Automated Drug Design

Project Proposal

Claire Fielden
University of Cape Town
FLDCLA001@myuct.ac.za

Sibusiso Dlamini
University of Cape Town
DLMSIB054@myuct.ac.za

1. Introduction

Automated Drug Design (ADD) is the application of computational methods to the process of drug discovery. ADD can be envisioned as a search problem, where chemical space in the range of 10^{26} and 10^{60} must be traversed for potential candidates, or "hits"[20]. A desirable candidate will affect the regulatory action of the target receptor. The mechanism of binding between ligand and receptor is customarily a combination of hydrogen bonding, electrostatic attractions, and van der Waals interactions.

Designing a novel ligand is a complex process in which chemical compounds are identified for having *potential* as therapeutic agents. Currently, machine learning algorithms are being used to investigate how *in vitro* experiments can be accelerated. This proposal will put forward prospective solutions to the current problems being faced in ADD.

1.1. Sparse Intrinsic Rewards

Physical properties such as LogP and Quantitative Estimate of Drug-Likeness (QED) can be calculated directly from the molecular structure of a ligand in a computational environment. However, in this same environment, binding affinity (BA) between a novel compound and receptor can only be determined after molecular docking occurs. In addition to this, specific bioactivity for binding is only expressed by a small number of molecules. Based on these two aspects, the Sparse Rewards Problem inhibits the RL agent from learning an adequate policy.

In Maccallam's work [16], a Double Deep Q Network (DDQN) was used to train an agent to generate novel ligands. The agent was given 5000 episodes, in which each episode consisted of 40 steps. At each step, the agent chose to add or remove a bond from its current molecular graph. The agent was provided a reward, in the form of a Docking Score, only at the end of the 40-step rollout trajectory. This proved to be an insufficient mechanism for the extraction of a useful policy. Henceforth, the Docking Score (DS) of the generated ligands failed to increase within 5000 episodes. In fact, the DS displayed a decline and comparably inferior performance than that achieved when the agent chose actions at random. To rectify this issue, sparse rewards were replaced with a random walk in which certain functional groups were prioritized. However, convergence time, as well as the molecule the agent converged on, were still not satisfactory when using this method.

Alternatively, reward shaping was applied to distribute the terminal state reward across the preceding transitions. This created dense reward signals from sparse extrinsic rewards for each rollout trajectory. Subsequently, there was a significant improvement in the convergence time of the agent, as well as higher DS rewards. However, the graphs generated by the agent contained a substantial number of features that would not be applicable to *in vitro* scenarios. This is due to the fact that the reward function considered DS in isolation, with no consideration being made for QED or Synthetic Accessibility (SA).

1.2. Inaccurate Docking Score

Maccallum [16] observed, through the use of Autodock-GPU, a weak correlation between DS and IC50 due to a faulty scoring function. More precisely, docking seemed incapable of accurately assessing BA due to the fact that ligands and target receptors were both represented as rigid bodies. This issue is addressed in MolAICal [1], where binding is more than just a thermodynamic endeavor, but one that must sample solvent and off-target interactions with endogenous biomolecules as well [6]. This has rendered the docking score unreliable, which may be the match in the powder barrel.

Another limiting factor was the amount of time docking between ligand and receptor consumed. For each ligand produced, the agent has a minimal amount of time to be trained, which reduced the amount of exploration it could perform. Thus, one

can confidently conclude that the lowest-energy conformer for a given ligand was most probably not obtained.

Modern Drug discovery approaches have shifted towards using Computer Aided Drug Design methodologies such as generative models and virtual screening with the aim of lowering the astronomical costs of drug discovery campaigns. Both Virtual Hit Screening and generative Artificial Intelligence in *De Novo* Drug Design rely on the accuracy of scoring functions for their productivity. Scoring functions are responsible for predicting the binding pose of a protein-ligand complexes and predicting their Binding Affinity. The binding pose prediction problem has been solved [23] but Binding Affinity prediction remains one of the most difficult challenges in computational chemistry. Due to the computational constraints of classical scoring functions, their progress has stalled and has been overtaken by data-driven ML-based scoring functions [28]. An example of this is the NNScore2.0 scoring function; A Neural Network-based scoring function which uses energy terms computed by classical scoring functions and binding characteristics from the BINANA algorithm as its descriptors [8]. The BINANA algorithm in particular uses summations of ligand-protein atoms to compute electrostatic interactions. Not only is this form of representation based on the incorrect assumption that energy terms of ligand-protein atoms are additive [3] but they also lose some structural information about protein-ligand complexes.

