JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2018, 58, 1194−1204

pubs.acs.org/jcim

Reinforced Adversarial Neural Computer for *de Novo* Molecular Design

Evgeny Putin,[*,†,‡] Arip Asadulaev,[‡] Yan Ivanenkov,[†,¶,§] Vladimir Aladinskiy,[†,¶] Benjamin Sanchez-Lengeling,[∥] Alán Aspuru-Guzik,[∥,⊥] and Alex Zhavoronkov[*,†,#]

[†]Pharma.AI Department, Insilico Medicine, Inc, Baltimore, Maryland 21218, United States

[‡]Computer Technologies Lab, ITMO University, St. Petersburg 197101, Russia

[¶]Moscow Institute of Physics and Technology (State University), 9 Institutskiy lane, Dolgoprudny City, Moscow Region, 141700, Russian Federation

[§]Institute of Biochemistry and Genetics Russian Academy of Science (IBG RAS), Ufa Scientific Centre, Oktyabrya Prospekt 71, 450054, Ufa, Russian Federation

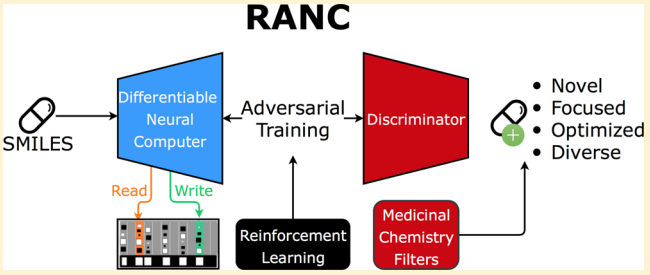
[∥]Chemistry and Chemical Biology Department, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02143, United States

[⊥]Biologically-Inspired Solar Energy Program, Canadian Institute for Advanced Research (CIFAR), Toronto, Ontario M5S 1M1, Canada

[#]The Buck Institute for Research on Aging, 8001 Redwood Boulevard, Novato, California 94945, United States

Supporting Information

ABSTRACT: *In silico* modeling is a crucial milestone in modern drug design and development. Although computer-aided approaches in this field are well-studied, the application of deep learning methods in this research area is at the beginning. In this work, we present an original deep neural network (DNN) architecture named RANC (Reinforced Adversarial Neural Computer) for the *de novo* design of novel small-molecule organic structures based on the generative adversarial network (GAN) paradigm and reinforcement learning (RL). As a generator RANC uses a differentiable neural computer (DNC), a category of neural networks, with increased generation capabilities due to the addition of an explicit memory bank, which can mitigate common problems found in adversarial settings. The comparative results have shown that RANC trained on the SMILES string representation of the molecules outperforms its first DNN-based counterpart ORGANIC by several metrics relevant to drug discovery: the number of unique structures, passing medicinal chemistry filters (MCFs), Muegge criteria, and high QED scores. RANC is able to generate structures that match the distributions of the key chemical features/descriptors (e.g., MW, logP, TPSA) and lengths of the SMILES strings in the training data set. Therefore, RANC can be reasonably regarded as a promising starting point to develop novel molecules with activity against different biological targets or pathways. In addition, this approach allows scientists to save time and covers a broad chemical space populated with novel and diverse compounds.



INTRODUCTION

Early stages of modern drug design and development (DDD) are commonly based on three crucial scientific disciplines: *in silico* modeling, combinatorial organic synthesis, and high-throughput biological screening (HTS).[1] A huge number of novel small-molecule drug molecules [molecule, structure, compound, and molecular structure are used interchangeably in the manuscript] with high diversity in structure have been discovered via this cumulative approach.[2−6]

At the modeling and HTS stages, *in silico* techniques are utilized to characterize desired molecular qualities for a focused target,[7] such as activity, selectivity, pharmacokinetic profile, toxicity, solubility, stability, and synthetic accessibility. Molecules are also filtered by expert opinion, to be finally considered as potential lead compounds for organic synthesis, to be further investigated in an appropriate protein or cell-based assay.

Based on their performance, structures and motifs are incorporated into subsequent HTS cycles where other properties might be the focus of the optimization: selectivity profile, metabolic stability, surrogate ADME properties, potential toxicity (for instance, hERG channels inhibition), solubility, permeability, and stability. The best candidates are

Received: December 2, 2017
Published: May 15, 2018

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

incorporated to preclinical in vivo evaluation programs and clinical trials.

Recently due to advances in machine learning (ML) and artificial intelligence (AI) techniques, *de novo* structure generation has been adapted as a data-driven approach to provide novel good molecules either targeted on a particular target or on a class of targets with the ultimate goal of being able to suggest structures smartly, finding interesting compounds suitable for further developement as an expert medicinal chemist would.

While these approaches can be quite fast and generate a large amount of samples, the complete drug discovery cycle is many orders of magnitude slower and more expensive. For instance, from development to market, a new drug can take 10 years and cost 2.6 billion dollars.[8]

For a known target, a medicinal chemist is usually faced with a very tight chemical space around the reported candidate molecule, it is estimated the patent landscape of the pharmaceutical industry covers up to 70% of all possible structural modifications and analogues.[9] It should be noted that the space of all organic synthetically available drug-like molecules is estimated to be $10^{60}-10^{100}$ compounds.[10] Coverage of candidates should take into account desired features and be clear of intellectual property (IP) conflicts. Moreover, diversity and novelty to existing chemistry should be taken into account. However, these challenges can be effectively tackled with joint *in silico* techniques and ML.

AI-based methods and particularly deep neural networks (DNNs) recently have made tremendous progress in pattern recognition, nature language processing,[11] biomedicine,[12] bioinformatics,[13,14] and particularly in DDD.[15−19] We have recently described the application of DNNs for drug repurposing and for the prediction of key pharmacological properties.[20] In addition, DNNs have been actively used to generate chemical structures *de novo*, thereby providing a really extensive chemical space filled with a number of various molecules.[21−24] Nevertheless, as mentioned above, such networks still need to be able to tackle the constrains for real world applications. Current models need to be able to recognize interesting and attractive scaffolds, easily synthesizable products and potential patterns relevant to medicinal chemistry. Several attempts are currently underway to address these issues;[17,24] however, this problem has not been solved yet.

One of the most successful attempts to address the aforementioned issues is an ORGANIC (Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry) architecture[25] that combines adversarial training[26] and deep reinforcement learning (RL). Based on the SMILES string representation of the molecules, ORGANIC trains a generator model to create molecular structures that are penalized or rewarded by a discriminator model and a reward function which quantifies desired properties such as drug-likeliness. The discriminator attempts to classify proposed molecules as fake or real, based on a data distribution, essentially incentivizing the generator to create realistic samples. This allows scientists to generate targeted molecules with predefined desired properties what can potentially be very useful in *in silico* DDD process.

On the other hand, generative adversarial networks (GANs) can be problematic to train, often suffering from mode-collapse, when the generator ends up exploiting repetitious patterns in samples. Or from the perfect discriminator problem,[27] when a discriminator overwhelms the generator, preventing the generator from learning or improving its capacity. This potential problem can arise in the adversarial component in the ORGANIC paradigm, leading to a collapse during training, and thus becoming a naive RL approach.[24] When generating SMILES, the perfect discriminator problem might translate to producing structures with much shorter string lengths when compared to the initial training set. This can lead to chemical characteristics such as molecular weight, the number of atoms, logP, and TPSA that differ substantially from the initial distribution of the training set. Such a degenerative process in the generated molecules is undesirable and can prevent discovery of novel, effective, and diverse compounds.

