increased. After 50 epochs, the median NLL value for multi-target test compounds was lowered relative to the median for the other training compounds and the difference further increased during fine-tuning. Similarly, the median NLL for multi-target training compounds consistently decreased during fine-tuning, as monitored in Fig. 1. After 200 epochs, the median NLL approached a value of 8. Taken together,

these observations indicated that the model increasingly learned structural features shared by multi-target training and test compounds and discriminated single- and no-target compounds, consistent with the underlying design idea.

Generating multi-target compounds

Of note, the NLL values were exclusively calculated for the canonical SMILES of each compound. Since a compound may also be represented by a variety of non-canonical SMILES strings, the likelihood of generating a compound is expected to be underestimated by the NLL calculated on the basis of its canonical SMILES representation. Therefore, to more comprehensively monitor the ability of the generative model to create multi-target compounds, 1,000,000 SMILES were sampled randomly after each epoch of fine-tuning, canonicalized, and filtered for PubChem compounds. The results are shown in Fig. 2.

Prior to fine-tuning, the model generated 3% of known compounds across the three different sets, with no significant difference between multi-, single-, and no-target compounds. After 25 epochs of fine-tuning, 20% of the multi-target training set, 10% of the multi-target test set, and 4% of both the single-target and no-target test sets were retrieved. Throughout fine-tuning, the number of reproduced multi-target training and test compounds increased. After 200 epochs, 85% of the multi-target training and 21% of the test set were reproduced, in contrast to only 4% of the single-target and 3% of the no-target test set. The increase in the number of multi-target test set compounds provided firm evidence that the model recognized structure-promiscuity patterns in the training set and used these patterns to preferentially generate multi-target compounds.

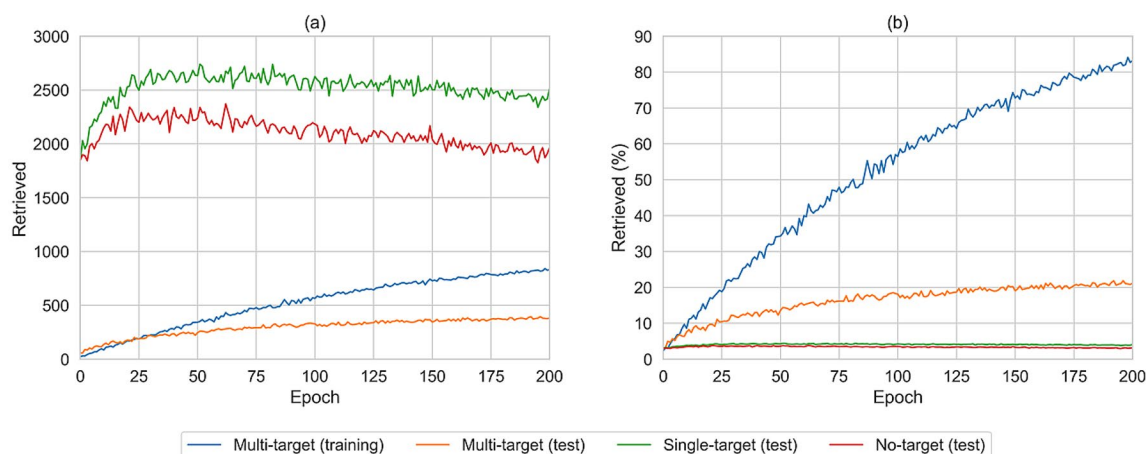


Fig. 2 Compound retrieval during fine-tuning. Reported is the number of retrieved multi-, single-, and no-target compounds among 1,000,000 SMILES sampled after each fine-tuning epoch. **a** shows the absolute number of retrieved compounds and **b** the percentage for each data set