2. Research Questions

2.1. Can Curriculum Learning (CL) improve the current RL agent's convergence time whilst maintaining the synthesizability of the produced ligand?

The goal of introducing a new reward shaping algorithm is to identify whether the agent can converge on an optimal ligand in a shorter amount of time. A machine learning model converges when additional training will not improve its prediction capabilities.

The implementation of curriculum objectives to measure drug-likeness (QED) and SA will be discussed in Section 3.1.1. These objectives will guide the agent to generate graphs of compounds containing less unsynthesizable features. The goal is to achieve a QED score of 0.872, and a DS of -16kcal/mol, using fewer epochs and thus less time. Further discussion of implementation can be found in section 6.1.

2.2. Will extending the descriptors of an existing Neural Network-based scoring function to include the properties of protein targets yield a more accurate binding affinity predictions than the original scoring function?

Recent studies have shown ML-based scoring functions (MLSFs) to be superior to their classical counterparts due to the growing availability of (Protein-Ligand complEx with known Binding Affinity of Structure) PLEXBAS data. Hence the decision to focus on Neural Network based scoring functions. This project aims to improve Neural Network based scoring functions based on the following theories and assumptions.

The most accurate binding affinity predictors use various representation strategies to enable their models to learn more relationships between properties of protein-ligand complexes and their Binding Affinity [14, 22]. Geometric and structural representation strategies such as

The increasingly available PLEXBAS data should improve the Scoring Functions' performance. State-of-the-art predictive tools are trained on millions of data points whereas the CASFv2018 data set contains 4,463 complexes in its refined set. NNScore2.0. which was only developed in 2011 only had access to CASFv2009 which had a total of 1,741 complexes in its refined set [2]. Therefore a new scoring function trained on a larger data set would have an advantage over older models and produce more accurate Binding Affinity predictions. This should further confirm the ability of Neural Networks to learn from larger data sets [3, 23].

The last theory is that training on a diverse set of PLEXBAS data will improve the predictive power of ML-based Scoring functions. A study conducted by Li et al. [15] experimented with the RFScore scoring function by training it on increasingly less diverse test sets which saw significant improvement in the Binding Affinity predictions of scoring functions. However, restricting the training set according to similarity cutoffs reduces the available PLEXBAS data points that a scoring function can be trained on. Therefore this theory contradicts the previous one since training on fewer data should produce less accurate scoring functions. This training tactic was tested on a Random Forrest implementation of a scoring function and thus it remains to be seen whether Li et al.'s [25] studies are transferrable to Neural Network-based scoring functions.

3. Procedures and Methods

3.1. Curriculum Learning

CL consists of two main components, namely:

1. Curriculum Progression Criteria
2. Production Objective

The agent must satisfy a specified score for an objective before moving on to the subsequent, more strenuous task. The curriculum must be learnt before it can progress to the Production Phase, where the agent is to achieve its ultimate goal.

3.1.1. Curriculum Phase

The number of curriculum epochs for each curriculum objective will not be limited. Curriculum epochs are computationally inexpensive and thus do not affect the overall training time of the agent.

Maccallam et al. [16] benchmarked their findings against the Half maximal inhibitory concentration (IC50) values obtained between Napthyridine, Aminopyridine (SFK40), Imidazopyridazine (SFK52) and the adenosine triphosphate binding site of plasmodium falciparum phosphatidylinositol 4-kinase. Therefore, as the first curriculum objective, the agent can be guided to produce compounds that display a high Tanimoto Similarity to these already-existing ligands.

Objective One in the agent's Curriculum Progression Criteria will be to achieve a Tanimoto Score of 0.75. Objective Two will be to achieve a QED score of 0.872, to be measured using a module provided by RDKit. These values were obtained from better-performing agents in previous work by Guo et al. [11] and Maccallam et al.[16].

3.1.2. Production Phase

The agent will be given a maximum number of production episodes, 4500. Each episode will start by training the agent on the Curriculum Progression Criteria and end with a molecular docking simulation using AutoDock GPU. The DS provided at the terminal state will act as the agent's final, extrinsic reward.