There are many ongoing efforts to study and improve the convergence properties in GANs. Some rely on modifying the loss functions;[28] other paths of improvement rely on larger training sets or altering the discriminator or generator network. Our approach lies in utilizing a more powerful generator network to make better artificial samples to prevent the discriminator model from overcoming the generator during training. Previous efforts have utilized less powerful architectures such as long short-term memory (LSTM)[29] and gated recurrent units in recurrent neural networks (RNNs) for sequence generation. With this purpose we utilize a differentiable neural computer (DNC)[30] architecture which incorporates a memory bank to account for short and long-term patterns of sequences. It should be noted that Popova et al.[31] have also previously utilized memory-augmented LSTM networks coupled with RL.

The introduced architecture called RANC (Reinforced Adversarial Neural Computer) meets the mentioned problem and demonstrates performance superior to ORGANIC in several metrics key to drug discovery. Our engine generates more unique and diverse structures as well as clusters with the lengths close to the reference samples, keeping the distributions of key molecular descriptors as in the training sets. As a result, many of the resulting structures meet the crucial criteria used in medicinal chemistry of today and are able to pass medicinal chemistry filters (MCFs). Such analysis has also been performed for two training data sets to estimate how the quality of the reference molecules impacts the properties of the generated structures. Our approach is the first that combines a DNC architecture with RL and adversarial training.

Our contribution in the work is as follows:

1. We tackle current problems with adversarial training by proposing a generator architecture with external memory. An analysis of the generated molecules showed that RANC molecules match the distributions of the lengths and the key molecular descriptors of the training molecules. We compare with ORGANIC and report improvements in generation with these parameters.
2. Second, we perform extensive chemical analysis of the training molecules as well as of molecules generated by the ORGANIC and RANC models which had not been previously reported. By this analysis we assess how a prefiltered training set influences the quality of the generated structures and compare molecules generated by both models from the medicinal chemistry perspective.

METHOD

The generation of discrete data using RNNs, in particular, LSTMs[29] with maximum likelihood estimation (MLE), has been shown to work well in practice. However, this often

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

suffers from the so-called *exposure bias*, lacking some of the multiscale structures or salient features of the data.

Graves et al.[30] proposed an extension of the neural Turing machine[32] called DNC. DNC comprises a controller which could be any type of DNN, with external memory trained in the end-to-end manner. In graph experiments, after almost two million training samples, LSTM reached an average of only 37%, while DNCs reached up to 98.8% accuracy after one million training samples. During synthetic question answering trials using the bAbI data set,[33] DNC has also outperformed LSTM by a significant margin.[30] Coupled with RL, DNC are able to read and write sequence structures that are relevant for the reward.

As well as in question answering task, in automatic molecule generation, the capacity and strength of the model play key roles and serve as the main argument for the selection of the DNC architecture for the generation of molecular structures.

The RANC (see Figure 1) method is based on the ORGANIC paradigm, where DNC is applied instead of an

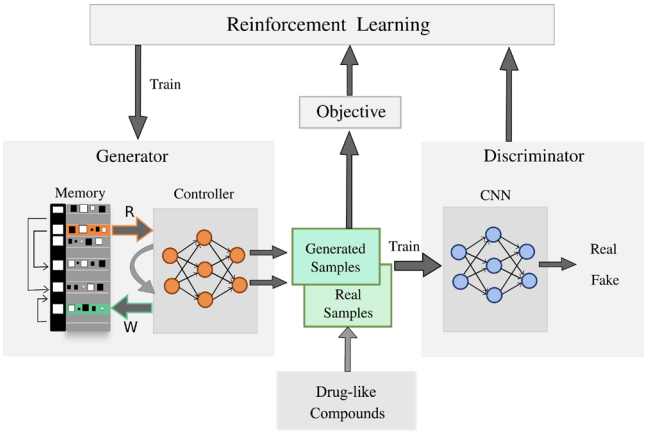


Figure 1. Schematic view of the RANC model. The generator's controller generates samples and also produces vectors that parametrize one write head (green) and read heads (orange). The write and erase vectors, defined by the write head, are used to edit the $N{\cdot}W$ external memory matrix. Then, generated and real samples are used to train the discriminator. After training, the discriminator returns rewards which are estimated as a likelihood of fooling the discriminator, and the sum of these rewards multiplied by the lambda coefficient, applies to the training of the generator.

LSTM. The use of DNC as a SMILES-based θ-parametrized molecule generator G_θ has several advantages, all of them are rooted in the availability of differentiable external memory and attention mechanisms. The first advantage is the ability of DNC to remember very complex sequences. The second one is that DNC allows generation of much longer sequences $Y_{1:T} = (y_1, ..., y_T)$ compared to LSTM, where T is the length of the sequence.

We used the MLE for pretraining the generator G_θ[34] and RL for training it via in an adversarial setting. The objective of the RANC generator G_θ is to simultaneously fool ϕ-parametrized discriminator D_ϕ and maximize the objective reward function O, based on the policy gradient technique.[35]

$$J(\theta) = \sum_{y_1 \in Y} G_\theta(y_1|s_0)\cdot Q(s_0, y_1) \tag{1}$$

where $Q(s, a)$ is an expected total reward starting from state s, taking action a. We used a convolutional neural network (CNN) as discriminator D_ϕ, because it showed high performance in sequence classification tasks.[36] The objective of D_ϕ is to minimize the cross entropy between the ground truth and the generated sequences. We also applied discriminator pretraining with the fixed generator G_θ after MLE pretraining, where parameter ζ regulates for how many epochs the discriminator should be pretrained only on valid SMILES sequences Y'.

$$\nabla_\phi J(\phi) = \begin{cases} \min_\phi E_{Y\sim p_{\text{data}}(Y)}[\log D(Y)] + E_{Y\sim p_{G_\theta}(Y)}[\log(1 - D(Y))], & \text{if epoch} < \zeta \\ \min_\phi E_{Y'\sim p_{\text{data}}(Y')}[\log D(Y')] + E_{Y\sim p_{G_\theta}(Y')}[\log(1 - D(Y'))], & \text{if epoch} > \zeta \end{cases} \tag{2}$$

To calculate $Q(s_0, y_1)$ for partial sequences at intermediate states and also to avoid long-term reward problem, N-time Monte Carlo search MC^{G_θ} has been applied. Following the roll-out policy G_β, MC^{G_θ} starts to roll out from some state until it reaches the end of the sequence.

$$\text{MC}^{G_\theta}(Y_{1:t}; N) = \{Y^1_{1:T}, ..., Y^N_{1:T}\} \tag{3}$$

Then we are able to compute $Q(s, a)$ for every intermediate state using MC^{G_θ}, and thus, we can present an approximation of the rewards for the state s and the action a per every generated sample in a batch.

$$Q(s = Y_{1:t-1}, a = y_t) = \begin{cases} \frac{1}{N}\sum_{n=1}^{N} R(Y^n_{1:T}), Y^n_{1:T} \in \text{MC}^{G_\theta}(Y_{1:T}; N), & \text{if } t < T \\ R(Y_{1:T}), & \text{if } t = T \end{cases} \tag{4}$$

where, following the original ORGANIC concept, the reward is a sum of the discriminator probability $D_\phi(Y_{1:T})$ (which measures how close the generated sequences $Y_{1:T}$ are to the ground truth sequences) and the objective reward functions O (which is equal to 1, when the sequence meets to the requirements of the filter, and 0 otherwise).

$$R(Y_{1:T}) = \lambda{\cdot}D_\phi(Y_{1:T}) + (1 - \lambda){\cdot}O(Y_{1:T}) \tag{5}$$

Finally, the DNC generator G_θ parameters θ can be derived following SeqGAN[34] unbiased estimation.

$$\nabla_\theta J(\theta) \simeq \frac{1}{T}\sum_{t=1}^{T} E_{y_t\sim G_\theta(y_t|Y_{1:t-1})}[\nabla_\theta \log G_\theta(y_t|Y_{1:t-1})\cdot Q(Y_{1:t-1}, y_t)] \tag{6}$$

After the generated sequences become closer to the original ones, the discriminator D_ϕ should be retrained using eq 2. The overall training procedure of the RANC model is summarized in Algorithm 1.

