On the value of using 3D-shape and electrostatic similarities in deep generative methods

Giovanni Bolcato[a], Esther Heid[b] and Jonas Boström*[,c]

[a]Molecular Modeling Section, University of Padova, 35131 Padova, Italy. [b]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, 02139 Massachusetts, United States [c]Medicinal Chemistry, Early CVRM, BioPharmaceuticals R&D, AstraZeneca, 431 50 Mölndal, Sweden.

ABSTRACT

Multi-parameter optimization, the heart of drug design, is still an open challenge. Thus, improved methods for automated compounds design with multiple controlled properties are desired. Here, we present a significant extension to our previously described fragment-based reinforcement learning method (DeepFMPO) for the generation of novel molecules with optimal properties. As before, the generative process outputs optimized molecules similar to the input structures, now with the improved feature of replacing parts of these molecules with fragments of similar 3D-shape and electrostatics. We developed and benchmarked a new python package, ESP-Sim, for the comparison of electrostatic potential and molecular shape, allowing the calculation of high-quality partial charges (e.g., RESP with B3LYP/6-31G**) obtained using the quantum chemistry program Psi4. By performing comparisons of 3D-fragments, we can simulate 3D properties while overcoming the notoriously difficult step of accurately describing bioactive conformations. The new improved generative (DeepFMPO v3D) method is demonstrated with a scaffold-hopping exercise identifying CDK2 bioisosteres. All code is open-source and freely available.

INTRODUCTION

A crucial task in all drug discovery projects is designing molecules against multiple, often contradictory objectives.[1] Much of today's drug hunters' time is therefore spent on attempting to find an optimal compromise where all desirable properties are satisfied in a single molecule. The use of sophisticated computational methods, leveraging high-quality datasets to help solve this task, is thus conceptually very attractive.

Recent advances in artificial intelligence (AI) and machine learning (ML) have given rise to an immense popularity of inverse design,[2] and the field shows little signs of slowing down.[3] In inverse design, desired properties are specified *a priori*, and such methods generate compounds fitting that description.[4] Significant progress has been made in this area and a plethora of approaches for deep learning in molecular design has been published the last few years.[2] Many methods include reinforcement learning[5-7] to generate molecules, most often in the form of SMILES strings.[8] Other popular methods include generative methods such as recursive neural networks, generative adversarial networks or variational autoencoders, which are sometimes steered with reinforcement learning to control the molecular properties. The SMILES format in itself is nothing but amazing.[9] Using SMILES is also convenient for the AI algorithm, since a string is trivial to manipulate and transform. In addition, there are success stories of using SMILES in the area of generative design.[10] However, all molecules are 3D objects and a conservative modification to a SMILES string may cause a large effect in their 3D structure. Examples include the removal of brackets denoting substitution, such that a Y-shaped compound becomes linear, or the removal or changing of ring-closing locants. Therefore, optimization of molecular structures cannot be smooth in the space of 3D properties even though the SMILES strings change by only small amounts from iteration to iteration of the AI algorithm. We have previously presented a fragment-based generative approach (DeepFMPO) that addressed these modifications to the structure issue, albeit as 2D descriptions.[11] Here, we introduce a significant extension to DeepFMPO, using detailed descriptions of 3D-properties to represent molecules more accurately.

Shape and electrostatic properties of molecules are primary determinants of molecular recognition and should consequently be the method of choice when comparing the similarity of molecules encountered at various stages in drug design. Even though these methods have been used to achieve major impacts in related areas (e.g., virtual screening leading to the discovery of novel and unexpected chemotypes[12-14]), they have been largely unexplored in the context of *de novo* generative methods, although promising attempts have been made.[15,16] One reason for the reluctance of using 3D methods is the challenge of obtaining accurate descriptions of molecules bioactive conformations. In this work, we reduce this notoriously difficult step by using 3D-fragments of the complete target compounds.

Moreover, much of the work in the generative *de novo* design area has been focused on the development of maximally expressive methods whose purpose is to explore the entire chemical space. Our approach is different in this regard, since it specifically rewards the generation of molecules that are similar to known lead compounds. Another such method is the recently published MolDQN method, which maximizes a "drug-likeness" (QED) score while also maintaining similarity to the original molecule.[17] Virtual screening is related approach to generative methods,[18] and can also be a powerful method for finding hits and lead compounds with desired properties. However, a virtual screen is limited in regard to what is in the queried databases (*i.e.* it is not possible to find something that is not there).

Here, we describe a new open-source python package, ESP-Sim, for calculating shape and electrostatic similarities, and its implementation in DeepFMPO.[11] We highlight its usefulness with a scaffold-hopping study that identifies bioisosteres for a set of CDK2 kinase inhibitors.[20]

METHODS

The DeepFMPO method is based on an actor–critic model for reinforcement learning.[11] It is a fragment-based generative method that learns how to modify compounds and improve them. That is, molecules are split into fragments, and these fragments are replaced with other similar fragments in the (deep) learning process of generating novel molecules with optimal properties. Technically, the fragments are encoded into binary strings, and similar fragments are assigned similar encodings. This is achieved by constructing a balanced binary tree. In the process of assembling the tree, similarities between all fragments are calculated. Fragments are paired in a greedy bottom-up manner, where the two most similar fragments are paired first. The joining procedure is repeated until all fragments are put together in a single tree. Subsequently this information is used to generate encodings for all fragments. The paths from the root to the leaves defines the encoding for each fragment. For every branch in the tree a one ("1") is added to the encoding when going to the left and a zero ("0") is added when going to the right, see Figure 1. The rightmost character in the encoding corresponds to the branching closest to the fragment. In this process, the pairwise similarity between all fragments is calculated. There are many ways to calculate chemical similarities, and the most used approaches currently employ 2D fingerprints.

Here we present a new implementation of DeepFMPO utilizing a 3D-based molecular alignment method, where the ElectroStatic Potential (ESP) similarity between pairs of fragments is calculated. To this aim, we developed an open-source python package, ESP-Sim, which calculates the overlap integrals of the electrostatic potentials (generated from Coulomb potentials) of two molecules or fragments. Within DeepFMPO, the computation of ESP similarities can be broken down into six steps for each fragment pair (see Figure 2a) and is described in more detail below. Steps 2, 3 and 6 correspond to function calls of the ESP-Sim package, whereas steps 1 and 5 are innate to DeepFMPO. It is worth noting that this fragment alignment approach eliminates the challenging step of generating bioactive conformations for complete molecules, as well as alleviates the issue of aligning them correctly.

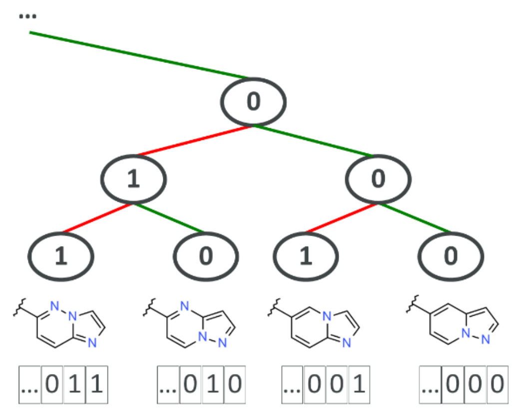

Figure 1. A snippet of the balance binary tree used in DeepFMPO. Fragments that are similar are placed close to each other. The encoding of a fragment is determined by the path from the root to the leaf. Every branching to the left adds a "1" to the end of the encoding and a branching to the right adds a "0".

