De novo design with deep generative models based on 3D similarity scoring

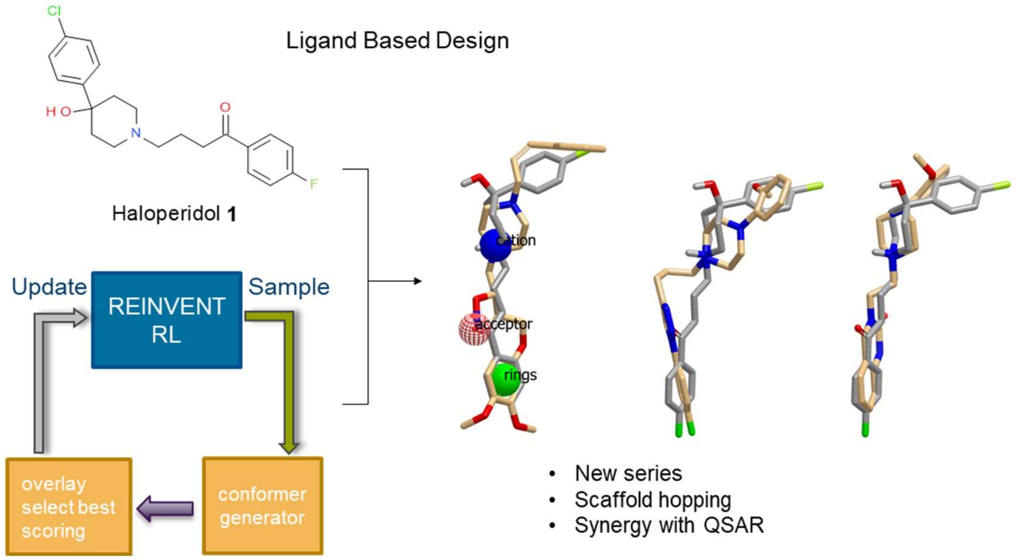
Kostas Papadopoulos[a,*], Kathryn A. Giblin[b], Jon Paul Janet[c], Atanas Patronov[a], and Ola Engkvist[a]

[a] Molecular AI, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden
[b] Medicinal Chemistry, Research and Early Development, Oncology R&D, AstraZeneca, Cambridge, UK
[c] Medicinal Chemistry, Research and Early Development, Cardiovascular, Renal and Metabolism (CVRM), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Abstract

We have demonstrated the utility of a 3D shape and pharmacophore similarity scoring component in molecular design with a deep generative model trained with reinforcement learning. Using Dopamine receptor type 2 (DRD2) as an example and its antagonist haloperidol **1** as a starting point in a ligand based design context, we have shown in a retrospective study that a 3D similarity enabled generative model can discover new leads in the absence of any other information. It can be efficiently used for scaffold hopping and generation of novel series. 3D similarity based models were compared against 2D QSAR based, indicating a significant degree of orthogonality of the generated outputs and with the former having a more diverse output. In addition, when the two scoring components are combined together for training of the generative model, it results in more efficient exploration of desirable chemical space compared to the individual components.

Graphical abstract



Keywords

Deep learning, Generative models, Reinforcement learning, DRD2, QSAR, 3D similarity, shape similarity

Abbreviations

1. Introduction

The generation of new promising lead compounds and their subsequent optimisation towards potential drug candidates is crucial for success in drug discovery. Virtual screening (VS) of large compound collections is one of the main computational methods to identify novel lead molecules [1]. While the number of compounds as well as the amount of computational resources available for VS make it possible to screen libraries with sizes in the magnitude of 10^{10} [2], they still fall far too short of the size of the available drug-like chemical space, estimated to be in the range of 10^{30} to 10^{60} [3], making any brute force approach infeasible. *De novo* generation of molecules provides an alternative solution where in principle there should be no restriction to the (implied) chemical space that is accessible. Deep generative models, building on the ground-breaking advances in deep learning [4], have demonstrated promising results in the past few years [5][6] including the successful case of delivery of novel bioactive compounds as discoidin domain receptor 1 (DDR1) inhibitors in 21 days from idea generation, to synthesis and biological testing [7].

Optimisation in medicinal chemistry is inherently multiobjective [8], [9] and one of the most successful ways so far for deep generative models to achieve multiobjective optimisation (MOO) has been through Reinforcement learning (RL) [10]. In this artificial intelligence (AI) approach, an agent is trained by acquiring rewards for different states and actions when interacting with its environment. In the context of molecular generation, 'states' can correspond to molecular structures either complete or not, and 'actions' to ways of building a structure depending on molecular representation. For example, possible actions to build a molecule would be adding a bond, or an atom, as in [11]. To build a molecule represented as a SMILES string, actions can be the addition of valid SMILES characters as in [12][13][14]. More abstract representations of molecules have also been used, for example in [7], as vectors in a latent space learned by a variational autoencoder (VAE) where actions were manipulations of these vectors in continuous space. In any case, the reward can be calculated by a scoring function where multiple scoring components that encode the desired optimisation objectives are combined, in a linear or non-linear way. It is also possible to differentiate the importance of the objectives by applying weights. See for example Eq. (1) in Section 2.5 for a specific formulation of a scoring function.

QSAR/QSPR predictive models are routinely used as scoring components to bias the generative output toward compounds of desired predicted activity or other property of interest. Multiple predictive models as scoring components can be used in cases of more complex objectives, for example to maximise selective target potency against one or more off-targets or when activity and ADMET properties such as solubility, metabolism, hERG inhibition [15] *etc*. need to be optimised together. Recent advances in property prediction [16] driven by progress in machine learning and deep neural networks (DNN) can lead to more efficient data-driven predictive models as scoring components.

However there are many issues with their use in the context of molecular generation. The fact that predictive models are data driven, means that they are dependent on the size and quality of available data. It is very common for drug discovery projects, especially in the early stages, that small amounts of data or noisy data (e.g. from HTS) are available. Even with larger training datasets, the resulting models fail to make low error predictions outside their applicability domain [17]. This is a general problem [18] not restricted to the prediction of molecular properties. It can be detrimental for molecular generation since the chemical space that generative models can explore is restricted by the applicability domain of the predictive model scoring component. This has been demonstrated in [19] where it was shown that molecule generation was biased towards the training set of the predictive model. A further complication that can affect the prediction error in QSAR/QSPR models is that activity / property spaces are not smooth but contain discontinuities also known as 'activity cliffs' [20] where a pair of highly similar molecules can have very different activity. Since QSAR models most commonly employ 2D descriptors or fingerprints representations of molecules, they fail to capture activity related molecular patterns in three dimensions. Another criticism of data driven models is that they generally provide low interpretability which although it does not directly affect the efficiency of molecule generation, it offers little justification and thus confidence on the validity of the generated ideas. Whilst QSAR models might have their place in exploitation scenarios akin to the lead optimisation problem, they do not provide a valid solution to the diverse exploration of chemical space in a new series generation scenario.

For these reasons, the use of physics-based [21] methods as scoring components could be investigated. In principle, these methods should not suffer from the issues of data availability and model applicability domain where, depending on the method, it can be possible to incorporate three-dimensional information and also provide higher interpretability of the results. Of course there is a good reason that they are not widely used and this has to do mainly with computational cost. Most methods require lengthy molecular dynamics (MD) simulations, or high cost quantum mechanics (QM) calculations in order to generate useful results. This makes them unsuitable for RL based optimisation by deep generative models which typically have to go through many iterations (*e.g.* 3000 in this work) of training. Still, the increasing availability of more computational power and the development of smarter algorithms opens up a space of physics-based methods that are computationally efficient to be used in RL based molecule generation. There are recent reports of using a docking scoring component such as in [11] where novel and diverse compounds were generated with predicted activity against two different targets, the domain receptor 1 kinase (DDR1) and the D4 dopamine receptor (DRD4).

In this work, we investigate the use of a 3D shape and pharmacophore similarity scoring component using ROCS [22]. We have chosen a ligand-based design case study and the dopamine receptor D2

(DRD2) as the biological target of interest. This target has been widely used in *de novo* generation case studies by us [12], [23]–[25] and other groups [26]–[28]. As the generative model, we used REINVENT 2.0 [29] which is publicly available as open access software [30]. 3D similarity together with 2D similarity scoring are routinely employed in VS and there is evidence of complementarity in the generated hit-lists either when the scoring is applied sequentially or in parallel [31]–[33]. We attempted to determine the degree of complementarity in RL based molecular generation between a 3D similarity component and a QSAR scoring component by evaluating them either together or as single components of the RL scoring function. In the same time we were interested to compare the two scoring modes against various generative model performance metrics under the hypothesis that the 3D similarity scoring component should result in a more diverse output of molecules compared to QSAR scoring, based on the distinction between physics-based and data-driven models.

Additionally we describe three use cases very close to medicinal chemistry optimisation practice and with retrospective evaluation of the generated molecules:

1. A ligand-based design case where only a single DRD2 active ligand is known (we use haloperidol **1**) but no further information is available about the bioactive conformation of **1** and of any additional molecules with DRD2 activity.
2. Based on the same case as above but with availability of DRD2 activity labelled data, which means that training a QSAR model is possible, we assess the synergistic effect of using both 3D similarity and QSAR scoring for training of the generative model as opposed to using only 3D similarity scoring.
3. We demonstrate the potential of a 3D similarity scoring component for scaffold hopping.

The use of a ROCS scoring component in RL based generative models has been reported before [34] with the authors suggesting that both QSAR and 3D-shape based similarity approaches produce significantly different design ensembles compared to 2D-similarity scoring components, however they do not explicitly compare 3D-shape against QSAR scoring nor do they provide further implementation details of both components. Molecular generation with a 3D shape and pharmacophore similarity optimisation objective has been reported in [35] but with a completely different architecture based on a 3D convolutional neural network (CNN) coupled with a shape variational autoencoder and without reinforcement learning (as in our case) or any other algorithm to enable multiobjective optimisation. In addition, molecules are generated by decoding a latent representation of a reference 3D shape in a data-driven fashion where in our case we obtain 3D shapes for generated molecules by applying a physics based method using a conformer generator with sampling of the conformational space and subsequently ranking these molecules by similarity to the reference shape.

Our study provides a comprehensive evaluation of a 3D similarity scoring component in the context of reinforcement learning with the RNN based generative model REINVENT in comparison to 2D QSAR scoring while highlighting its utility in ligand based design and notably as a tool for scaffold hopping.

2. Methods

2.1. Datasets

A set of DRD2 active compounds **D2ACTIVE** (*N*=4791) was extracted from ChEMBL 25 [36] as follows: Molecules with activity against the dopamine DRD2 receptor (CHEMBL217) were retrieved from a local copy of the ChEMBL database, filtered for standard types IC50, Ki and EC50, standard relation '=', grouped by 'Molecule CHEMBL ID' and aggregated by median 'pChEMBL value'. Only molecules with pChEMBL>=6.0 were selected with further filtration by molecular weight < 750 to afford the final set.