Maccallam et al. [16] implemented a potential-based reward shaping algorithm, in which the training curve of a better-performing agent progressed from a -6 kcal/mol DS to -16 kcal/mol within 2500 episodes. In order to observe the efficiency of CL in improving convergence time, these metrics must be met or improved.

3.2. Improvement to Docking Score

3.2.1. Neural Network Scoring function

The "Scoring power" of a scoring function is a scoring function's correlation with experimentally determined Binding Affinity data. This correlation is obtained by calculating the Pearson correlation coefficient [13] (Rp). A Pearson correlation coefficient takes on values between 0 and 1 which denote low and high accuracies respectively. Experimentally determined Binding Affinity values that the predictions will be correlated with will be obtained from the PDBbind database [14]. The test set will be the popular CASFv2016 core set [7] as this will allow for comparisons for a range of scoring functions. The NNScore2.0. will be used as a control and the goal will be to create a scoring function that produces a Pearson Correlation Coefficient of 0.7 Rp or more. The NNScore2.0 function will be extended to include protein descriptors and its hyperparameters will be tuned. Three separate training sets of the scoring functions will be developed and each of them will be distinguished by their similarity cutoffs and total size. A total of 4 scoring functions will be evaluated and their results will be analysed to draw comparisons.

4. Ethical, Professional and Legal Issues

No forseeable legal issues as all modules to be used such as RDKit and AutoDock GPU, are open source. There is no need for human involvement in these experiments and the forseeable outcomes of the experiment do not pose any ethical issues. Furthermore, the Open Source Software guidelines will be adhered to, as outlined by Grodzinsky et al. [10] and as such will use all above-mentioned frameworks in a manner that is ethical and professional.

5. Related Work

5.1. Sparse Reward Amelioration Strategies

A common way to predict the binding affinity of a novel ligand is through quantitative structure-activity relationship (QSAR) models trained on historical data [12] for a protein of interest. Training a network to optimize the potency of generated molecules against a desired receptor will produce inactive molecules in most cases. Under these conditions, the agent fails to maximize the BA of generated ligands.

Korshunova et al. [12] highlighted that the Sparse Rewards Problem presents an agent with difficulties in exploring chemical space appropriately. Therefore, a promising molecule with high bioactivity for a protein of interest is usually undiscoverable. This study found that, in the absence of rewards from active molecules, one of three optimizations could be applied. These are, namely, real-time reward shaping, fine-tuning by transfer learning, as well as experience replay.

Unlike bandit problems [17], which balance actions which are immediately rewarding, RL settings require planning over several time steps.

This, however, does not mean bandit problems are completely useless. As seen by AIRS [27], intrinsic rewards are incorporated to guide the agent towards the ultimate extrinsic reward. Singh et al. [26] combines a variety of extrinsically rewarded tasks that can be learned as the agent develops skills to achieve novel events. The key aspect of this research is that intrinsic reward is only generated by unexpected salient events.

Other previous work, such as that by Bellemare et al. [4] leverages a density model to approximate the frequency at which the agent enters certain states, thus defining the intrinsic reward as inversely proportional to the pseudo-count. Alternatively, Puthak et al. [18] uses curiosity-driven exploration to utilize prediction error as an intrinsic reward. An inverse-forward dynamics model is used to learn the representation of a state space, where the intrinsic reward is based on the prediction error of the encoded next-state. Similarly, RIDE [19] uses curiosity by setting the intrinsic rewards as the difference between two consecutive encoded states. The agent is thus encouraged to choose actions that result in significant state changes. This is useful when choosing the next Molecular Fingerprint from a given state, as it provides aggressive exploration incentives.

Guo et al. [11] has identified the above-mentioned costs when using Reinforcement Learning for DNDD. CL, however, is a suitable alternative as the agent may decompose the complex objective of a high BA into simpler constituent objectives. This is predicted to accelerate convergence time as corresponding gradients from sequential simpler tasks are more effective at traversing the optimization landscape.

5.2. Scoring functions

Binding Affinity, a measure of the strength between two binding molecules, is affected by several factors including changes in free energy. Predicting changes in free energy is one of the most challenging chemistry problems [24]. There is generally a trade-off between computational efficiency and the accuracy of scoring functions. Computational efficiency is required to screen the plethora of compounds in high throughput screening and in generative models within a reasonable amount of time. Accuracy is required to ensure that the best drug candidates are correctly identified as "hits", and to ensure that generative models that optimise BA produce true strong binding candidates.