The Molecular Alignment of Fragments

All single bonds in a molecule that extend from a ring atom are broken in the DeepFMPO process, creating the molecular fragments. The attachment atoms (previously connected with a single bond) are labeled in this step. To calculate ESP similarities the fragments must be aligned in 3D. Here, a conformational search is conducted generating an ensemble of low-energy conformers for all fragments containing rotatable bonds, using the ETKDG method[21] as implemented in RDKit.[22] As default a maximum of ten conformations of each fragment is generated. An anchor group is connected to the fragments' attachment atom and serves as a template in the alignment procedure. The coordinates of the anchor group are fixed in 3D space. The rationale for this step is that ligands containing related fragments typically bind in a similar orientation,[23,24] and these fragments will frequently make similar ligand-protein interactions. Consequently, to ensure that the fragments are aligned as accurately as possible, an anchor group is attached to the fragments and used in the molecular alignment step. The anchor group was arbitrarily chosen to be a hexazine ring with a methylene linker subunit. This group is of reasonable size for a template and highly unique (i.e., hexazines are never present in drug-like molecules) for easy identification and removal downstream in the process. A few experiments were conducted with other types of structural fragments as anchors to gauge possible conformational effects (*vide infra*). For each pair of fragments, the pair of conformations with the best shape overlay in terms of the highest shape Tanimoto value are stored. The anchor is then replaced with a hydrogen (see Figure 2). In cases where fragments include several labeled atoms, these are replaced with a methyl group. In this manner, all labelled atoms are replaced by a methyl which may be considered neutral in terms of electrostatic similarities. Finally, the ESP Tanimoto value is calculated between the pair of conformers with the best shape alignment (see the "*Electrostatic Similarity Calculations*" section below).

Electrostatic Similarity Calculations

The presented ESP-Sim method uses the cheminformatics toolkit RDKit[22] to generate different conformations of a molecule with (or without) a constrained anchor or core group, and computes shape and electrostatic potential similarities between pairs of conformers. Alternatively, ESP similarities can be computed on pre-aligned molecules. The electrostatic potential similarity is computed via the overlap integral of the Coulomb potentials of two molecules, as well as their respective self-overlap integrals as either Tanimoto[25,26] or Carbo similarity.[27,28] The Coulomb potential $V(\boldsymbol{r})$ describes the electric potential at a point $\boldsymbol{r}$ as a sum of potentials of point charges q_i at points $\boldsymbol{r}_i$ as

$$V(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0}\sum_i \frac{q_i}{|\boldsymbol{r} - \boldsymbol{r}_i|}$$

where ϵ_0 is the vacuum permittivity. Since analytic integration of the Coulomb potential at $\boldsymbol{r} = 0$ is not possible, we provide options to either approximate each potential with a sum of three Gaussian functions and integrate the fit function analytically analogous to Good *et al.*[29] or to perform a Monte-Carlo integration over the space outside of the van-der-Waals radii of each atom, and inside a user-defined margin. Partial charges can either be supplied by the user, calculated via RDKit (Gasteiger or MMFF94 charges), predicted via a recent machine-learning (ML) model,[30] or computed using the open-source quantum chemistry program Psi4,[31] with the option of using Restrained Electrostatic Potential (RESP) charges.[32] There are a range of different methods and basis sets available in Psi4. For example, the often-recommended combination of using the B3LYP method and the 6-31G** basis set, although using those can be computational demanding. Within DeepFMPO, Dask,[33] a library for parallel computing in Python, is used to speed up the process. It should be noted that the RESP/Psi4 method is not parametrized for atoms beyond the atomic number of Argon. To allow for larger atoms (e.g., bromine), their van der Waals (vdW) radius needs to be specified separately. In this code version, we set the vdW radii for bromo to 1.8 (file: resp/vdw_surface.py), following the GAMESS scheme[34] derived from the Merz-Kollman-Singh publication.[35] In addition to electrostatic similarities, ESP-Sim can furthermore output the shape Tanimoto similarity of molecules, describing the volume overlap. For DeepFMPO, we used the Tanimoto similarity of electrostatic potentials obtained via fitting to Gaussian functions (ESP-Tanimoto). We furthermore provide an option to add the volumetric shape score resulting in an ESP-TanimotoCombo score.

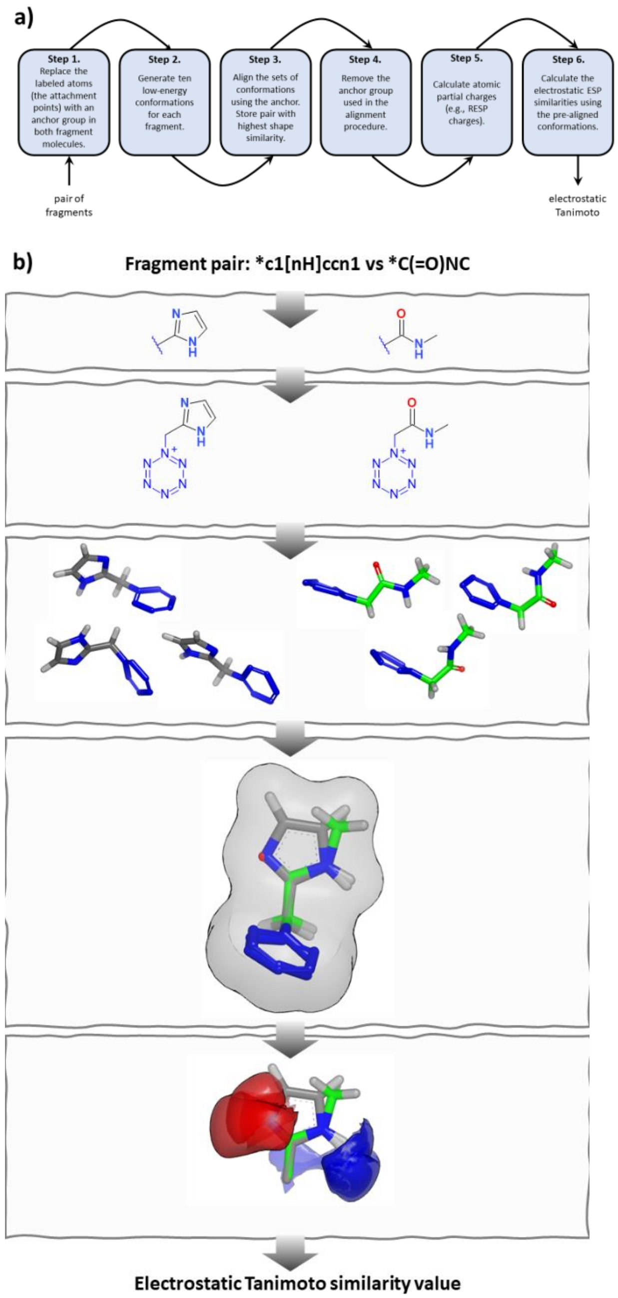

Figure 2. a) The stepwise procedure to obtain the electrostatic shape potential similarity values for pairs of fragments. b) An example of the corresponding procedure in graphics.

RESULTS

In the following, we showcase the performance of ESP-Sim on a variety of benchmark tasks. We then perform a retrospective case study, where we aim to demonstrate the value of using shape and electrostatic similarities in scaffold-hopping exercises. Scaffold hopping is a method for identifying bioisosteric replacements[36,37] with the intention of retaining biological activity of analog compounds but also improving other relevant molecular properties. It can also be used as a design strategy for intellectual property (IP) reasons.

ESP-Sim Benchmark studies

To evaluate the influence of the employed partial charge distribution on the observed scores within ESP-Sim, ESP similarities were computed for the same molecule at the same geometry but with different partial charges. As ground truth, quantum-mechanically (QM) obtained RESP charges at a high level of theory (MP2 with a polarizable PVTZ basis set) were used [30], on a selection of about 3000 neutral molecules. RESP charges are specifically designed to reproduce the electrostatic potential of a molecule, so that a comparison of electrostatic potentials obtained from different charge distributions to the QM RESP charges allows for a detailed assessment of the quality of each approach for similarity comparisons. We evaluated Gasteiger[38] (default in RDKit), MMFF94,[39] and AM1-BCC[40,41] partial charges, as well as a machine learning model (ML).[42] The ML partial charge model is provided with the ESP-Sim package on Github. Table 1 provides an overview of the observed mean absolute deviations of the respective partial charges from the RESP charges, as well as the ESP similarities evaluated via Carbo or Tanimoto similarities. We find that AM1-BCC charges reproduce the QM electrostatic potentials best, followed by the deep learning model, MMFF, and Gasteiger.