A decoy set of inactive compounds, **D2INACTIVE** was created after retrieving 2000 DRD2 inactive compounds from ExCAPE-DB [37] by random selection and with pXC50<5 since the activity threshold in ExCAPE and the DRD2 target is pXC50=5.

Generative models were pre-trained using two different datasets: i. the **STD** dataset obtained from the ChEMBL 25 dataset following filtration rules as in [12] with a size of *N*=1,435,546 and ii. To investigate a scenario where fewer task-relevant molecules where included in the prior, we formed the **AGN** dataset (*N*=1,431,348), obtained from **STD** by removing (a) all molecules (*N*=401) that contain substructure **2** derived from haloperidol **1** and (b) all (D2 active) molecules in common with **D2ACTIVE** (*N*=3857). We constructed the dataset **prior 100K** by random selection of 100,000 molecules from **STD** as a more computationally accessible representative subset of **STD**.

The dataset for modelling activity prediction **D2QSAR** (*N*=347,079) was obtained as the union of the **D2ACTIVE** set with a set of DRD2 inactive compounds that were retrieved from ExCAPE-DB [37] by random selection of 342288 inactive compounds (pXC50<5). Molecules were stripped from stereochemical information and duplicates were removed by considering only the highest activity. Molecular representations were created using RDKit [38] Morgan fingerprints of radius 3 as 2048-dimensional bit vectors. They were labelled from {0, 1} as 0=inactive, 1=active.

We constructed the validation set **D2TEST** from known DRD2 active (pChEMBL>=6, *N*=1164) derived from **STD** and inactive compounds (pXC50<5, *N*=237) from ExCAPE-DB. The molecules were scored i. with the QSAR scoring function (Section 2.2) and labelled 'higher': QSAR score>=0.8 or 'lower': QSAR score<0.5 and ii. with the ROCS scoring function (Section 2.4) and labelled as 'higher': ROCS score>=0.7 or 'lower': ROCS score<0.5

For the generation of the reference query for 3D similarity scoring we used a collection of known DRD2 agonists as their SMILES representations including the following: chlorprothixene, olanzapine, eticlopride, dopamine, apomorphine, nemonapride, risperidone, haloperidol and chlorpromazine. OpenEye's QUACPAC v.1.7.0.2 was used to assign tautomeric forms and protonation states at pH 7.4, followed by OMEGA v.3.0.1.2 to generate 3D conformations for the molecules in the set, using the default parameters in 'classic' mode with a maximum number of 200 conformers for each molecule resulting into 3024 conformers in total as the **D2ROCS** dataset.

Descriptions of all datasets used in the text can be found in Table S1 and Venn diagrams in Figure S1.

2.2. Conventions and notation

In this text we use 'ROCS', 'QSAR' in capitals to refer either to the scores obtained by the respective methods or to the scoring components as parts of a Reinforcement Learning (RL) training scoring function. We use 'rocs', 'qsar' in small letters to refer to the generative models trained with a ROCS or a QSAR scoring component respectively. We also use the notation 'rocs+qsar' for the generative model trained with both ROCS and QSAR scoring components. Generative models that have only been pre-trained with prior data (*e.g.* from STD or AGN) and have not been subjected to RL training (or transfer learning [29]) are referred to as priors or prior agents. STD and AGN are used interchangeably to refer either to the datasets themselves or to the priors resulting from pre-training with the respective dataset.

2.3. QSAR scoring component

The primary objective of the QSAR model in this work is to provide scoring feedback to the RL agent of REINVENT for the generation of new molecules which makes necessary to train the model on the maximum possible amount of data. In addition to that, the same QSAR model was used as an oracle in the use case described in Section 3.4.3. We employed the scikit-learn v.0.21.3 [39] implementation of the random forest classifier with the "out of bag" (oob) error functionality activated and with the parameter `class_weight='balanced'` which applies weights inversely proportional to the two class frequencies. The performance of the model was estimated by 10-fold stratified cross validation (0.95 Confidence Intervals in brackets): Accuracy: 0.992 [0.991 0.994], Matthews correlation coefficient: 0.70 [0.63 0.77] and ROC AUC: 0.83 [0.77 0.88]. The output of the scikit method `predict_proba()` which is an uncalibrated estimate of the probability for a given molecule to be active, was used as the corresponding QSAR component score during REINVENT training.

2.4. 3D similarity query

From **D2ROCS**, each of the generated conformers of haloperidol **1** was aligned and scored for 3D similarity against the rest of the molecules in the dataset with Tanimoto Combo scoring using OpenEye's ROCS v.3.2.2.2. The best haloperidol conformer was selected on the basis of lower strain energy as recorded by OMEGA, higher Tanimoto Combo scores overall against all other molecules in **D2ROCS** and visual inspection. This conformer was used to create a ROCS shape query by keeping the default shape feature and selecting three pharmacophoric (or else 'colour') features as shown in **Error! Reference source not found.Error! Reference source not found.**. We did not attempt any further refinement in the context of the receptor using any of the structural information available for the receptor and the bound ligand [40], as this work is meant to be purely a showcase of ligand based design. It's useful to note that 3D similarity scoring in VS has been reported [22][41] to show robustness using a conformation for the reference molecule not necessarily identical to the bioactive conformation.

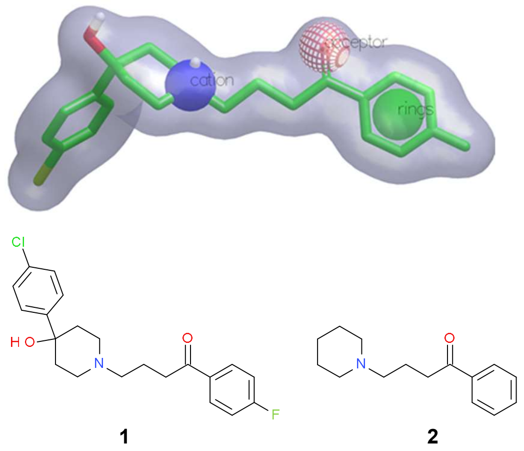


Fig. 1. (top) ROCS query showing colour features as spheres: green for ring systems, red for H-bond acceptors and blue for cations. (bottom) structures haloperidol **1** and substructure **2** used in analysis.

2.5. REINVENT

REINVENT 2.0 code [30] was adapted to implement the 3D similarity scoring functionality based on the OpenEye python toolkit v.2019.10.2 [42] which includes QUACPAC, OMEGA and ROCS. REINVENT cannot represent stereochemical information however *de novo* 3D structures can be obtained after stereoisomeric enumeration and overlay. The implementation with examples will be included in the upcoming release of the new version 3.0 of REINVENT. As shown in **Fig. 2**, SMILES strings sampled by the agent during reinforcement learning (RL) are first corrected for protonation state and tautomeric form with QUACPAC. Then 3D conformers are generated with OMEGA with stereoisomers enumeration enabled for a maximum of 3 stereocenters. Alignment to the reference query followed by overlay and similarity scoring with ROCS results in the selection of the best scoring 3D conformer (ranking by ComboScore). This pose is (optionally) saved while the ROCS score value is fed back to the RL agent. The ROCS score defined here as the output of the ROCS scoring component of the generative model scoring function, is calculated as the average of the shape and colour RefTversky similarity score values, obtained by the API functions `oeshape.GetRefTversky()` and `oeshape.GetRefColorTversky()` respectively.

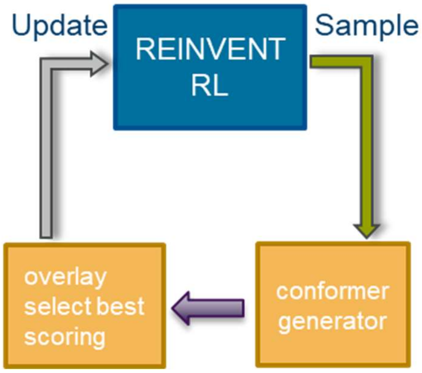


Fig. 2. REINVENT RL training flow diagram

To explore the effects of the scoring function composition and the prior, all combinations of the respective parameters shown in Table 1 were considered. Each run with activated diversity filter (DF) [23], [29] was repeated three times to evaluate the stochastic effect of the neural network training. Two additional runs without a DF were included: i. ROCS scoring and AGN prior ii. QSAR scoring and AGN prior. There were 20 REINVENT training runs in total. For each run the agent was trained for 3000 steps with a learning rate of 0.0001 a batch size of 128 and a product type [29] scoring function, where the score is calculated as a weighted geometric mean of the scoring components $c_1, \dots, c_n$ with weights $w_1, \dots, w_n$ and a custom alerts penalty CA:

$$total\_score = \mathrm{CA}\left(\prod_{i=1}^{n} c_i^{w_i}\right)^{\frac{1}{\sum_{i=1}^{n} w_i}} \tag{1}$$

The CA scoring component is penalising the presence of undesired substructures defined by SMARTS [43] strings, contained in the generated molecules, as described in [29]. We used CA to penalise the following functionalities: rings of size greater than 8, peroxides, sulphides, hydrazines, thioethers, acetals/aminals/hemiaminals and carbocations.

Throughout each RL training run, generated structures were collected with a *total_score* greater than 0.4 and checkpoints of the RL agent state were saved every 50 steps, resulting into 60 saved checkpoints in total. The combined results from sampling during RL training were collected in the dataset **SAMPLE_PRE** (*N*=5,724,859).

For each of the 20 runs (Table 1) and for each of the 60 checkpoints, the corresponding saved agent was sampled to a size of N_{all}=10000 SMILES strings, resulting in 1200 10,000-batches of SMILES strings that were combined into the dataset **SAMPLE_POST** (*N*=10,857,843). The structures in this set were evaluated for validity, uniqueness, diversity, novelty and frequency of matches with haloperidol derived substructure **2** (See following Section 2.6)

Table 1. Description of REINVENT runs

Scoring	Scoring function composition	Prior	DF[a]
qsar	D2 QSAR model / MW<550; 3:1 weighting	STD, AGN	on/off[b]
rocs	ROCS RefTversky similarity, 1:1 colour-shape / MW <550; 3:1 weighting	STD, AGN	on/off[b]
rocs+qsar	ROCS RefTversky similarity, 1:1 colour-shape / DRD2 QSAR model / MW<550; 3:3:1 weighting	STD, AGN	on

[a] DF=Diversity Filter.
[b] DF inactive only for the run with AGN prior

2.6. Evaluation

To evaluate the generative models we used the following metrics:

Validity as the ratio N_{valid}/N_{all} where N_{valid} is the number of sequences that are successfully parsed by RDKit to yield valid SMILES representations of molecules from a total of N_{all} generated sequences.

