Deep Generative Models for Drug Design and Response

Karina Zadorozhny and Lada Nuzhna

Northwestern University

June 2021

Abstract

Designing new chemical compounds with desired pharmaceutical properties is a challenging task and takes years of development and testing. Still, a majority of new drugs fail to prove efficient *in vivo*. Recent success of deep generative modeling holds promises of generation and optimization of new molecules. In this review paper, we provide an overview of the current generative models, and describe necessary biological and chemical terminology, including molecular representations needed to understand the field of drug design and drug response. We present commonly used chemical and biological databases, and tools for generative modeling. Finally, we summarize the current state of generative modeling for drug design and drug response prediction, highlighting the state-of-art approaches and limitations the field is currently facing.

1 Introduction

The path from initial screening to marketable drugs takes around 20 years [1], with cost being anywhere from $0.5 to $2.6 billion dollars [2]. Given the remarkable impediments associated with drug design, it is necessary to develop more precise strategies to propose drug candidates. Computational chemistry approaches to drug design, often called computer-aided drug design, received a lot of attention in the past 30 years. The growing availability of high-throughput screening (HTS) data holds a great potential to accelerate the development of efficient new compounds.

Pharmacologically active drug compounds act by binding to their target (which is most commonly a protein) and altering their function by either stimulating or inhibiting it. Broadly speaking, the current drug design approaches can be divided into two categories. The first one is ligand-based (or indirect) approaches that rely on existing molecules known to be a binding partner of some target and try to mimic their properties. These approaches include molecular fingerprint similarity search, pharmacophore modeling, or quantitative structure-activity relationship (QSAR). The second approach is structure-based (or direct) where a 3D structure of a desired target (or a target that is highly similar to the desired one) is known. New molecules are then designed to bind to the target based on its structural properties. The methods that fall in this category include molecular docking or molecular dynamic simulations.

Deep generative models (DGM) have attracted a lot of attention in recent years. Given their success in image and text generation [3, 4], there is an ongoing effort to apply these approaches to the medical field as well. Current applications include medical image analysis, data augmentation, generation of new DNA, RNA, or protein sequences with specific properties and missing regions in known protein structures [5].

Despite the optimism, designing novel drugs is particularly difficult due to the graphical structure of molecules, chemical validity, and feasibility of synthesis of generated molecules. Several major simplifications have been used so far to reduce the complexity of the task. For example, instead of working with and generating full molecular structures, molecules can be described by their properties in binary strings. However, to fully automate the drug design process, it is necessary to incorporate complex information, such as chirality and structural conformations, into the molecular representations.

Strikingly, 75-97% of novel drugs fail human clinical trials. The primary reason for failing a trial is a lack of demonstrated efficiency and the second most common reason is safety concerns [6]. Therefore, optimizing for efficient physiological response and low toxicity in addition to favorable chemical properties is at the core of the approaches that could lead to successful drug

design. This is encompassed in the field of drug response prediction.

In this review article, we describe the current DGM approaches for drug design with the aim of bridging the fields of drug design and drug response prediction to allow for designing more efficient and more powerful pipelines. We first briefly introduce the most common neural architectures that are used for drug design and drug response prediction. We also list the most common databases and tools, and compare current molecular data representations. We describe state-of-the-art techniques for small molecule design, gene therapy, and protein sequence generation. Finally, we conclude our review with remarks on the most challenging aspects of drug design and considerations about the direction of future research.

2 Overview of generative models

Deep generative models have achieved great success ranging from generation of high-quality images to large corpora of meaningful text. Formally, deep generative models are neural networks trained to model a distribution of training samples and are able to sample from the learned distribution to generate new samples. Explicit generative models specify the full data distribution and allow to express likelihoods that a given sample comes from the distribution. On the other hand, implicit generative models only specify a procedure of drawing samples to generate new data but do not provide a way to express a full distribution.

2.1 GANs

Generative Adversarial Networks (GANs) are implicit density models introduced in 2014 [7]. The idea behind generating samples with GANs is to have two networks - one is a generator model that creates new samples from a random vector and the other is the discriminator (or the critic) that aims to distinguish newly generated samples from the real ones. These two parts are competing with each other. The generator gets better at generating realistically looking samples while the discriminator gets better at classifying those that are real and those that are fake. After both networks have been trained, only the generator is used to create realistic inputs from a random starter vector.

2.2 Autoencoders

A standard autoencoder (AE) has two parts - an encoder and a decoder. The inputs are passed through an encoder that outputs some representation of these inputs to the latent space. From there, the decoder can reconstruct the representation back such that it is as close as possible to the original input. Most commonly, the latent space is chosen such that it has fewer dimensions than the inputs. This leads to compression of the samples and extraction of the most useful information. The encoded latent space, which contains all the information that is necessary to retrieve the original input, can then be used for downstream tasks. For example, a classifier can be fitted on top of a learned latent space. Compared to Principal Component Analysis (PCA), autoencoders can perform nonlinear dimensionality reduction. Autoencoders can also be used for denoising where a noisy input is reconstructed without the unwanted noise.

