

Computational Drug Discovery

unlocking the future of medicine

featuring interview with
Dr Pablo Acera Mateos
from CCI



SOCIETY NEWS

INTERVIEW WITH
DR PABLO ACERA MATEOS

COMPUTATIONAL DRUG DISCOVERY

CODE YOUR WAY

Society News



BINFSOC

Recent Events.

- Paint & Sip
- Barbeque
- Networking Night

Upcoming Events.

- CPM Workshop

Term 2 Week 8



PAINT & SIP



On Thursday of Week 2, UNSW Bioinformatics society collaborated with UNSW Biotechnology and Biomolecular Sciences Student Society to host the BABSOC x BINFSOC Paint & Sip event at the Biosciences Building. Accompanied by Bob Ross's soothing voice, students took the opportunity to showcase their artistic talents while sipping on their favourite beverages. Regardless of their level of experience, all participants were able to enjoy themselves and connect with fellow peers in an evening filled with laughter and creativity.





NETWORKING NIGHT



On Wednesday, June 12th, we had the pleasure of hosting our Annual Networking Night, a much-anticipated flagship event at UNSW Bioinformatics society. The evening featured eight rounds of dynamic roundtable discussions with leading industry experts in the bioinformatics discipline, followed by open-floor networking. This setup allowed students to engage directly with each speaker, creating valuable connections and insights. A delightful assortment of refreshments and desserts were served throughout the event, keeping all guests energised during their professional networking interactions.

We want to thank the organising team and the students who attended, making for a lively evening. We extend our heartfelt gratitude to our distinguished guests for your participation and enthusiasm: thank you for making it special!

- Aravind Venkateswaran
- Allegra Angeloni
- Ignatius Pang
- Joseph Copty
- Ksenia Skvortsova
- Laurence Wilson
- Pablo Acera Mateos
- Raymond Louie
- Sam El-Kamand

Your contributions enriched our discussions, and the shared experiences and stories will undoubtedly assist us in navigating the industry and addressing entry barriers into various initiatives.

The enthusiasm and engagement at the event embody the spirit of collaboration we aim to foster at UNSW Bioinformatics society. We welcome and deeply appreciate the support of sponsors to help elevate our impact. Please reach out to us if your organisation is interested.





SAUSAGE SIZZLE



Here at UNSW Bioinformatics society, we warmed up the chilly weather of Wednesday Week 4 with a delightful sausage sizzle at The Quad. From 12PM to 2PM, we witnessed an impressive turnout that significantly enhanced our society's presence on campus, with queues of students eagerly lining up to partake.

The barbecue featured a variety of sausages including both vegetarian and non-vegetarian options, to be enjoyed with bread as well as a selection of toppings including onions, mustards, ketchup, BBQ sauce and more! Refreshing beverages were also available to complement the savoury food. Attendees enjoyed not only a feast for the taste buds, but also a relaxing opportunity to socialise and bond. We thank all participants for joining us and look forward to hosting similar events in the future.



Interview with Dr Pablo Acera Mateos

Application of computational
drug discovery in curing
childhood cancer



“

Our group leverages computational biology to uncover new therapeutic targets in childhood cancers and to find novel or existing drugs that can specifically address these targets.

Author: Pablo Acera Mateos

Most common paediatric cancers lack targeted drug therapies, often forcing reliance on chemotherapy, which can lead to long-term debilitating effects in children. There's a significant need for safer and more effective treatments for these young patients. Our group leverages computational biology to uncover new therapeutic targets in childhood cancers and to find novel or existing drugs that can specifically address these targets. Targeted therapies have the potential to minimise damage to healthy cells, reducing adverse side effects significantly.

One of the limitations of the development of new disease therapies is limited by the proteome's low druggability, estimated at 15%. This highlights the urgent need for novel therapeutics beyond traditional pro-

tein-targeting approaches. The advent of RNA-targeting small molecules marks an exciting opportunity for novel therapeutic strategies and potentially broadens the target landscape. Recently, splicing modulator compounds (SMCs) have been used to reduce the expression of target genes by introducing poison exons in a sequence-specific manner. Our group is developing a drug discovery initiative focused on leveraging splicing modulator compounds to disrupt undruggable gene targets by introducing deleterious cryptic exons.