Uniqueness, as the ratio N_{uniq}/N_{valid} where N_{uniq} is the number of unique compounds in a list of N_{valid} valid molecules. We calculate N_{uniq} from the canonical representation of the molecules as SMILES strings using the RDKit function `Chem.MolToSmiles()`

As a measure of chemical diversity, for a set of N_{uniq} molecules, we follow the definition by Li *et al*. [27] With the modifications of using 2048-bit Morgan fingerprints of radius 3 as vector representations $\{\mathbf{x_i}\}_{i=1}^{N}$ and an equivalent formula to their unbiased estimator, *internal diversity* is calculated as:

$$D = 1 - \binom{N_{uniq}}{2} \sum_{1 \le i < j \le N_{uniq}} k_{Tan}\left(\mathbf{x_i}, \mathbf{x_j}\right) \tag{2}$$

where k_{Tan} is the Tanimoto similarity function.

For a set of N_{uniq} generated molecules we calculate the ratio of matches to the corresponding prior dataset (STD or AGN) as a measure of *novelty* in a similar way to the metric used in the GuacaMol benchmark.[44]

Similarity to a nearest neighbour SNN as defined in MOSES [45] but calculating Tanimoto similarity k_{Tan} using 2048-bit Morgan fingerprints of radius 3 instead. More formally, for a set $\mathcal{M}$ of molecules and a reference set $\mathcal{S}$ we calculate:

$$SNN = \frac{1}{|\mathcal{M}|} \sum_{x \in M} \max_{y \in \mathcal{S}} k_{Tan}(x, y) \tag{3}$$

Additionally, we calculate the ratio of the generated molecules that match the haloperidol substructure **2**. By construction, the AGN prior agent has not been exposed to any molecules containing this substructure during pre-training, so this metric can provide an estimate of *generalisation*, the ability of the RL agent to learn beyond pre-training data.

Synthetic accessibility is one of the most important properties for molecular generative models [46] but in the same time the most challenging to estimate especially in the case of large collections of molecules. In this work we use the retrosynthetic accessibility score (RAscore) introduced by Thakkar *et al*. [47] and its "SA score" implementation as part of REINVENT 2.0. [30] This is essentially a binary classification predictive model and the score is the estimated probability of finding a synthetic route for a given compound.

Learning by the generative model can be assessed by how likely the model is to generate high scoring molecules from an external validation set and also how unlikely the generation of lower scoring molecules is. In general, the probability of an agent to generate a given molecule can be estimated by sampling sufficiently large samples to minimize the error of estimation of this probability statistic. In the case of REINVENT the probability can be directly accessed from the output of the RNN for a SMILES sequence $S = s_1 s_2 \dots s_T$ as:

$$P(S) = \prod_{t=1}^{T} P_{RNN}\left(s_{t+1} | s_t, s_{t-1}, \dots, s_1\right) \quad (4)$$

For each molecule m_i all of its R_i possible SMILES representations S_i need to be considered. For this purpose distinct SMILES strings representations were generated non exhaustively following a variation of the procedure described by J. Arús-Pous *et al.* [48]. In the case of the **D2TEST** 649,434 SMILES representations were obtained in total. It follows that $P(m_i) = \sum_{j=1}^{R_i} P(S_{i_j})$ and thus for a validation set of molecules $M = \{m_i\}_{i=1}^{N}$, the probability of the agent generating at least one molecule from the set is $P(M) = \sum_{i=1}^{N} P(m_i)$. After normalising for set size, we calculate:

$$\bar{P}(M) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{R_i} P\left(S_{i_j}\right) \quad (5)$$

For clustering of molecular datasets, the Bemis-Murcko (BM) scaffold [49] is calculated for each molecule in the set using RDKit, resulting in clusters with common BM scaffold. This approach is preferred for computational efficiency with large datasets and for being reasonably chemically meaningful (See also the discussion in Section 3.4).

3. Results and Discussion

3.1. Efficiency of learning by the model

We assessed the ability of the generative models to optimise the objectives encoded in their scoring function throughout training. The dataset **SAMPLE_PRE** was used which is obtained by sampling during training by reinforcement learning. For all scoring modes (rocs, qsar and rocs+qsar) and with the diversity filter activated, there is a steady increase in *total_score*, faster in the first 500 steps and slower in later steps with no significant effect from the prior (**Fig. 3**A). The rocs scoring mode model converges to a lower *total_score*, which is expected given the dynamic range of the ROCS scoring function (See also **Fig. 5**A for the distribution of ROCS scores for the set of DRD2 actives). Removing the DF results in higher variance, especially for the qsar scoring component, although the maximum score is higher possibly due to favouring exploitation by not penalising excessive generation of molecules with the same Murcko scaffold (Figure S2). To look into the individual scoring components, we also obtained *post hoc* ROCS scores for the qsar based model generated molecules and QSAR scores for the molecules generated by the rocs based model. Fig. 3B shows that the rocs based models generate compounds with low QSAR score and thus predicted to be inactive whereas the qsar and qsar+rocs based model efficiently learn to generate predicted active compounds. This behaviour shows that the rocs models explore a different chemical space compared to qsar based models. The low predicted activity of the rocs derived molecules for the agent trained on QSAR could be attributed to falling outside the applicability domain of the QSAR predictive model, although further evidence would be required to support this. Interestingly, the mixed rocs+qsar models perform well in both tasks, optimizing for ROCS score (Fig. 3C) and QSAR score, supporting the complementarity argument.

The generative models were also scored for synthetic accessibility using the SA score (Section 2.6). It needs to be emphasised that the SA score was not an optimisation objective and consequently it was

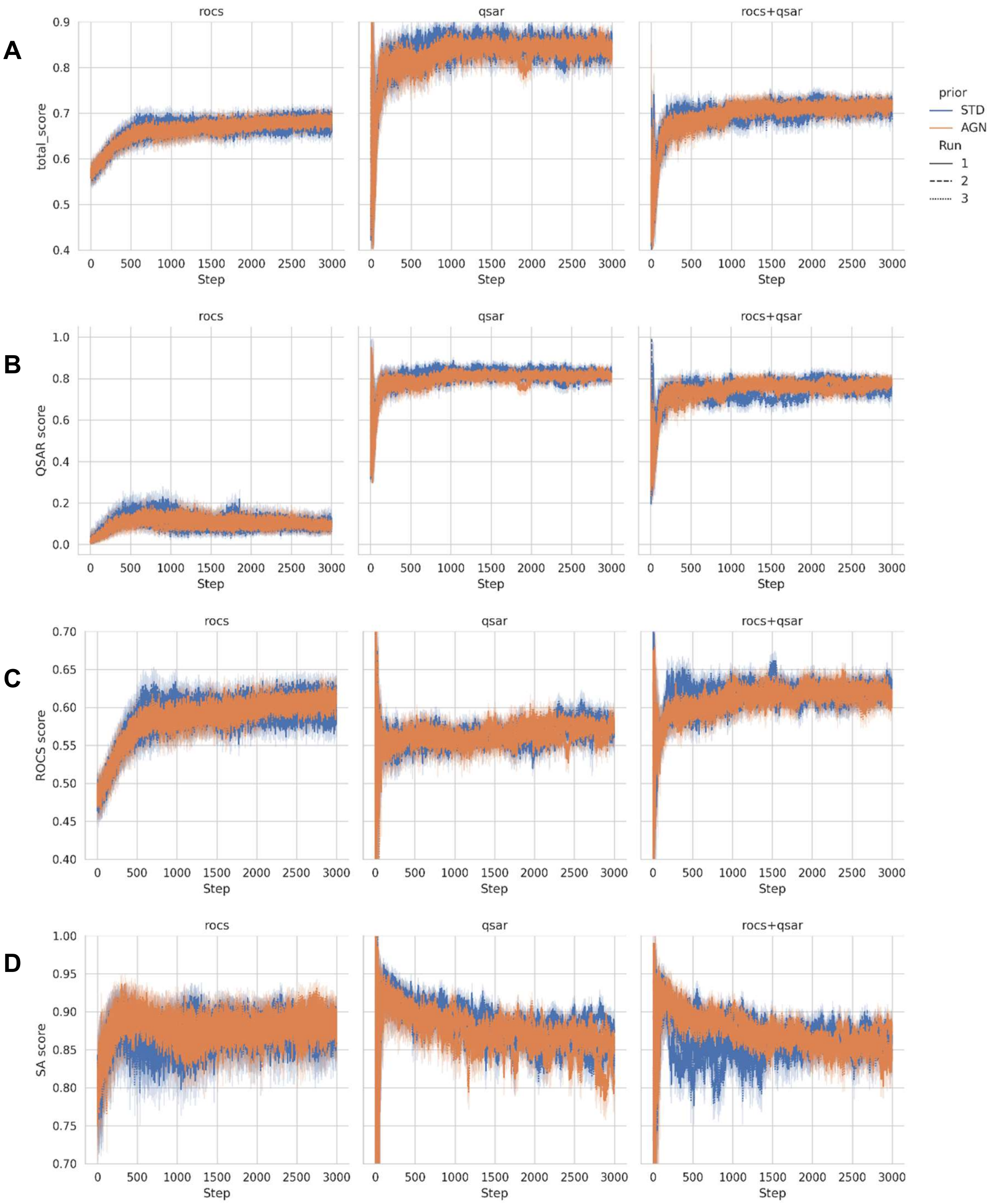


Fig. 3. Scoring of the output from generative models during training: **A**. total_score, **B**. QSAR score (probability of being active), **C**. ROCS score and **D**. SA score (synthetic accessibility, see Section 2.6). Score values refer to molecules that were collected during training. SA score, ROCS score for qsar based models and QSAR score for rocs based models were calculated *post hoc* and were not part of the respective scoring function for RL training. The diversity filter was activated for all runs described here. Individual line plots depict mean and 95% CI over the values calculated for generated molecules in each step.

not a scoring component for any of the generative models but was calculated *post hoc*. Despite this, the models converged to molecules that are estimated to be synthesizable with an aggregate probability ~0.85 (Fig. 3D). However the qsar based models show a steady drop of SA until the end compared to the rocs based models indicating that the former generate increasingly more complex structures to optimise the QSAR score but within the constraint of its applicability domain whereas in the latter models, ROCS scoring is more permissive allowing the rocs models to explore and optimise without increased complexity cost. The mixed rocs+qsar based models show a behaviour very close to qsar based models.