2.3 VAEs

Variational Autoencoders (VAEs) are regularized probabilistic autoencoders [8]. Similar to a standard autoencoder, they contain an encoder and a decoder. However, instead of encoding each sample to some point a latent space, the encoder defines a whole distribution. From this distribution, vectors are being sampled and fed into a decoder that reconstructs them back to be as close as possible to the original input. To give the latent space distribution some specific properties, it is regularized during training. Most commonly, we force the latent space to follow a normal distribution since that provides a closed form for many equations and is simple to work with. The loss function for VAEs thus contains two terms – a reconstruction error and a regularization term which is a Kullback-Leibler divergence between the learned latent distribution and a specified prior (normal distribution).

2.4 AAEs

Adversarial Autoencoders (AAEs) are regularized autoencoders that adapt GANs training technique [9]. Similar to VAEs, they impose certain properties on the learned latent space such as being normally distributed but they use a different regularization technique. Apart from an encoder and a decoder, there is an auxiliary adversary network which is trained to distinguish samples from the latent space and samples drawn from a chosen distribution. The encoder tries to fool the adversary by producing samples that are indistinguishable from the chosen dis-
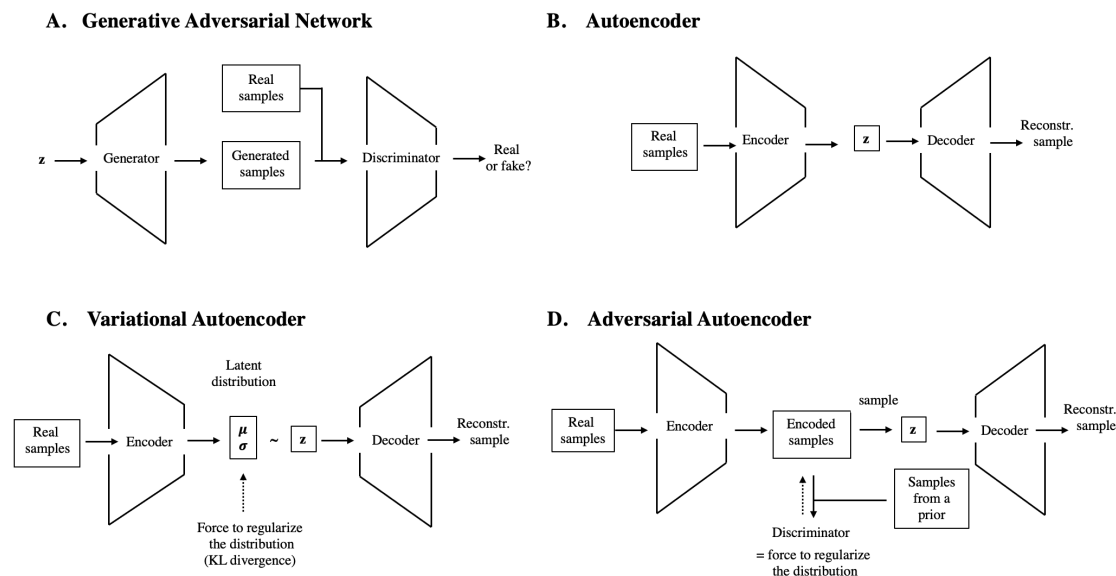
Figure 1: Generative Models. A. Generative Adversarial Network (GAN). B. Autoencoder. C. Variational Autoencoder (VAE). D. Adversarial Autoencoder (AAE).

tribution and this serves as regularization of the latent space. One of the benefits of the AAEs compared to VAEs is that the chosen prior distribution that the encoder tries to mimic can be any distribution, even a non-parametric one. Depending on the choice of the encoder, the AAE can be either a deterministic model that maps all the inputs to the same space in the latent space or have a probabilistic encoder that outputs parameters for a full latent posterior distribution.
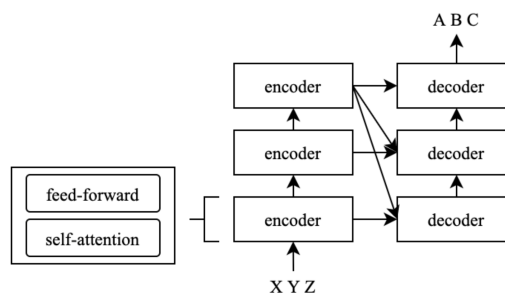
2.5 Transformers

Transformer is an attention-based architecture that gained popularity in recent years due to its successful use in language models [10, 11]. Transformer has two functional parts: a block of stacked encoders and a block of decoders (Figure 2). Encoding part produces a representation of the input that tells us how different parts of the input sequence are related to each other. "Relevancy" is determined by the attention formula. This approach allows us to establish connections between parts of the sequence that can be far from each other. This makes it attractive for use in protein structure prediction, where functional groups at different sides of the sequence can interact with each other, dictating the 3D conformation of the sequence.