Of the available classical scoring functions, the physics-based scoring functions have shown the most potential to reliably predict BA [**Liu2015**]. Unsurprisingly they are the most computationally expensive category of classical scoring functions. MLSFs have and continue to show promise. The rapid advances in AI and increasing availability of data MLSFs further highlight its potential.

5.3. Benchmark metrics

"Docking power", "Ranking power" and "Scoring power" are the standard metrics used to assess a Scoring Functions ability [8]. Docking power is a scoring function's ability to correctly identify the native binding pose of a protein-ligand complex among a set of decoys. A molecule's Binding Affinity changes depending on its binding pose, so a scoring function should be able to correctly identify the binding pose with the highest binding affinity. Ranking power is a scoring function's ability to rank multiple ligands for a given receptor according to their binding affinities. Scoring power is a measure of a scoring function's predictive power. Along with computational efficiency, these metrics are the standard for evaluating comparisons and gauging a scoring function's viability.

6. Anticipated Outcomes

6.1. System and Design Challenges

6.1.1. Curriculum Learning

As per the DDQN framework designed by Maccallam et al. [16], the agent will be implemented as a pair of fully connected deep ANNs. In order to accommodate for experience replay, a replay memory buffer will be allocated to the agent. As the replay memory is not adaptive, it will not provide optimal performance, however in this context, its minuscule size is unlikely to cause critical performance issues.

For the sake of computational efficiency, the generative model will work with Molecular Fingerprints during training. These graphs are utilized as *.mol* files and very rarely exceed 4KB. Thus, space requirements are not an issue. Each accessible state is considered relative to the number of steps the agent has left in which to modify its molecular graph. These steps are limited to 40 as in previous work [29].

In the production phase, AutodockGPU will be utilized to return the ligand's binding affinity (BA). One molecular docking simulation takes, on average, one minute [21]. Therefore, running Autodock simulations will be the most computationally expensive aspect when CL experiments take place.

6.1.2. Scoring Functions

Evaluating scoring functions will be automated with the help of various tools. The test sets will

be downloaded from the PDBbind database to obtain proteins in *.pdb* format and ligands in *.mol2* format. AutoDock MGLtools will be used to prepare the files to produce *.pdbqt* files. The *.pdbqt* format is an extension of the *.pdb* format which includes the partial charges of each atom, the atom types, assigned rotatable bonds and the number of torsional degrees of freedom. The structural similarity measurement programs and various scoring functions are written in different programming languages and therefore will be automated with a shell script.

6.2. Impact

The first improvement to current work is the acceleration of convergence. Complex reward functions often result in minima that are difficult to find. Therefore, the resulting small gradients elicit minimal change to the agent policy. Policy-based RL can be infeasible for complex MPO objectives, leading to suboptimal allocation of computational resources. Curriculum Learning will be used to encourage exploration and improve productivity such that the agent samples from diverse local minima.

The second improvement to the current work is the augmentation of molecules identified for synthesis. A recurring issue amongst the literature, particularly identified by Maccallam [16], is the generation of infeasible atomic arrangements. Having a multivariate reward function is necessary to filter out unrealistic features which occur in a macromolecule when DS is rewarded in isolation. Tanimoto Similarity, as used by PaccMannRL [5] and Zhenpeng et al. [29], combined with QED, is a favourable mechanism for identifying utilizable functional analogues.

6.3. Success Factors

- Improvement of synthesizability of ligands produced.
- Acceleration of agent convergence, henceforth improvement of productivity.
- Increase in scoring power and or ranking power of the newly implemented scoring function relative to its original.

7. Project Plan

7.1. Risks

The project risks are given in the risk matrix in Appendix A.

7.2. Timeline

The project will run over eight project weeks, beginning 24th July and ending with the submission of the web page on the 16th of October. The scheduling and division of work over this time along with the dates of important milestones are detailed in the Gantt chart found in Appendix B.

7.3. Resources Required

The following resources will be required to complete this project.

- Open-source scoring functions (AutoDock suite).
- Protein-ligand databases with known binding affinity data (PDBbind, DUD-E).
- Structural similarity measurement software (TM-align)
- Protein family classifier tool (Interpro).
- Open-source Python libraries (RDKit).
- Personal laptop and programming software.