It is also useful to examine the distributions of the ROCS and QSAR score values for the generated molecules. It needs to be stressed that these distributions as shown in **Fig. 5**A describe molecules generated throughout the training run and so include molecules generated in earlier steps which are potentially suboptimal. As references for comparison, the **D2ACTIVE**, **D2INACTIVE** and **prior 100K** datasets were used (Section 2.1) and the ROCS and QSAR scores were calculated. In the case of ROCS score values, (**Fig. 5**A) the rocs based generative model shows clear enrichment compared to **prior 100K** and with a bimodal distribution whose lower mode aligns with the mode in **prior 100K** and includes early-step suboptimal structures closet to the (pre-)training set. It is worth noting that the **D2ACTIVE** set contains many DRD2 active molecules that do not match the ROCS reference query (**Error! Reference source not found.**A) explaining the higher spread of the distribution towards lower values. The qsar based model also shows enrichment in higher ROCS score molecules even though it did not include ROCS score as an optimisation objective. However the rocs-based models do not follow the same pattern (**Fig. 5**B) with the majority of generated molecules populating the low end of the QSAR score scale. We hypothesise that due to the applicability domain (AD) of the QSAR model being a fraction of the one for ROCS scoring, it is expected that a large amount of ROCS optimised molecules will fall outside the AD of the QSAR model. These molecules would receive an even lower QSAR score because of using an unbalanced training set skewed towards inactives and/or because of uncalibrated

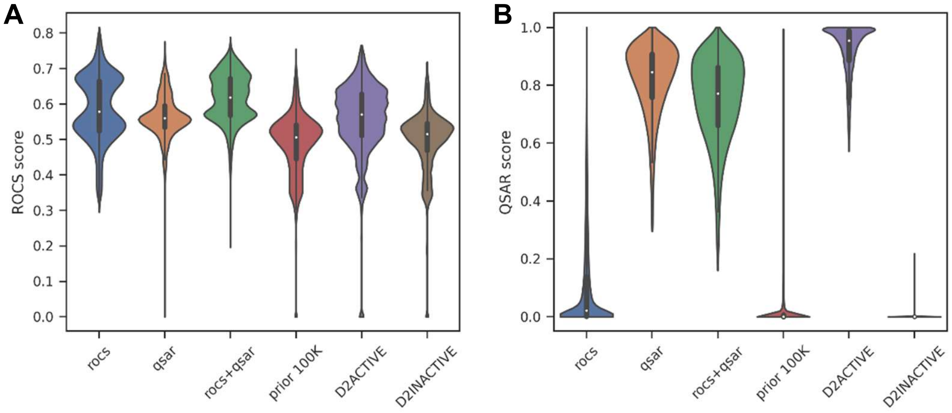


Fig. 4. Distributions of ROCS (A) and QSAR (B) scores for generated molecules compared to reference dataset molecules

probabilities produced by the training algorithm. However this was not investigated any further in this study.

Finally, one of the main observations from examining the score distributions in Fig. 4 is that the combined rocs+qsar based model shows top performance for both objectives (ROCS or QSAR score), exploring an optimal chemical subspace which is not identical to the subspaces of each one of the single components, supporting the hypotheses of: a) the ROCS scoring component working efficiently in the context of multiobjective optimisation with REINVENT and b) orthogonality with QSAR scoring in the generated output.

3.2. Evaluation of performance of the generative models

We evaluated a number of performance metrics that were calculated on the **SAMPLE_POST** dataset which contains multiple 10K batches sampled *post hoc* from saved checkpoints (60) of the models every 50 steps (Fig. 5). Validity (A) is high for all models, with higher values for those with no DF (A2). However the performance of no DF models collapses when evaluated for uniqueness (B2) and internal diversity (C2) with a steep drop in the early stages of training, presumably when the model discovers a high scoring solution and then generates identical or very similar molecules, a state also known as mode collapse. The drop is slightly less sharp but still significant for the rocs based models showing the rocs scoring component to be more resilient to mode collapse, most likely due to its intrinsic ability for exploration. DF activation successfully circumvents the problem in agreement with the results from [23]. The rocs based models generate significantly more diverse sets of molecules. Diversity drops for all models after the earlier stage of training around step 500 which reflects the transition from a generic untrained to a specialised trained model. The same ranking of models but with less significant differences, at least after the earlier stage of training, can be seen for uniqueness. Novelty shows a steady increase from the early untrained prior-like state towards the trained late stage for all models with activated DF (D1). Removing the DF results in lower novelty with a decreasing trend on further training (D2).

We have measured the frequency of generating molecules that contain substructure **2** (Fig. 5E). As mentioned before, compounds with this substructure do not appear in the AGN prior dataset. Considering that **2** is part of the ROCS reference query, molecules that contain **2** are very likely to score highly with the ROCS and the QSAR scoring functions and thus this frequency metric can help, in some part at least, to evaluate the ability of a generative model to generalise and produce output beyond the data (prior) used for pre-training. Indeed the rocs based model with AGN prior seems to learn the substructure later than the STD pre-trained rocs model with the latter achieving significantly higher frequency values earlier but which tend to decrease in later stages possibly because of engagement of

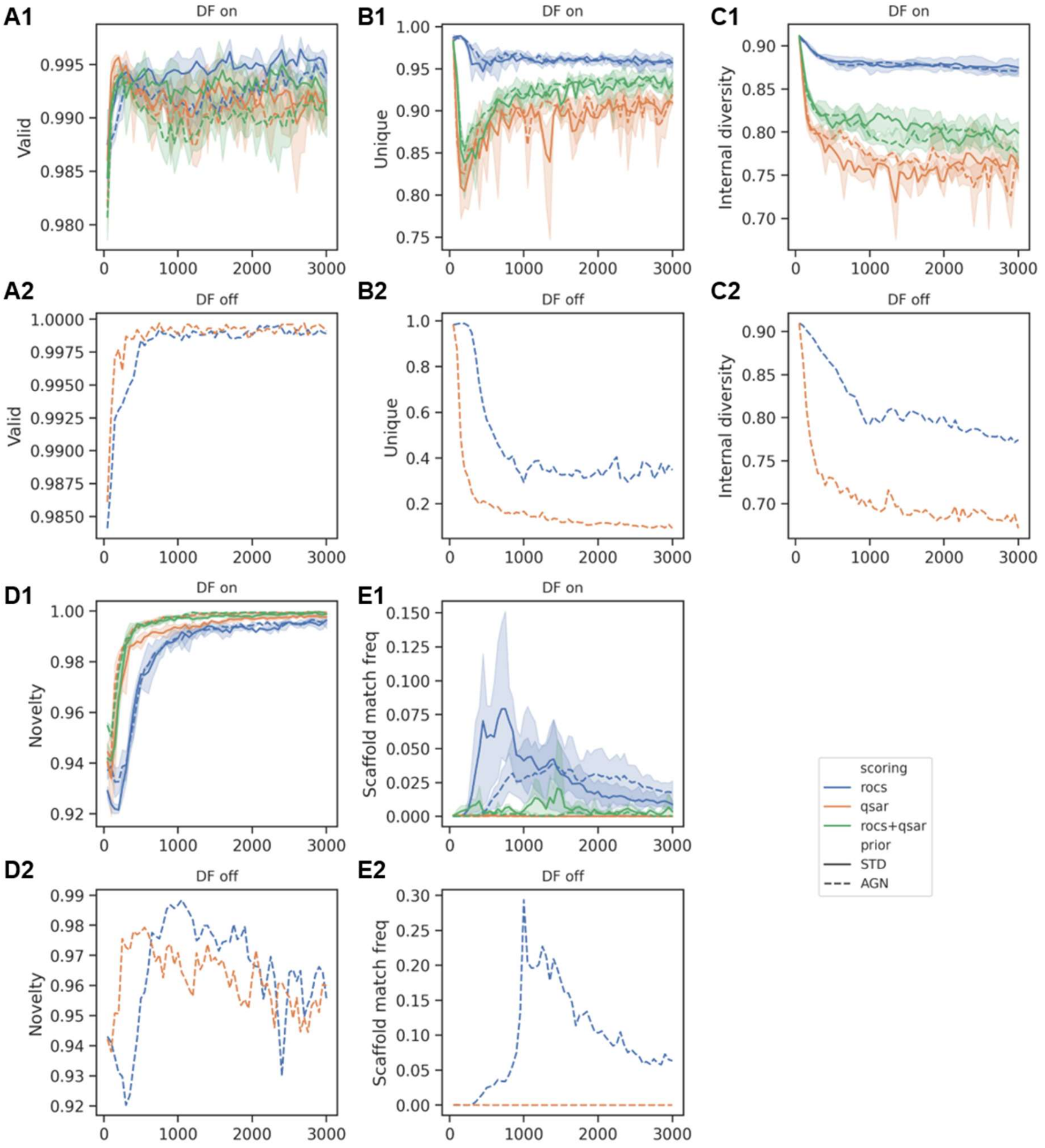


Fig. 5. Performance metrics with progression during training, including (A) validity, (B) uniqueness, (C) internal diversity, (D) novelty and (E) frequency of matching substructure **2**. Both cases of models with activated DF (A1-E1) and no DF (A2-E2) are shown. Metrics evaluated on molecules from the **SAMPLE_POST** dataset. Mean values and CI at 0.95 over 3 re-runs are displayed for the models with activated DF only. DF=Diversity Filter.

the Diversity Filter. Remarkably, qsar based models seem to fail to learn substructure **2** although all DRD2 active molecules containing **2** were part of the training set of the QSAR model.

In relation to internal diversity, uniqueness and novelty, the combined rocs+qsar based models show performance which lies in between the individual rocs and qsar based ones however they appear to lean more towards a qsar-like behaviour. This bias could be the result of the QSAR component generating molecules that score high with both QSAR and ROCS scoring functions whereas rocs generated molecules tend to score very low with the QSAR model, even when they achieve high ROCS scores (see Fig. 4 and the score distributions of the individual components). In effect, the agent for the rocs+qsar based model learns that it can maximize its reward by prioritising the QSAR component over the ROCS one. The QSAR component is more likely to lead the agent to a chemical space with higher scores for both objectives.