Our group leverages computational tools to streamline the drug discovery process, focusing primarily on early stages such as target identification. This involves determining the specific genes, proteins, or biological pathways to target for therapy-

utic interventions in various diseases. Because we focus on identifying new splicing modulators, we work and develop computational tools to analyse, mine and understand transcriptomics and splicing. One of our key strategies related to target identification involves identifying cryptic exons throughout the human genome, which are potential targets for splicing modulator compounds. We have crafted a method that merges deep learning [1] techniques with extensive high-throughput RNA sequencing data from the GTEx [2] project, leading to the discovery of thousands of previously unannotated cryptic exons. Additionally, some of these discovered cryptic exons can be translated into new protein motifs. With recent advances in large language models for proteins [3], we can now predict the impact of these new motifs on protein functionality.

One of the challenges, yet also opportunities, in the field of RNA-targeting splicing modulation compounds is its emerging stage, with only a few examples of these molecules available. Additionally, there is scarce data on the precise mechanisms of action of some of these

compounds. Furthermore, the interactions between RNA and small molecules are significantly different from those between proteins and small molecules that have been studied deeply. RNA's limited structural diversity, compared to proteins, makes it sometimes difficult for splicing modulator compounds to achieve high specificity, and it also makes it more difficult to predict where they are going to act. This lack of specificity could potentially affect the toxicity profile of these compounds.

We believe that splicing modulator compounds hold significant potential and the latest developments [4,5] signify a growing recognition of RNA splicing as a viable target for therapeutic intervention, expanding the arsenal of strategies against potential undruggable gene targets and diseases caused by splicing defects.

[1] <https://doi.org/10.1016/j.cell.2018.12.015>

[2] <https://gtexportal.org/home/>

[3] Zeming Lin et al., Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379, 1123–1130 (2023). DOI: [10.1126/science.adc2574](https://doi.org/10.1126/science.adc2574).

[4] Ratni H, Scalco RS, Stephan AH. Risiplam, the First Approved Small Molecule Splicing Modifier Drug as a Blueprint for Future Transformative Medicines. *ACS Med Chem Lett.* 2021 Jan 28;12(6):874–877. doi: [10.1021/acsmmedchemlett.0c00659](https://doi.org/10.1021/acsmmedchemlett.0c00659). PMID: 34141064; PMCID: PMC8201486.

[5] <https://www.nature.com/articles/s41467-022-28653-6>



Pablo Acera Mateos

Postdoctoral Researcher

CHILDREN'S
CANCER INSTITUTE



unlocking the future of medicine

Author: Gavin Li

Compiled: Yvonne Huang

Edit: Yvonne Huang

computational drug discovery

As a transformative modern revolution in the pharmaceutical industry, computational drug discovery leverages the explosion in biomolecule data and the power of advanced computer algorithms to design and optimise new therapeutic compounds. This innovative approach expedites the drug discovery process, reduces costs, and improves the precision of targeting specific diseases. Additionally, computational approaches can shorten the research cycle and mitigate ethical implications associated with extensive animal testing and late-stage failures.

What is computational drug discovery?

Traditionally, novel drug discovery and development is time-consuming, risky, and costly, often taking around 14 years and costing between \$0.8 to \$1.0 billion USD. Following recent advancements in combinatorial chemistry and high-throughput screening technologies, several approaches have been developed to improve the efficiency and success rate of the drug discovery process. One of the most effective techniques currently available is computer-aided drug design (CADD). CADD is an umbrella term used to represent the use of computational tools to store, manage, analyse, and model compounds. This method has been implemented to accelerate many

