Target2DeNovoDrug : a novel programmatic tool for deep learning based de novo drug design for a target of interest

Rafal Madaj[b], Ben Geoffrey*[a], Akhil Sanker[c], Pavan Preetham Valluri[d]

[a] *University of Madras, Chepauk, Chennai 600 005, India*
[b] *Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Poland*
[c] *SRM University, Tamil Nadu 603203, India*
[d] *PSG College of Technology, Coimbatore, India*
*Corresponding author email : bengeof@gmail.com

Abstract

The on-going data-science and AI revolution offers researchers with fresh set of tools to approach structure-based drug design problems in the computer aided drug design space. A novel programmatic tool that can be used in aid of *in silico*-deep learning based de novo drug design for any target of interest has been reported. Once the user specifies the target of interest, the programmatic workflow of the tool generates novel SMILES of compounds that are likely to be active against the target. The tool also performs a computationally efficient *In-Silico* modeling of the target and the newly generated compounds and stores the results in the working folder of the user. A demonstrated use of the tool has been shown with the target signatures of Tumor Necrosis Factor-Alpha, an important therapeutic target in the case of anti-inflammatory treatment. The future scope of the tool involves, running the tool on a High Performance Cluster for all known target signatures to generate data that will be useful to drive AI and Big data driven drug discovery. The code is hosted, maintained and supported at the GitHub repository given in link below

https://github.com/bengeof/Target2DeNovoDrug

Introduction

The on-going data-science and AI revolution offers researchers with fresh set of tools to approach structure-based drug design problems in the computer aided drug design space[1-11]. In the pre-AI era, one approached structure-based drug design using structure-activity and in silico modelling approaches that were available in the cheminformatic, bioinformatics and computational biology space[12-17]. However with the availability of Big Data in the aforesaid areas, it is now possible to apply data-driven, machine/deep learning approaches to these areas[18-20]. In our previous effort, Target2Drug [21], we have implemented an automated computer aided drug discovery protocol which incorporates machine/deep learning approached in addition to approaches available from the bioinformatics and computational biology space. The tool Target2Drug in an automated drug discovery tool which identifies in silico best drug candidates from the chemical space of all PubChem compounds for a particular drug target specified by the user. In our present work, Target2NewDrug we incorporate a deep learning based de novo drug design protocol to generate new compounds as drug candidates that are not presently known and reported on PubChem for a particular drug target specified by the user. Currently, one observes from literature two categories of approaches in deep learning based de novo drug design : Generative graph neural networks that use molecular graph representations and generative recurrent and adversarial neural network that use string representation of molecules in the form of SMILES for training and generation of new molecules. Owing to the computational and the memory usage efficiency offered by the latter approach, we incorporate a generative LSTM neural network approach for deep learning based de novo drug design in our Target2NewDrug tool which in also in keeping with our philosophy to be computational light as possible as it was the case in Target2Drug.

Methods

The algorithmic workflow and architecture of Target2DeNovoDrug is shown Fig.1 below. The workflow begins with the collection of target signatures in the form of amino acid sequence or it's corresponding nucleotide sequence. The BLAST protocol of Target2Drug identifies known targets similar to the input target signature provided by the user and also with PubChem data linked to the target required to perform AutoQSAR based drug lead generation. QSAR studies involve establishing linear and non-linear statistical correlation between experimental activity and molecular descriptors of active compounds. A total of 8 molecular descriptors for each active compound against a target is downloaded from PubChem and a QSAR model is built with all possible combinations and selections of descriptors where ${}^{n}C_{r}$ = 256, n = 8 and r = 2,3,4,5,6,7. The QSAR model with high statistical quality is chosen for prediction by choosing the QSAR model with R^2 value being close to 1. The predicted compounds satisfy the Lipinski's drug likeness criteria. The SMILES of the PubChem compounds with predicted activity against the target from the Target2Drug tool is collected and is used to train a generative LSTM neural network which is an adaptation of Gupta, Anvita et al [22] to generate SMILES of new compounds that are not presently known and registered on PubChem but are likely to be active against the target by virtue of the training data of the neural network being SMILES of compounds active against the target. Further, the tools performs a fast *In Silico* modelling of the interaction of the drug target and compounds predicted by the tool to be active against the target and stores the results in the working folder (Fig.2) To demonstrate the use of the tool, a demonstrated use has been carried out with the target signatures of Tumor Necrosis Factor-Alpha, a drug target of importance for anti-inflammatory disease[23]. The code is hosted, maintained and supported at the GitHub repository given in link below

