

BITS :: Call for Abstracts 2024 - Oral communication

<i>Type</i>	Oral communication
<i>Session</i>	Bioinformatics AI, Models and Tools
<i>Title</i>	TumFlow: An AI Model for Predicting New Anticancer Molecules
<i>All Authors</i>	Davide Rigoni (1), Sachithra Yaddehige (2), Nicoletta Bianchi (3), Alessandro Sperduti (4), Stefano Moro (1) and Cristian Taccioli (2)

Affiliation

(1) Molecular Modelling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences, Via Francesco Marzolo 5, 35131, University of Padova, Italy
(2) Department of Animal Medicine, Production and Health, Viale dell'Università 16, 35020, University of Padova, Italy
(3) Department of Translational Medicine, Via Luigi Borsari 46, 44121, University of Ferrara, Italy
(4) Department of Mathematics "Tullio Levi-Civita", Via Trieste 63, 35121, University of Padova, Italy

Motivation

Melanoma is the fifth most common cancer in the United States. It stands as the most severe type of skin cancer due to its potential to metastasize to other body parts if not detected and treated promptly. Conventional drug discovery methods face limitations due to the inherently time-consuming and costly. However, the advent of Artificial Intelligence (AI) has opened up new possibilities for simulating and evaluating numerous drug candidates, thereby mitigating the requisite time and resources. In this context, deep generative models, employing machine learning techniques to create new molecular structures, hold promise for accelerating the discovery of effective anticancer therapies. These models aim to capture the underlying patterns and structures in the training data, enabling them to generate novel samples that share characteristics with the original data. In the field of new drug generation, various approaches based on Variational Autoencoders, Generative Adversarial Networks, Normalizing Flows and Diffusion Models have been explored.

This research proposes TumFlow, a novel normalizing flow approach for generating molecular graphs for cancer therapeutics. TumFlow, building on the foundational work of MoFlow, a pioneering model in the field of models applied to graph structures, adapts and enhances these capabilities specifically to address melanoma treatment challenges.

Methods

TumFlow aims to learn the unknown probability distribution that has generated the chemical structures of the molecules in the dataset, with the purpose of using the learned distribution to generate new novel chemical structures that should convey similar substructures and similar anticancer activities. Therefore, it is developed to generate new antitumor molecules against melanoma, addressing all the unique challenges and requirements of melanoma treatment. TumFlow is trained on the comprehensive NCI-60 dataset, made public by the National Cancer Institute, which encompasses thousands of molecules tested across a broad spectrum of tumour cell lines, with a specific focus on the SK-MEL-28 melanoma cell line.

TumFlow generates new molecules adopting an approach that closely resembles structure optimization. Starting from a molecule with high antitumoral efficacy (i.e., low GI50 value), TumFlow modifies its structure to exhibit better antitumoral effects. The generation of new molecules is performed following two different approaches: (i) in the first, the starting point consists of molecules with higher antitumoral efficacy appearing in the training set; (ii) in the second, the starting point consists of molecules known for their efficacy in clinical treatments for melanoma.

Results

The submitted Figure reports some novel molecules generated by TumFlow starting from the molecule appearing in the training set. The score reported under each generated molecule represents the TumFlow predicted GI50 score, while the colour conveys the similarity score of the newly generated structure with the starting molecule structure. The score on the molecule top right box reports the normalized Synthetic Accessibility Score, a metric that estimates the ease (0) or difficulty (1) of synthesizing the molecule. All the presented novel molecules are not available on PubChem.

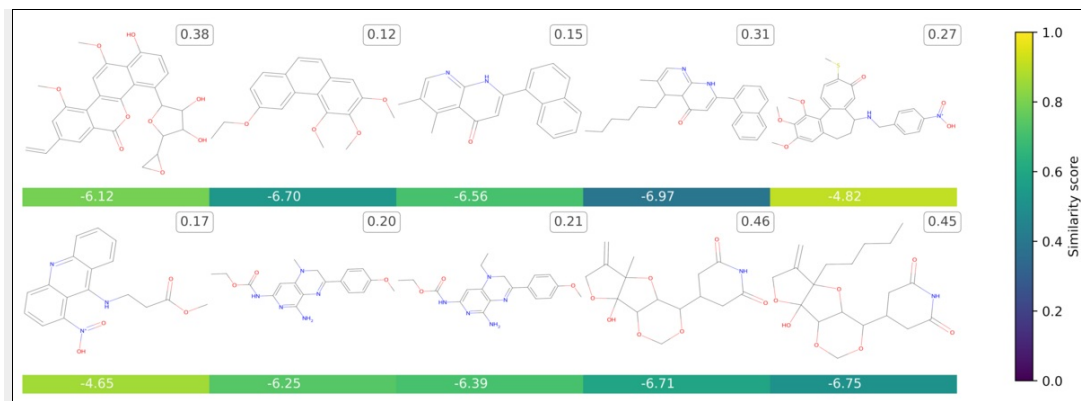
Tumflow, when adopted to generate new molecules starting from clinical molecules, was able to generate a previously studied compound against cancer. The generation of a novel molecule, previously studied for its anticancer properties and not present in the training set, underscores TumFlow potential to explore chemical space beyond the confines of existing datasets. This capability suggests that TumFlow has the capacity to propose compounds with therapeutic relevance that might not have been part of the original training data, stressing its potential to contribute to the discovery of compounds with valuable properties.

Info

<https://www.biorxiv.org/content/10.1101/2024.02.06.579053v1.supplementary-material>

filename grid_of_molecules_vitro_random.png

Figure



Availability <https://www.biorxiv.org/content/10.1101/2024.02.06.579053v1.abstract>

Dissemination Material

Social

<https://www.linkedin.com/in/drigoni/>

Summary

My research interest is in the application of Artificial Intelligence and Deep Learning in Medicinal Chemistry. I am currently working on generative models for de-novo drug design. Deep Generative models in deep learning are like chemistry artists, designing completely new molecules from scratch. By analyzing large amounts of data about existing molecules and their properties, these models learn the fundamental rules of molecular structure and behavior. With this knowledge, they can design new compounds that could potentially become new drugs. It's a bit like having a virtual chemist who supports the drug design process by drawing molecules that would otherwise never have been imagined.

Corresponding Author

<i>Name, Surname</i>	Davide, Rigoni
<i>Email</i>	davide.rigoni.1@unipd.it
<i>Submitted on</i>	23.04.2024

Società Italiana di Bioinformatica

C.F. / P.IVA 97319460586

E-mail bits@bioinformatics.it

Sede legale Viale G. Mazzini, 114/B - 00195 Roma

Website bioinformatics.it