Table 1. Mean absolute deviations between Gasteiger, MMFF, ML or AM1-BCC partial charges q compared to RESP charges, as well as similarities of electrostatic potentials compared to RESP evaluated either via Carbo or Tanimoto similarity

Partial charges	MAE q [e]	ESP-Sim (Carbo)	ESP-Sim (Tanimoto)
Gasteiger	0.16	0.78	0.60
MMFF	0.17	0.80	0.64
ML	0.17	0.85	0.61
AM1-BCC	0.12	0.88	0.78

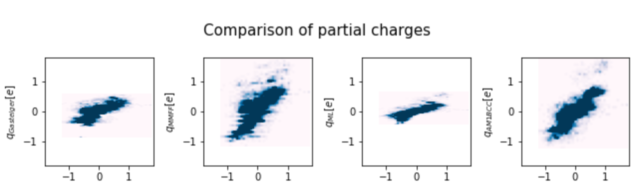

Figure 3. Heatmap of quantum mechanical RESP partial charges compared to Gasteiger, MMFF, ML or AM1-BCC partial charges.

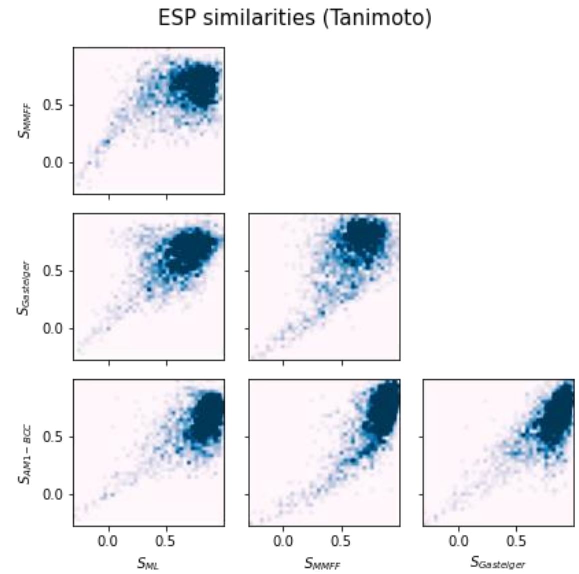

Figure 4. Electrostatic potential similarities between molecules with RESP partial charges to molecules with Gasteiger, MMFF, ML or AM1-BCC partial charges (at the exact same geometries evaluated via Tanimoto similarity. An analogous figure for Carbo similarity is given in the Supporting Information.

Figure 3 depicts a heatmap of the QM atomic charges compared to Gasteiger, MMFF, ML and AM1-BCC charges. Although there is no perfect correspondence of QM charges to any of the evaluated charges, we find highest agreements for AM1-BCC charges. Notably, Gasteiger and ML charges lead to narrower ranges than MMFF and AM1-BCC charges. This is also reflected in the Carbo and Tanimoto ESP similarities in Table 1. The Carbo metric, which is largely insensitive toward the magnitude of a function,[43] yields more favorable scores for ML and AM1-BCC, in contrast to the Tanimoto metric, which is more sensitive to the absolute magnitudes, thus ranking MMFF better than ML. These differences are small, though, and a comparison of observed similarities via different charge distribution in reference to QM RESP shows similar trends between all options, Figure 4. We can therefore assume that even Gasteiger charges lead to a fair depiction of the electrostatic potential for most molecules. In fact, benchmarking of ESP-Sim on protein-docking databases shows little dependence of ranking metrics on the employed partial charge distribution, as detailed in the following.

We furthermore compared ESP-Sim scores to similarities obtained via the state-of-the-art tool EON[44] for about 450 fragments generated by DeepFMPO for various partial charge distributions. We find a strong correlation, with Spearman correlation coefficient of about 0.8. A detailed analysis is given in the Supporting Information. In addition, we assessed the ability of ESP-Sim scores to identify potential ligands to protein targets. We compared the performance of ESP-Sim electrostatic and shape similarities to a set of re-scoring functions[45-51] on the dopamine D4 receptor, for which experimental data on active and inactive compounds is known.[52] We furthermore benchmarked ESP-Sim on the 102 DUD-E targets[53] and compare its performance against a variety of ligand-based approaches.[54-59] For both comparisons, we find that ESP-Sim electrostatic and shape similarities perform very well. Details on these benchmarks are given in the Supporting Information.

Assessing Various Molecular Similarity Measures

A frequently occurring scenario is that a drug hunting team has identified a promising compound, from an internal lead generation effort or from the literature, that needs optimization. For the sake of argument, compound **1**[20] in Figure 5 is such a compound.

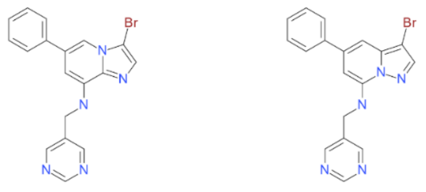

Figure 5. Two equipotent CDK2 kinase inhibitors. CDK2 inhibitors containing the related bicyclic heterocycles imidazopyridine (**1**) and pyrazolopyridine (**2**) were discovered through high-throughput screening by Fischmann *et al.*[20] and here used as a scaffold-hopping example.

With compound **1** at hand, the design question is then "*which compound should we make next?*". The optimization task usually includes improving molecular properties (e.g., permeability, solubility, clearance, selectivity, etc.) and perhaps also IP related issues. A common scenario then is for the project team, to try to come up with ideas of novel central rings to be introduced as scaffold replacements. In this context, it should be noted that heterocyclic rings are often considered special and typically end up in different patent applications.[60] Also, with regards to calculating molecular properties (e.g. lipophilicity), many 2D-based methods are not adequately parametrized and have difficulties in assessing heterocyclic compounds accurately.[61] So, how can break-through ideas for novel central rings be generated and which methods can be used to do it? Here, compound **2** (Figure 5) is one answer to the question "*what to make next?*". It is equipotent to compound **1** and importantly, contains a different but related central scaffold. That is, the bicyclic heterocycle in compound **1** (imidazo(1,2-a)pyridine) and compound **2** (pyrazolo(1,5-a)pyridine) are both 9-membered ring-systems with identical substituents.

To investigate how different methods predict the similarity of these kind of central bicyclic heterocyclic scaffolds we first generate a dataset of fragments containing the same framework and similar substitution pattern. Thus, the ChEMBL v28 database[62] was queried for compounds including a 9-membered bicyclic ring system, with three substituents, using SMARTS matching.[22] For comparison reasons, the substituents were subsequently removed providing 30 different scaffolds, see Figure 6. In this manner we identified an extensive list of 9-membered bicyclic heterocyclic scaffolds present in drug-like molecules that potentially could act as replacements for the pyrazolopyrimidine in compound **1**.

All 30 bicyclic systems were subsequently subjected to pair-wise comparisons using a range of standard 2D similarity measures, together with the 3D-based ESP-Sim measure. A summary of the results obtained from each method is reported in Table 2. For completeness, the results using four different anchor fragments (hexazine, carboxylic acid, piperidine and iodine) are shown in Figure 7. The heat-maps are essentially the same indicating that the method is not dependent on the choice of anchor fragment.

Table 2. Rankings for the **1** vs **2** fragment pair, among pair-wise comparison of 30 different heterocyclic rings. The rankings, and Tanimoto value, using a range of different 2D similarity methods available through RDKit and the new ESP-Sim measure are reported. Hexazine was used as anchor fragment.