https://github.com/bengeof/Target2DeNovoDrug

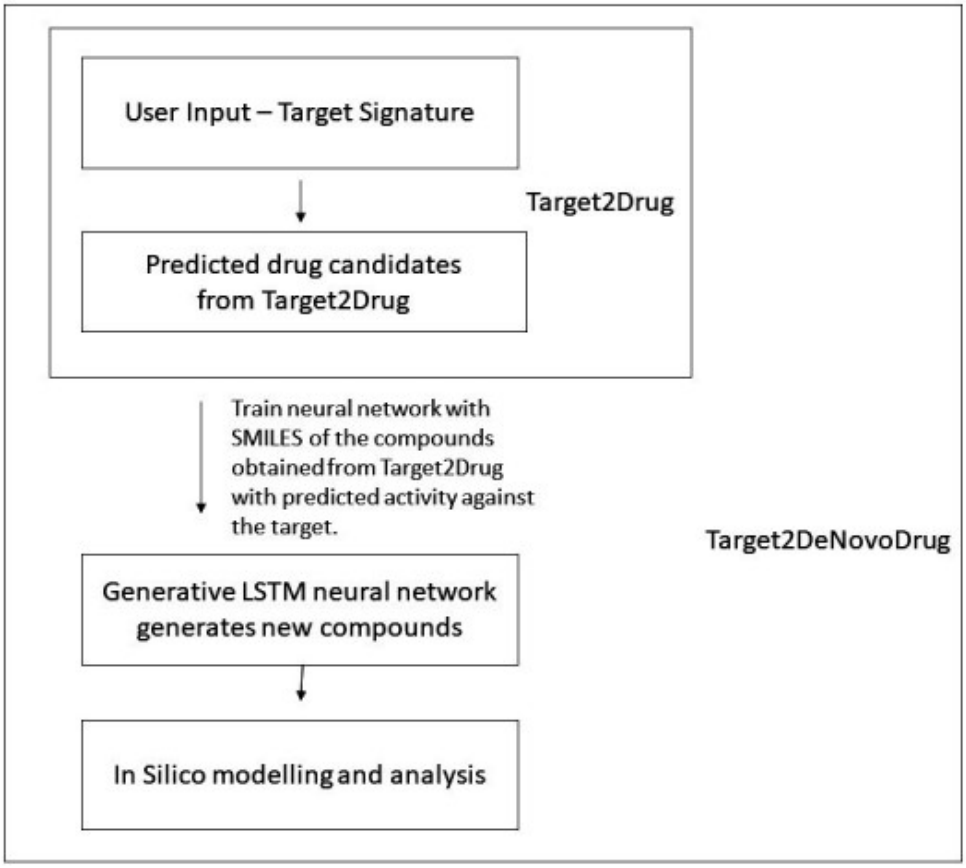


Fig. 1 – Block diagrammtic representation of the algorithm of the tool

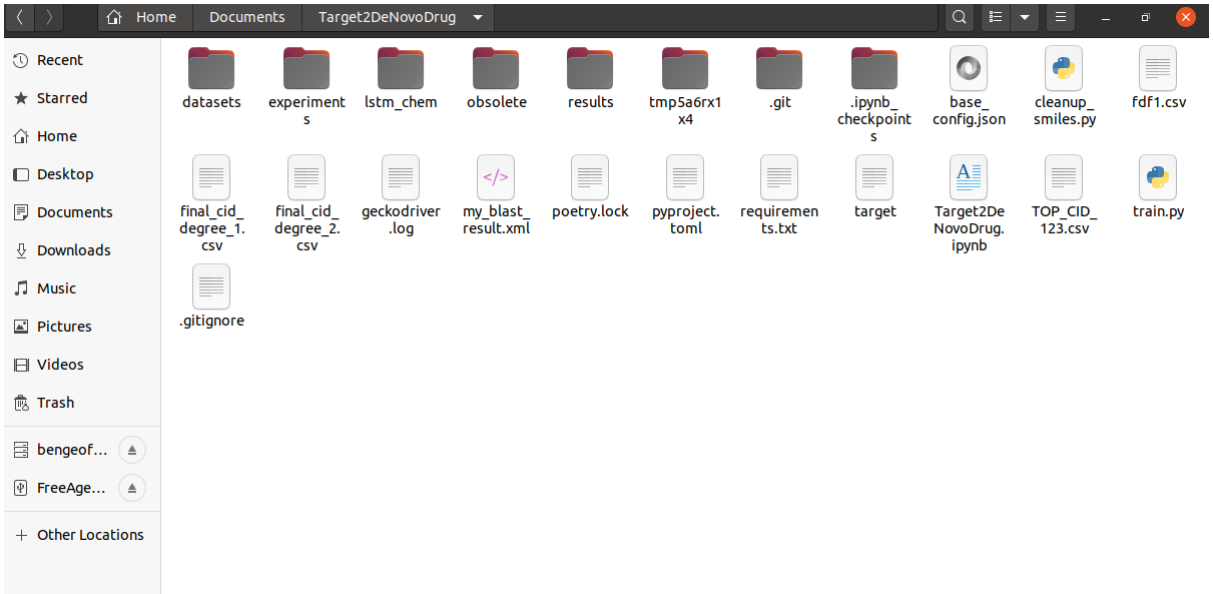
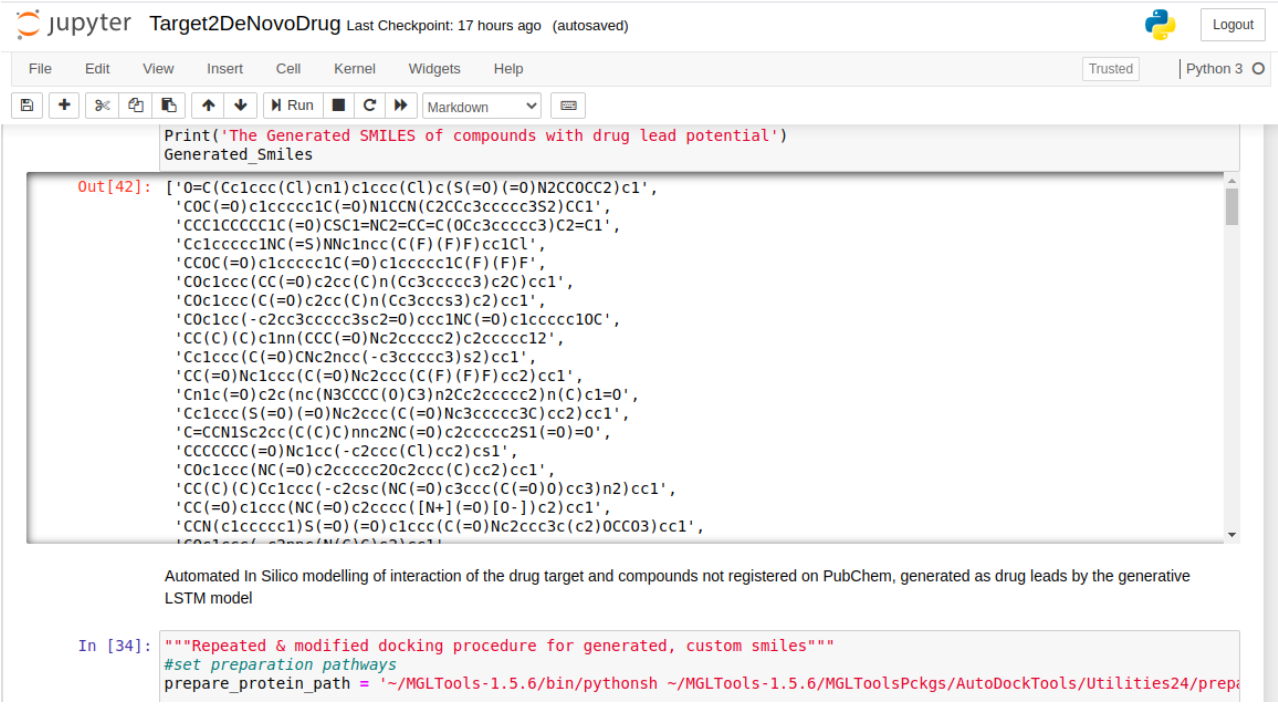


Fig. 2 – Working Folder

Results and Discussion

The target signature of Tumour Necrosis Factor Alpha was given as input to the tool. The target signature can be identified with NCBI protein ID P01375. The Target2Drug tool identified data linked with target available on PubChem required to perform AutoQSAR based drug lead generation. The SMILES of the compounds generated as drug leads were used to train the generative LSTM neural network which was used to generate new compounds currently not known and registered on PubChem but are likely to be active against the target. The 500 novel SMILES were generated out of which 182 were valid SMILES. A computationally efficient *In Silico* modeling was carried out to model the interaction of the target and PubChem compounds generated as drug leads by the Target2Drug protocol and De Novo generated SMILES of the deep learning based DeNovo drug design protocol. The newly Generated SMILES is shown in Fig.3