Figure 2: Transformer architecture.

3 Databases and tools

In this section, we highlight some of the most useful databases and tools for drug design and drug response prediction.

3.1 Databases of Chemical Compounds

Table ?? shows the most commonly used databases of binding affinities, bioactivity, and 3D structures of chemical compounds.

3.2 Databases of Cell Line Screens

In Table ??, we present open source databases of drug response (sensitivity) data on cancer cell lines as well as gene expression data collected using high-throughput screenings for cells and tissues.

3.3 Tools

QSAR is a computational method for modeling relationships between structural properties of chemical compounds and their biological properties (such as enzyme activity, minimum effective dose, toxicity). This can often be used to validate the results achieved by deep nets and for benchmarking. However, QSAR is sometimes at risk of giving inaccurate predictions.

RDKit open-source chemoinformatics library (available from http://www.rdkit.org) can be used for a number of 2D and 3D molecular operations, from stereochemistry identification to detecting chemistry problems.

4 Data Representations

Given the diversity of problems and screenings that drug discovery can cover, data representations also differ significantly. We present some of the most common ones. For a detailed review of the molecular representations, see David et al. [12]. We stress that there exist molecules that cannot be described using a standard notation.

4.1 SMILES and Its Variants

- Simplified Molecular Input Line Entry System (SMILES). It is used for representing molecules as strings. To date, it is the most popular string notation for molecules. The strings are created by traversing a molecular graph in a specific order (RDKit uses a depth-first search algorithm). SMILES strings are not unique and one molecule can have several representations. Therefore, a canonical string is often chosen for a given molecule.

- SMILES Arbitrary Target Specification (SMARTS). This representations is an extension of SMILES that allows to encode more information about bond types, connectivity, and atomic properties in a string.

- SMIRKS. It is a hybrid between SMILES and SMARTS that allows to express reactants and products of a generic chemical reaction.

4.2 Molecular Fingerprints

- Binary encoding. It is the simplest fingerprint that indicates presence of absence of individual substructures.

- Circular Fingerprints. Circular fingerprint considers connectivity of atoms up to a certain distance to solve the problem of

a) Formula: $C_2H_5NO_2$ b) SMILES   c) amino acid encoding

[$([$([NX3H2,NX4H3+]),$([NX3H](C)(C))][CX4H2][CX3](=[OX1])[OX2H,OX1-,N])]
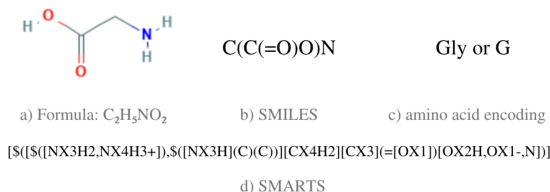
d) SMARTS

Figure 3: Example of molecular representations. Amino acid glycine represented in different formats.

non-uniqueness of representations. Every atom receives a descriptor that indicates its closest neighbors. Examples of circular fingerprints include extended-connectivity fingerprints (ECFPs) such as ECFP4 or ECFP6 where the number indicates the radius around an atom to be considered. Other examples are Multilevel Neighborhoods of Atoms (MNA) or MolPrint2D Format (MPD).

- Spectrophores. A fingerprint that describes molecules' surface properties. Molecules that have similar shapes and 3D properties will retains a similar representations.

4.3 Peptide and Protein Representations

- Amino Acid Sequences. This is the simplest way of representing peptides of proteins. Individual amino acids which are building blocks of peptides are assigned a letter and the final peptide is represented a string.

- Distance and Contact Maps for protein structures. Protein contact maps represent distances between all pairs of amino acids in a 3D protein structure in a matrix.

4.4 Graph Representations

- Adjacency Matrix. To represent a presence of a bond in a graph, a standard adjacency matrix.

- Node Feature Matrix. A larger matrix can be used that indicates atom types, formal charges, or other properties of each atom.

- Edge Feature Matrix. Each row represents on edge in a graph and is a binary vector that indicates properties of that such (such as the type of a bond).