Neighbors of multi-target compounds

Nearest neighbor relationships were previously found to play an important role in distinguishing between different types of promiscuous and non-promiscuous compounds using supervised machine learning [33–35]. The influence of nearest neighbor relationships indicated that promiscuous compounds were typically more similar to other promiscuous than non-promiscuous compounds and vice versa [35]. To analyze structural neighbors of all generated compounds, Tanimoto similarity to other training and test compounds was calculated (excluding exactly reproduced compounds). As a neighbor criterion, an ECFP6 similarity threshold of 0.6 was applied, thus focusing on closely related compounds. To account for the difference in size between the multi-target test set and the single- and no-target test sets, the number of detected neighbors was normalized relative to the size of each set. The results are shown in Fig. 3 (and essentially parallel the observations made in Fig. 2). Prior to fine-tuning, 3% of the generated compounds were structural neighbors of PubChem training and test compounds. Over the first 30 epochs of fine-tuning, the number of generated neighbors increased for the training and test sets. After epoch 50, the absolute number of generated neighbors for the no-target test set decreased to 38,000 compounds but remained constant at 48,000 compounds for the single-target set. For the multi-target training and test set, the number of neighbors increased throughout fine-tuning to 22,000 and 10,000 compounds, respectively (corresponding to 22 neighbors per training and five neighbors per test compound). The large number of neighbors generated for multi-target compounds provided further evidence for the ability of the fine-tuned

model to recognize characteristic structural patterns and create structural analogs.

Figure 4 shows examples of newly generated multi-target candidate compounds and their nearest neighbors from the training and test set. In all three instances, structural modifications compared to the nearest neighbor from the training set produced candidate molecules that closely resembled test set compounds, hence illustrating the ability of the fine-tuned model to sample chemical space populated by multi-target compounds. The generation of such analogs complemented the capacity of the model to reproduce known multi-target compounds, which was monitored as a quality criterion.

Scaffold analysis

In addition to nearest neighbor analysis, we also assessed the similarity between the newly generated compounds, training set, and test set compounds on the basis of Bemis and Murcko (BM) scaffold composition [54]. From each compound, the BM scaffold was extracted and scaffolds of newly generated, training, and test set compounds were compared. The multi-target training set was found to contain 869 unique BM scaffolds and the multi-target test set 1463 BM scaffolds, 1252 of which (86%) were not present in the training set. Furthermore, the single-target and no-target test sets yielded 33,977 and 34,024 BM scaffolds, respectively, 335 and 245 of which were present in the multi-target training set, respectively. We then determined scaffolds from each data set that were generated during fine-tuning. The results are shown in Fig. 5. Prior to fine-tuning, the model retrieved ~41% of the BM scaffolds from the training and test sets. Over the first 25 epochs of fine-tuning, the total

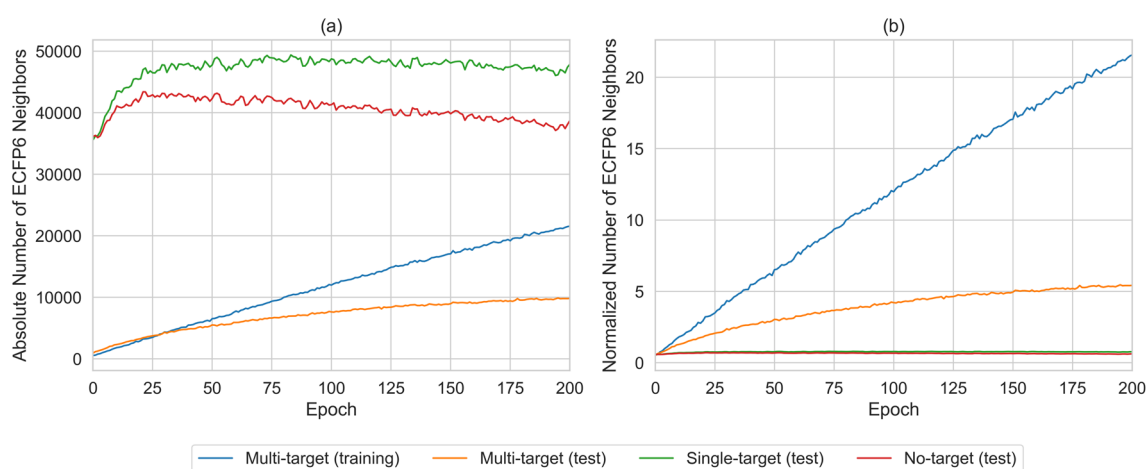


Fig. 3 Distribution of structural neighbors. Reported is the number of generated (fingerprint) neighbors of multi-, single-, and no-target compounds among 1,000,000 SMILES sampled after each fine-tuning epoch. Duplicated canonical SMILES were removed and only

unique, valid, and novel SMILES were considered. **a** shows the absolute number of generated neighbors and **b** the normalized number of generated neighbors per known compound