The binding affinity of the target and PubChem compounds generated as drug is tabulated in Table 1 and the De Novo generated drug leads are shown in Table 2. A selection of drug lead compounds with Binding Affinity lower than -8 kcal/mol is tabulated and the pymol visualization of the target-compound interaction is shown in Fig.4a, b, c, d & e

Table 1 – In Silico Modelling Target2Drug

S.No.	Ligand (PubChem CIDs)	Target	Binding Affinity (Kcal/mol)
1	356283	1a8m	-8.4
2	11547797		-8.1
3	70681601		-8.6
4	129780132		-8.3
5	144048489		-8.4

Table 2 – In Silico Modelling Target2DeNovoDrug

S.No.	Ligand (De Novo generated SMILES)	Target	Binding Affinity (Kcal/mol)
1-SMI	Cc1cccc(NC(=O)Nc2nc3ccccc3nc2NS(=O)(=O)c2ccccc2)c1	1a8m	-8.1
2-SMI	Cc1ccc(NS(=O)(=O)c2ccc(C(=O)NC(=O)c3ccc(C)cc3)cc2)cc1		-8.1
3-SMI	Cc1ccc(S(=O)(=O)Nc2cccc(C(F)(F)F)c2)cc1		-8
4-SMI	CC(C)CNC(=O)c1ccccc1C(=O)c1ccccc1C(=O)c1cccs1		-8.3
5-SMI	O=C(Nc1cccc2cccnc12)Nc1cccc2cccnc12		-8