5 Drug Design

5.1 Small Molecule Design

With many variations of generative architectures published in recent years, there are a number of ways to enforce properties onto generated structures. Desired drug features, such as growth inhibition of tumors (or GI50), can be used as input to encoder in Conditional Autoencoders [13]. Some work was also done to frame drug design as a 'translation' problem, applying transformer architecture for converting amino-acid sequence to SMILES representation of the molecule that can bind it [14]. However, even with faster screening methods, docking validation still remains a bottleneck in this type of candidate generation. Moreover, when mapping discrete representations into continuous space, "grammar" of chemical syntax is hard to preserve. This can lead to unrealistic aromatic systems and incorrect atomic valences. Because of this, SMILES-generated molecules often require additional "correcting" architectures.

One possible approach for correction is sequence-to-sequence learning with attention, where the input is a wrong sequence and the target is the corrected one [15]. Grammar VAE (GVAE) [16] was also used to generate SMILES that follow syntactic constraints given by a context-free grammar. Finally, syntax-directed VAE (SDVAE) [17]) can make use of attribute grammar to enforce syntactic and semantic constraints on generated SMILES.

While focusing on SMILES strings or molecular fingerprints simplifies generation of new molecules, a lot of information is lost using these approaches. For instance, chemically similar molecules can have very different SMILES strings (Figure 4. This representation prevents learning meaningful smooth embedding.
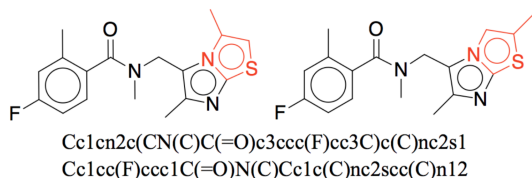

Cc1cn2c(CN(C)C(=O)c3ccc(F)cc3C)c(C)nc2s1
Cc1cc(F)ccc1C(=O)N(C)Cc1c(C)nc2scc(C)n12

Figure 4: SMILES strings for these two very similar molecules show only 50.5% similarity. Adapted from [18] .

Graph neural networks that capture graph structures by message passing between the nodes are attractive methods for generation of full molecular structures [19]. Given that graphs are discrete structures with arbitrary connectivity, graph generation is a difficult domain and rela-tively young domain of generative modeling.

One approach is to generate graph nodes and edges one by one. This was done using an LSTM-based autoregressive model [20]. Simonovsky and Komodakis, [21] used GraphVAE to directly generate nodes (atoms) and predict an adjacency matrix indicating edges (bonds) of a graph. Another approach, which enhances chemical validity of generated molecular graphs, is to use smaller subgraphs as building blocks of a final molecular graph. Jin et al., [18] used a model called Junction Tree VAE (JT-VAE), which first generated a junction tree that serves as a scaffold for the molecule and then specifies the subgraphs as nodes of the junction tree. Their method of coarse-to-fine generation of molecules significantly outperformed previous SMILES-based approaches such as GrammarVAE [16], Character VAE [22], or graph generating GraphVAE in validity of generated molecules. Another model , DeLinker, which is a gated graph neural network (GGNN) that links two molecular fragments together to form more complex compounds [23] also incorporates 3D structural information to constrain the linking process. However, DeLinker's performance was not compared to previous methods.

5.2 Gene Therapy

Generative models have a particular promise for the area of gene therapy, where adeno-associated virus (AAV) capsids serve as vectors for gene delivery. However, due to wide-spread immunity against common AAVs, capsids often need to be diversified such that antigens on their surface aren't recognized by the human immune system. Previous approaches, like directed evolution or DNA shuffling between serotypes, often produced unstable results (Bryant et al., 2021; Ogden et al., 2019). Earlier this year Dyno Therapeutics used variational autoencoders to create novel and stable adeno-associated viral capsids [24]. This case also emphasizes that, compared to supervised methods, unsupervised generation can potentially take advantage of evolutionary information.

5.3 Proteins and Peptides

Peptides are short sequences of amino acids that can be used as therapeutics. There are over 60 peptides approved for clinical use by the FDA. While many naturally-occurring peptides are approved for medical use, they often end up having some weaknesses, such as low membrane permeability, chemical instability and short half-life. Synthetic peptides generated by computa-

tional methods can overcome these weaknesses [25].

Das et al. [26], compared usage of Wasserstein Autoencoder (use Mean Maximum Discrepancy as a regularization in their loss) and Beta Variational Autoencoders (Kullback-Leibler divergence as regularization) for the screening of peptides with antimicrobial properties. WAE managed to learn meaningful latent representation that captured correlation between evolutionary similarity and distance in the z-space (Beta Variational Autoencoders didn't have the same correlation in their latent space but showed comparatively good BLEU score, reconstruction error and perplexity). VAEs can also be used for generation of functional variants of peptides: Hawkins-Hooker et al., 2021, showed evidence of conditional VAEs being able to increase solubility of luxA sequence from P. luminescens without losing core properties of the wild-type. Given the complexity of de novo design, diversification of already functional peptides can allow for a more reliable design of therapeutics.