RESULTS

Data. We have used two different data sets:

1. Drugs data set: a subset of 15 000 drug-like compounds from the ZINC[37] database of 35 million commercially available compounds. The selected samples were

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

Algorithm 1 Training procedure of the RANC model

Require: DNC generator policy G_θ ; roll-out policy G_β; Discriminator D_ϕ; objective O; set S of sequences of length T;
Initialize G_θ, D_ϕ with random weights θ, ϕ; Initialize λ, ζ;
Pre-train G_θ using MLE on S; Pre-train D_ϕ by minimizing cross entropy via generated negative samples by G_θ or verified negative samples, and positive S;

```
repeat
  for g-steps do
    Generate a sequence Y_{1:T} = (y_t, ..., yT) ~ G_θ; additionally writing and reading from the external memory.
    for t in 1 : T do
      Compute Q(s = Y_{1:t-1}, a = y_t); by Eq.(4) by calculating rewards using Eq.(5)
    end for
    Update generator parameters θ via Eq.(6)
    for d-steps do
      Train D_φ for k epochs by Eq.(2) via only verified generated negative samples using G_θ and positive S;
    end for
  end for
until model coverage.
```

identical in structure to that used by Sanchez-Lengeling and colleagues.[25]

2. ChemDiv (CD) data set: a collected subset of 15 000 drug-like molecules available within the ChemDiv [http://www.chemdiv.com/] stock collection.

One of the most reliable ways to generate more adequate structures appropriate for subsequent synthesis and biological evaluation, in terms of rationality and drug discovery success, is a training database that contains a sufficient number of reference molecules to approximate and cover the chemical space we wish to optimize. These samples should meet several criteria important for further development. One of our goals is to estimate how a prefiltered training set influences the quality of the generated structures.

As an exploratory analysis, we studied chemical statistical features of the training data sets. For each data set, we have calculated internal diversity, the number of unique diverse fragments, the number of clusters, singletons, the number of unique heterocycles, average cluster size, the number of compounds that have not passed MCFs, number of molecules that satisfied drug-likeness criteria, for example, Lipinski rule of five (RO5).

One of the most important statistical features for medicinal chemistry is the presence or absence of some substructures or fragments in a compound. For this case the *Fragments* metric looks at the number of unique fragments that are identified as diverse in a set. Details for selection are describe by Trepalin et al.[38] Additionally, MCFs allow scientists to eliminate compounds that contain "alert" substructures. For instance, the presence of 1,4-Michael acceptors fragment in the compound could potentially lead to toxicity[39] and cause various side effects. Herein, we have applied 150 MCFs (for more details see the work of Kalgutkar et al.[40]): 1,4-Michael acceptors, electrophilic spices, strained heterocycles, undesirable elements (P, Si, B, etc.), metabolically unstable compounds, alkylators and intercalators, sugars, nucleosides, hemiacetal and aminals, Schiff bases, reactive groups, chelators.

The internal diversity of the compounds included in the reference data set was calculated based on extended-connectivity fingerprints[41] using Tanimoto similarity.[42] Clustering analysis was performed as follows. Molecules were distributed in different clusters by the similarity in structure using a routine Tanimoto approach. Compounds with a similarity coefficient over 0.6 were assigned to the same cluster. The minimum number of structures per cluster (5) was selected empirically to reduce the number of singletons and to adequately assort compounds among clusters thereby minimizing overlapping. All the calculations were performed using proprietary ChemoSoft Software by the ChemDiv company. The results of such analysis are presented in Table 1.

Table 1. Key Chemical Statistical Parameters for the CD and Drugs Data Sets

data set	CD	Drugs
size	15000	15000
diversity	0.87	0.86
fragments[a]	8510	21773
clusters	793	223
singletons	3035	13554
heterocycles[b]	644	1052
cluster size[c]	18.9	67.2
MCF (%)	0.14	32.6
RO5 (%)	77.9	89.7

[a]Number of unique fragments considered to be diverse. [b]Number of unique heterocycles. [c]Average cluster size. Percentages are taken over the entire data set.

As shown in Table 1, the CD and Drugs data sets have a similar diversity coefficient; however, the remaining features are notably different. Thus, the number of structural clusters in the CD data set is almost four times higher in contrast to the Drugs data set. On the contrary, the number of singletons is four times lower vs the CD data set. Keeping in mind that the diversity of both data sets is almost identical, this clearly indicates that the CD molecules are distributed more normally and the related chemical space is more structured and well-organized to avoid undesired outliers. However, heterocyclic diversity of the CD data set is slightly lower than that of the Drugs data set. In our opinion, from the medicinal chemistry perspective, the number of molecules, which successfully passed MCFs, is one of the crucial milestones for the development of novel drug molecules with low rate of toxicity and side-effects. To our satisfaction, CD pool is much more abundant with good compounds in contrast to Drugs molecules. This issue can be of critical significance for the generation of novel structures using deep learning engines.

Settings. To perform an objective comparison between RANC and ORGANIC all the trials were performed with almost the same hyper-parameters and settings, except for the number of training epochs, because these algorithms train at different rates. As an input for the RANC/ORGANIC model we use a SMILES string encoded as a one-hot vector. For each data set, we create a dictionary based on unique characters that are used for encoding the SMILES. The sizes of the dictionaries for the CD and Drugs data sets were 29 and 34, accordingly.

The DNC generator G_θ in the RANC model consisted of LSTM controller with 256 neurons, with one read head R, and memory size of $N{\cdot}W$, where N is a length of train data dictionary, and W is a max length (90) of train data sequences. A depth search of 16 was utilized for the Monte Carlo policy rollout. RMSprop was applied to optimize the controller with 1×10^{-3} leaning rate and 0.9 momentum. Coefficient λ for the reward function was set to optimal value of 0.5 and coefficient ζ was set to 0.8. The settings of the discriminator D_ϕ remained the same as in the ORGANIC model. The batch size was set to 64 for both the generator and the discriminator.[23]

The RANC model was implemented in Tensorflow.[43] We used the ORGANIC implementation as made available by August 2017 [available at https://github.com/aspuru-guzik-group/ORGANIC]. RDKit software was used to preprocess

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

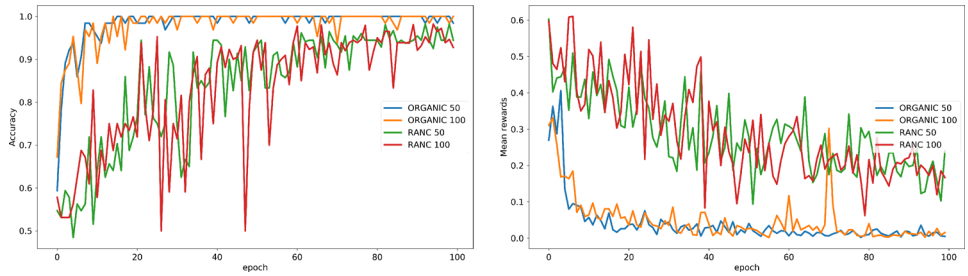


Figure 2. (left) Classification accuracies of the discriminators by pretraining epochs. (right) Mean values of the adversarial rewards of the generators by pretraining epochs. Numbers after names of the models in the legends correspond to pretraining epochs of the generators.

and analyze molecules.[44] Training of the models was conducted using NVIDIA Titan Pascal GPUs with 128 Gb of RAM.