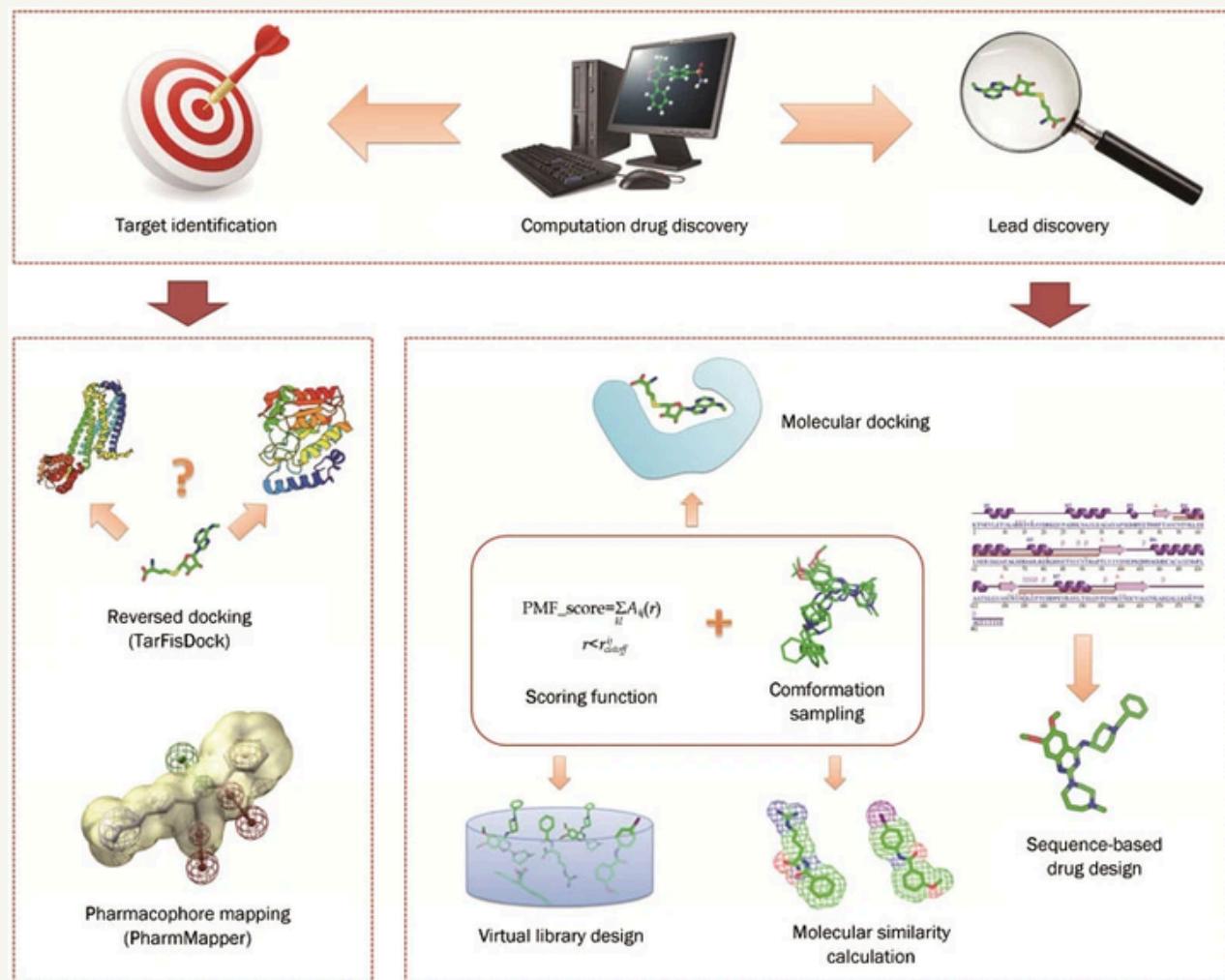
aspects of the drug discovery process, ranging from compound design to systematic lead candidate assessment. With the unprecedented increase of known biomacromolecule and small molecule data in the post-genomic era, computational tools now apply to almost every stage of drug discovery and development, potentially cutting costs by up to 50%.

There are several categories of popular CADD approaches, including structure-based drug design (SBDD), ligand-based drug design (LBDD), and sequence-based methods. SBDD relies on the structural knowledge of target macromolecules, whilst LBDD makes use of quantitative structure-activity relationships (QSAR) and pharmacophore modelling to predict interactions between drug targets and ligands. On the other hand, sequence-based approaches are used to identify potential targets and leads when structural data is unavailable. They harness the power of bioinformatics to predict 3D structures of unknown proteins, either through a comparison of sequence similarity against proteins with known structures (homology modelling), or through domain analysis that makes functional inferences based on the motifs and domains present.

Due to the specificity of each method, no single method meets all practical needs in the broad field of drug discovery, so combinational and hierarchical strategies that integrate multiple computational approaches are often used. The success of these methods relies heavily on technical factors like conformation generation and sampling, scoring functions, optimisation algorithms, and molecular similarity calculations.