Most drugs target receptors and channels (both are classes of proteins) on the surface of the cell or soluble (not part of a membrane) proteins. Given this, knowing the exact conformation of the target can be of much help when designing ligands. Moreover, proteins can serve as a therapeutics themselves. A huge milestone in deriving 3D structure from an amino acid sequence was achieved by DeepMind in protein structure annual competition, Critical Assessment of Protein Structure Prediction (CASP) [27]. While their earlier architecture in CASP13, AlphaFold-1, relied on ResNet-like architecture to predict amino acid residue distances [28], a major breakthrough was achieved with AlphaFold-2 in CASP14, when the team transitioned to transformer-like architecture [27]. Previously, to predict the structure of a protein, one would have to undergo expensive and laborious X-ray crystallography or cryo-electron microscopy. Combination of generative methods with the knowledge of precise target shape can profoundly cut time for drug screening (Callaway, Ewen, 2020).

6 Drug response

Apart from designing novel pharmaceutically active compounds, another important task is to be able to predict their efficiency. Ultimately, incorporating drug response directly as one of the optimization goals for generation of new molecules can lead to a more powerful drug design pipeline.

Recent research efforts have produced large-scale publicly available drug screening profiles (Table **??**). Most of them contain high-throughput screenings for anti-cancer drugs against cancer cell lines. These cell lines are most commonly immortalized human or animal cells which are grown outside an organism, on a Petri dish. Despite the tremendous value that cell lines provide to our understanding of disease progression and treatment [29] there are important drug response differences compared to full organisms. Therefore, there has been an ongoing effort to transfer findings from cell lines to actual application in humans where the amount of available data on drug responses is very limited.

The input features that are used for drug response prediction on cell lines consist of one or multiple of the following: expression levels or different genes, gene methylation status, and protein abundance. Less common are DNA mutation profiles (single nucleotide variations) and copy number variations. The nature of the problem – expression profiles of hundreds or thousands of genes – creates a high feature to sample ratio that complicates the prediction task. Initial approaches for drug response prediction included linear regression and simple clustering algorithms, SVM, and Random Forests. Traditional feature selection and dimensionality reduction approaches achieved some limited success [30, 31]. While the task itself is not generative in nature, DGM models such as VAEs, conditional VAEs, and AAEs, have been used to perform dimensionality reduction, deconfounding or domain adaptation for better response prediction.

The first study to apply generative models to drug response predictions was described by Dincer et al. [32]. In this work, a VAE model has been used to learn relevant features from gene expression data of cancer cells treated with a drug. Compared to drug response data, there is a vast amount of unlabeled gene expression results which their model leveraged. First, the VAE was trained on gene expression data of AML patients and then, a much smaller sample size for which responses to various drugs were available were passed through the encoder. On the compressed data, a linear classifier was fit to predict a binary target if patients responded to therapy or not. This method improved performance over no feature reduction.

In another approach, a semi-supervised VAE called Dr.VAE divided the task of predicting drug response into two parts – first, it learns how a drug perturb gene expression of cells, and second, it relates gene expression changed to drug sensitivity [33]. This means that instead of using only post-treatment gene expres-

sion data (as done above), the data is considered in pairs of pre- and post-treatment and there is a learnable function that describes this change. Dr.VAE aims to learn a joint distribution of gene expression pairs, latent representations, and a categorical variable that corresponds to the drug response. The marked improvement of this approach over baselines could be potentially even greater if combined with supplying the model with chemical properties of tested drugs. Combining gene expression data with embedded structure of chemical compounds was done by Kuenzi et al. [34] which used a simple feed-forward network.

Generative models have also been used for deconfounding in drug response prediction. As mentioned above, there is often a mismatch between drug efficiency tested in cell lines and in humans. When using predictive models, they can detect cell-line specific patterns. If an autoencoder is applied directly, it will mainly learn to encode these non-relevant signals as they are usually the main source of variation in the dataset. Adversarial Deconfounding Autoencoder (AD-AE) was used to extract only the relevant signal from this data (i.e., expression patterns of cancerous cells independent of their source) [35]. This model is most similar to AAE, but the adversary network tries to predict a confounder, such as sex, biological age or origin of cell, from embeddings. This forces the latent space to contain only confounder-invariant representations. AD-AE was also better at domain adaptation to different datasets which indicates that it learned more meaningful cancer responses to drugs. Given that there is much more gene expression data on cancer cell lines than on tumor cells, AD-AE has also been applied for domain adaptation from cancer cells to tumors [36].