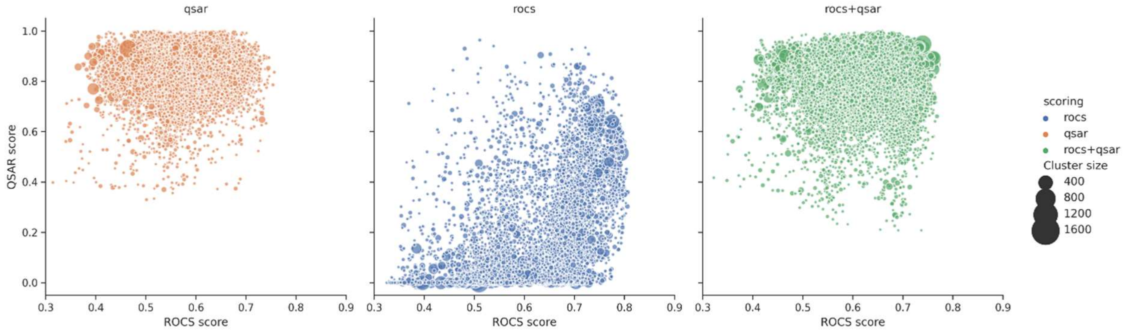


Fig. 6. Scatter plot of QSAR against ROCS scores. Marker sizes correspond linearly to cluster size. Only clusters with size greater than 10 are shown. All models were pretrained with the STD prior. Results after combining 3 re-runs for each model

Molecules in the **SAMPLE_PRE** dataset were clustered by their Bemis-Murcko (BM) scaffold as described in Section 2.6. For each cluster, ROCS and QSAR scores are aggregated by the respective median values. Fig. 6 shows the rocs+qsar based model to perform best producing clusters in the upper right 'sweet spot' area of higher ROCS and QSAR scores. Fig. 7A shows that the rocs based models generate a significantly higher number of BM scaffolds confirming the higher chemical diversity for this

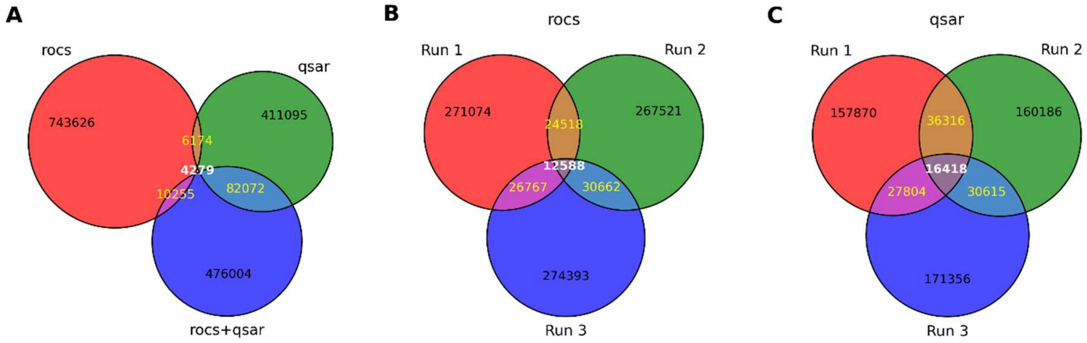


Fig. 7. Number of Bemis-Murcko (BM) scaffolds for all models (A) and: for rocs (B) and qsar (C) over 3 re-runs. All runs with STD prior and DF on. Venn circles represent sets and thus unique BM scaffolds. All areas correspond within some approximation error to set size.

method and with minimal overlap with either qsar or rocs+qsar generated clusters further supporting complementarity between rocs and qsar models.

To assess the effect of stochasticity in training of the generative models we compared their output after repeating 3 times each training run. Stochasticity can appear for example from random initialization of the weights of the generative deep neural network. Fig. 7B and C show small overlap between the 3 runs for the rocs based models which is significantly lower compared to the qsar based models after normalising for the number of BM scaffolds for each run. A practical implication of this observation is that whenever access is required to a generative output covering a larger volume of chemical space, then one can simply try to multiple re-runs of training the model (we only tried 3 re-runs)

3.3. Comparison with test set

We calculated the metric $\bar{P}$ for the molecules in the reference test **D2TEST** using Eq. (5) from Section 2.6. The $\bar{P}$ metric can be thought as the average probability for a given NN model to generate a molecule of a reference set. The calculation is based on the probability of formation of a molecule as obtained by the RNN output of REINVENT and not by statistical estimation *e.g.* by sampling. We used the **D2TEST** dataset and the 'higher', 'lower' ranking labels based on ROCS or QSAR scores (Section

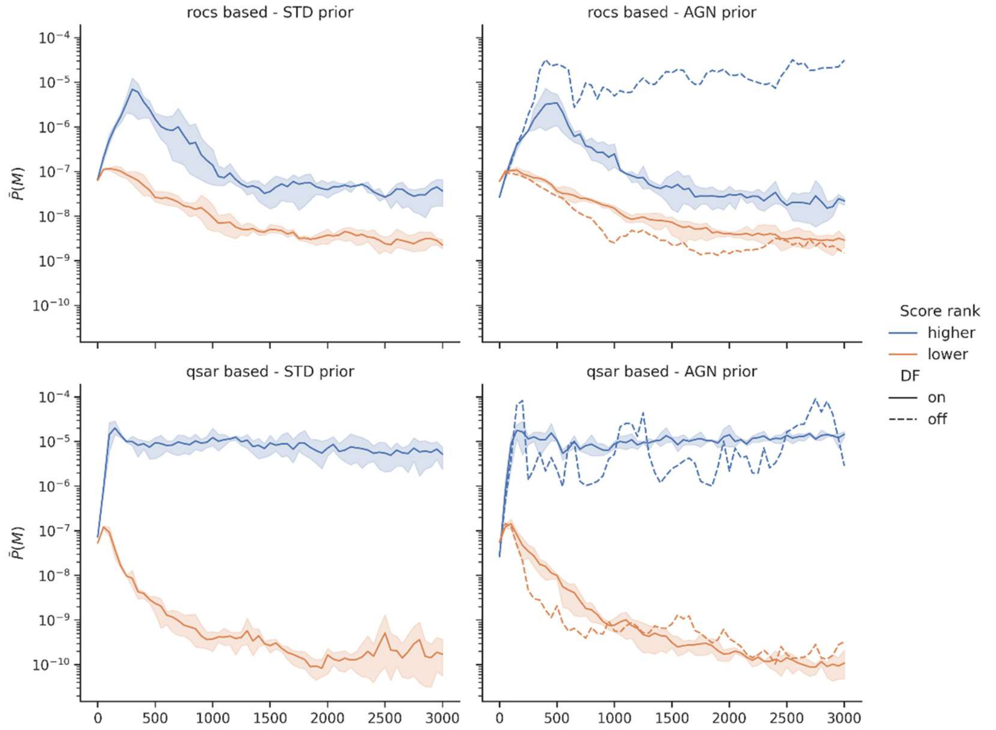


Fig. 8. Evaluation of the $\overline{\mathrm{P}}$ metric for the rocs (Top) and qsar (Bottom) based models. The ranking labels are set as 'higher': ROCS score>=0.7 or QSAR score>=0.8 and 'lower': ROCS score<0.5 or QSAR score<0.5. CI at 0.95 over 3 re-runs are displayed

2.1). We did not use their already known DRD2 activity labels since the optimisation objective is to generate high scoring molecules which are not necessarily DRD2 active. Fig. 8-Top shows that the rocs based model learns to generate 'higher' labelled (ROCS score>=0.7) molecules from the reference set at least up to around training Step 500. The subsequent drop is most likely due to engagement of the DF as it is not observed in the case of the rocs model with no DF. $\overline{P}$ is strictly decreasing for the 'lower' labelled (ROCS score<0.5) molecules showing that the model learns to avoid them. $\overline{P}$ at Step 0, ($\overline{P_0}{\sim}10^{-7}$) corresponds to the unoptimized model that has only been pre-trained with the prior (AGN or STD). It can be considered as a baseline value with values $\overline{P} > \overline{P_0}$ indicating learning of optimal molecules and $\overline{P} < \overline{P_0}$ indicating "un-learning" of suboptimal molecules. In the case of the qsar based models (Fig. 8-Bottom) a similar behaviour is observed with rocs that shows selective learning but this time with the gap between the $\overline{P}$ values of 'higher' (QSAR score>=0.7) and 'lower' (QSAR score<0.5) labels even more pronounced and additionally with no significant differentiation upon DF application.

3.4. Relevance of generated structures

3.4.1 Visual inspection

Despite lacking in objectivity, visual inspection is a crucial step in virtual screening campaigns [32] and equally important in assessing the utility of molecules produced by generative models. We examine the output from the rocs and the combined rocs+qsar models after selecting the highest scoring molecules with *total_score*>0.8 then clustering by common BM scaffold and finally removing clusters with size less than 10. For each cluster $\mathcal{M}$ we calculated the similarity metric SNN from Eq. (3) using **D2ACTIVES** as the reference set $\mathcal{S}$. Fig. 9 shows representative examples of high scoring and high SNN similarity clusters. In many cases such as for **4**, **6**, **7**, **8**, **9** and **10** the generative model reproduces a known active molecule from **D2ACTIVES** otherwise it generates molecules very similar to their NN, a known DRD2 active compound.

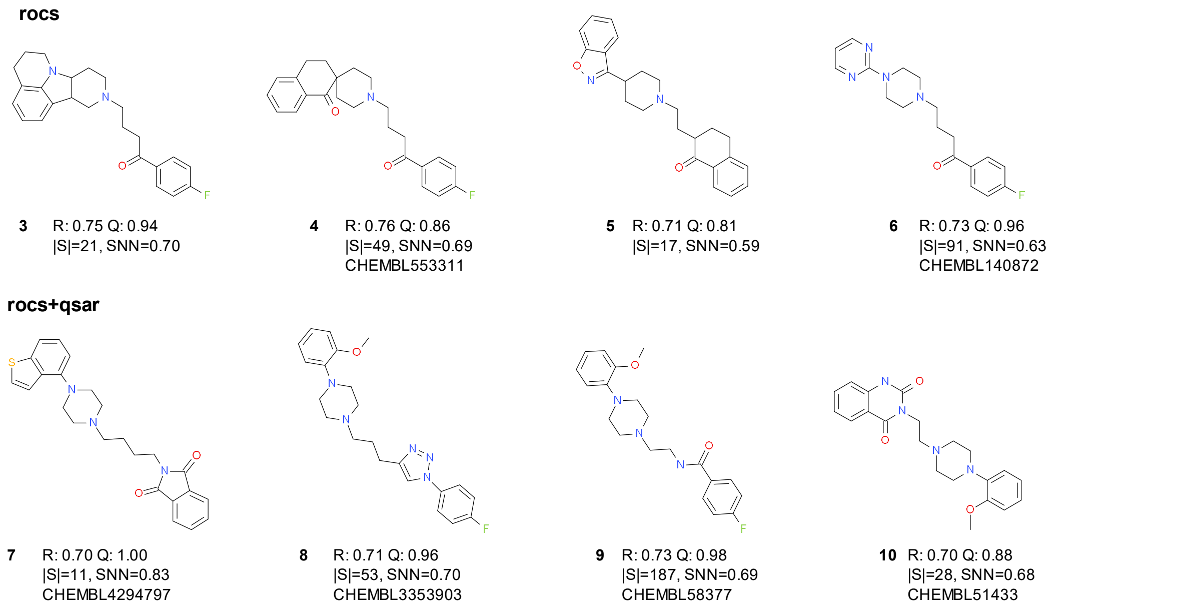


Fig. 9. Representatives of clusters with highest mean similarity to **D2ACTIVES** for rocs and rocs+qsar based models. QSAR scores for the rocs based model were calculated after generation. We used combined output from models with AGN or STD prior, DF on, 3 runs each. R: ROCS score and Q: QSAR score for the representative molecule; |S|: cluster size; SNN: mean NN similarity over the molecules in the cluster to their NNs in the **D2ACTIVES** set. CHEMBL IDs for identical molecules (ignoring stereochemistry) in ChEMBL 25

While it is reassuring that the model can produce identical or close analogues to known actives, the expectation for a generative model is to create solutions that span the largest possible volume of chemical space. In practice, this translates to novel chemical series with new features to differentiate from already known chemical space. For example, bioisosteric replacement and scaffold hopping [50], [51] are used for lead optimisation or to access back-up or a new lead series whereas fast follower approaches aim to escape public or patent protected chemical space [52], [53]. For that reason we examine the high scoring and high populated BM clusters but this time with the lowest SNN similarity values. We consider BM clustering as an approximation to the chemical series definition by medicinal chemists and the SNN value as a metric of novelty of the BM derived series. Fig. 10 shows hand-picked BM clusters with their respective best scoring representative structures. Both the rocs and the rocs+qsar based models and within the restriction of the 3D query (Fig. 1) seem to be capable of generating complex or unusual but still reasonable substructure motifs. For example all rocs generated molecules in Fig. 10-Top except for **13** introduce a ring containing substructure not present in the **D2ACTIVES** set. More interestingly, substructures shown in bold for **12**, **14**, **16** and **18** from rocs and **20** and **25** from rocs+qsar are not present in any molecules in the STD prior dataset and thus were not seen by the generative models during initial pre-training.