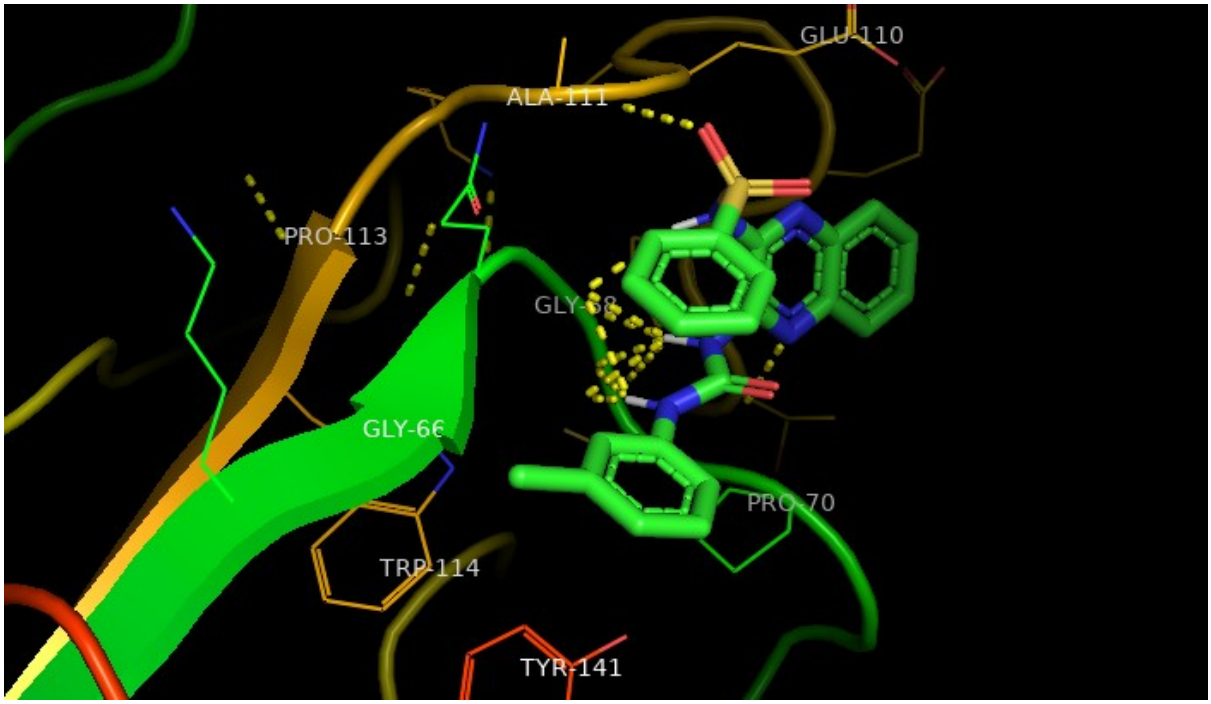


Fig.4a - Target-compound interaction of the De Novo generated SMILES 1-SMI

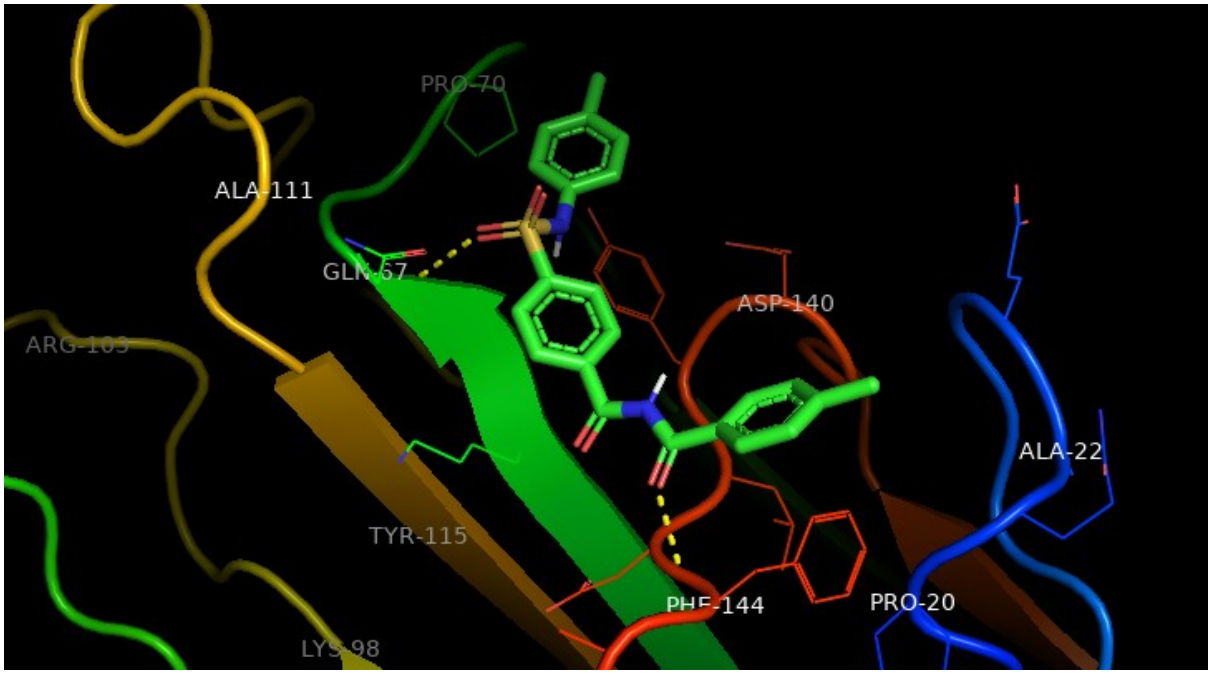


Fig.4b - Target-compound interaction of the De Novo generated SMILES 2-SMI

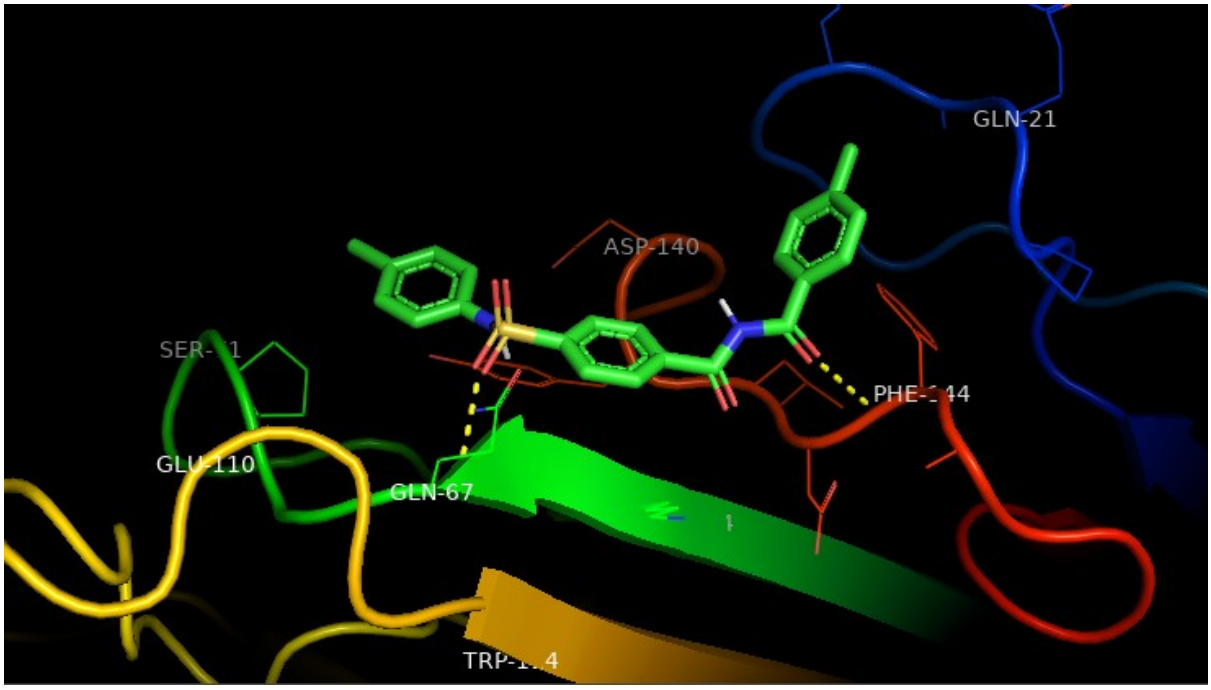


Fig.4c - Target-compound interaction of the De Novo generated SMILES 3-SMI

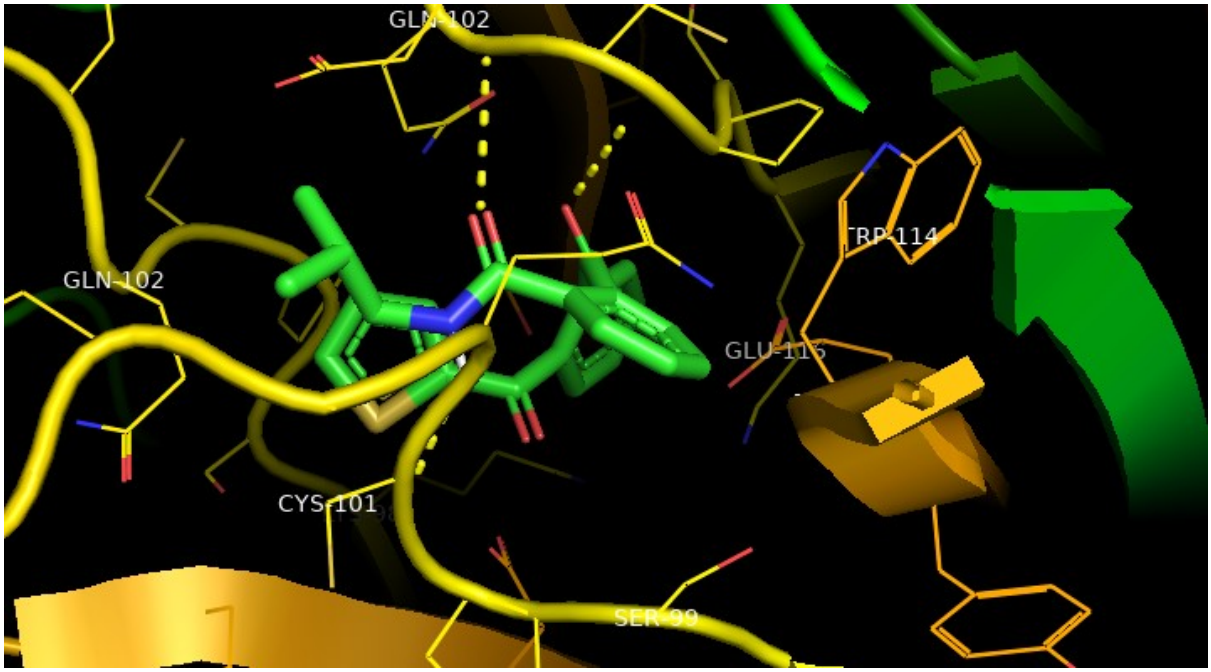


Fig.4d - Target-compound interaction of the De Novo generated SMILES 4-SMI