Finally, drug response can serve as a direct optimization target for generating new molecules. While conditional VAE (CVAE) has been used in the past for drug design with specific physicochemical properties (see section Drug Design; [37]), Joo et al., used a conditional VAE used to generate new drug candidates with the desired property of inhibiting cancer cell growth [13]. The model takes in molecular fingerprints and a one-hot vector which indicates the binarized concentration of a drug needed to inhibit 50% of cells in a population (GI50). During inference, a label with desired GI50 value is passed to the decoder along with a random sample from a latent space. Generated molecular fingerprints did indeed resemble similarity to those of FDA-approved drugs for breast cancer indicating that the model learned the relationship between the structural properties encoded in the fingerprints and efficiency of drugs. However, generating a molecular fingerprint is not the same as generating a full molecular structure. In the future, more direct approaches that generate full molecular graphs directly can be used with similar optimization targets.

7 Challenges and Future Directions

With exceptionally high failure rates and a large amount of unlabeled data, drug design and drug response prediction represent attractive areas for applying generative models. However, as emphasized by some authors in the field, reduction in cost and time for drug screening using AI often does not lead to improved decisions [38].

Generation of new pharmaceutically active compounds is challenging for several reasons. A lot of information is lost by using molecular representations. Many of the available data formats (such as SMILES or amino acid strings), when used alone, ignore the complexity of biological systems when trying to predict drug performance. This complexity includes but is not limited to secondary targets, drug metabolites and concentration-dependent effects.

Generating a molecular fingerprint of a drug or a SMILES string is far from generating a full molecular structure. Moreover, *chiral* molecules (molecules that are not identical to their mirror images) often have very different properties. Drug receptors can recognize only one *enantiomer* (one mirror image) but not the other. Sometimes, they can even have opposing properties with one being an inhibitor while the other being an activator of the same receptor [39]. This information is completely lost in SMILES strings. Better molecular representations or modelling of full molecular graphs could therefore substantially advance the field in the upcoming years.

AlphaFold represents an excellent example of how emphasizing relevance of data (protein structure) can help solve the most complex biological problems [28]. In fact, Alphafold's transformer achieved state-of-art results in protein folding with just training 170,000 samples.

Another problem of some studies is a lack of proper benchmarking and evaluation compared to previously published work. Several papers provided only proof-of-the-concept examples of validity of their approaches. Having rigorous, standardized, and well-known comparison standards for drug design, similar to benchmarks for image generation on CIFAR-10 or Im-

ageNet, would enable more transparent evaluation of newly proposed methods. Examples of proposed benchmarks include GuacaMol [40] or MOSES [41].
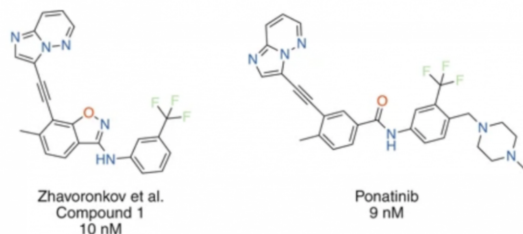

Figure 5: Compound 1 is DDR1 kinase inhibitor produced by the model and Ponatinib is a compound from dataset most similar to it. Adapted from

Furthermore, the field of generation is particularly vulnerable to IP issues. For example, DDR1 kinase inhibitors recently produced by the generative pipeline of Insilico [42] were claimed to be "the AlphaGO moment of drug design". However, later a few researchers pointed out that the drug is almost identical to the one that appeared in the dataset of patented ligands they used, drug Iclusig (compound ponatinib) marketable by ARIAD Pharmaceuticals (Figure 5) [43]. To prevent this from happening, the future research in de novo candidate generation should publish molecules from the training set that are most similar to the generated candidates. This heavy reliance on data and previous research, however, points to one other problem. In recent years, there has been growing evidence on lack of reproducibility in biological research. Recent investigation into oncology drug design showed that two-thirds of published data on drug targets is not reproducible [44]. Given wrong data, even state-of-the-art models won't be able to generate meaningful candidates.

Conclusions

In this paper, we present the current advances in the fields of drug design and drug response prediction using deep generative models. While there is a clear intuition for why these methods can be applied to drug discovery, most of the progress is yet to be made. Only a small fraction of research discussed in this review went beyond proof-of-concept and was used in industry pipelines. Generative models face limitations similar to previous approaches used in drug discovery and drug response prediction, (like QSAR), namely, little understanding of actual mechanism of action and risk of inaccurate pre-

dictions. Even though large databases of unlabeled molecules are attractive for unsupervised generations, to really increase success rate of drug discovery, future research should also make an emphasis on relevancy and quality of data used in the training.

References

[1] Steven M. Paul, Daniel S. Mytelka, Christopher T. Dunwiddie, Charles C. Persinger, Bernard H. Munos, Stacy R. Lindborg, and Aaron L. Schacht. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, 9(3):203–214, March 2010. Number: 3 Publisher: Nature Publishing Group.