7.4. Deliverables

The main deliverable for this project is the final research paper, which will explore whether the research objectives stated in this proposal have been achieved. In addition to this, the other deliverables are:

- Two literature reviews
- Project proposal
- Project progress demonstration
- Draft of final paper
- Final project report
- Final code submission
- Final project demonstration
- Project website
- Project poster

7.5. Milestones

Milestones be found on the Gantt chart in Appendix B. The milestones of most significance are the completion and submission of the Literature Reviews and Project Proposals in the Project Planning phase, and the Project Code and Final Paper in the Project Execution phase.

7.6. Work Allocation

This project has been siloed into two sub-projects in which both team members will focus on achieving their own objectives with no reliance on each other. Thus, Claire will be responsible for implementing a curriculum learning algorithm while Sibu trains and evaluates scoring functions.

Appendix A RISK MATRIX

.

Risk Condition	Consequence	Prob	Impact
The integration of the text representation and sequence model of the Curriculum Learning framework cannot be successfully integrated with the molecular graph implementation of the original work's DDQN.	The framework at hand will be insufficient and inoperable.	Moderate	High
The available computational resources cause docking simulations to take an extended period of time	The project is hindered by lack of time, causing appropriate experiments to not be completed on time.	Low	Critical
Curriculum objectives are insufficient in enabling the agent to achieve comparable docking scores to previous work.	The agent does not receive the training necessary to complete the design task of hit-to-lead optimization	Moderate	Medium
A member of our team is unable to contribute to the project in a meaningful capacity.	Project deliverables will depend entirely on a single team member.	Moderate	Low.
Software incompatible i.e. What if one of software suites are incompatible with Mac?	A non-functional product that is unable to be made use of.	Low	Critical
Current computer resources are unable to process large amounts of data at an acceptable rate.	Poor runtime performance will impact the users experience negatively.	High	Medium
Docking simulations take too long to perform adequate training of ANN.	Late delivery/inability to deliver a finished product.	Moderate	High

Table 1: Risk Probability and Impact

Risk Condition	Mitigation	Monitoring
The integration of the text representation and sequence model of Curriculum Learning framework cannot be successfully integrated with the molecular graph implementation of the original work's DDQN.	A new framework must be designed to accommodate both Curriculum Learning and a DDQN.	Define a table of experiments with expected outcomes and employ debugging techniques to ensure the framework is operating as expected.
The available computational resources cause docking simulations to take an extended period of time	Request access to HPC Cluster	Create a schedule that budgets time appropriately for the experiments and ensure experiments that must be completed are done within the allocated timeframe.
Curriculum objectives are insufficient in enabling the agent to achieve comparable docking scores to previous work.	Research and redefine the curriculum objectives, or add more curriculum objectives, that may enhance the agent's ability to explore the optimization landscape.	Test the agent's convergence time starting with fewer episodes to ensure that the given curriculum can decrease convergence time before spending too much time on docking simulations.
A member of our team is unable to contribute to the project in a meaningful capacity.	Schedule to get as much work done as early as possible to avoid last-minute failure, thus leaving enough time for the other member to complete the task if this does happen.	Team members will keep an open line of communication so that if anything occurs, everyone is prepared for a larger workload.
Software incompatible i.e. What if one of software suites are incompatible with Mac?	Thorough investigation of the programming modules being used, ensuring that it is compatible with MacOS.	Prototyping as the application develops.
Current computer resources are unable to process large amounts of data at an acceptable rate.	Poor runtime performance will impact the users experience negatively.	Test application on multiple, varying platforms.