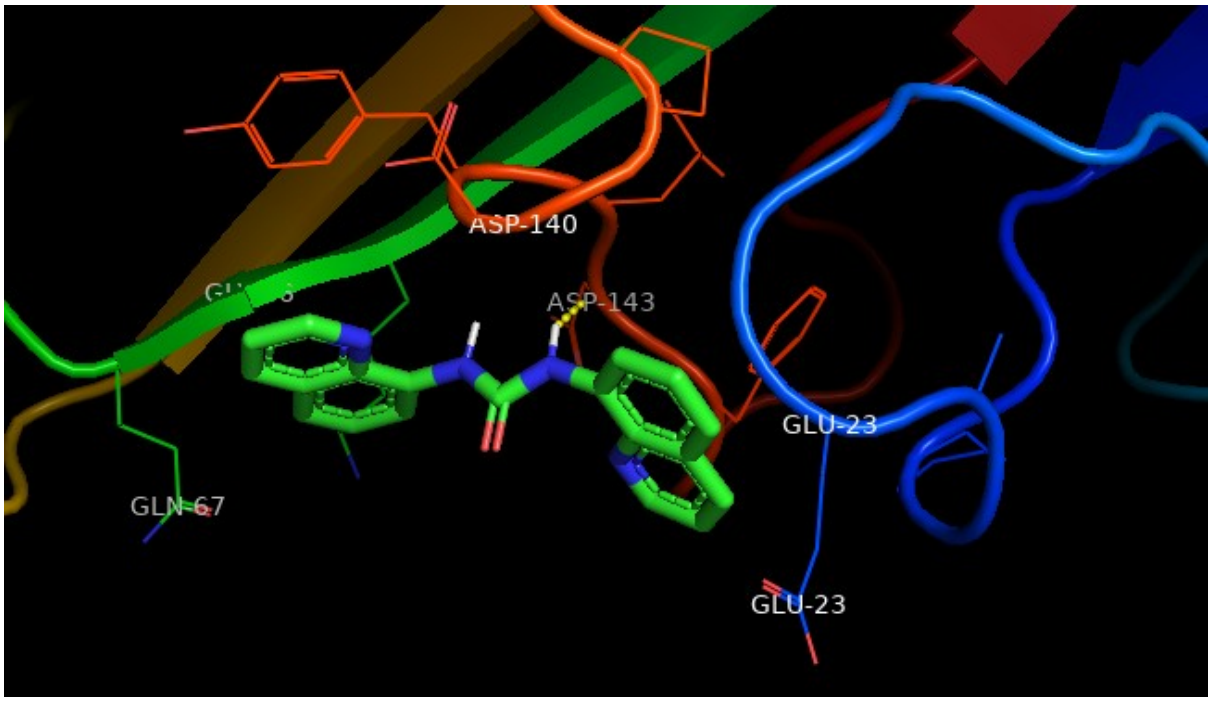


Fig.4e - Target-compound interaction of the De Novo generated SMILES 5-SMI

Thus a demonstrated use of the tool *in silico*-deep learning based de novo drug design for a target has been shown.

Conclusion and future outlook

The on-going data-science and AI revolution offers researchers with fresh set of tools to approach structure-based drug design problems in the computer aided drug design space. A novel programmatic tool that can be used in aid of *in silico*-deep learning based de novo drug design for any target of interest has been reported. A demonstrated use of the tool has been shown with the target signatures of Tumor Necrosis Factor-Alpha, an important therapeutic target in the case of anti-inflammatory treatment. The future scope of the tool involves, running the tool on a High Performance Cluster for all known target signatures to generate data that will be useful to drive AI and Big data driven drug discovery