Method	Rank (max = 30)
ESP-Sim (B3LYP/6-31G**)	1
Morgan fingerprint (radius 2)	5
Morgan fingerprint (radius 3)	5
MACCS keys fingerprints	17
MCS Tanimoto	21
Topological fingerprints	22

The **1** vs **2** fragment pair is top-ranked when using the ESP-Sim (B3LYP/6-31G**) metric, but not by the 2D-based methods. The Morgan fingerprints rank the **1** vs **2** pair among the top five (Table 2), which is reasonably high. However, given the challenges and resource investments required to establish new synthetic routes, our experience is that very few alternative ring analogs are explored in real-life projects. Typically, only a couple of ring replacements are made and tested, essentially enforcing that only top-ranked scaffolds would be followed-up. Two other observations provide further support for the use of the ESP-Sim method. First, the MACCS keys fingerprint resulted in very similar values for many scaffolds (e.g., the Tanimoto similarity value for five scaffolds against scaffold of compound **1** show identical values – 0.87), suggesting that the MACCS keys similarity metric is not sufficient for capturing such subtle differences. Second, there are a couple of clearly structurally dissimilar fragments in Figure 6 (e.g. 1,4,6-trimethylpyrazolo[5,4-b]pyridine vs 2,4,7-trimethylimidazo[2,1-f][1,2,4]triazine) that are ranked low when using ESP-Sim (as they most probably should), but top-ranked when using Morgan 2D-fingerprints.

As a final observation, deriving ESP-similarities with methods of lower theory for calculating the underlying partial charges (Gasteiger, MMFF and HF/3-21G, data not shown) also yielded the **1** vs **2** pair as top-ranked, suggesting that such partial charges may be sufficient and a cost-effective alternative for this data set. We recommend using a higher level of theory, although computationally more demanding. For example, RESP partial charges derived using the B3LYP method and the 6-31G** basis set or AM1-BCC partial charges, which were found to reproduce the QM electrostatic potentials best in our benchmark.

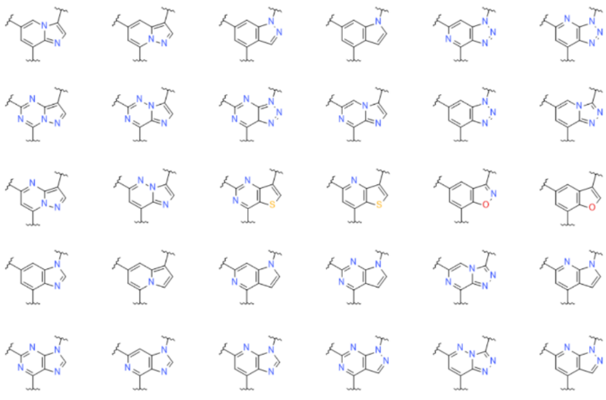

Figure 6. Bicyclic heterocyclic scaffolds in ChEMBL compounds matching the SMARTS pattern "*[A][cH0]1[c,n][c,n]([A])[c,n]2[c,n][c,n][c,n]([A])[c,n]2[c,n]1*".

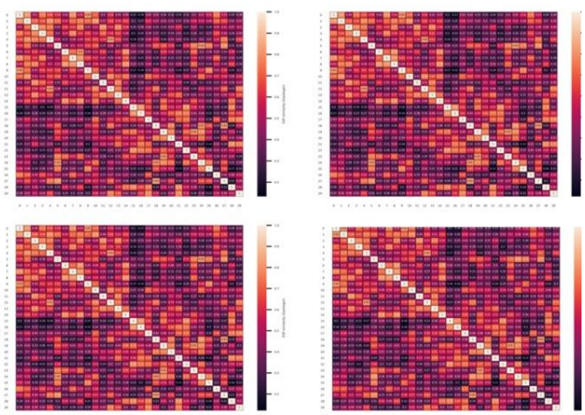

Figure 7. All-against-all comparison experiments were conducted with four structurally different anchor fragments (top-left: hexazine, top-right: carboxylic acid, bottom-left: piperidine, bottom-right: iodine). The different anchors give essentially the same results.

Generating "Sweet Spot" Molecules

Having established the value of using the ESP-Sim measure, the next step was to include it in the generative (DeepFMPO) method. An experiment was set-up to mimic a real-world scenario, where a set of lead compounds is optimized toward sweet spot criteria through a multi-parameter optimization process. Three different calculated properties (Molecular Weight, Polar Surface Area[22] and clogP[63]), commonly used in the optimization of leads to candidate drugs, were selected for this purpose. It should be noted that the choice of molecular properties was also selected for practical reasons facilitating reproducibility. Namely, there are methods to calculate them using RDKit.[22] The aim of the set-up was to bias the generation of compounds to fulfill the criteria for the three calculated properties, while also maintaining their similarity in shape and electrostatics towards a known set of lead compounds. The agent in the reinforcement learning method was rewarded for producing valid molecules and got a higher reward when generating molecules with properties in the targeted ranges. Since this was a scaffold-hopping exercise, with the goal of identifying a new bioisosteric scaffold, the minimum and maximum target values for the three properties were centered around the corresponding values for compound **1** (i.e., 320<MW<420, 2.3<clogP<4.3 and 45<PSA<65).

The library of input fragments was generated from a set of structurally diverse compounds known to exhibit inhibitory effects against kinase targets, including compounds that have shown activity against the specific biological target of interest (CDK2). The data set was extracted from the ChEMBL database (version 28) using simple text searches, resulting in a set of 557 fragments (including the ones in Figure 6), as obtained from 1059 compounds. The lead series compounds were obtained by a substructure search using the (imidazo(1,2-a)pyridine) central scaffold of compound **1** on the surechembl website (https://www.surechembl.org/search) and yielded 138 close analogs, which is a typical number to what a drug hunting program would have access to. The data sets are available online (https://github.com/giovanni-bolcato/deepFMPOv3D). The calculation for this dataset requires 8h on a i9-9820x CPU, using the 20 cores.

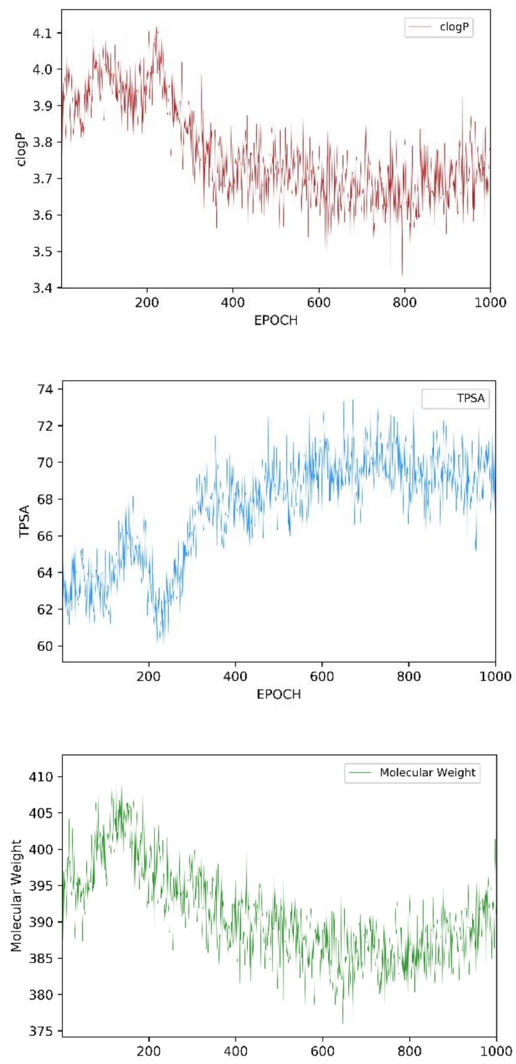

Figure 8a-c. Graphs showing how molecular weight, logP and TPSA values change during the epochs, as the mean value of all the compounds for each epoch.