Perfect Discriminator Problem. In adversarial training, the goal is to obtain a generator distribution that is able to match the real data distribution perfectly then the discriminator will be maximally confused, predicting 0.5 for all inputs. A perfect discriminator would be able to discriminate between generated and real data perfectly. In this scenario the loss of the discriminator approaches zero. As a result, the gradient of the generator vanishes, making it almost impossible for the generator to continue improving.

We investigated this issue in the ORGANIC and RANC models on the CD data set. Generators were pretrained using MLE loss and two instances of the generators were fixed at 50 and 100 epochs. Discriminators were pretrained for 100 epochs to detect fake samples. Mean accuracies and mean adversarial rewards are shown in Figure 2.

Figure 2 shows that accuracy for the ORGANIC discriminators is very close to perfect after 15 pretraining epochs. After the MLE pretraining of the generators, ORGANIC generates a relatively low percent (24%) of the valid SMILES strings and the generated molecules contain noncharacteristic substructures from the training data set. As a result, after few pretraining epochs of the discriminator, it can easily detect fake samples. This could be an issue of a weak generator or a too powerful discriminator. This greatly affects the quality of RL; the rewards from the discriminator become much smaller than the objective-reinforced rewards (they are 1 if the molecules pass RO5 and 0 otherwise), converting the process to naive RL. However, the ORGANIC model has a λ parameter, which regulates the contribution of the rewards, but even when the λ parameter is set to 0.9, the rewards of the discriminators can be smaller than the objective-reinforced rewards (see Figure 2).

Meanwhile, the RANC discriminators do not reach accuracies that are close to 1 even on the hundredth epoch of the pretraining. Consequently, the adversarial rewards of the generators are much bigger compared to ORGANIC. Such behavior is explained by the fact that the DNC generators have a high rate (76%) of generation valid SMILES strings. In addition to strengthening the generator and improving the quality of the generation, we found it useful to apply a special parameter ζ (see Algorithm 1) which controls how many epochs the discriminator should be pretrained only on the valid SMILES strings.

For molecule generation that retains the properties of the reference molecules it is important to supply adversarial rewards for the generator. And from the above analysis we claim that the RANC model does not suffer from the perfect discriminator problem.

Training and Evaluation. Here the RANC and ORGANIC models were trained on the CD and Drugs data sets with RO5 as objective-reinforced reward function. To perform an honest comparison of the models we restricted the length of the training SMILES strings up to 90 characters. For ORGANIC, we used 240 and 100 epochs to pretrain the generator and the discriminator, accordingly, and 100 epochs to train the model. All other settings of the model remained as reported by Guimaraes et al.[23] Meanwhile, the RANC model was trained with only 20 and 15 generator and discriminator pretraining epochs, correspondingly, and solely with 50 RL epochs.

At the end of each epoch, ORGANIC sampled 6400 molecules as well as RANC. However, ORGANIC sampled 32 000 molecules every tenth epoch. Thus, after 100 RL epochs of training ORGANIC and 50 epochs of RANC training, we got 896 000 and 320 000 generated SMILES strings of molecules, respectively. The statistical analysis of the generated structures by both models on the CD and Drugs data sets is shown in Table 2.

Table 2. Statistics for the Generated SMILES Strings by RANC and ORGANIC Models on the CD and Drugs Data Sets

	CD		Drugs	
data set approach	RANC	ORGANIC	RANC	ORGANIC
ave. length	46	23	40	23
valid (%)	58	87	76	92
unique (%)	48	18	76	24

RANC generates SMILES stings almost twice as long as ORGANIC (see Table 2). Furthermore, RANC has three times higher percentage of the unique generated SMILES stings than ORGANIC. Although ORGANIC has very good generation rate of the valid SMILES strings on both data sets, such circumstances are explained by the fact that it is much easier for the model to generate short valid SMILES stings than long ones. RANC is able to preserve the average length of the training SMILES strings.

Figure 3 shows the changes observed in the average lengths of generated SMILES strings depending on the training epochs of the models and the distributions of the lengths of the molecules generated by the models in comparison with training molecules. RANC stably generates SMILES strings which are close in length to the training molecules. In contrast, ORGANIC generates much shorter SMILES strings after several RL epochs. This more evidence that the ORGANIC model suffers from the perfect discriminator problem. Gradients of the discriminator quickly vanishe, and only the

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

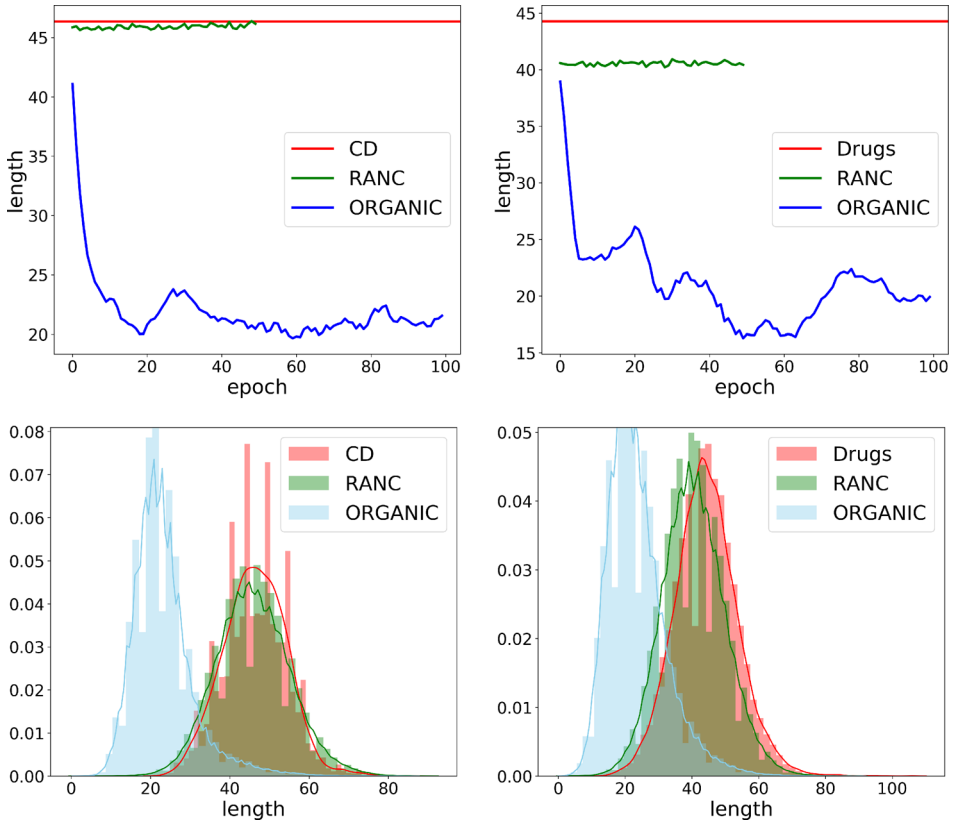


Figure 3. (top) Mean lengths of the generated SMILES strings on each training epoch by ORGANIC and RANC models for the CD and Drugs data sets using RO5 as objective reward function. Red lines are means in the original data sets. (bottom) Length distribution of the generated SMILES strings by both models on the CD and Drugs data sets.

