ICR The Institute of Cancer Research

PhD Project Proposal

Funder details

Studentship funded by:	**MRC iCASE**

Project details

Project title:	**Using Grand canonical Monte Carlo to predict cryptic pockets for drug binding**

Supervisory team

Primary Supervisor:	Prof Swen Hoelder
Associate Supervisor(s):	Dr Mike Bodnarchuk (industrial supervisor at Astra Zeneca)
Secondary Supervisor:	Dr Rob van Montfort

Divisional affiliation

Primary Division:	Cancer Therapeutics
Primary Team:	Medicinal Chemistry Team 4
Site:	Sutton

Project background

Small molecule drug discovery relies on enclosed pockets on the targeted protein that offer sufficient opportunity for potent interaction with a drug-like compound. These compounds then either bind and block critical functions of the protein or more recently, can be used to recruit E3 ligases to the target leading to its degradation. However, several well validated cancer targets do not feature pockets e.g. in crystal structures and hence are considered undruggable.

A wealth of data shows the proteins can have so called cryptic pockets that can be explored for drug discovery. These pockets are not present in the experimentally observed conformation of the proteins but are present in other, less populated conformations. So far, these cryptic pockets and inhibitors binding to them have been found serendipitously for example by high throughput screening and solving the crystal structure of identified inhibitors.

Computational methods capable of predicting cryptic pockets would greatly extend our ability to explore targets deemed undruggable and have thus a very significant potential for patient benefit.

The overall aim of this project is to explore and validate computational methods for identifying cryptic pockets and compounds that bind to them. We will particularly focus on Grand Canonical Monte Carlo (GCMC) simulations due to their enhanced sampling ability in and around protein cavities.

Project aims

- The overall aim is to establish and validate computational approaches to identify cryptic proteins pockets that can be explored for small molecule drug discovery. We will particularly focus on Grand Canonical Monte Carlo simulations.
- The first aim is to establish the necessary code and software to run the simulation. We propose to use the GCMC modules in the open-source package OpenMM to do this
- We will then explore the established computational tools to conduct a retrospective study. In this retrospective study we will validate that our computational approach is able to predict cryptic fragment binding sites that have been observed experimentally. We expect this to lead to a first publication.
- We will then select a cancer target for a prospective study to discover new, cryptic binding sites. We will validate these sites through biochemical testing and crystallography.

Research proposal

Two of the grand challenges in computational chemistry are understanding and predicting the location of small molecules binding to proteins, and the prediction of allosteric/cryptic pockets. Whilst techniques such as Molecular Dynamics (MD) can predict the location of water molecules and small fragments, MD itself is limited in its application due to an inability to efficiently sample so-called occluded regions (parts of the protein inaccessible to solvent). As such, methods have been sought to overcome this deficiency, the gold-standard of which is Grand Canonical Monte Carlo (GCMC).

GCMC works by inserting or deleting small molecules in a predefined region of the protein. Historical uses of GCMC looked at the absorption of gases onto zeolites, however more recent applications have explored locating water molecules in and around protein binding sites. To obtain physiologically relevant molecular ensembles a bias is applied to the simulation, which is often set to the binding free energy of water. As a result, a comprehensive picture of solvation around the protein can be obtained which can be used to target areas of the binding site where the desolvation penalty might be lower and lead to boosts in binding affinities of lead molecules. Importantly, we have successfully applied GCMC to predict water networks in an ongoing year 2 PhD project supervised by the same supervisory team.

In principle, the concept of insertion and deletion of molecules can also be extended to small molecule inhibitors, particularly fragments. Whilst there are still no formal publications detailing the theory or application, numerous conference abstracts have begun to show that reliable predictions can be made when GCMC utilises water and fragments. In the first phase of the project, we propose to implement, utilise, and evaluate the fragment-water GCMC method on simple test cases. The successful candidate will perform a literature search to find examples where small molecules/fragments have shown to bind to both allosteric and orthosteric binding sites, and then assess the potential for this new approach to retrospectively find the binding sites and molecule locations. The analysis performed using GCMC will be directly compared to other prediction techniques (such as Mix-Solvent MD and FTMap) to assess whether the anticipated sampling efficiency afforded by GCMC is warranted.

Having established the utility of GCMC in locating small molecules in orthosteric and allosteric binding sites, the focus of the project will switch to finding cryptic pockets. There exists no established best-practice for finding such pockets, although techniques such as metadynamics (a MD-based method) are typically utilised. The first goal within this project phase will be another literature search to find examples for cryptic pockets (ideally with crystallographic evidence) and choose a variety of systems to simulate. We then propose to combine metadynamics simulations (to enhance the sampling of protein conformations) with the new GCMC technique (to sample within the protein configurations afforded by metadynamics). To our knowledge this is an unprecedented approach, and it is envisaged that this will give a high chance of locating established cryptic pockets.

Successfully finding literature cryptic pockets will yield multiple high-impact publications and open the door to applying this methodology on novel systems of interest to both ICR and AstraZeneca.

Supervision of the project

The student will be supervised by Dr Andrea Scarpino and Prof Swen Hoelder (ICR) and Dr Mike Bodnarchuk (Astra Zeneca). At the ICR, the student will be situated both in the *in silico* chemistry team and the Medicinal Chemistry Team 4 at the Centre for Cancer Drug Discovery (CCDD) and benefit from the significant experience in drug design and medicinal chemistry in both teams. In addition, in the cause of the project we will have to opportunity to obtain input from other experts at The ICR, for example Dr. Rob van Montfort (crystallography) and Dr Gary Newton (medicinal chemistry).

In addition, the student will have the opportunity to do some of this research at the Astra-Zeneca research site.

Benefit of conducting this project in the context of a collaboration with Astra Zeneca

Astra Zeneca is one of the leading pharmaceutical companies with a strong emphasis on small molecule discovery. Astra Zeneca also has strong track record in structure-based design and *in silico* methods. This expertise will benefit the project in a number of ways for example by contributing insights and learning about identifying cryptic pockets observed within Astra Zeneca projects as well as giving the opportunity to include Astra Zeneca data and drug target systems in our analysis. In addition, given the strong emphasis on small molecule drug discovery, the outcome of this project will also be relevant to in silico chemist at Astra Zeneca.

Candidate profile

Note: the ICR's standard minimum entry requirement is a relevant undergraduate Honours degree (First or 2:1).

Pre-requisite qualifications of applicants:

Candidates must have a first class or upper second class honours BSc Honours/MSc in chemistry or computational science or related disciplines.

Intended learning outcomes:

- Expertise in computational chemistry, in particular detailed knowledge and experience with using Grand Canonical Monte Carlo simulations for drug design, and a strong understanding of the features which influence protein-ligand affinities.
- Structure-based design.

There is also the opportunity to apply and gain insights into some of the following techniques:

- Docking and pharmacophore modelling
- Molecular dynamics
- Biological testing and crystallography
- Organic synthesis
- Machine learning

Advertising details

Project suitable for a student with a background in:

- ☐ Biological Sciences
- ☐ Physics or Engineering

☒ Chemistry

☐ Maths, Statistics or Epidemiology

☒ Computer Science