DeepFMPO with the ESP-Sim measure generated a total of 6359 unique molecules, when terminated at 1000 epochs. About two-thirds of those were sweet spot compounds. Hence, the agent generated compounds that have all three properties within the desired ranges. This number (ca 4000) is lower than when using a standard generative method facilitating the selection process, and a result of intentionally biasing using 3D-similarity. The evolution of the percentage of generated molecules that demonstrate properties within the target ranges during the training process is shown in Figure 8a-c, displaying evidence of learning. A significant number of the generated compounds include the central scaffold of compound 2, and a number of those show near identical substitution pattern to compound **1**. These bioisosteric compounds were observed in early epochs. One other nine-membered scaffold (3) was also represented among the generated, see Figure 9. When performing the same experiment but with simpler standard similarity measures (Morgan fingerprints, MACCS keys and Topological fingerprints), no compounds with the central scaffold of compound 2 (or 3) were generated. This provides an incentive for the use of DeepFMPO with ESP-Sim in scaffold-hopping exercises.

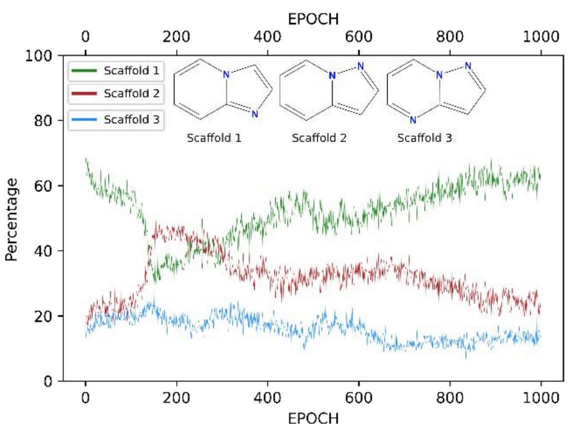

Figure 9. A graph showing the frequency of occurrence of compounds including the central fragment of compound 1, **2** and **3**. The y-axis represents the total number of compounds each epoch (percentage).

DISCUSSION

In the current work we set out to explore the use of a sophisticated similarity metric in generative methods. The power of rewarding compounds that are similar in 3D aspects, in addition to other molecular property constraints, is often underappreciated. It is a challenging task, due to the issues involved with conformer generation and molecular alignments. Nonetheless, this is a design strategy that we believe should be given more attention and we discuss why below.

Molecular representations in deep generative methods

Deep generative models typically use non-3D methods to represent molecules. Text-based methods and the use of SMILES strings is still the most prevalent representation. The reason for this is probably because SMILES can be massively expressive, and that it is trivial to manipulate and transform strings. However, there are some drawbacks with using SMILES strings.[11,64] A significant problem is that a conservative change can have a huge change in the 3D structure of a molecule. This is important since all molecules are 3D objects. Here, we have addressed this issue by extending the fragment based DeepFMPO method, where molecules are built from similar fragments, instead of sequences of letters (as is the case for SMILES based methods). Fragment-based methods are considered intuitive and often mimic the way medicinal chemists think and design. The approach was recently described by Meyers *et al.* as a method "*offer an appealing compromise between molecular expressivity and practicality*" [64]. Hence, a common medicinal chemistry design strategy is to work on molecular series, swapping fragments and substituents in one part of the molecule, while keeping other parts of the compounds unaltered. This is often a challenge for generative methods working on SMILES strings,[64,65] leading us to the next topic of discussion.

Deep generative methods can generate many compounds

Most generative AI methods produce tens of thousands of unique and diverse high-scoring compounds when used without stringent filters. This is related to Brenner's underdetermined inverse problem stating that available data does not uniquely specify systems.[66] Also, although there may be nothing chemically wrong with AI generated molecules (i.e., all atoms in common valences and charge states), some can be exotic,[67] and an experienced medicinal chemist would reject them upfront. The issue of such unwanted molecules is manageable from a technical perspective. For example, one can enforce substructure rules and penalize the existence of undesired moieties (e.g., radicals, peroxides, anhydrides, strained and chemically unstable systems) in the reward functions, or as post-filters.

A more difficult problem to address is how to prune down the very many generated compounds to the few worth making. In reinforcement learning, a scoring function is used for this purpose. A complicating factor here is that drug discovery is complex and not all factors used in decision-making are easy to capture and thus not readily converted into rules that the AI methods can use in their rewards system. For example, a compound with several stereocenters is usually difficult to make (and resolve) and should consequently get a low reward score unless its building blocks are already available on the shelf. Also, absorption is a critical parameter for the optimization of oral drugs. Permeability over Caco-2 cells is often used as a surrogate when assessing absorption. A complicating factor here is that the uptake over the Caco-2 cells can be hampered by efflux, and in the case of high cell permeation the efflux is less relevant. A reward function handling such scenarios would require several "if-then-else" statements. These can be included in reward scores but are not always trivial to define and set-up for edge cases. In addition, multiparameter optimization becomes increasingly challenging when there are many constraints to fulfill.[68] In brief, the biggest challenge of deep generative methods is to define relevant reward scores, and this is unfortunately less studied.

Simple drug-likeness rules, multivariate methods for DMPK properties (solubility, permeability, clearance, etc.) and safety, as well as docking scores are typically included in reward scores as filters. However, several thousands of compounds will inevitably still pass those filters. This is related to the common lack of sufficient high-quality data, and the fact that we still often struggle with making predictions to the required accuracy. Prediction of biological activity is an extremely hard problem since many phenomena involved are difficult to quantify precisely. Standard docking scores are most often not sufficient. Although, at times, methods such as Free Energy Perturbation (FEP) can improve the scoring accuracy for small perturbations of one structure into another, but not for major structural changes.[69] The use of FEP combined with active learning is gaining traction and is showing promise.[70,71] Nonetheless, when the output contains many structurally diverse molecules, as frequently is the case for expressive SMILES-based generative methods, current methods' accuracies are not sufficient to filter down many compounds to a selected few. Despite the increasing prevalence of physics-based models in generative modelling, bioaffinity prediction remains very challenging.

Here, we propose shape and electrostatic potential matching as a strategy to bias generative models to propose compounds with different fragments (that are likely bioisosteres) of known lead compounds. The tool is designed to generate novel molecules with optimized properties. One example usage is scaffold hopping. Here it should be noted that there are many other scaffold hopping tools available[72] – ranging from CAVEAT,[73] which is one of the early 3D database searching programs, to the more recent BROOD.[74] In the context of generative methods, Langevin *et al.* recently described a new RNN-based algorithm, named SAMOA (Scaffold Constrained Molecular Generation), to perform scaffold-constrained molecular design.[65] Generative methods benefit from the associated reinforcement learning methods, allowing multi-objective molecular design optimization, while only exploring the relevant chemical space.

Using similarity as a design strategy

As mentioned above, current generative AI methods generally suffer from the lack of prediction accuracy. Thus, learning from past drug hunting experiences, we deliberately bias the AI method to generate compounds that are similar to active molecules already discovered. We approach this problem by relying on the similarity principle,[75] which states that similar molecules tend to have similar properties.[76] Some advantages to this approach are discussed below. First, by generating molecules similar to the initial set available in the project, confidence in the predictions can be high because they remain in the applicability domain of the model. This is contrary to expressive methods that are designed to fully explore chemical space and generate structurally diverse compounds, which are consequently also the most uncertain to predict. Second, for similar compounds, the same chemical intermediates and established synthetic routes can often be re-used, facilitating speedy progress. Third, sometimes certain structural fragments (e.g., "privileged structures"[77]) are difficult to replace without severe drops in potency due to specific ligand-protein interactions.