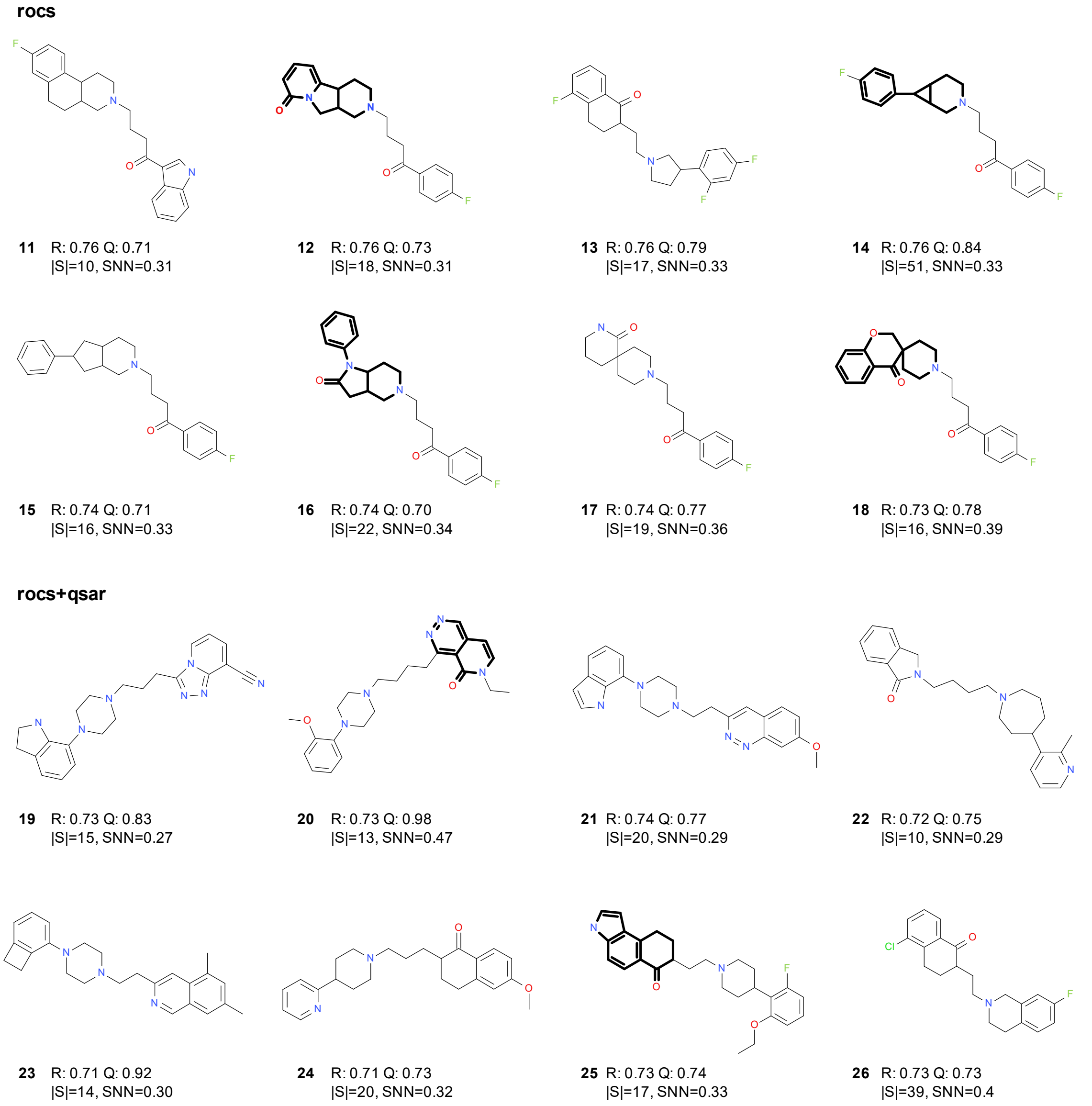


Fig. 10. As in Fig. 9 except for showing cluster representatives with lowest average similarity. Bold emphasis indicates substructures not present in the STD prior.

3.4.2 Scaffold hopping

We demonstrate the utility of the rocs and rocs+qsar based generative models in scaffold hopping by showing that these models are capable of: 1) Generating scaffolds to replace parts of the initial query molecule **1** resulting into new molecules with high 3D overlay with **1**. Compounds **12**, **14**, **16** and **18** (rocs) and **20** and **25** (rocs+qsar) contain novel scaffolds, not present in the data used for pre-training of the generative models. 2) Generating scaffold replacements that result in molecules identical to known DRD2 actives. We show in Fig. 9 the examples of generated molecules **4** and **6** (rocs) and **8** and **10** (rocs+qsar) which are identical to known DRD2 actives in ChEMBL 25 (CHEMBL IDs are also shown). We confirmed that all of them were generated by models pre-trained with the AGN prior. We remind that the AGN prior does not include any known DRD2 active molecules and additionally does not include any molecules containing substructure **2**. Further evidence that the models learn to generate those four molecules as opposed to memorizing them during pre-training can be seen in Table 2 that shows significant increase in the probability of formation $P(m_i)$ (Section 2.6) for each one of them, between the states of the corresponding generative models before and after training. The overlays with the reference molecule **1** as were generated by the ROCS scoring component during the RL training stage are shown in Fig. 11. Example **4** indicates ability of the agent to learn to generalise beyond pre-training data by producing substructure **2** not present in the prior. Examples **4** and **6** show replacements of the 4-phenyl-piperidine-4-ol scaffold in **1**. Structures **8** and **10** show alternative H-bond acceptors as part of new 1,2,3-triazole or a quinazolodione scaffolds respectively and adjustment of the length of the alkyl chain linker to achieve the right geometry.

Table 2. Probability of formation before and after training for **4**, **6** (rocs based model) and **8**, **10** (rocs+qsar based model)

Compound	Probability ($\times 10^{-9}$)	
	Prior (AGN)	Trained agent
4	0.41	37518
6	161.2	806435
8	3.15	41855
10	931	53954

While the RNN in REINVENT encodes SMILES string representations of molecules, it excludes tokenization of stereochemistry specific SMILES characters such as '@', '/' and '\' therefore generated molecules lack any stereochemistry. However during model scoring by the ROCS component, the SMILES strings sampled from the RNN acquire stereochemical information through the steps of stereo-

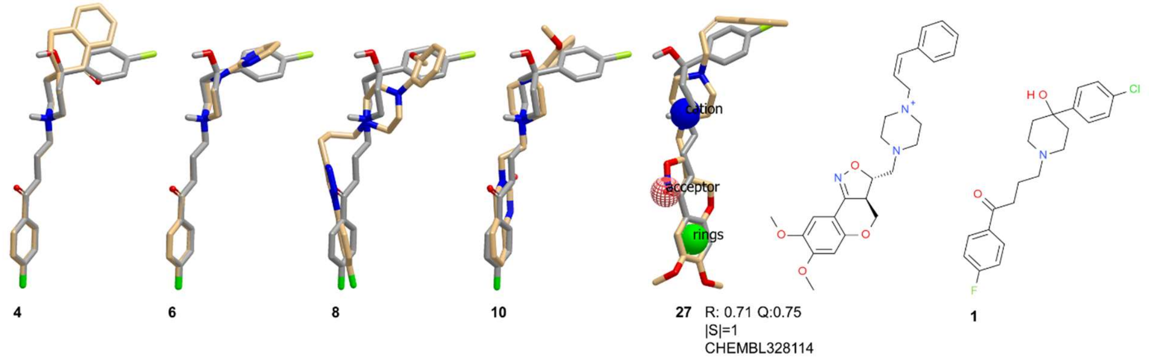


Fig. 11. Overlays with ROCS query reference **1** (grey). **4**, **6**, **8** and **10** are examples of scaffold hopping reproducing known D2 active molecules not used for pre-training of the generative model. **27** was generated with stereochemistry and charge information as shown.

enumeration by OMEGA and then selection of the best overlay by ROCS as described in Section 2.5. Fig. 11 shows **27** as an example of a generated molecule with the optimum stereochemistry to match the query. Furthermore, **27** is an additional example of a known DRD2 active molecule (CHEMBL3281114) that was generated by the rocs based model pre-trained with AGN and thus serving as a showcase of a new active chemical series derived from **1**.

3.4.3 Ligand based design case 1

We consider a use case of optimisation where only a single active ligand is known that can be used as a starting point to generate novel and diverse active molecules. More specifically, our starting point is haloperidol **1** and under a purely ligand-based design scenario there is no further information about the biological target or the bioactive conformation of **1**. This would eliminate the possibility of training a QSAR model in the absence of data while the use of 2D similarity scoring would only yield molecules with inadequate differentiation from **1**. We assess the performance of the 3D ROCS similarity scoring component in this scenario using the output of the rocs based generative model from the **SAMPLE_PRE** dataset. We evaluate the generated molecules utilising our QSAR model which in this case is serving only as an oracle considering as a 'hit' any molecule with QSAR score greater or equal to 0.7. Following the discussion so far, this is a less optimal choice of an oracle since its scoring ability is diminished outside its applicability domain and possibly further affected, as noted in Section 3.1, by the imbalance in the training set and the use of uncalibrated probabilities. While this would mean that the hit rate% values in Table 3 are underestimated, their relative values and thus the enrichment ratio should be sufficiently reliable for our purpose of showing enrichment. We observe *predicted actives* enrichment compared to the **prior 100K** data for rocs based generative models pre-trained with either of the STD or AGN prior datasets.