Table 2: Risk Mitigation and Monitoring

Appendix B GANTT CHART

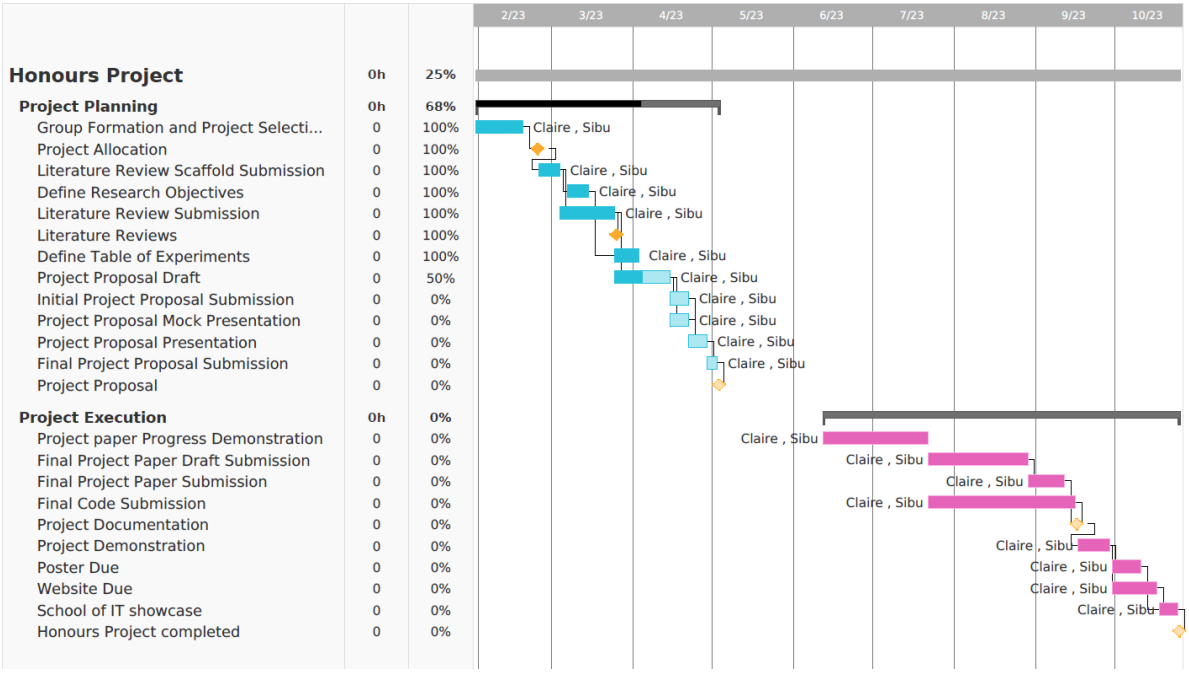

Figure 1: Milestones and Schedule

Glossary

a priori Latin: "from what comes before" / "from first principles, before experience". Refers to reasoning and deductions which follow from theory rather than empirical data

in vitro Latin: "in the glass". Refers to experiments conducted outside of a biological context, usually in some form of assay

binding affinity The extent to which a drug binds to a target receptor at any given concentration, otherwise known as the "firmness" with which the drug binds to the receptor

convergence When the loss of a machine learning model settles within a near-zero range during training. At this point, there are little to no changes in the model's learning rate

Curriculum Learning A reward shaping algorithm[11] that optimizes the training process of an RL agent, specifically by providing curriculum objectives that become more difficult as the agent becomes less naive

Deep Neural Network A neural network that selects the best action for an agent to take based on the maximum Q-value of the next state. The Q-Network is optimized towards a target network that is periodically updated with the latest weights every (k) steps, where (k) is a hyperparameter

Docking Score The scalar value returned from Autodock-GPU which provides an estimate of the () between a ligand and target receptor

Double Deep Q Network A Deep Neural Network that implements Double Q-Learning

Double Q-Learning An algorithm in which two different action-value functions, Q and Q', are used as estimators. It is preferable to basic Q-learning as it solves the problem of over-estimations of action value

electrostatic attractions The attractive force defined by Coulomb's law occurring between oppositely charged nueclei, atoms or molecules

episode A sequence of actions taken from an initial state, ending in a terminal state, in which the agent receives a reward. In a single episode, the agent interacts with the environment according to a particular policy

functional analogues Molecules which share a structure that is similar enough that they can be expected to exhibit similar chemistry

functional group A moiety in a molecule that causes characteristic chemical reactions

Half maximal inhibitory concentration An indication of how much drug is needed to inhibit a biological process by half. A measure of ligand potency and efficacy

hydrogen bonding A weak bond between two molecules resulting from an electrostatic attraction between a proton in one molecule and an electronegative atom in the other

ligand In the context of drug design, this is a molecule which binds to a receptor. Note that this is slightly different to the way the term is used in coordination chemistry, where it refers to a molecule which binds to a metal atom forming a coordination complex