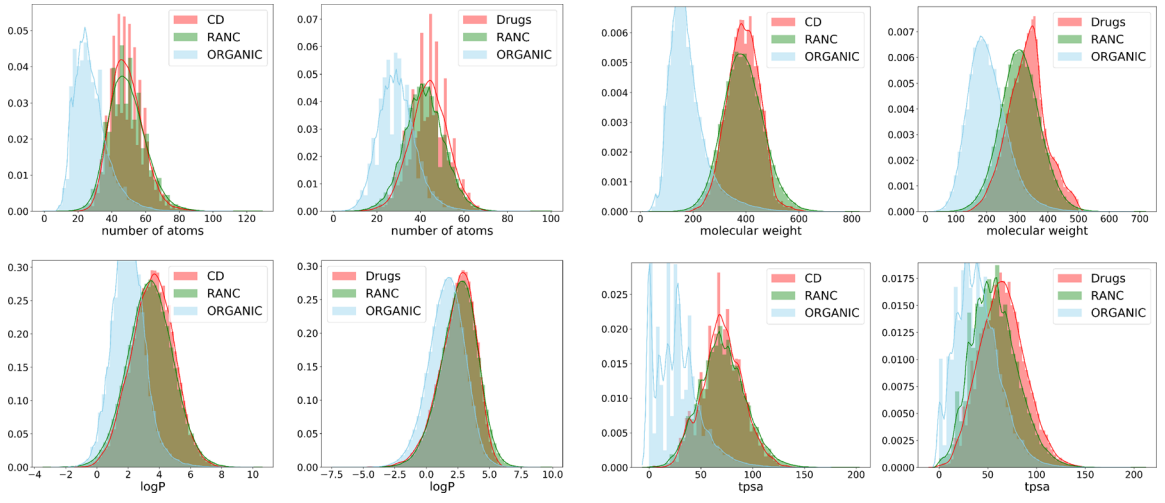


Figure 4. Distributions of the representative molecular descriptors calculated for the generated molecules.

RL component sufficiently contributes to the learning process of the generator.

In this scenario, it is not very hard for ORGANIC to get some predetermined reward from the RL component by generating molecules that could successfully pass RO5, because the rule does not have lower boundaries for all the chemical properties that are calculated for a molecule. If a sample satisfies the rule, the generator will get a predefined reward, no matter if this molecule is a short or long SMILES string. Thus, in general, shorter SMILES strings would improve the generator better than longer ones, because it will constantly get the reward from them.

The same behavior was also observed when using the Muegge metric,[45] a selection criteria for drug-like compounds that is stronger than RO5 since it utilizes more information about the compounds (see SI Figure 1 and SI Table 1). Potentially, optimizing with harsher constraints, could help to alleviate this issue. Further research into the interplay between the complexity of the rewards in the RL setting and generated samples is needed to understand if this degradation can be avoided in other ways.

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

Nevertheless RANC improves over several of the problems found in adversarial training, displaying very stable and consistent behavior during training while generating structures that match the distributions of the training molecules in several key descriptors: length of SMILES strings, molecular weight, logP, TPSA, etc. (see Figure 4). This observation is evidence that the generative component is working, since the samples that are created have properties similar to the initial data distribution. It is of particular importance in a modern DDD process.

Chemical Analysis of Generated Molecules. To further understand the applicability of RANC and ORGANIC in modern *in silico* molecular design we investigated the generated molecules by both models. We have randomly selected 20 000 structures from each pool and for each data set that met to the Muegge drug-likeness filter. Instead of using RO5 as a selection criteria we decided to use more rigid one that imposes more restrictions on the molecules. For instance, a molecule with molecular weight less than 200 will not pass the Muegge filter but will successfully pass RO5. Thus, the purpose of using the Muegge filter was to further refine molecules based on medicinal chemistry principals.

The results of the performed analysis are summarized in Table 3. The structures generated by RANC are more attractive in terms of DDD since the number of compounds successfully passed the MCFs is almost five and 1.7 times higher for the CD and Drugs data sets than ORGANIC, respectively. The more favorable results have been obtained using CD training data set. The structures generated by RANC are well clustered with an acceptable number of singletons. Although in some cases ORGANIC does have higher number of singletons or clusters.

Table 3. Key Chemical Statistics for the Generated Molecules by RANC and ORGANIC on the CD and Drugs Data Sets

	CD		Drugs	
data set approach	RANC	ORGANIC	RANC	ORGANIC
size	20000	20000	20000	20000
diversity	0.86	0.89	0.85	0.86
fragments[a]	256000	428000	341000	324000
clusters	704	559	434	697
singletons	12797	14202	16330	11340
heterocycles[b]	2286	3805	1277	1215
cluster size[c]	10.2	10.4	8.5	12.4
MCF (%)	7.1	34.7	23.7	39.6

[a]Number of unique fragments considered to be diverse. [b]Number of unique heterocycles. [c]Average cluster size. Percentage values are taken over entire data set. Shaded values represent best values for parameters that are directly comparable.

Next, we have studied whenever or not, common diversity and heterocyclic diversity of the generated molecules by both models are comparable (see SI Figure 2). The slope and distribution of the structures, for example, in SI Figure 2a are different from that depicted in 2b. Thus, for 10 000 structures (50%) the diversity coefficients are 0.893 and 0.926, respectively (dotted line). The solid line box highlights the number of compounds with a half of diversity scale. Considering the presented distribution profile, we conclude that, in general, structures generated by ORGANIC using the CD data set have higher diversity in structure in contrast to RANC. However, we also speculate that this high diversity is mainly achieved by compounds which have not passed MCFs (6952 compounds, 34.7%). In the case of the Drugs data set, a more similar distribution is observed.

Additionally we have investigated the distribution of the generated structures by similarity to the compounds from the training data sets (see Figure 5 and SI Figure 3). ORGANIC has generated less similar virtual compounds in contrast to RANC. Thus, only 0.2% and 0.3% of the structures have similarity index more than 0.7 for both CD and Drugs training sets. The ORGANIC model has generated about 25% and 15% of structures matching the similarity of 0.5−0.7% for the CD and Drugs data sets. In contrast, with the same similarity index range, the RANC model has generated 45% and 48% of structures, respectively. The number of structures produced by RANC with a similarity index of 0.5−0.7 which apparently exceeds the ORGANIC results. It should be noted that the structures from this range are especially interesting in the case of a target-specific generation. On the one hand, compounds with moderate similarity index retain some crucial structural patterns of known molecules (privileged scaffolds and their isosteres); in other words, they are sufficiently different from the reference compounds.

Moreover, Figures 5 and SI Figure 3 show the quality of the generation in orthodox medicinal chemistry terms. The number of structures which did not pass MCFs are also highlighted. Obviously, the RANC model has generated virtual structures which have more appropriate substructures for drug discovery. Although the cheminformatics parameters listed in Table 3 do not allow us to clearly define strengths and weaknesses of the models, a comparative analysis of the similarity distributions (see Figure 5 and SI Figure 3) has revealed key differences described above. It is also worth mentioning that the profiles of similarity distribution for ORGANIC remain almost the same if considering two distinguished training sets, while profiles for RANC are rather different. This is clearly demonstrated by the number of structures with a similarity index range of 0.7−1. It can be explained by differences in the training sets. For instance, in the case of RANC the average length of produced SMILES is 46 and 40 symbols for the CD and Drugs data sets, respectively, while for ORGANIC it is the same (see Table 2). Likewise, another representative example is the percentage of valid and unique SMILES (see Table 2). Finally, it can be speculated, that the RANC model is more sensitive to the differences in training sets in comparison to ORGANIC in terms of diversity.

Additionally, we also looked at QED (Quantitative Estimate of Drug-likeliness), as an alternative metric for drug-like compounds. QED is a measure ranging from 0 to 1 that captures the abstract notion of aesthetics in medicinal chemistry,[46] penalizing molecular patterns commonly found in common drugs and penalizing patterns that lead to toxicity or difficulty in synthesizability. We computed QED scores for the molecules from both models and analyzed their distributions, as can be observed in Figure 6. On the CD set, we can see that both models are able to improve on the initial distribution of the data and get comparable performance. The starkest difference lies in the Drugs set on which RANC was able to emulate the same distribution of QED scores with higher mass around the larger values, while ORGANIC has a visibly lower-value shift in its distribution. Overall RANC demonstrates a larger count of molecules with high QED scores in both settings.