Table 3. Performance of rocs based deep generative models pre-trained with STD and AGN priors using a QSAR predictive model as an oracle. Only considered results from models with activated DF. Hits=molecules with QSAR score >= 0.7. Enrichment is calculated as the ratio $\frac{\text{Hit rate\%}_{\text{generative model}}}{\text{Hit rate\%}_{\text{Prior 100K}}}$

	rocs STD	rocs AGN	**Prior 100K**
Hit rate %	1.12	0.99	0.25
Enrichment	4.50	3.99	1

3.4.4 Ligand based design case 2

In the next use case of practical value, we investigate the advantage of using a scoring function that combines both a ROCS and a QSAR scoring component. We consider a scenario similar to the one in Section 3.4.3 using **1** as a starting point but this time DRD2 activity labelled data are available allowing us to use the same QSAR predictive model as a scoring component for RL training. We have already established in Section 3.1 the increased efficiency of the combined rocs+qsar based model to optimize for both objectives (ROCS score and QSAR score) compared to the single component generative models either rocs or qsar (Fig. 4A and B). In the context of this test case, we demonstrate improvement of a combined rocs+qsar based generative model over a single rocs based model by comparing the numbers of experimentally measured DRD2 active molecules (contained in the **D2ACTIVES** dataset) recovered by the two models. While recovery of active molecules can be a poor or even misleading performance metric for molecular generative models [19], we consider here their ratio to be informative for comparing the 2 models. Table 4 shows *true actives* enrichment for the rocs+qsar based model compared to the single rocs model for both AGN and STD priors .

Table 4. Number of known DRD2 actives recovered by the generative models and enrichment ratio for the combined rocs+qsar generative model against the rocs model. Only models with activated DF are considered.

	STD prior		AGN prior	
	#Actives[a]	Enrichment	#Actives[a]	Enrichment
rocs	64.7 (12.9)	1	39.7 (3.5)	1
rocs+qsar	158.7 (4.6)	2.45	110.0 (8.7)	2.77

[a] Mean value and standard deviation over 3 identical runs

4. Conclusions

This study was designed to mimic a ligand based drug discovery project where no structural information about the receptor or the bioactive conformation of the reference ligand exist. We have shown that a 3D shape and pharmacophore similarity scoring function (ROCS) can be used as a scoring component to train a RL based generative model (REINVENT) resulting in enrichment of the generated output compared with the prior. We found this output to be more chemically diverse compared to a QSAR based generated output supporting our initial hypothesis based on the argument that physics-based scoring components allow for a significantly larger coverage of the chemical space compared to QSAR predictive models with an applicability domain restricted by their training dataset. The two scoring methods are orthogonal by construction and we have shown that there is a high degree of complementarity for their generated outputs. The two scoring components can be combined together with the resulting trained model generating molecules optimised for both objectives.

The relevance of the structures generated by the generative models rocs and rocs+qsar was demonstrated by their ability to generate identical or very similar molecules to known DRD2 actives, not just as singletons but as members of larger clusters of common BM scaffolds. In many cases those molecules with high similarity or identical to known actives were generated by models that were pre-trained with the AGN prior which does not include any known actives, highlighting the ability of the models to generalise. In the same time, the rocs and rocs+qsar based models showed that they can generate high scoring clusters of molecules with novel chemotypes, highly dissimilar to known DRD2 actives and often times not even included in the priors. We have also shown examples (Fig. 11) where the definition of novelty is extended to pharmacophores (ROCS colour), for example the replacement of an O acceptor of a carbonyl group in **1** with a triazole N in **8**. We consider these examples to conform to the scaffold hopping definition as scaffold hops from **1** either into known DRD2 'privileged' scaffolds or into novel scaffolds contained in high scoring clusters of generated compounds and in both cases scaffolds that were not encountered by the generative model during pre-training. We thus demonstrated the potential of the ROCS scoring component in scaffold hopping.

Furthermore, we considered two use cases of ligand based design with applicability in practical medicinal chemistry optimisation using the results obtained from our computational study and retrospective evaluation of the generated structures for DRD2 activity. We were able to confirm in the first case, starting from a single lead compound only, that a rocs based generative model achieves enrichment in DRD2 (predicted) actives. In the second case, with additional information in the form of a dataset of molecules labelled with DRD2 activity, we showed that use of a combined rocs+qsar model is more efficient in recovering known DRD2 active compounds compared to a single rocs model. These results demonstrate the ability of the rocs based model to generate new leads with minimal available information but also synergy between the ROCS and QSAR scoring components in the presence of relevant activity data.

In summary, we have shown the utility of a 3D similarity scoring component for *de novo* molecular generation. Even if this study was designed around a ligand based design case, it should be possible to apply the same methods in structure based design (SBD) cases where the bioactive conformation of the cognate ligand can be obtained. However we can make no claims on the generalisability of the results of this study to other biological targets and 3D shape and pharmacophore queries. Current work includes investigation of a 3D electrostatic similarity scoring component as well as the use of a 3D similarity scoring component together with a docking scoring component in a SBD scenario, either in the same scoring function or sequentially.

References

[1] D. Stumpfe, P. Ripphausen, and J. Bajorath, “Virtual compound screening in drug discovery,” *Future Medicinal Chemistry*, vol. 4, no. 5. Future Science Ltd London, UK , pp. 593–602, 29-Apr-2012.

[2] C. Gorgulla *et al.*, “An open-source drug discovery platform enables ultra-large virtual screens,” *Nature*, vol. 580, no. 7805, pp. 663–668, Apr. 2020.

[3] P. G. Polishchuk, T. I. Madzhidov, and A. Varnek, “Estimation of the size of drug-like chemical space based on GDB-17 data,” *J. Comput. Aided. Mol. Des.*, vol. 27, no. 8, pp. 675–679, Aug. 2013.

[4] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, 27-May-2015.

[5] D. C. Elton, Z. Boukouvalas, M. D. Fuge, and P. W. Chung, “Deep learning for molecular design - A review of the state of the art,” *Molecular Systems Design and Engineering*, vol. 4, no. 4. Royal Society of Chemistry, pp. 828–849, 01-Aug-2019.

[6] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, “The rise of deep learning in drug discovery,” *Drug Discovery Today*, vol. 23, no. 6. Elsevier Ltd, pp. 1241–1250, 01-Jun-2018.

[7] A. Zhavoronkov *et al.*, “Deep learning enables rapid identification of potent DDR1 kinase inhibitors,” *Nat. Biotechnol.*, vol. 37, no. 9, pp. 1038–1040, 2019.

[8] C. A. Nicolaou and N. Brown, “Multi-objective optimization methods in drug design,” *Drug Discovery Today: Technologies*, vol. 10, no. 3. Elsevier, pp. e427–e435, 01-Sep-2013.

[9] S. J. Lusher, R. McGuire, R. C. Van Schaik, C. D. Nicholson, and J. De Vlieg, “Data-driven medicinal chemistry in the era of big data,” *Drug Discovery Today*, vol. 19, no. 7. Elsevier Ltd, pp. 859–868, 01-Jul-2014.

[10] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.

[11] W. Jeon and D. Kim, “Autonomous molecule generation using reinforcement learning and docking to develop potential novel inhibitors,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Dec. 2020.

[12] M. Olivecrona, T. Blaschke, O. Engkvist, and H. Chen, “Molecular de-novo design through deep reinforcement learning,” *J. Cheminform.*, vol. 9, no. 1, p. 48, Sep. 2017.

[13] M. Popova, O. Isayev, and A. Tropsha, “Deep reinforcement learning for de novo drug design,” *Sci. Adv.*, vol. 4, no. 7, p. eaap7885, Jul. 2018.

[14] X. Liu, K. Ye, H. W. T. van Vlijmen, A. P. IJzerman, and G. J. P. van Westen, “An exploration strategy improves the diversity of de novo ligands using deep reinforcement learning: A case for the adenosine A2A receptor,” *J. Cheminform.*, vol. 11, no. 1, pp. 1–16, 2019.

[15] M. C. Sanguinetti and M. Tristani-Firouzi, “hERG potassium channels and cardiac arrhythmia,” *Nature*, vol. 440, no. 7083. Nature Publishing Group, pp. 463–469, 23-Mar-2006.

[16] W. P. Walters and R. Barzilay, “Applications of Deep Learning in Molecule Generation and Molecular Property Prediction,” 2020.

[17] R. P. Sheridan, “The Relative Importance of Domain Applicability Metrics for Estimating Prediction Errors in QSAR Varies with Training Set Diversity,” *J. Chem. Inf. Model.*, vol. 55, no. 6, pp. 1098–1107, 2015.

[18] A. D’Amour *et al.*, “Underspecification Presents Challenges for Credibility in Modern Machine Learning,” *arXiv Prepr. arXiv2011.03395*, 2020.

[19] P. Renz, D. Van Rompaey, J. K. Wegner, S. Hochreiter, and G. Klambauer, “On failure modes in molecule generation and optimization,” *Drug Discov. Today Technol.*, vol. 32–33, no. xx, pp. 55–63, 2019.

[20] D. Stumpfe, H. Hu, and J. Bajorath, “Evolving Concept of Activity Cliffs,” *ACS Omega*, vol. 4, no. 11, pp. 14360–14368, 2019.

[21] J. Liu and R. Wang, “Classification of current scoring functions,” *J. Chem. Inf. Model.*, vol. 55, no. 3, pp. 475–482, Mar. 2015.

[22] P. C. D. Hawkins, A. G. Skillman, and A. Nicholls, “Comparison of shape-matching and docking as virtual screening tools,” *J. Med. Chem.*, vol. 50, no. 1, pp. 74–82, 2007.

[23] T. Blaschke, O. Engkvist, J. Bajorath, and H. Chen, “Memory-assisted reinforcement learning for diverse molecular de novo design,” *J. Cheminform.*, vol. 12, no. 1, pp. 1–17, 2020.

[24] P.-C. Kotsias, J. Arús-Pous, H. Chen, O. Engkvist, C. Tyrchan, and E. J. Bjerrum, “Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks,” *Nat. Mach. Intell.*, vol. 2, no. 5, pp. 254–265, May 2020.

[25] J. Arús-Pous *et al.*, “SMILES-based deep generative scaffold decorator for de-novo drug design,” *J. Cheminform.*, vol. 12, no. 1, pp. 1–18, 2020.

[26] J. Horwood and E. Noutahi, “Molecular Design in Synthetically Accessible Chemical Space via Deep Reinforcement Learning,” *ACS Omega*, 2021.

[27] Y. Li, J. Hu, Y. Wang, J. Zhou, L. Zhang, and Z. Liu, “DeepScaffold: A Comprehensive Tool for Scaffold-Based de Novo Drug Discovery Using Deep Learning,” *J. Chem. Inf. Model.*, vol. 60, no. 1, pp. 77–91, 2020.