LogP The partition coefficient of a molecule between aqueous and lipophilic phases usually considered as octanol and water, defining a drug's solubility

molecular docking An *a priori* simulation of binding between a target macromolecule and a ligand, using a Docking Score to rank candidate dockings

Molecular Fingerprint The bitstring representation of a chemical structure.

policy A policy π(s) comprises the suggested actions that the agent should take for every possible state s$\in$ S. It is thus a strategy the agent uses to pursue goals

productivity By exploiting knowledge retention, the agent is able to find favourable areas of chemical space at a faster rate. Thus, more favourable compounds can be sampled in a reduced amount of time

Quantitative Estimate of Drug-Likeness The qualitative measure of drug-likeness, directly proportional to the desirability of a molecule. It reflects the underlying distribution of molecular properties such as molecular weight, LogP, and the presence of unwanted chemical functionalities.

random walk A stochastic process in which the current observation is equal to the previous observations with a random modification

receptor The region within a macromolecule where another, smaller molecule will bind thus operating as a transducer of biological signals

Reinforcement Learning A feedback-based Machine learning technique in which an agent learns to behave in an environment by taking actions that will maximize a pre-defined reward.

replay memory buffer A data structure that temporarily saves the agent's observations, allowing experiences to be updated multiple times whilst the agent is trained.

reward shaping In order to ameliorate the Sparse Rewards Problem, small intermediate rewards are created in order to accelerate algorithm

rollout trajectory Smaller sequences of experience that are stored in an experience replay buffer, an accumulation of which make up a episode

Sparse Rewards Problem An issue usually associated with Reinforcement Learning algorithms, in which an agent does not receive a sufficient amount of feedback from its environment when taking an action.

synthesizability The measure Synthetic Accessibility based on structure complexity and similarity, as well as synthetic pathways[9]

Synthetic Accessibility The measure of the ease of synthesis of a chemical compound.

Tanimoto Similarity The measure of how similar two molecules are, given their Molecular Fingerprint, based on common bits and the resulting value of the Tanomoto coefficient.

target receptor A biological target to which a ligand or drug binds, resulting in a change in its behaviour

terminal state The final step in an agent's rollout trajectory

van der Waals interactions The attractive or repulsive forces occurring between molecules due to their polarized electron clouds

Acronyms

ADD Automated Drug Design

BA binding affinity

CL Curriculum Learning

DDQN Double Deep Q Network

DNDD De Novo Drug Design

DS Docking Score

IC50 Half maximal inhibitory concentration

MPO Multi-Parameter Optimization

QED Quantitative Estimate of Drug-Likeness

QSAR quantitative structure-activity relationship

RL Reinforcement Learning

SA Synthetic Accessibility

References

[1]Q. Bai, S. Tan, T. Xu, H. Liu, J. Huang, and X. Yao, *Molaical: a soft tool for 3d drug design of protein targets by artificial intelligence and classical algorithm*, 3 (2020), 10.1093/bib/bbaa161.

[2]P. J. Ballester and J. B. O. Mitchell, *A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking*, 9 (2010), pp. 1169–1175.

[3]B. Baum, L. Muley, M. Smolinski, A. Heine, D. Hangauer, and G. Klebe, *Non-additivity of functional group contributions in protein–ligand binding: a comprehensive study by crystallography and isothermal titration calorimetry*, 4 (2010), pp. 1042–1054.

[4]M. G. Bellemare, S. Srinivasan, G. Ostrovski, T. Schaul, D. Saxton, and R. Munos, *Unifying count-based exploration and intrinsic motivation* (2016).

[5]J. Born, M. Manica, A. Oskooei, J. Cadow, G. Markert, and M. Rodríguez Martínez, *Paccmann-nrl: de novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning*, 4 (2021), p. 102269, 10.1016/j.isci.2021.102269.

[6]C.-e. A. Chang, *Understanding ligand-receptor non-covalent binding kinetics using molecular modeling*, 6 (2017), pp. 960–981, 10.2741/4527.

[7]T. Cheng, X. Li, Y. Li, Z. Liu, and R. Wang, *Comparative assessment of scoring functions on a diverse test set*, eng, 4 (WASHINGTON, 2009), pp. 1079–1093.