As a related example, the strategy of molecular optimization using similarity was recently applied by Zhavoronkov and coworkers. They reported that deep learning enabled rapid identification of potent DDR1 kinase inhibitors.[78] Walters and Murcko analyzed the Zhavoronkov *et al.* study and reported that the AI-generated compound **B** (Figure 9) shared a common substructure with an already marketed multi-kinase inhibitor (Ponatinib, Figure 9), which was indeed included in the training set.[79] In some more detail, they ring-closed a benzamide carbonyl into an isoxazole moiety to yield an equipotent and unique compound.[78] These two compounds are very similar with regards to shape and electrostatics, see Figure 10. Thus, Zhavoronkov's AI method successfully mimicked typical medicinal chemistry behavior, keeping certain parts fixed and making minor modifications to other.

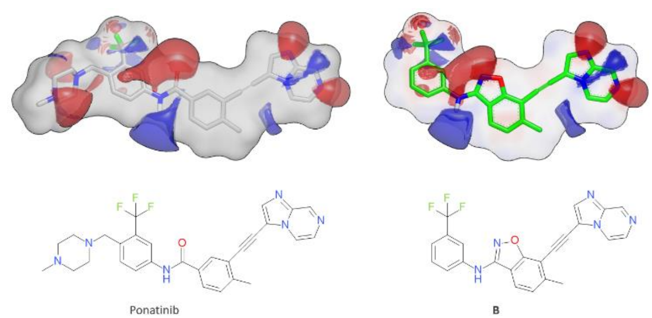

Figure 10. Designing similar compounds can be a good tactic in drug discovery. Here illustrated with two potent DDR1 kinase inhibitors. The AI-generated compound **B** by Zhavoronkov *et al.* and ponatinib, a marketed multi-kinase inhibitor.[78] The compounds share a rather large common substructure. The hydrogen-bond acceptor and donor functionalities are visualized with electrostatic contours (red: negative, blue: positive). The ESP-Sim Tanimoto value is 0.81 for this pair.

It is sometimes believed that computer-aided design (CAD) methods need to provide radically "non-intuitive" different compounds to merit its use. However, believing that CAD approaches should surprise us and produce results that we would not have expected is a tall order. In this context, the scoring functions used in generative methods for reinforcement learning are not designed to extrapolate, and do not account for all aspects involved in designing drugs process. Thus, the power of current AI's lies more in pattern recognition than in creative discovery.

Palazzesi and Pozzan recently reported a list of over 100 deep generative methods published in the literature between 2017–2020.[80] The methods are innovative and perform well in benchmark studies that measure models' ability to, for example, reproduce property distributions, and generate valid, diverse, and novel molecules.[81] One may thus conclude that generative modelling is essentially a solved problem – given a reward function, we now have the methods to generate molecules that satisfy it. Despite this success, biology and drug discovery remain immensely complex, and it is our viewpoint that current generative methods best serve to augment drug design. To take the next step (full autonomy), calculated predictions need ultra-high accuracy and for that we need to develop a broader understanding of human biology. The state of AI in drug design may be seen as analogous to the automotive industry. While the future of autonomous vehicles is promising and exciting, we are not near fully autonomous cars yet. Candidate drugs, as well as cars, still require human attention, given the complexity involved and the vast amount of edge cases that are non-trivial to code up efficiently. Thus, humans (with domain knowledge) are still very much needed in the process: to steer the tools and triage the output. In this context, we would like to highlight the Gruenif.ai tool where the user can provide feedback interactively while molecules are generated.[82] Such "human-in-the-loop" methods can be very effective. Future versions of DeepFMPO will include such functionalities.

CONCLUSIONS

The use of sophisticated computational methods for de novo design is attractive and deep generative methods have gained a lot of attention. Significant progress has been made when it comes to *generating* molecules. However, *scoring* them accurately remains a major challenge. Real-life project experience informs us that *in silico* predictions (e.g., synthesis, potency, properties) are constantly improving, but they are generally not accurate enough to prioritize a handful compounds for synthesis from a long list of high-scoring AI generated molecules. Thus, what really needs solving is being able to do ultra-accurate predictions to advance the field to the next level. Until then, the approach of biasing molecular design towards compounds similar to known actives will remain as one pragmatic and fruitful way to success.

Here we present a 3D fragment-based reinforcement learning approach for the generation of novel molecules with optimized properties, called "DeepFMPO v3D". We furthermore developed a python package, ESP-Sim, for calculating molecular shape and electrostatic similarities. We benchmarked ESP-Sim on a variety of tasks including the evaluation of detailed 3D similarities, protein-ligand docking and re-scoring of docked ligands and report competitive performances. The inclusion of ESP scores into DeepFMPO promotes the generation of compounds similar to existing lead molecules, towards desirable sweet spot properties. The proposed method allows the calculation of high-quality partial charges (e.g., RESP with B3LYP/6-31G**) obtained using the quantum chemistry program Psi4. In a scaffold-hopping case-study we show that our approach of using shape and electrostatics similarities performs well. DeepFMPO v3D ranks known equipotent scaffolds higher and generates them earlier (*i.e.,* speedier). The way the 3D method is implemented makes the approach essentially alignment-independent (on a molecular level) and does not require knowing the bioactive conformation. Both DeepFMPO v3D and ESP-Sim are freely available online.

AUTHOR INFORMATION

Corresponding Author

* Telephone number: +46 31 7065251. E-mail: dr.jonas.bostrom@gmail.com.

ACKNOWLEDGMENTS

The helpful code advice from Niclas Ståhl, and the constructive feedback on earlier versions of the manuscript from colleagues is gratefully acknowledged. E.H. acknowledges support from the Austrian Science Fund (FWF), project J-4415, and the Machine Learning for Pharmaceutical Discovery and Synthesis Consortium (MLPDS).

Data and Software Availability

The ESP-SIM code can be found here: https://github.com/hesther/espsim. The repository also contains the data sets (3000 molecules with QM RESP charges, 451 molecules from DeepFMPO, instructions to download D4 and DUD-E) and scripts to reproduce all benchmarks. The DeepFMPO_v3D code version together with the associated data sets used in this study are here: https://github.com/giovanni-bolcato/deepFMPOv3D.

Supporting Information Available

Supplementary information associated with this article can be found online.

AUTHOR CONTRIBUTIONS

E.H. developed and implemented the ESP-Sim method. G.B. implemented ESP-Sim into DeepFMPO and performed the machine learning experiments. The project was devised by J.B. All authors contributed to the writing of the manuscript.

REFERENCES

1. Hann, M. M., Keserü, G. M. Finding the Sweet Spot: The Role of Nature and Nurture in Medicinal Chemistry. *Nat. Rev. Drug Discov.* 2012, 11, 355–365.

2. Elton, D.C., Boukouvalas, Z., Fugea, M.D., Chunga, P. W. Deep Learning for Molecular Design — A Review of the State of the Art. *Mol. Syst. Des. Eng.* 2019, 4, 828–849

3. Merz Jr., K.M., De Fabritiis, G., Wei, G. JCIM Special Issue on Generative Models for Molecular Design. *J. Chem. Inf. Model.* 2020, 60, 1072.