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

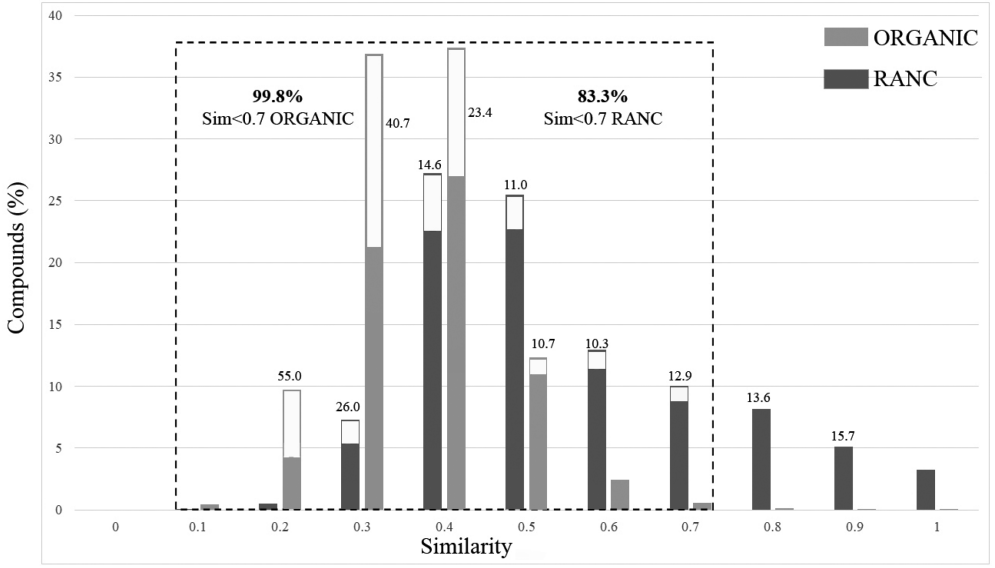


Figure 5. Distribution of the generated structures by similarity to the compounds included in the CD training data set; the colorless area in bars corresponds to compounds which did not pass MCFs.

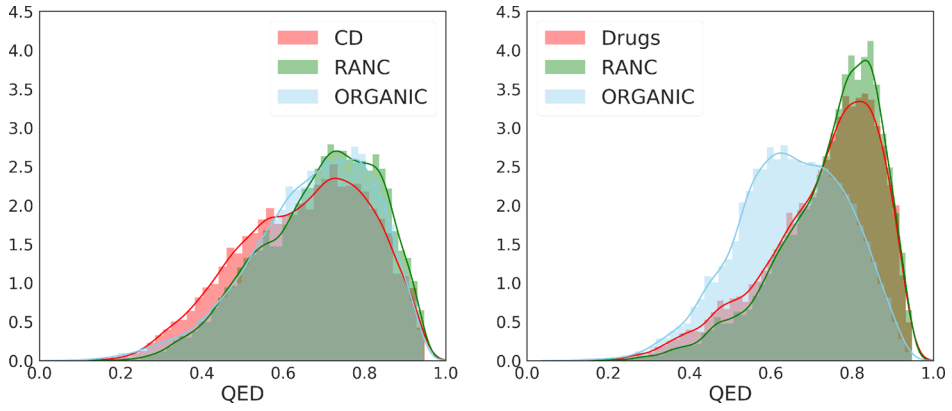


Figure 6. QED scores of the generated molecules with respect to the training data sets.

The representative examples of the structures generated by the models on the CD and Drugs data sets are shown in Figures 7 and 8, respectively.

As clearly shown in Figures 7 and 8, the structures generated by RANC (**1a**−**10a** and **11a**−**20a**) are more complex and contain very attractive moieties, e.g. sp3-rich fragments, spiro-, and PPI frames. These include: annelated benzodiazepinone (**12a**) which can be attributed, for instance, to β-turn mimetics category, spiro-dipyrrolidine equipped by Trp-like side-chain (**18a**), or 2,3-dihydroindole (**16a**) which can be classified as α-helix mimetic, as well as compounds with basic nitrogen (pH = 7.4) and aromatic moiety (e.g., **1a**−**3a**, **17a**, **19a**) which can be regarded as potential CNS-active molecules, regulators of ion channel activity, or biogenic amine receptors binders, like androgen or serotonin ligands. Anyhow, we, therefore, speculate that the RANC model produces pretty structures quite suitable for organic synthesis and subsequent biological evaluation.

CONCLUSION

In this work, we introduced RANC architecture for *de novo* molecular design. Utilizing a DNC controller as a generator model helps to balance the generator and discriminator during adversarial training. Our approach improves over ORGANIC, showing very efficient and stable learning using SMILES strings without any significant losses in the length of the generated structures meanwhile taking into account the distributions of the underlying chemical features of the training molecules. In addition, RANC excels at generating unique and adequate structures.

An extensive analysis of relevant metrics to DDD has revealed that many structures generated by RANC successfully passed MCFs and include a considerable number of unique heterocycles and privileged substructures. These parameters are vital for modern DDD process, as they assist a medicinal chemist to select the most attractive molecules and find out if they contain toxic or undesirable fragments, and what about their solubility in the water or DMSO, etc. Therefore, RANC can be used as a starting point to discover novel small-molecule organic molecules with potential biological activity.

There are many future research directions to improve our framework and tackling DDD challenges in general. One is integrating more complex reward functions in a multiobjective constrained optimization fashion. Another is exploring additional metrics such as natural product-likeliness, synthesizability, toxicity, or PAINS scores. There are some better

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

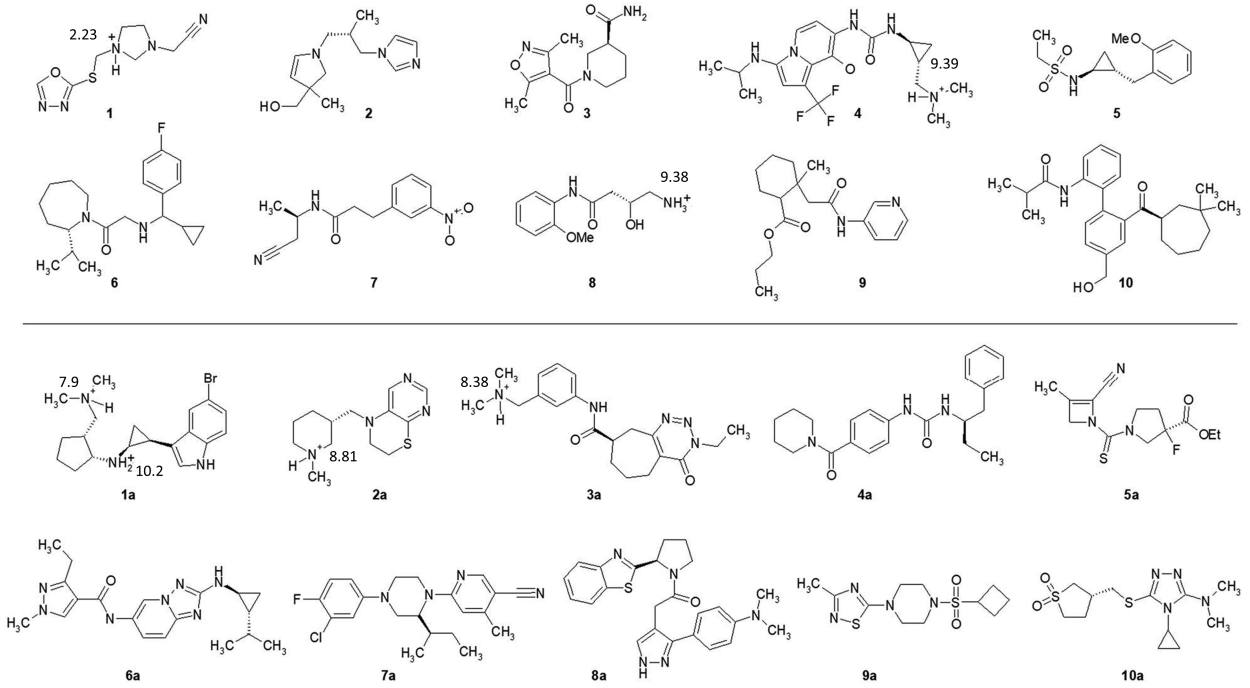


Figure 7. Structures generated using Drugs data set with ORGANIC model (structures **1**−**10**) and RANC model (structures **1a**−**10a**).