References

1. Bajorath, Jürgen, Steven Kearnes, W. Patrick Walters, Nicholas A. Meanwell, Gunda I. Georg, and Shaomeng Wang. "Artificial intelligence in drug discovery: Into the great wide open." (2020): 8651-8652
2. David, Laurianne, Amol Thakkar, Rocío Mercado, and Ola Engkvist. "Molecular representations in AI-driven drug discovery: a review and practical guide." *Journal of Cheminformatics* 12, no. 1 (2020): 1-22.
3. Baskin, Igor I. "The power of deep learning to ligand-based novel drug discovery." *Expert Opinion on Drug Discovery* (2020): 1-10.
4. Smalley, Eric. "AI-powered drug discovery captures pharma interest." (2017): 604.
5. Chan, HC Stephen, Hanbin Shan, Thamani Dahoun, Horst Vogel, and Shuguang Yuan. "Advancing drug discovery via artificial intelligence." *Trends in pharmacological sciences* 40, no. 8 (2019): 592-604.
6. Batool, Maria, Bilal Ahmad, and Sangdun Choi. "A structure-based drug discovery paradigm." *International journal of molecular sciences* 20, no. 11 (2019): 2783.
7. Sellwood, Matthew A., Mohamed Ahmed, Marwin HS Segler, and Nathan Brown. "Artificial intelligence in drug discovery." (2018): 2025-2028.
8. Yang, Xin, Yifei Wang, Ryan Byrne, Gisbert Schneider, and Shengyong Yang. "Concepts of artificial intelligence for computer-assisted drug discovery." *Chemical reviews* 119, no. 18 (2019): 10520-10594.
9. Mak, Kit-Kay, and Mallikarjuna Rao Pichika. "Artificial intelligence in drug development: present status and future prospects." *Drug discovery today* 24, no. 3 (2019): 773-780.
10. Stephenson, Natalie, Emily Shane, Jessica Chase, Jason Rowland, David Ries, Nicola Justice, Jie Zhang, Leong Chan, and Renzhi Cao. "Survey of machine learning techniques in drug discovery." *Current drug metabolism* 20, no. 3 (2019): 185-193.
11. Yang, Yongliang, S. James Adelstein, and Amin I. Kassis. "Target discovery from data mining approaches." *Drug discovery today* 17 (2012): S16-S23.
12. Song, Chun Meng, Shen Jean Lim, and Joo Chuan Tong. "Recent advances in computer-aided drug design." *Briefings in bioinformatics* 10, no. 5 (2009): 579-591.
13. Macalino, Stephani Joy Y., Vijayakumar Gosu, Sunhye Hong, and Sun Choi. "Role of computer-aided drug design in modern drug discovery." *Archives of pharmacal research* 38, no. 9 (2015): 1686-1701.

14. Yu, Wenbo, and Alexander D. MacKerell. "Computer-aided drug design methods." In *Antibiotics*, pp. 85-106. Humana Press, New York, NY, 2017.
15. Taft, Carlton A., and Vinicius Barreto Da Silva. "Current topics in computer-aided drug design." *Journal of pharmaceutical sciences* 97, no. 3 (2008): 1089-1098.
16. Van Drie, John H. "Computer-aided drug design: the next 20 years." *Journal of computer-aided molecular design* 21, no. 10-11 (2007): 591-601.
17. Huang, Hung-Jin, Hsin Wei Yu, Chien-Yu Chen, Chih-Ho Hsu, Hsin-Yi Chen, Kuei-Jen Lee, Fuu-Jen Tsai, and Calvin Yu-Chian Chen. "Current developments of computer-aided drug design." *Journal of the Taiwan Institute of Chemical Engineers* 41, no. 6 (2010): 623-635.
18. Brown, Nathan, Jean Cambruzzi, Peter J. Cox, Mark Davies, James Dunbar, Dean Plumbley, Matthew A. Sellwood et al. "Big data in drug discovery." In *Progress in medicinal chemistry*, vol. 57, pp. 277-356. Elsevier, 2018.
19. Chen, Hongming, Ola Engkvist, Yinhai Wang, Marcus Olivecrona, and Thomas Blaschke. "The rise of deep learning in drug discovery." *Drug discovery today* 23, no. 6 (2018): 1241-1250.
20. Barrett, S. J., and W. B. Langdon. "Advances in the application of machine learning techniques in drug discovery, design and development." In *Applications of Soft Computing*, pp. 99-110. Springer, Berlin, Heidelberg, 2006.
21. Geoffrey AS, Ben, Rafal Madaj, Akhil Sanker, Pavan Preetham Valluri, Judith Gracia, and Harshmeet Singh. "Target2drug: A Novel Programmatic Workflow to Automate in Silico Drug Discovery." (2020).
22. Gupta, Anvita, Alex T. Müller, Berend JH Huisman, Jens A. Fuchs, Petra Schneider, and Gisbert Schneider. "Generative recurrent networks for de novo drug design." *Molecular informatics* 37, no. 1-2 (2018): 1700111.
23. Strieter, Robert M., STEVEN L. Kunkel, and ROGER C. Bone. "Role of tumor necrosis factor-alpha in disease states and inflammation." *Critical care medicine* 21, no. 10 Suppl (1993): S447-63.