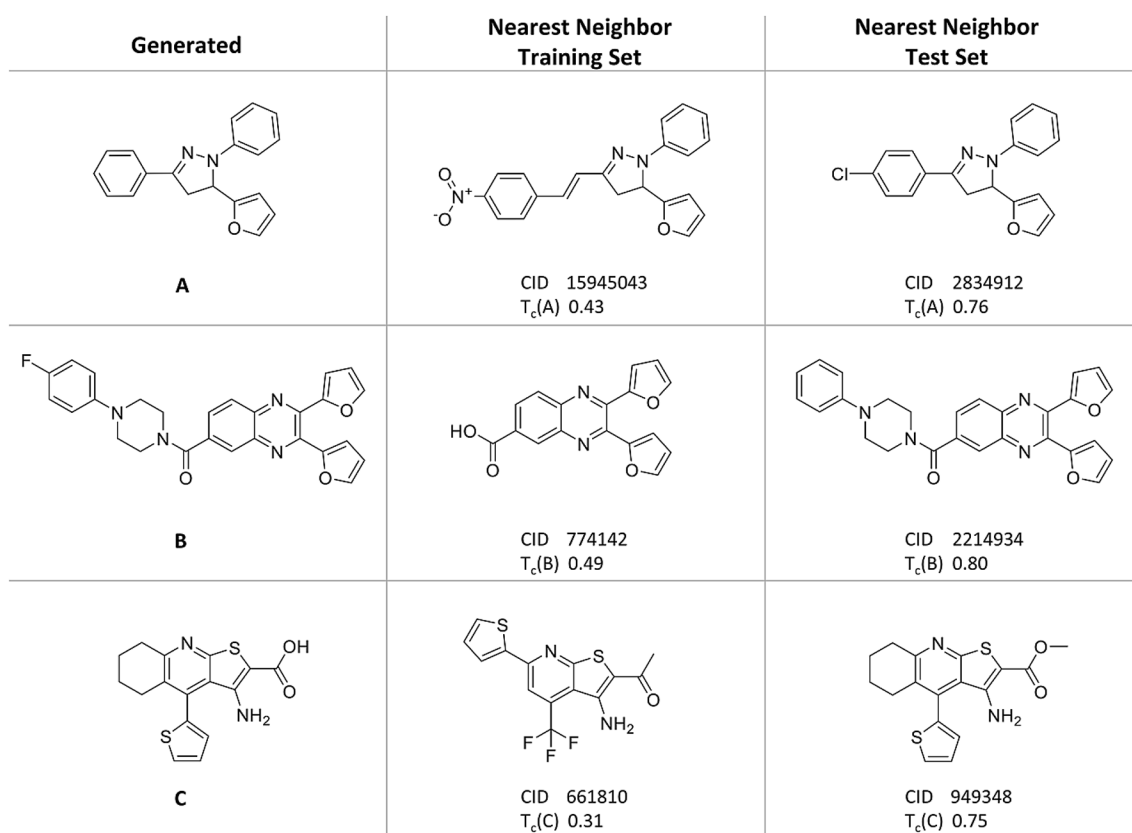


Fig. 4 Exemplary compounds. Shown are three examples of newly generated multi-target compounds together with their nearest neighbors from the multi-target training and test set, respectively. In each case, the calculated ECFP6 Tanimoto coefficient (T_c) is reported