Methodologies

Various methodologies drive CADD, each addressing a distinct stage of the drug development process. Two of the most important goals of many CADD tools are target identification and lead discovery. Target identification, a pivotal first step, aims to identify efficacious and safe targets that play an important role in the disease process and, above all, are ‘druggable’, meaning that the targets are accessible by the putative drug molecule. Lead discovery on the other hand, ensues to find candidate compounds (called leads) that can interact with the identified target, and



Different methodologies and platforms in the computational drug discovery field. Image: Ou-Yang et al., 2012, Nature.

“

Out of every **8,000 compounds** the companies screen for medicinal use, only one reaches the market. The computer should help lower those odds...

This means that chemists will not be tied up for **weeks**, sometimes **months**, painstakingly assembling test drugs that a computer could show to have little chance of working.

The potential saving to the pharmaceutical industry: **millions of dollars** and **thousands of man-hours**.

Designing Drugs with Computers
Discover magazine
August 1981

have the potential of progressing into preclinical and finally clinical development.

An example of a widely used tool that can be used to assist with the target identification process is TarFisDock. TarFisDock is a web server that employs the reverse docking strategy: the process of screening a given small molecule against a library of protein structures to yield the set of all possible binding proteins (candidates with the highest binding affinities). This platform is supported by the potential drug target database (PDTD), which documents over 1100 protein 3D structures obtained from the Protein Data Bank. Apart from target identification, TarFisDock has a number of other important applications. For instance, in constructing drug target networks for systematic study of drug-target interactions, or in identifying targets that could be potentially responsible for the side effects of an existing drug (off-target effects).

Complementing reverse docking approaches, pharmacophore modelling and mapping is another popular technique implemented for target identification. The pharmacophore of a molecule describes the set of features (e.g., spatial, geometrical, physiological) that defines its proper functioning within the cellular environment. In other words, it associates the 3D structure of molecules with their biological

or pharmacological activities. Pharmacophore modelling is the process of identifying, representing and analysing these essential features as means of studying their interactions with other molecules. This approach excels in identifying the optimal mode of interaction between the potential target candidates and the small molecule probes. PharmMapper is one such example that enhances target prediction via 'reverse' pharmacophore mapping using a vast database of models. This web-based tool leverages the power of a large, in-house database of over 7000 pharmacophore models annotated with their target information: PharmTargetDB.

There are also a range of computational approaches designed to enhance the lead discovery process. An example is docking-based virtual screening, which targets and optimises for one or more objectives, such as binding energy, shape complementarity, and chemical complementarity. GAsDock for instance, is a tool that uses entropy-based genetic algorithms to optimise ligand binding poses to macromolecules. Multi-Objective Scoring Function Optimization Methodology (MOSFOM) is another technique developed to further improve hit rates by optimising both energy and contact scores simultaneously.

In summary, CADD has seen huge success in integrating diverse methodologies from target identification to virtual library cons-



Image: Chris Gash

truction and sequence-based design to enhance efficiency and accuracy in drug discovery processes.

So, how is CADD used?

Computational drug discovery methods have significantly advanced target identification and lead discovery in various therapeutic areas. In target identification, TarFisDock coupled with the PDTD database has identified potential targets for bi-

oactive compounds with unknown *in vivo* targets, such as inhibitors against *Helicobacter pylori* PDF discovered through reverse docking. Similarly, the anti-cancer mechanism of [6]-gingerol, a natural component of ginger, by targeting and inhibiting LTA4H was revealed through the use of TarFisDock.

In lead discovery, docking-based virtual screening facilitated the discovery of RhoA inhibitors from the SPECS database, demonstrating significant inhibitory effects in cardiovascular diseases. Pharmacophore-based screening combined with molecular docking identified novel thiazolidine-2,4-dione analogues as potent IGF-1R inhibitors, promising for cancer therapy.

These applications illustrate the versatility and efficacy of CADD methods in accelerating the identification of novel drug targets and the development of potent therapeutic leads across various disease contexts.

What's next?