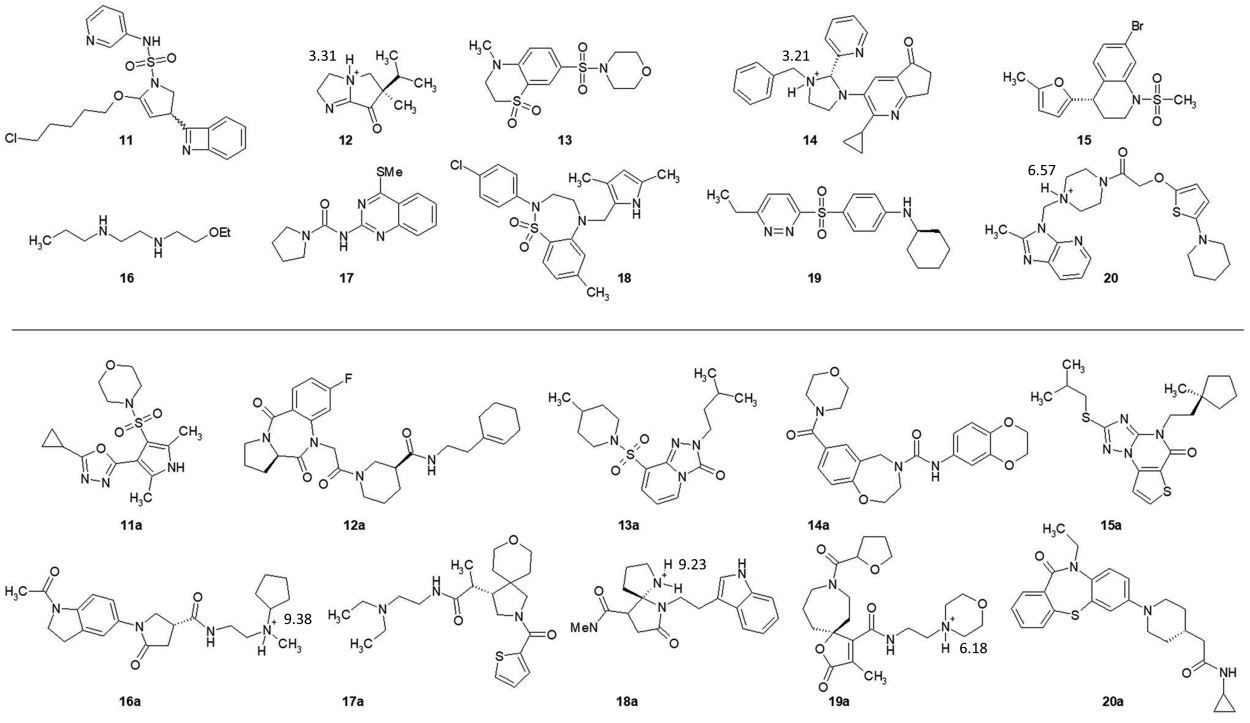


Figure 8. Structures generated using CD data set with ORGANIC model (structures **11**−**20**) and RANC model (structures **11a**−**20a**).

methods in understanding and stabilizing convergence of adversarial training; some work already lies in Wasserstein distances and actor-critic methods. New RL paradigms such as evolutionary strategies might allow increasing our capacity and specificity in the optimization of properties. From a medicinal chemistry perspective, it would be very interesting to investigate how the results of modeling depend on the carefully selected reference compounds as well as the application of the model for the generation of structures with biological activity toward specific targets.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.7b00690.

Figures on SMILES average string length, RO5 and Muegge metrics, diversity plots, and analysis of similarity (PDF)

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

AUTHOR INFORMATION

Corresponding Authors

*E-mail: putin@insilicomedicine.com (E.P.).
*E-mail: alex@insilicomedicine.com (A.Z.).

ORCID

Evgeny Putin: 0000-0002-3012-9708
Alán Aspuru-Guzik: 0000-0002-8277-4434
Alex Zhavoronkov: 0000-0001-7067-8966

Notes

The authors declare the following competing financial interest(s): E.P., A.A., Y.I., V.A., and A.Z. are associated with the company, Insilico Medicine, Inc, engaged in drug discovery and aging research.

ACKNOWLEDGMENTS

The authors would like to acknowledge Government of the Russian Federation (grant 074-U01), Ministry of Education and Science of the Russian Federation, government grant 20.9907.2017/VU (for valuable expert opinion, discussion and manuscript preparation), and Russian Science Foundation 17-74-30012, IBG RAS Ufa (for training sets preparation and analysis). B.S.-L. and A.A.-G. thank Tata Steel and Dr. Anders Froseth for support.

ABBREVIATIONS

AI, artificial intelligence; CD, ChemDiv data set; CNN, convolutional neural network; DDD, drug design and development; DNC, differential neural computer; DNN, deep neural network; Drugs, ZINC data set; GAN, generative adversarial network; HTS, high-throughput screening; LSTM, long short-term memory; MCF, medicinal chemistry filter; MLE, maximum likelihood estimation; ML, machine learning; ORGANIC, Objective-Reinforced GAN for Inverse-design Chemistry; QED, Quantitative Estimate of Drug-likeliness; RANC, Reinforced Adversarial Neural Computer; RL, reinforcement learning; RNN, recurrent neural network; RO5, Lipinski's Rule of Five; SMILES, Simplified Molecular Input Line Entry System