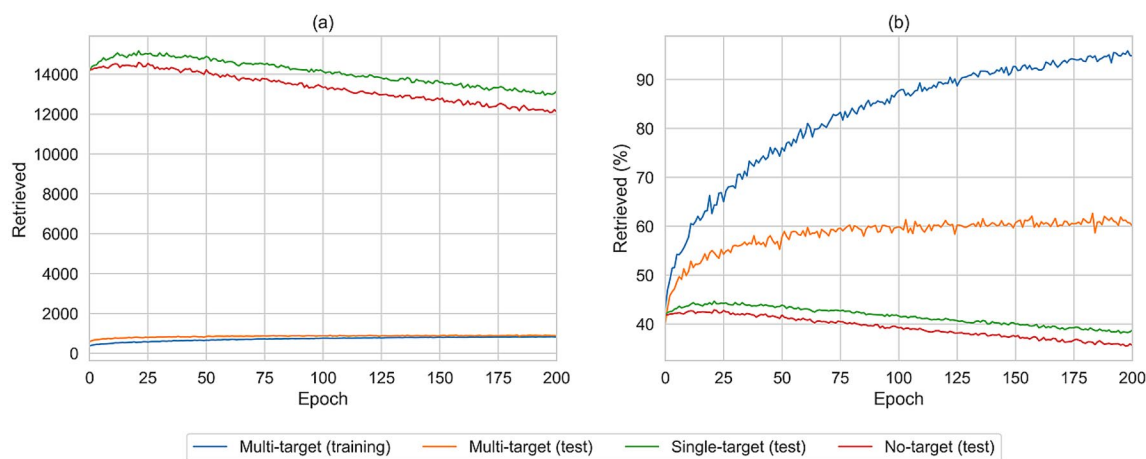


Fig. 5 Scaffold retrieval during fine-tuning. Reported is the number of multi-, single-, and no-target BM scaffolds detected in 1,000,000 SMILES sampled after each fine-tuning epoch. **a** shows the total number of BM scaffolds and **b** the percentage for each data set

number of retrieved BM scaffolds increased for the training and all test sets. After epoch 25, the number of retrieved scaffolds decreased for the no-target and single-target test set to 12,119 (36%) and 13,052 (38%), respectively. For the multi-target training and test set, the number of retrieved

BM scaffolds increased throughout fine-tuning to 833 (96%) and 905 (62%), respectively. Remarkably, the majority of the retrieved BM scaffolds for the multi-target test set (698 of 905; 77%) were not present in the training set. Hence, scaffold analysis provided further evidence for the ability of the

fine-tuned model to recognize structural characteristics of multi-target compounds.

Compound classification

To further explore newly generated compounds, we trained a decision tree ensemble classifier using the gradient boost algorithm from the XGBoost library [55]. The classifier was built to distinguish multi-target compounds from single- and no-target compounds. It was derived using the multi-target training set (positive class label) and combined random subsets of 30,000 single- and no-target compounds each (negative class label). Using the remaining screening compounds as a test set, the classifier reached a ROC AUC score of 0.82, a Matthews correlation coefficient (MCC) of 0.30, and recall of 0.32, hence confirming reasonable accuracy.

Applying the classifier, prior to fine-tuning, 2.4% of the generated compounds were labeled as multi-target compounds. During fine-tuning, the fraction of compounds classified as multi-target compounds steadily increased. After 200 epochs, 26.6% of the newly generated compounds were predicted to be multi-target compounds. Thus, compound classification also supported the ability of the fine-tuned model to preferentially generate multi-target compounds.

Conclusion

In this work, we have attempted to fine-tune a deep generative model originally trained on bioactive compounds for de novo design for recognizing and producing multi-target compounds. Therefore, high-confidence data sets of multi-, single-, and no-target (inactive) screening compounds were assembled considering positive and negative assay results. Using a subset of known multi-target compounds, the publicly available REINVENT model was fine-tuned using transfer learning, and its ability to re-generate known multi-, single-, and no-target compounds was evaluated on the basis of NLL analysis. Consistent with its derivation, the original REINVENT model was tailored towards the generation of single-target compounds, but also recognized no-target compounds. However, fine-tuning via unsupervised transfer learning systematically increased the likelihood of generating multi-target compounds, while decreasing the likelihood of producing single- or no-target compounds. During fine-tuning, the model regenerated known multi-target test compounds at increasing rates, in contrast to single- or no-target compounds. Moreover, the analysis of structural neighbors of training and test compounds, scaffold assessment, and compound classification studies further supported the ability of the fine-tuned model to particularly generate multi-target candidate compounds. Taken together, the results also provided evidence for the presence of structure-promiscuity