[2] Jerry Avorn. The \$2.6 Billion Pill — Methodologic and Policy Considerations. *New England Journal of Medicine*, 372(20):1877–1879, May 2015. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMp1500848.

[3] Ceren Guzel Turhan and H. Bilge. Recent Trends in Deep Generative Models: a Review. September 2018.

[4] Alankrita Aggarwal, Mamta Mittal, and Gopi Battineni. Generative adversarial network: An overview of theory and applications. *International Journal of Information Management*, page 100004, January 2021.

[5] Lan Lan, Lei You, Zeyang Zhang, Zhiwei Fan, Weiling Zhao, Nianyin Zeng, Yidong Chen, and Xiaobo Zhou. Generative Adversarial Networks and Its Applications in Biomedical Informatics. *Frontiers in Public Health*, 8, May 2020.

[6] Chi Heem Wong, Kien Wei Siah, and Andrew W Lo. Estimation of clinical trial success rates and related parameters. *Biostatistics (Oxford, England)*, 20(2):273–286, April 2019.

[7] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, June 2014. arXiv: 1406.2661.

[8] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv:1312.6114 [cs, stat]*, May 2014. arXiv: 1312.6114.

[9] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial Autoencoders. *arXiv:1511.05644 [cs]*, May 2016. arXiv: 1511.05644.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*, May 2019. arXiv: 1810.04805.

[11] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. *arXiv:2005.14165 [cs]*, July 2020. arXiv: 2005.14165.

[12] Laurianne David, Amol Thakkar, Rocío Mercado, and Ola Engkvist. Molecular representations in AI-driven drug discovery: a review and practical guide. *Journal of Cheminformatics*, 12(1):56, September 2020.

[13] Sunghoon Joo, Min Soo Kim, Jaeho Yang, and Jeahyun Park. Generative Model for Proposing Drug Candidates Satisfying Anticancer Properties Using a Conditional Variational Autoencoder. *ACS Omega*, 5(30):18642–18650, August 2020. Publisher: American Chemical Society.

[14] Daria Grechishnikova. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *bioRxiv*, page 863415, June 2020. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[15] Maksym Druchok, Dzvenymyra Yarish, Oleksandr Gurbych, and Mykola Maksymenko. *Towards Efficient Generation, Correction and Properties Control of Unique Drug-like Structures*. October 2019.

[16] Matt J. Kusner, Brooks Paige, and José Miguel Hernández-Lobato. Grammar Variational Autoencoder. *arXiv:1703.01925 [stat]*, March 2017. arXiv: 1703.01925.

[17] Hanjun Dai, Yingtao Tian, Bo Dai, Steven Skiena, and Le Song. Syntax-Directed Variational Autoencoder for Structured Data. *arXiv:1802.08786 [cs]*, February 2018. arXiv: 1802.08786.

[18] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv:1802.04364 [cs, stat]*, March 2019. arXiv: 1802.04364.

[19] Dejun Jiang, Zhenxing Wu, Chang-Yu Hsieh, Guangyong Chen, Ben Liao, Zhe Wang, Chao Shen, Dongsheng Cao, Jian Wu, and Tingjun Hou. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *Journal of Cheminformatics*, 13(1):12, February 2021.

[20] Yujia Li, Oriol Vinyals, Chris Dyer, Razvan Pascanu, and Peter Battaglia. Learning Deep Generative Models of Graphs. *arXiv:1803.03324 [cs, stat]*, March 2018. arXiv: 1803.03324.

[21] Martin Simonovsky and Nikos Komodakis. GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders. *arXiv:1802.03480 [cs]*, February 2018. arXiv: 1802.03480.

[22] Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science*, 4(2):268–276, February 2018. Publisher: American Chemical Society.

[23] Fergus Imrie, Anthony R. Bradley, Mihaela van der Schaar, and Charlotte M. Deane. Deep Generative Models for 3D Linker Design. *Journal of Chemical Information and Modeling*, 60(4):1983–1995, April 2020. Publisher: American Chemical Society.

[24] Sam Sinai, Nina Jain, George M. Church, and Eric D. Kelsic. Generative AAV capsid diversification by latent interpolation. *bioRxiv*, page 2021.04.16.440236, April 2021. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[25] Keld Fosgerau and Torsten Hoffmann. Peptide therapeutics: Current status and future directions. *Drug Discovery Today*, 20, October 2014.