REFERENCES

(1) Holenz, J.; Mannhold, R.; Kubinyi, H.; Folkers, G. *Lead Generation, Vol. 68: Methods and Strategies*; John Wiley & Sons, 2016; Vol. 2.
(2) Zhang, R.; Xie, X. Tools for GPCR Drug Discovery. *Acta Pharmacol. Sin.* **2012**, *33*, 372−384.
(3) Fradera, X.; Babaoglu, K. Overview of Methods and Strategies for Conducting Virtual Small Molecule Screening. *Current Protocols in Chemical Biology* **2017**, 196−212.
(4) Yu, H.-b.; Li, M.; Wang, W.-p.; Wang, X.-l. High Throughput Screening Technologies for Ion Channels. *Acta Pharmacol. Sin.* **2016**, *37*, 34−43.
(5) Mannhold, R.; Kubinyi, H.; Folkers, G. *Drug Metabolism Prediction*; John Wiley & Sons, 2014; Vol. 63.
(6) Balakin, K. V. *Pharmaceutical Data Mining: Approaches and Applications for Drug Discovery*; John Wiley & Sons, 2009; Vol. 6.
(7) Balakin, K.; Ivanenkov, Y.; Savchuk, N. Compound Library Design for Target Families. *Methods Mol. Biol.* **2009**, *575*, 21−46.
(8) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *Journal of health economics* **2016**, *47*, 20−33.
(9) Ivanenkov, Y. A.; Aladinskiy, V. A.; Bushkov, N. A.; Ayginin, A. A.; Majouga, A. G.; Ivachtchenko, A. V. Small-Molecule Inhibitors of Hepatitis C Virus (HCV) Non-Structural Protein 5A (NS5A): A Patent Review (2010−2015). *Expert Opin. Ther. Pat.* **2017**, *27*, 401−414.
(10) Schneider, G.; Fechner, U. Computer-Based De Novo Design of Drug-Like Molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649.
(11) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436−444.
(12) Mamoshina, P.; Vieira, A.; Putin, E.; Zhavoronkov, A. Applications of Deep Learning in Biomedicine. *Mol. Pharmaceutics* **2016**, *13*, 1445−1454.
(13) Min, S.; Lee, B.; Yoon, S. Deep Learning in Bioinformatics. *Briefings Bioinf.* **2016**, bbw068.
(14) Pastur-Romay, L. A.; Cedrón, F.; Pazos, A.; Porto-Pazos, A. B. Deep Artificial Neural Networks and Neuromorphic Chips for Big Data Analysis: Pharmaceutical and Bioinformatics Applications. *Int. J. Mol. Sci.* **2016**, *17*, 1313.
(15) Zhang, L.; Tan, J.; Han, D.; Zhu, H. From Machine Learning to Deep Learning: Progress in Machine Intelligence for Rational Drug Discovery. *Drug Discovery Today* **2017**, *22*, 1680.
(16) Gawehn, E.; Hiss, J. A.; Schneider, G. Deep Learning in Drug Discovery. *Mol. Inf.* **2016**, *35*, 3−14.
(17) Gupta, A.; Müller, A. T.; Huisman, B. J.; Fuchs, J. A.; Schneider, P.; Schneider, G. Generative Recurrent Networks for De Novo Drug Design. *Mol. Inf.* **2018**, *37*, 1700111.
(18) Yuan, W.; Jiang, D.; Nambiar, D. K.; Liew, L. P.; Hay, M. P.; Bloomstein, J.; Lu, P.; Turner, B.; Le, Q.-T.; Tibshirani, R.; Khatri, P.; Moloney, M. G.; Koong, A. Chemical Space Mimicry for Drug Discovery. *J. Chem. Inf. Model.* **2017**, *57*, 875−882.
(19) Korotcov, A.; Tkachenko, V.; Russo, D. P.; Ekins, S. Comparison of Deep Learning With Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Datasets. *Mol. Pharmaceutics* **2017**, *14*, 4462.
(20) Aliper, A.; Plis, S.; Artemov, A.; Ulloa, A.; Mamoshina, P.; Zhavoronkov, A. Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. *Mol. Pharmaceutics* **2016**, *13*, 2524−2530.
(21) Bjerrum, E. J.; Threlfall, R. Molecular Generation With Recurrent Neural Networks (RNNs). *arXiv.org* **2017**, No. 1705.04612v2.
(22) Gómez-Bombarelli, R.; Duvenaud, D.; Hernández-Lobato, J. M.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *arXiv.org* **2016**, No. 1610.02415.
(23) Guimaraes, G. L.; Sanchez-Lengeling, B.; Farias, P. L. C.; Aspuru-Guzik, A. Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models. *arXiv.org* **2017**, No. 1705.10843.
(24) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular De-Novo Design Through Deep Reinforcement Learning. *J. Cheminf.* **2017**, DOI: 10.1186/s13321-017-0235-x.
(25) Sanchez-Lengeling, B.; Outeiral, C.; L, G.; Aspuru-Guzik, A. Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC). *ChemRxiv.org* **2017**, 5309668.
(26) Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Proc. Syst.* **2014**, 2672−2680.
(27) Arjovsky, M.; Bottou, L. Towards Principled Methods for Training Generative Adversarial Networks. *arXiv.org* **2017**, No. 1701.04862.
(28) Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein gan. *arXiv.org* **2017**, No. 1701.07875.
(29) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural computation* **1997**, *9*, 1735−1780.
(30) Graves, A.; Wayne, G.; Reynolds, M.; Harley, T.; Danihelka, I.; Grabska-Barwińska, A.; Colmenarejo, S. G.; Grefenstette, E.; Ramalho, T.; Agapiou, J.; Puigdomémech Badia, A.; Moritz Hermann, K.; Zwols, Y.; Ostrovski, G.; Cain, A.; King, H.; Summerfield, C.; Blunsom, P.; Kavukcuoglu, K.; Hassabis, D. Hybrid Computing Using a Neural

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204

Network With Dynamic External Memory. *Nature* **2016**, *538*, 471−476.
(31) Popova, M.; Isayev, O.; Tropsha, A. Deep Reinforcement Learning for De-Novo Drug Design. *arXiv.org* **2017**, No. 1711.10907.
(32) Graves, A.; Wayne, G.; Danihelka, I. Neural Turing Machines. *arXiv.org* **2014**, No. 1410.5401.
(33) Weston, J.; Bordes, A.; Chopra, S.; Rush, A. M.; van Merriënboer, B.; Joulin, A.; Mikolov, T. Towards Ai-Complete Question Answering: A Set of Prerequisite Toy Tasks. *arXiv.org* **2015**, No. 1502.05698.
(34) Yu, L.; Zhang, W.; Wang, J.; Yu, Y. SeqGAN: Sequence Generative Adversarial Nets With Policy Gradient. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*; 2017; pp 2852−2858.
(35) Sutton, R. S.; McAllester, D. A.; Singh, S. P.; Mansour, Y. Policy Gradient Methods for Reinforcement Learning With Function Approximation. *Adv. Neural Inf. Process. Syst.* **2000**, 1057−1063.
(36) Zhang, X.; LeCun, Y. Text Understanding From Scratch. *arXiv.org* **2015**, No. 1502.01710.
(37) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757−1768.
(38) Trepalin, S. V.; Gerasimenko, V. A.; Kozyukov, A. V.; Savchuk, N. P.; Ivaschenko, A. A. New diversity calculations algorithms used for compound selection. *Journal of chemical information and computer sciences* **2002**, *42*, 249−258.
(39) Mulliner, D.; Wondrousch, D.; Schüürmann, G. Predicting Michael-Acceptor Reactivity and Toxicity Through Quantum Chemical Transition-State Calculations. *Org. Biomol. Chem.* **2011**, *9*, 8400−8412.
(40) Kalgutkar, A. S.; Gardner, I.; Obach, R. S.; Shaffer, C. L.; Callegari, E.; Henne, K. R.; Mutlib, A. E.; Dalvie, D. K.; Lee, J. S.; Nakai, Y.; O'Donnell, J.; Boer, J.; Harriman, S. A Comprehensive Listing of Bioactivation Pathways of Organic Functional Groups. *Curr. Drug Metab.* **2005**, *6*, 161−225.
(41) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(42) Zhang, B.; Vogt, M.; Maggiora, G. M.; Bajorath, J. Design of Chemical Space Networks Using a Tanimoto Similarity Variant Based Upon Maximum Common Substructures. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 937−950.
(43) Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; Kudlur, M.; Levenberg, j.; Monga, R.; Moore, S.; G Murray, D.; Steiner, B.; Tucker, P.; Vasudevan, V.; Warden, P.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: A System for Large-Scale Machine Learning. *arxiv.org* **2016**, No. 1605.08695.
(44) Landrum, G. *RDKit: Open-Source Cheminformatics*. http://www.rdkit.org (accessed 2012).
(45) Muegge, I. Selection Criteria for Drug-Like Compounds. *Med. Res. Rev.* **2003**, *23*, 302−321.
(46) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012**, *4*, 90.

DOI: 10.1021/acs.jcim.7b00690
J. Chem. Inf. Model. 2018, 58, 1194−1204