relationships that were detected, learned, and utilized by the model, consistent with earlier findings. Notably, corresponding structural patterns were captured by randomized SMILES of multi-target compounds used for the fine-tuning and recognized in an unsupervised manner. Taken together, our findings provide proof-of-concept for generative de novo multi-target compound design. As a part of our study, the data sets and custom code generated for our analysis have been made freely available [56].

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Schneider G, Fechner U (2005) Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov* 4:649–663. <https://doi.org/10.1038/nrd1799>
2. Chen H, Engkvist O, Wang Y et al (2018) The rise of deep learning in drug discovery. *Drug Discov Today* 23:1241–1250. <https://doi.org/10.1016/j.drudis.2018.01.039>
3. Chen H, Engkvist O (2019) Has drug design augmented by artificial intelligence become a reality? *Trends Pharmacol Sci* 40:806–809. <https://doi.org/10.1016/j.tips.2019.09.004>
4. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Model* 28:31–36. <https://doi.org/10.1021/ci00057a005>
5. Arús-Pous J, Blaschke T, Ulander S et al (2019) Exploring the GDB-13 chemical space using deep generative models. *J Cheminform* 11:20. <https://doi.org/10.1186/s13321-019-0341-z>
6. Olivecrona M, Blaschke T, Engkvist O, Chen H (2017) Molecular de-novo design through deep reinforcement learning. *J Cheminform* 9:48. <https://doi.org/10.1186/s13321-017-0235-x>
7. Segler MHS, Kogej T, Tyrchan C, Waller MP (2018) Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent Sci* 4:120–131. <https://doi.org/10.1021/acscentsci.7b00512>
8. Gupta A, Müller AT, Huisman BJH et al (2018) Generative recurrent networks for de novo drug design. *Mol Inform* 37:1700111. <https://doi.org/10.1002/minf.201700111>
9. Blaschke T, Engkvist O, Bajorath J, Chen H (2020) Memory-assisted reinforcement learning for diverse molecular de novo design. *J Cheminform* 12:1–17. <https://doi.org/10.1186/s13321-020-00473-0>
10. Kotsias P-C, Arús-Pous J, Chen H et al (2020) Direct steering of de novo molecular generation with descriptor conditional