4. Sanchez-Lengeling, B., Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* 2018, 361, 360–365.

5. Schneider, G., Fechner, U. Computer-Based De Novo Design of Drug-Like Molecules. *Nat. Rev. Drug Discov.* 2005, 4, 649–663

6. Gupta, A., Müller, A.T., Huisman, B.J.H., Fuchs, J.A., Schneider, P., Schneider, G. Generative Recurrent Networks for De Novo Drug Design. *Mol Inform.* 2018, 37, 1700111

7. Popova, M., Isayev, O., Tropsha, A. Deep Reinforcement Learning for De Novo Drug Design. *Sci. Adv.* 2018, 4, eaap7885

8. Weininger. D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comp. Sci.* 1988, 28, 31–36.

9. Dalke, A. Weininger's Realization (accessed October 10, 2021) http://www.dalkescientific.com/writings/diary/archive/2016/12/02/Weiningers_realization.html

10. Merk, D.; Friedrich, L.; Grisoni, F.; Schneider, G. De Novo Design of Bioactive Small Molecules by Artificial Intelligence. *Mol. Inf.* 2018, 37, 1700153

11. Ståhl, N., Falkman, G., Karlsson, A., Mathiason, G., Boström, J. Deep Reinforcement Learning for Multiparameter Optimization in De Novo Drug Design. *J. Chem. Inf. Model.* 2019, 59, 3166–3176.

12. Muchmore, S., Souers, A. J., Akritopoulou-Zanze, I. The Use of Three-Dimensional Shape and Electrostatic Similarity Searching in the Identification of a Melanin-Concentrating Hormone Receptor 1 Antagonist. *Chem. Biol. Drug Des.*, 2006, 67, 174–176.

13. Churchill, G.C. Identification of a Chemical Probe for NAADP by Virtual Screening. *Nat. Chem. Biol.* 2009, 4, 220–226.

14. Boström, J. Grant, J.A.; Fjellström, O.; Thelin, A.; Gustafsson D. Potent Fibrinolysis Inhibitor Discovered by Shape and Electrostatic Complementarity to the Drug Tranexamic Acid. *J. Med. Chem.* 2013, 56, 8, 3273–3280.

15. Skalic, M.; Jimenez, J.; Sabbadin, D.; De Fabritiis, G. Shape-Based Generative Modeling for De-Novo Drug Design. *J. Chem. Inf. Model.* 2019, 59, 1205–1214.

16. Imrie, F., Bradley, A. R., van der Schaar, M. Deane, C. M Deep Generative Models for 3D Linker Design. *J. Chem. Inf. Model.* 2020, 60, 1983–1995.

17. Zhou, Z., Kearnes, S., Li, L. Zare, R. N., Riley, P. Optimization of Molecules via Deep Reinforcement Learning. *Sci Rep* 9, 10752 (2019). https://doi.org/10.1038/s41598-019-47148-x

18. Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening – An Overview. Drug Discovery Today 1998, 3, 160– 178.

19. FastROCS; OpenEye Scientific Software Inc.: Santa Fe, NM, U.S., https://www.eyesopen.com/molecular-modeling-fastrocs (accessed: 2021-12-17)

20. Fischmann, T.O.; Hruza, A.; Duca, J.S.; Ramanathan, L.; Mayhood, T.; Windsor, W.T.; Le, H.V.; Guzi, T.J.; Dwyer, M.P.; Paruch, K.; Doll, R.J.; Lees, E.; Parry, D.; Seghezzi, W.; Madison, V. Structure-Guided Discovery of Cyclin-Dependent Kinase Inhibitors. *Biopolymers*, 2008, 89, 372–379.

21. Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Comp. Sci.* 2015, 55, 2562-74.

22. Landrum, G.A. RDKit: Open-Source Cheminformatics Software, version 2021.03.1 http://www.rdkit.org

23. Hare, B.J.; Walters, W.P.; Caron, P.J.; Bemis, G.W. CORES: An Automated Method for Generating Three-Dimensional Models of Protein/Ligand Complexes. *J. Med. Chem.* 2004, 47, 19, 4731–4740.

24. Boström, J., Hogner, A. Schmitt, S. Do Structurally Similar Molecules Bind in a Similar Fashion. *J. Med. Chem.* **2006**, 49, 6716–6725.

25. Tanimoto, T.T. An Elementary Mathematical Theory of Classification and Prediction, IBM Report (November, 1958), cited in: G. Salton, Automatic Information Organization and Retrieval (McGraw-Hill, 1968) pp. 238-249.

26. Willett, A. P.; Barnard J.M.; Downs, G.M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* 1998, 38, 983-996.

27. Carbó, R.; Leyda, L.; Arnau, M. How Similar Is a Molecule to Another? An Electron Density Measure of Similarity Between Two Molecular Structures. *Int. J. Quantum Chem.* 1980, 17, 1185-1189.

28. Carbó, R.; Domingo, L. LCAO-MO Similarity Measures and Taxonomy. *Int. J. Quantum Chem.* 1987, 32 517-545.

29. Good, A. C.; Hodgkin, E. E.; Richards, W. G. (1992). Utilization of Gaussian Functions for the Rapid Evaluation of Molecular Similarity. *J. Chem. Inf. Comput. Sci.* 1992, 32, 188-191.

30. Heid, E., Fleck, M., Chatterjee, P., Schröder, C., and MacKerell Jr, A. D. Toward Prediction of Electrostatic Parameters for Force Fields That Explicitly Treat Electronic Polarization. *J. Chem. Theory Comput.*, 2019, *15*, 2460-2469.

31. Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: An Open-Source Ab Initio Electronic Structure Program. *WIREs Comput. Mol. Sci.* 2012, 2, 556–565.

32. The RESP code is a plug-in to the Psi4 quantum chemistry package available from https://github.com/cdsgroup/resp (Sherrill Research Group).

33. Dask is a flexible library for parallel computing in Python. https://docs.dask.org/en/latest/ (accessed June 22, 2021).

34. GAMESS parameters https://github.com/streaver91/gamess/blob/master/source/prplib.src

35. Singh, U.C.; Kollman P.A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comp. Chem.* 1984, 5, 129–145.

36. Ertl P. In Silico Identification of Bioisosteric Functional Groups. *Curr. Opin. Drug Discov. Develop.* 2007, 10, 281–288.

37. Meanwell NA. Synopsis of Some Recent Tactical Application of Bioisosteres in Drug Design. *J. Med. Chem.* **2011**, 54, 2529–2591.

38. Gasteiger, J., and Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron*, 1980, *36*(22), 3219-3228.

39. Halgren, T. A. Merck Molecular Force Field. II. MMFF94 Van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comp. Chem.*, 1996, *17*(5-6), 520-552.

40. Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comp. Chem.*, 2000, *21*(2), 132-146.

41. Jakalian, A., Jack, D. B., and Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comp. Chem.*, 2002, *23*(16), 1623-1641.

42. Guan, Y., Coley, C. W., Wu, H., Ranasinghe, D., Heid, E., Struble, T. J., Pattanaik L. Green, W. H. and Jensen, K. F. Regio-Selectivity Prediction With a Machine-Learned Reaction Representation and On-the-Fly Quantum Mechanical Descriptors. *Chem. Sci.*, 2021, *12*(6), 2198-2208.

43. Good, A. C. The Calculation of Molecular Similarity: Alternative Formulas, Data Manipulation and Graphical Display. *J. Mol.Graph*, 1992, *10*(3), 144-151.

44. EON: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com

45. Martin, L. J. D4-Rescore. https://github.com/ljmartin/d4-rescore.

46. Durrant, J. D., and McCammon, J. A. NNScore 2.0: A Neural-Network Receptor–Ligand Scoring Function. *J. Chem. Inf. Model*, 2011, *51*(11), 2897-2903.

47. Wójcikowski, M., Kukiełka, M., Stepniewska-Dziubinska, M. M., and Siedlecki, P. Development of a Protein–Ligand Extended Connectivity (PLEC) Fingerprint and Its Application for Binding Affinity Predictions. *Bioinf.*, 2019, *35*(8), 1334-1341.

48. Ballester, P. J., and Mitchell, J. B. A Machine Learning Approach to Predicting Protein–Ligand Binding Affinity With Applications to Molecular Docking. *Bioinf.*, 2010, *26*(9), 1169-1175.