In recent years, CADD has revolutionised both industry and academia, significantly accelerating the identification of potent drug candidates. This progress was driven by the development of advanced methodologies that construct high-quality datasets and libraries tailored for molecular diversity or similarity optimization. Distri-

buted computing has also gained popularity for large-scale virtual screening.

Despite these advancements, caution is necessary to avoid the "Garbage In-Garbage Out" phenomenon, emphasising the need to integrate computational findings with experimental validation. It is important to note that computational tools are designed to complement rather than replace traditional research methods in drug discovery processes.

Looking ahead, the integration of computational chemistry, biology, chemoinformatics, and bioinformatics is poised to define the emerging field of pharmacoinformatics. Future research will focus on incorporating genomic insights to elucidate gene product functions and discover new drug targets. CADD methods will continue to evolve, while designed small molecules will serve as crucial probes for functional studies, marking a dynamic future for computational drug discovery.

References and further reading:

- Ou-Yang, Ss., Lu, Jy., Kong, Xq. et al. Computational drug discovery. *Acta Pharmacol Sin* 33, 1131–1140 (2012). <https://doi.org/10.1038/aps.2012.109>.
- Tyagi, R., Singh A., Chaudhary K., Yadav., M. K., 2022, Chapter 17 - Pharmacophore modeling and its applications, Bioinformatics. <https://doi.org/10.1016/B978-0-323-89775-4.00009-2>.
- Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. *Br J Pharmacol.* 2011 Mar;162(6):1239-49. doi: 10.1111/j.1476-5381.2010.01127.x. PMID: 21091654; PMCID: PMC3058157.

Code Your Way



Welcome back to "Code Your Way"

In this issue, the BINFSOC team is pleased to bring you a guide to getting you started in GitLab. GitLab is what you will use in many of your courses that involve group and individual programming projects. We will provide you with all the details you need to help you get your repository setup so that you can begin collaborating on projects.

Your feedback is crucial to us – if you would like us to cover something, please do write to us at the address shared on the last page.

Author: Gavin Li Compiled: Rubin Roy Edit: Yvonne Huang



GitLab

Getting Started with GitLab

To set up your Git environment for the first time, follow these steps:

1. Sign Up or Install GitLab

You have two main options for using GitLab: you can either sign up for a GitLab.com account, which provides a cloud-based solution hosted by GitLab, or you can install GitLab on your own infrastructure.

- *GitLab.com*: Signing up for a GitLab.com account is the quickest way to get started. Simply visit the GitLab website and follow the prompts to create your account.
- *Self-Hosted*: If you prefer to host GitLab on your own infrastructure, you can download and install the GitLab Community Edition or GitLab Enterprise Edition from the GitLab website.

2. Create a New Project

Once you have signed up or installed GitLab, the next step is to create a new project.

To do this:

- Click on the "New Project" button.
- Choose a project name and visibility level (public, internal, or private).
- Optionally, add a project description and initialise the repository with a README file.

3. Set Up SSH Keys

To securely interact with your GitLab repositories, it's recommended to set up SSH keys. This involves generating an SSH key pair on your local machine and adding the public key to your GitLab account.

- Generate an SSH key pair using the `> ssh-keygen` command.
- Copy the contents of the public key (`~/.ssh/id_rsa.pub`) and add it to your GitLab account settings.

4. Clone Your Repository

To start working on your project locally, you'll need to clone the repository to your machine.

- Copy the SSH or HTTPS URL of your repository from GitLab.
- Run `> git clone <repository-url>` in your terminal to clone the repository to your local machine.

5. Start Collaborating

With your project set up on GitLab, you can now start collaborating with your team.

- Invite team members to join your project and grant them the appropriate access levels (e.g., Developer, Maintainer).
- Use GitLab's issue tracking, merge requests, and code review features to manage your development process efficiently.

Contact us



IF YOU HAVE ANY COMMENTS or feedback regarding BINFsights, please write to us at binfo@unswbinfsoc.com

We also encourage anyone to share with us anything you'd like us to take a look at, be it a bioinformatics tool that you have made or find useful; or news in the bioinformatics world that you'd like to see written about in future issues.



TO VIEW PAST AND PRESENT issues of BINFsights, check out our website at unswbinfsoc.com/binfo
Stay tuned on our Facebook page for updates regarding events and society news.

-- The BINFSOC Team

BINF
sights.