- recurrent neural networks. *Nat Mach Intell* 2:254–265. <https://doi.org/10.1038/s42256-020-0174-5>
11. Yang X, Wang Y, Byrne R et al (2019) Concepts of artificial intelligence for computer-assisted drug discovery. *Chem Rev* 119:10520–10594. <https://doi.org/10.1021/acs.chemrev.8b00728>
 12. McGovern SL, Caselli E, Grigorieff N, Shoichet BK (2002) A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J Med Chem* 45:1712–1722. <https://doi.org/10.1021/jm010533y>
 13. Morphy R (2009) Selectively nonselective kinase inhibition: striking the right balance. *J Med Chem* 53:1413–1437. <https://doi.org/10.1021/jm901132v>
 14. Aldrich C, Bertozzi C, Georg GI et al (2017) The ecstasy and agony of assay interference compounds. *J Med Chem* 60:2165–2168. <https://doi.org/10.1021/acs.jmedchem.7b00229>
 15. Stork C, Chen Y, Sicho M, Kirchmair J (2019) Hit Dexter 2.0: machine-learning models for the prediction of frequent hitters. *J Chem Inf Model* 59:1030–1043. <https://doi.org/10.1021/acs.jcim.8b00677>
 16. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* 5:993–996. <https://doi.org/10.1038/nrd2199>
 17. Anighoro A, Bajorath J, Rastelli G (2014) Polypharmacology: challenges and opportunities in drug discovery. *J Med Chem* 57:7874–7887. <https://doi.org/10.1021/jm5006463>
 18. Bolognesi ML, Cavalli A (2016) Multitarget drug discovery and polypharmacology. *ChemMedChem* 11:1190–1192. <https://doi.org/10.1002/cmde.201600161>
 19. Mei Y, Yang B (2018) Rational application of drug promiscuity in medicinal chemistry. *Future Med Chem* 10:1835–1851. <https://doi.org/10.4155/fmc-2018-0018>
 20. Zimmermann GR, Lehár J, Keith CT (2007) Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discov Today* 12:34–42. <https://doi.org/10.1016/j.drudis.2006.11.008>
 21. Ramsay RR, Popovic-Nikolic MR, Nikolic K et al (2018) A perspective on multi-target drug discovery and design for complex diseases. *Clin Transl Med* 7:3. <https://doi.org/10.1186/s40169-017-0181-2>
 22. Lavecchia A, Cerchia C (2016) In silico methods to address polypharmacology: current status, applications and future perspectives. *Drug Discov Today* 21:288–298. <https://doi.org/10.1016/j.drudis.2015.12.007>
 23. Zhang W, Pei J, Lai L (2017) Computational multitarget drug design. *J Chem Inf Model* 57:403–412. <https://doi.org/10.1021/acs.jcim.6b00491>
 24. Lagunin A, Stepanchikova A, Filimonov D, Poroikov V (2000) PASS: prediction of activity spectra for biologically active substances. *Bioinformatics* 16:747–748. <https://doi.org/10.1093/bioinformatics/16.8.747>
 25. Jenkins JL, Bender A, Davies JW (2006) In silico target fishing: predicting biological targets from chemical structure. *Drug Discov Today Technol* 3:413–421. <https://doi.org/10.1016/j.ddtec.2006.12.008>
 26. Nidhi, Glick M, Davies JW, Jenkins JL (2006) Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J Chem Inf Model* 46:1124–1133. <https://doi.org/10.1021/ci060003g>
 27. Nigsch F, Bender A, Jenkins JL, Mitchell JBO (2008) Ligand-target prediction using Winnow and Naive Bayesian algorithms and the implications of overall performance statistics. *J Chem Inf Model* 48:2313–2325. <https://doi.org/10.1021/ci800079x>
 28. Liu X, Ouyang S, Yu B et al (2010) PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res* 38:W609–W614. <https://doi.org/10.1093/nar/gkq300>
 29. Koutsoukas A, Simms B, Kirchmair J et al (2011) From in silico target prediction to multi-target drug design: current databases, methods and applications. *J Proteom* 74:2554–2574. <https://doi.org/10.1016/j.jprot.2011.05.011>
 30. Nickel J, Gohlke B-O, Erehman J et al (2014) SuperPred: update on drug classification and target prediction. *Nucleic Acids Res* 42:W26–W31. <https://doi.org/10.1093/nar/gku477>
 31. Reker D, Rodrigues T, Schneider P, Schneider G (2014) Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc Natl Acad Sci USA* 111:4067–4072. <https://doi.org/10.1073/pnas.1320011111>
 32. Jasial S, Gilberg E, Blaschke T, Bajorath J (2018) Machine learning distinguishes with high accuracy between pan-assay interference compounds that are promiscuous or represent dark chemical matter. *J Med Chem* 61:10255–10264. <https://doi.org/10.1021/acs.jmedchem.8b01404>
 33. Blaschke T, Miljković F, Bajorath J (2019) Prediction of different classes of promiscuous and nonpromiscuous compounds using machine learning and nearest neighbor analysis. *ACS Omega* 4:6883–6890. <https://doi.org/10.1021/acsomega.9b00492>
 34. Feldmann C, Yonchev D, Stumpfe D, Bajorath J (2020) Systematic data analysis and diagnostic machine learning reveal differences between compounds with single- and multitarget Activity. *Mol Pharm* 17:4652–4666. <https://doi.org/10.1021/acs.molpharmaceut.0c00901>
 35. Feldmann C, Yonchev D, Bajorath J (2020) Analysis of biological screening compounds with single- or multi-target activity via diagnostic machine learning. *Biomolecules* 10:1605. <https://doi.org/10.3390/biom10121605>
 36. Korcsmáros T, Szalay MS, Böde C et al (2007) How to design multi-target drugs. *Expert Opin Drug Discov* 2:799–808. <https://doi.org/10.1517/17460441.2.6.799>
 37. Wei D, Jiang X, Zhou L et al (2008) Discovery of multitarget inhibitors by combining molecular docking with common pharmacophore matching. *J Med Chem* 51:7882–7888. <https://doi.org/10.1021/jm8010096>
 38. Shang E, Yuan Y, Chen X et al (2014) De novo design of multitarget ligands with an iterative fragment-growing strategy. *J Chem Inf Model* 54:1235–1241. <https://doi.org/10.1021/ci50021v>
 39. Hiremathad A, Keri RS, Esteves AR et al (2018) Novel Tacrine-Hydroxyphenylbenzimidazole hybrids as potential multitarget drug candidates for Alzheimer's disease. *Eur J Med Chem* 148:255–267. <https://doi.org/10.1016/j.ejmech.2018.02.023>
 40. Blaschke T, Arús-Pous J, Chen H et al (2020) REINVENT 2.0: an AI tool for de novo drug design. *J Chem Inf Model* 60:5918–5922. <https://doi.org/10.1021/acs.jcim/0c00915>
 41. Wang Y, Bryant SH, Cheng T et al (2017) PubChem BioAssay: 2017 update. *Nucleic Acids Res* 45:D955–D963. <https://doi.org/10.1093/nar/gkw1118>
 42. Gaulton A, Hersey A, Nowotka M et al (2017) The ChEMBL database in 2017. *Nucleic Acids Res* 45:D945–D954. <https://doi.org/10.1093/nar/gkw1074>
 43. Gilson MK, Liu T, Baitaluk M et al (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res* 44:D1045–D1053. <https://doi.org/10.1093/nar/gkv1072>
 44. Kavlock RJ, Austin CP, Tice RR (2009) Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal* 29:485–497. <https://doi.org/10.1111/j.1539-6924.2008.01168.x>
 45. Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from