[28] W. Jin, K. Yang, R. Barzilay, and T. Jaakkola, “Learning Multimodal Graph-to-Graph Translation for Molecular Optimization,” *arXiv Prepr. arXiv1812.01070*, Dec. 2018.

[29] T. Blaschke *et al.*, “REINVENT 2.0: An AI Tool for De Novo Drug Design,” *J. Chem. Inf. Model.*, vol. 60, no. 12, pp. 5918–5922, Dec. 2020.

[30] “MolecularAI/Reinvent.” [Online]. Available: https://github.com/MolecularAI/Reinvent. [Accessed: 02-Mar-2021].

[31] G. Hu, G. Kuang, W. Xiao, W. Li, G. Liu, and Y. Tang, “Performance evaluation of 2D fingerprint and 3D shape similarity methods in virtual screening,” *J. Chem. Inf. Model.*, vol. 52, no. 5, pp. 1103–1113, 2012.

[32] D. M. Krüger and A. Evers, “Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors,” *ChemMedChem*, vol. 5, no. 1, pp. 148–158, 2010.

[33] T. Miyao, S. Jasial, J. Bajorath, and K. Funatsu, “Evaluation of different virtual screening strategies on the basis of compound sets with characteristic core distributions and dissimilarity relationships,” *J. Comput. Aided. Mol. Des.*, vol. 33, no. 8, pp. 729–743, Aug. 2019.

[34] C. Grebner, H. Matter, A. T. Plowright, and G. Hessler, “Automated de Novo Design in Medicinal Chemistry: Which Types of Chemistry Does a Generative Neural Network Learn?,” *J. Med. Chem.*, vol. 63, no. 16, pp. 8809–8823, 2020.

[35] M. Skalic, J. Jiménez, D. Sabbadin, and G. De Fabritiis, “Shape-Based Generative Modeling for de Novo Drug Design,” *J. Chem. Inf. Model.*, vol. 59, no. 3, pp. 1205–1214, 2019.

[36] A. Gaulton *et al.*, “The ChEMBL database in 2017,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D945–D954, Jan. 2017.

[37] J. Sun *et al.*, “ExCAPE-DB: An integrated large scale dataset facilitating Big Data analysis in chemogenomics,” *J. Cheminform.*, vol. 9, no. 1, p. 17, Mar. 2017.

[38] G. Landrum *et al.*, “rdkit/rdkit: 2019_09_1 (Q3 2019) Release,” Oct. 2019.

[39] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[40] L. Fan *et al.*, “Haloperidol bound D2 dopamine receptor structure inspired the discovery of subtype selective ligands,” *Nat. Commun.*, vol. 11, no. 1, pp. 1–11, 2020.

[41] T. Kaserer, V. Obermoser, A. Weninger, R. Gust, and D. Schuster, “Evaluation of selected 3D virtual screening tools for the prospective identification of peroxisome proliferator-activated receptor (PPAR) γ partial agonists,” *Eur. J. Med. Chem.*, vol. 124, pp. 49–62, 2016.

[42] “OEToolkits 2019.Oct — Toolkits -- Python.” [Online]. Available: https://docs.eyesopen.com/toolkits/python/releasenotes/releasenotes2019_Oct.html. [Accessed: 26-Feb-2021].

[43] “Daylight Theory: SMARTS - A Language for Describing Molecular Patterns.” [Online]. Available: https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html. [Accessed: 15-Mar-2021].

[44] N. Brown, M. Fiscato, M. H. S. Segler, and A. C. Vaucher, “GuacaMol: Benchmarking Models for de Novo Molecular Design,” *J. Chem. Inf. Model.*, vol. 59, no. 3, pp. 1096–1108, Mar. 2019.

[45] D. Polykovskiy *et al.*, “Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models,” *Front. Pharmacol.*, vol. 11, pp. 1–19, 2020.

[46] W. Gao and C. W. Coley, “The Synthesizability of Molecules Proposed by Generative Models,” *J. Chem. Inf. Model.*, vol. 60, no. 12, Dec. 2020.

[47] A. Thakkar, V. Chadimová, C. Chadimová, E. J. Bjerrum, O. Engkvist, and J.-L. Reymond, “Retrosynthetic accessibility score (RAscore)-rapid machine learned synthesizability classification from AI driven retrosynthetic planning †,” *Chem. Sci.*, 2021.

[48] J. Arús-Pous *et al.*, “Randomized SMILES strings improve the quality of molecular generative models,” *J. Cheminform.*, vol. 11, no. 1, pp. 1–13, 2019.

[49] G. W. Bemis and M. A. Murcko, “The properties of known drugs. 1. Molecular frameworks,” *J. Med. Chem.*, vol. 39, no. 15, pp. 2887–2893, Jul. 1996.

[50] S. R. Langdon, P. Ertl, and N. Brown, “Bioisosteric Replacement and Scaffold Hopping in Lead Generation and Optimization,” *Mol. Inform.*, vol. 29, no. 5, pp. 366–385, May 2010.

[51] H. J. Böhm, A. Flohr, and M. Stahl, “Scaffold hopping,” *Drug Discovery Today: Technologies*, vol. 1, no. 3. Elsevier Ltd, pp. 217–224, 01-Dec-2004.

[52] Y. Jiang, Z. Liu, J. Holenz, and H. Yang, “Competitive Intelligence–based Lead Generation and Fast Follower Approaches,” John Wiley & Sons, Ltd, 2016, pp. 183–220.

[53] D. G. Brown and J. Boström, “Where Do Recent Small Molecule Clinical Development Candidates Come From?,” *J. Med. Chem.*, vol. 61, no. 21, pp. 9442–9468, 2018.

Supplementary Material

Tables

Table S1. Description of datasets used in the text.

Dataset	Size	Description	Use
STD	1435546	Subset of ChEMBL25	Prior (pre-training)
AGN	1431348	Subset of STD after removing DRD2 actives and haloperidol analogues	Prior (pre-training)
D2ACTIVES	4791	D2 actives in ChEMBL25	
D2INACTIVES	2000	Random selection of DRD2 inactives from ExCAPE-DB	
prior 100K	100000	a 100K sample from STD	
D2TEST	1401	1164 actives from STD and 237 inactives from ExCAPE-DB	Calculate ROCS and QSAR scores and probability of generation
D2QSAR	347079	All DRD2ACTIVES and 342288 inactives from ExCAPE-DB	Training of a DRD2 activity prediction QSAR model
D2ROCS	9	3024 conformers generated with OMEGA	Obtain best haloperidol conformation to use for ROCS query
SAMPLE_PRE	5724859	Combined output *during* training, from 20 REINVENT runs: 3000 training steps, 1-3 repeats each run	
SAMPLE_POST	10857843	Combined output over 20 REINVENT runs from sampling *after* training: 60 checkpoints and 1-3 repeats each run; 10000 sized sample each checkpoint	

Figures

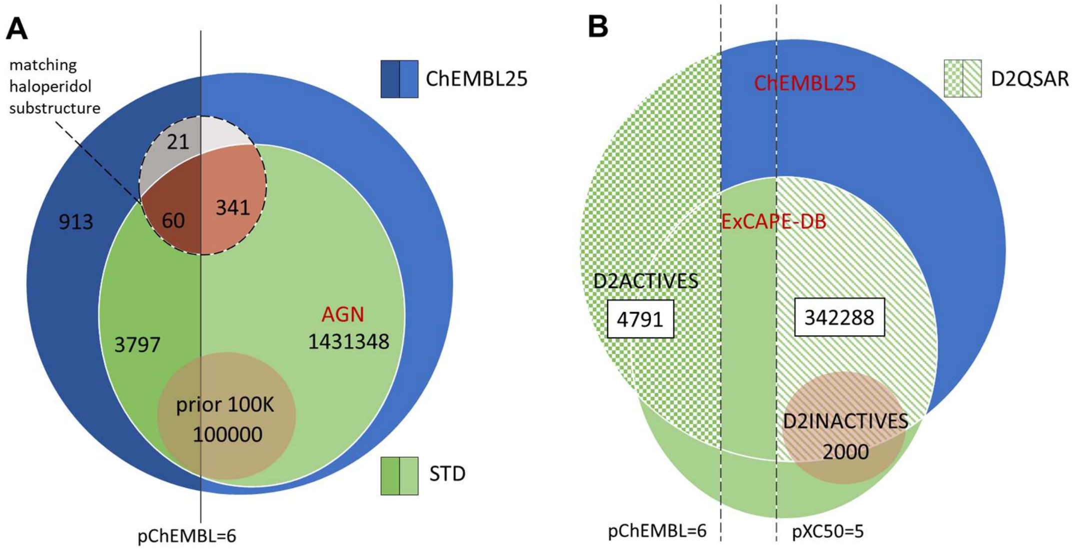


Figure S1. Construction of datasets used in the main text

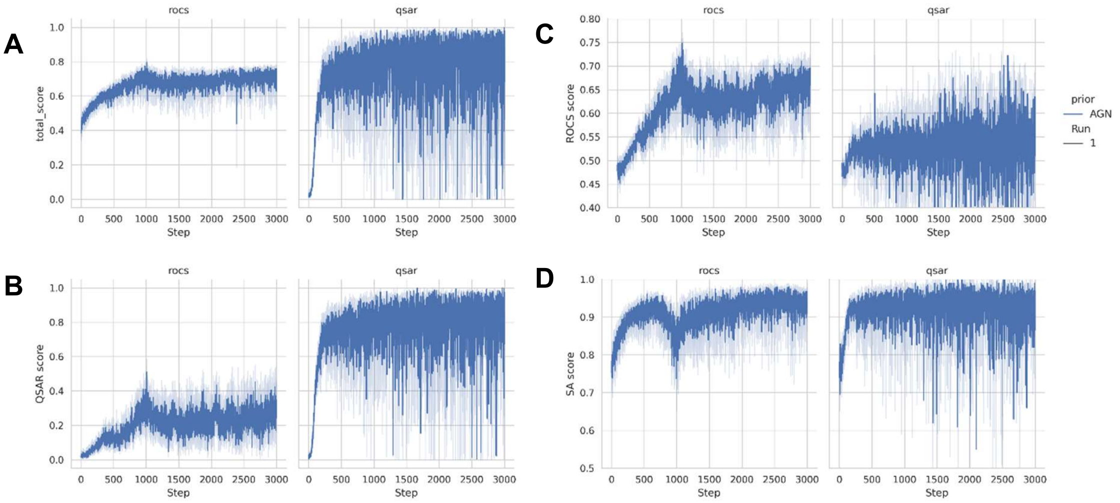


Figure S2. Scoring of the output from generative models during training: **A**. total_score, **B**. QSAR score, **C**. ROCS score and **D**. SA score. Score values refer to molecules that were collected during training. SA score, ROCS score for qsar based models and QSAR score for rocs based models were calculated *post hoc* and were not part of the respective scoring function for RL training. The diversity filter was *not* activated for all runs described here.