49. Koes, D. R., Baumgartner, M. P., and Camacho, C. J. Lessons Learned in Empirical Scoring With Smina From the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.*, 2013, *53*(8), 1893-1904.

50. Landrum, G. A., Penzotti, J. E., and Putta, S. Feature-Map Vectors: A New Class of Informative Descriptors for Computational Drug Discovery. *J. Comput. Aided Mol. Des.*, 2006, *20*(12), 751-762.

51. Wójcikowski, M., Ballester, P. J., and Siedlecki, P. Performance of Machine-Learning Scoring Functions in Structure-Based Virtual Screening. *Sci. Rep.*, 2017, *7*(1), 1-10.

52. Lyu, J., Wang, S., Balius, T. E., Singh, I., Levit, A., Moroz, Y. S., O'Meara, M. J., Che, T., Algaa, E., Tolmachova, K., Tolmachev, A. A., Shoichet B. K., Roth, B. L. and Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. *Nature*, 2019, *566*(7743), 224-229.

53. Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.*, 2012, *55*(14), 6582-6594.

54. Koes, D. R., and Camacho, C. J. Shape-Based Virtual Screening With Volumetric Aligned Molecular Shapes. *J. Comp. Chem.*, 2014, *35*(25), 1824-1834.

55. Puertas-Martín, S., Redondo, J. L., Ortigosa, P. M., and Pérez-Sánchez, H. OptiPharm: An Evolutionary Algorithm To Compare Shape Similarity. *Sci. Rep.*, 2019, *9*(1), 1-24.

56. Grant, J. A., Gallardo, M. A., and Pickup, B. T. A Fast Method of Molecular Shape Comparison: A Simple Application of a Gaussian Description of Molecular Shape. *J. Comp. Chem.*, 1996, *17*(14), 1653-1666.

57. Yan, X., Li, J., Liu, Z., Zheng, M., Ge, H., and Xu, J. Enhancing Molecular Shape Comparison by Weighted Gaussian Functions. *J. Chem. Inf. Model.*, 2013, *53*(8), 1967-1978.

58. Cleves, A. E., Johnson, S. R., and Jain, A. N. Electrostatic-Field and Surface-Shape Similarity for Virtual Screening and Pose Prediction. *J. Comput. Aided Mol. Des.*, 2019, *33*(10), 865-886.

59. von Behren, M. M., Bietz, S., Nittinger, E., and Rarey, M. mRAISE: An Alternative Algorithmic Approach to Ligand-Based Virtual Screening. *J. Comput. Aided Mol. Des*, 2016, *30*(8), 583-594.

60. Falaguera M.J.; Mestres J. Congenericity of Claimed Compounds in Patent Applications. *Molecules* 2021, 26, 5253.

61. Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of Log P Methods on More Than 96000 Compounds. *J. Pharm. Sci.* 2009, 98, 861– 893.

62. Bento, A.P.; Gaulton, A.; Hersey, A.; Bellis, L.J.; Chambers, J.; Davies, M.; Krüger, F.A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J.P. The ChEMBL Bioactivity Database: An Update (version 26) *Nucleic Acids Res.*, 42 (2014), pp. D1083-D1090.

63. Wildman, A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* 1999, 39, 868-873.

64. Meyers, J.; Fabian, J.B.; Brown, N. De Novo Molecular Design and Generative Models. *Drug Disc. Today*, 2021, 26, 2707-2715.

65. Langevin, M.; Minoux, H.; Levesque, M.; Bianciotto, M. Scaffold-Constrained Molecular Generation. *J. Chem. Inf. Model.* 2020, 60, 5637-5646.

66. Brenner, S. Sequences and Consequences. *Phil. Trans. R. Soc. B.*, 2010, 365, 207–212.

67. Gao, W., Coley, C.W. The Synthesizability of Molecules Proposed by Generative Models. *J. Chem. Inf. Model.* 2020, 60, 12, 5714–5723.

68. Jin,W., Barzilay, R., Jaakkola, T. Composing Molecules With Multiple Property Constraints https://arxiv.org/pdf/2002.03244v1.pdf

69. Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M.K., Greenwood, J. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.*, 2015, 137, 2695–2703.

70. Konze, K.D.; Bos, P.H.; Dahlgren, M.K.; Leswing, K.; Tubert-Brohman, I.; Bortolato, A.; Robbason, B.; Abel, R.; Bhat, S. Reaction-Based Enumeration, Active Learning, and Free Energy Calculations To Rapidly Explore Synthetically Tractable Chemical Space and Optimize Potency of Cyclin-

Dependent Kinase 2 Inhibitors. *J. Chem. Inf. Model.* 2019, 59, 3782-3793.

71. Ghanakota, P.; Bos, P.; Konze, K.; Staker, J.; Marques, G.; Marshall, K.; Leswing, K.; Abel, R.; Bhat, S. Combining Cloud-Based Free-Energy Calculations, Synthetically Aware Enumerations, and Goal-Directed Generative Machine Learning for Rapid Large-Scale Chemical Exploration and Optimization. *J. Chem. Inf. Model.* 2020, 60, 9, 4311–4325.

72. Böhm, H.-J.; Flohr, A.; Stahl, M. Scaffold Hopping. *Drug Disc. Today: Tech*. **2004**, 1, 217-224.

73. Lauri, G.; Bartlett, P.A. CAVEAT: A Program To Facilitate the Design of Organic Molecules. *J. Comput. Aided Mol. Des.*, **1994**, 8, 51-66.

74. BROOD; OpenEye Scientific Software Inc.: Santa Fe, NM, U.S., https://www.eyesopen.com/brood. (accessed: 2021-12-17).

75. Johnson, A. M.; Maggiora, G. M. (1990). *Concepts and Applications of Molecular Similarity*. New York: John Willey & Sons. ISBN 978-0-471-62175-1.

76. Martin, Y. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* 2002, 45, 19, 4350-4358.

77. Evans, B.E.; Rittle, K.E.; Bock, M.G.; DiPardo, R.M., Freidinger, R. M., Whitter, W. L., Lundell, G. F., Veber, D. F., Anderson, P. S., Chang, R. S. L., Lotti, V. J., Cerino, D. J., Chen, T. B., Kling, P. J., Kunkel, K. A., Springer, J. P., Hirshfield, J. Methods for Drug Discovery: Development of Potent, Selective, Orally Effective Cholecystokinin Antagonists. *J. Med. Chem.* 1988, 31, 2235-224.

78. Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V.; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; Volkov, Y.; Zholus, A.; Shayakhmetov, R.R.; Zhebrak, A.; Minaeva, L. I.; Zagribelnyy, B. A.; Lee, L. H.; Soll, R.; Madge, D.; Xing, L.; Guo, T.; Aspuru-Guzik, A. Deep Learning Enables Rapid Identification of Potent DDR1 Kinase Inhibitors. *Nat. Biotechnol*. 2019, 37, 1038–1040.

79. Walters, P., Murcko, M. Assessing the Impact of Generative AI on Medicinal Chemistry. *Nat. Biotechnol*. **2020**, 38, 143–145.

80. Palazzesi F., Pozzan A. (2022) Deep Learning Applied to Ligand-Based De Novo Drug Design. In: Heifetz A. (eds) Artificial Intelligence in Drug Design. Methods in Molecular Biology, vol 2390. Humana, New York, NY. pp 273-299.

81. Brown, N., Marco Fiscato, M., Segler, M.H.S., Vaucher, A. C. GuacaMol: Benchmarking Models for De Novo Molecular Design. *J. Chem. Inf. Model.* 2019, 59, 3, 1096-1108.

82. Winter, R., Retel, J., Noe, F., Clevert, D.-A, Steffen, A. Grünifai: Interactive Multiparameter Optimization of Molecules in a Continuous Vector Space. *Bioinformatics*, 36 (20202), 4093-4094.