- screening libraries and for their exclusion in bioassays. *J Med Chem* 53:2719–2740. <https://doi.org/10.1021/jm901137j>
46. Sterling T, Irwin JJ (2015) ZINC 15—ligand discovery for everyone. *J Chem Inf Model* 55:2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>
 47. RDKit (2020) RDKit: open-source cheminformatics and machine learning software. <https://www.rdkit.org/>. Accessed 1 Apr 2020
 48. Bruns RF, Watson IA (2012) Rules for identifying potentially reactive or promiscuous compounds. *J Med Chem* 55:9763–9772. <https://doi.org/10.1021/jm301008n>
 49. Irwin JJ, Duan D, Torosyan H et al (2015) An aggregation advisor for ligand discovery. *J Med Chem* 58:7076–7087. <https://doi.org/10.1021/acs.jmedchem.5b01105>
 50. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv:14126980. <http://arxiv.org/abs/1412.6980>
 51. Arús-Pous J, Johansson SV, Prykhodko O et al (2019) Randomized SMILES strings improve the quality of molecular generative models. *J Cheminf* 11:71. <https://doi.org/10.1186/s13321-019-0393-0>
 52. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. *J Chem Inf Model* 50:742–754. <https://doi.org/10.1021/ci100050t>
 53. Hu Y, Bajorath J (2017) Entering the ‘big data’ era in medicinal chemistry: molecular promiscuity analysis revisited. *Future Science OA* 3:F50179. <https://doi.org/10.4155/fsoa-2017-0001>
 54. Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. *J Med Chem* 39:2887–2893. <https://doi.org/10.1021/jm9602928>
 55. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, NY, USA, pp 785–794. <https://doi.org/10.1145/2939672.2939785>
 56. Blaschke T, Bajorath J (2021). Data set and source code for generative multi-target compound modeling. <https://github.com/tblaschke/reinvent-multi-target>. The composition of the deposition is detailed in a Data Note in Future Science OA (in press)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.