[26] Payel Das, Tom Sercu, Kahini Wadhawan, Inkit Padhi, Sebastian Gehrmann, Flaviu Cipcigan, Vijil Chenthamarakshan, Hendrik Strobelt, Cicero dos Santos, Pin-Yu Chen, Yi Yan Yang, Jeremy P. K. Tan, James Hedrick, Jason Crain, and Aleksandra Mojsilovic. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nature Biomedical Engineering*, pages 1–11, March 2021. Publisher: Nature Publishing Group.

[27] AlphaFold: a solution to a 50-year-old grand challenge in biology.

[28] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, January 2020.

[29] Peppino Mirabelli, Luigi Coppola, and Marco Salvatore. Cancer Cell Lines Are Useful Model Systems for Medical Research. *Cancers*, 11(8), August 2019.

[30] George Adam, Ladislav Rampášek, Zhaleh Safikhani, Petr Smirnov, Benjamin Haibe-Kains, and Anna Goldenberg. Machine learning approaches to drug response prediction: challenges and recent progress. *npj Precision Oncology*, 4(1):1–10, June 2020. Number: 1 Publisher: Nature Publishing Group.

[31] Delora Baptista, Pedro G Ferreira, and Miguel Rocha. Deep learning for drug response prediction in cancer. *Briefings in Bioinformatics*, 22(1):360–379, January 2021.

[32] Ayse Dincer, Safiye Celik, Naozumi Hiranuma, and Su-In Lee. DeepProfile: Deep learning of patient molecular profiles for precision medicine in acute myeloid leukemia. *bioRxiv*, page 278739, March 2018. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[33] Ladislav Rampášek, Daniel Hidru, Petr Smirnov, Benjamin Haibe-Kains, and Anna Goldenberg. Dr.VAE: improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics*, 35(19):3743–3751, October 2019.

[34] Brent M. Kuenzi, Jisoo Park, Samson H. Fong, Kyle S. Sanchez, John Lee, Jason F. Kreisberg, Jianzhu Ma, and Trey Ideker. Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell*, 38(5):672–684.e6, November 2020. Publisher: Elsevier.

[35] Ayse B. Dincer, Joseph D. Janizek, and Su-In Lee. Adversarial Deconfounding Autoencoder for Learning Robust Gene Expression Embeddings. *bioRxiv*, page 2020.04.28.065052, April 2020. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[36] Di He and Lei Xie. CODE-AE: A Coherent De-confounding Autoencoder for Predicting Patient-Specific Drug Response From Cell Line Transcriptomics. *arXiv:2102.00538 [cs, q-bio]*, January 2021. arXiv: 2102.00538.

[37] Jaechang Lim, Seongok Ryu, Jin Woo Kim, and Woo Youn Kim. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *Journal of Cheminformatics*, 10(1):31, July 2018.

[38] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? page 14, 2021.

[39] Richard B. Weiskopf, Carla Nau, and Gary R. Strichartz. Drug Chirality in Anesthesia. *Anesthesiology*, 97(2):497–502, August 2002.

[40] Nathan Brown, Marco Fiscato, Marwin H.S. Segler, and Alain C. Vaucher. GuacaMol: Benchmarking Models for de Novo Molecular Design. *Journal of Chemical Information and Modeling*, 59(3):1096–1108, March 2019. Publisher: American Chemical Society.

[41] Daniil Polykovskiy, Alexander Zhebrak, Benjamin Sanchez, Sergey Golovanov, Oktay Tatanov, Stanislav Belyaev, Rauf Kurbanov, Aleksey Artamonov, Vladimir Aladinskiy, Mark Veselov, Artur Kadurin, Sergey Nikolenko, Alán Aspuru-Guzik, and Alexander Zhavoronkov. *Molecular Sets*

(MOSES): A Benchmarking Platform for Molecular Generation Models. November 2018.

[42] Alex Zhavoronkov, Yan A. Ivanenkov, Alex Aliper, Mark S. Veselov, Vladimir A. Aladinskiy, Anastasiya V. Aladinskaya, Victor A. Terentiev, Daniil A. Polykovskiy, Maksim D. Kuznetsov, Arip Asadulaev, Yury Volkov, Artem Zholus, Rim R. Shayakhmetov, Alexander Zhebrak, Lidiya I. Minaeva, Bogdan A. Zagribelnyy, Lennart H. Lee, Richard Soll, David Madge, Li Xing, Tao Guo, and Alán Aspuru-Guzik. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9):1038–1040, September 2019. Number: 9 Publisher: Nature Publishing Group.

[43] W. Patrick Walters and Mark Murcko. Assessing the impact of generative AI on medicinal chemistry. *Nature Biotechnology*, 38(2):143–145, February 2020. Number: 2 Publisher: Nature Publishing Group.

[44] Florian Prinz, Thomas Schlange, and Khusru Asadullah. Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9):712–712, September 2011. Number: 9 Publisher: Nature Publishing Group.