[8]J. D. Durrant and J. A. McCammon, *Nnscore 2.0: a neural-network receptor-ligand scoring function.* 11 (Nov. 2011), pp. 2897–2903.

[9]W. Gao and C. W. Coley, *The synthesizability of molecules proposed by generative models*, 12, PMID: 32250616 (2020), pp. 5714–5723, 10.1021/acs.jcim.0c00174.

[10]F. S. Grodzinsky, K. W. Miller, and M. J. Wolf, *Ethical issues in open source software* (2003), pp. 193–205.

[11]J. Guo, V. Fialková, J. Arango, C. Margreitter, J. P. Janet, K. Papadopoulos, O. Engkvist, and A. Patronov, *Improving de novo molecular design with curriculum learning* (June 2022), pp. 1–9, 10.1038/s42256-022-00494-4.

[12]M. Korshunova, N. Huang, S. Capuzzi, D. S. Radchenko, O. Savych, Y. S. Moroz, C. Wells, T. M. Willson, A. Tropsha, O. Isayev, and et al., *A bag of tricks for automated de novo design of molecules with the desired properties: application to egfr inhibitor discovery* (2021), 10.26434/chemrxiv.14045072.

[13]H. Li, G. Lu, K.-H. Sze, X. Su, W.-Y. Chan, and K.-S. Leung, *Machine-learning scoring functions trained on complexes dissimilar to the test set already outperform classical counterparts on a blind benchmark*, 6 (2021).

[14]Z. Liu, Y. Li, L. Han, J. Li, J. Liu, Z. Zhao, W. Nie, Y. Liu, and R. Wang, *PDB-wide collection of binding data: current status of the PDBbind database*, 3 (Oct. 2014), pp. 405–412.

[15]J. Lu, X. Hou, C. Wang, and Y. Zhang, *Incorporating explicit water molecules and ligand conformation stability in machine-learning scoring functions*, 11 (2019), pp. 4540–4549.

[16]R. Maccallam, ≪Automatic hit-to-lead optimization≫, PhD thesis (2021).

[17]I. Osband, C. Blundell, A. Pritzel, and B. V. Roy, *Deep exploration via bootstrapped DQN* (2016).

[18]D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, *Curiosity-driven exploration by self-supervised prediction* (2017).

[19]R. Raileanu and T. Rocktäschel, *RIDE: rewarding impact-driven exploration for procedurally-generated environments* (2020).

[20]G. Schneider, *Automating drug discovery*, 2 (2017), pp. 97–113, 10.1038/nrd.2017.232.

[21]L. Solis-Vasquez, A. F. Tillack, D. Santos-Martins, A. Koch, S. LeGrand, and S. Forli, *Benchmarking the performance of irregular computations in autodock-gpu molecular docking* (2022), p. 102861, https://doi.org/10.1016/j.parco.2021.102861.

[22]M. Su, Q. Yang, Y. Du, G. Feng, Z. Liu, Y. Li, and R. Wang, *Comparative assessment of scoring functions: the casf-2016 update*, 2 (2019), pp. 895–913.

[23]B. Wang, Z. Zhao, D. D. Nguyen, and G.-W. Wei, *Feature functional theory–binding predictor (fft–bp) for the blind prediction of binding free energies*, 4 (2017), pp. 1–22.

[24]C. Wang and Y. Zhang, ≪Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest≫, Journal of computational chemistry **38**, 169–177 (2017).

[25]R. Wang, L. Lai, and S. Wang, *Further development and validation of empirical scoring functions for structure-based binding affinity prediction*, 1 (2002), pp. 11–26.

[26]R. J. Williams and D. Zipser, *A learning algorithm for continually running fully recurrent neural networks*, 2 (June 1989), pp. 270–280, https://doi.org/10.1162/neco.1989.1.2.270.

[27]M. Yuan, B. Li, X. Jin, and W. Zeng, *Automatic intrinsic reward shaping for exploration in deep reinforcement learning*, 2023.

[28]Z. Zhou, S. Kearnes, L. Li, R. N. Zare, and P. Riley, *Optimization of molecules via deep reinforcement learning*, eng, 1 (England, 2019), pp. 10752–10752.

[29]Z. Zhou, S. Kearnes, L. Li, R. N. Zare, and P. Riley, *Optimization of molecules via deep reinforcement learning*, 1 (2019), 10.1038/s41598-019-47148-x.