A Journey from Protein to Compound: Virtual Screening @ MDC

Kepa Burusco-Goni, Mouhamad Aboshokor, and Gemma L. Holliday

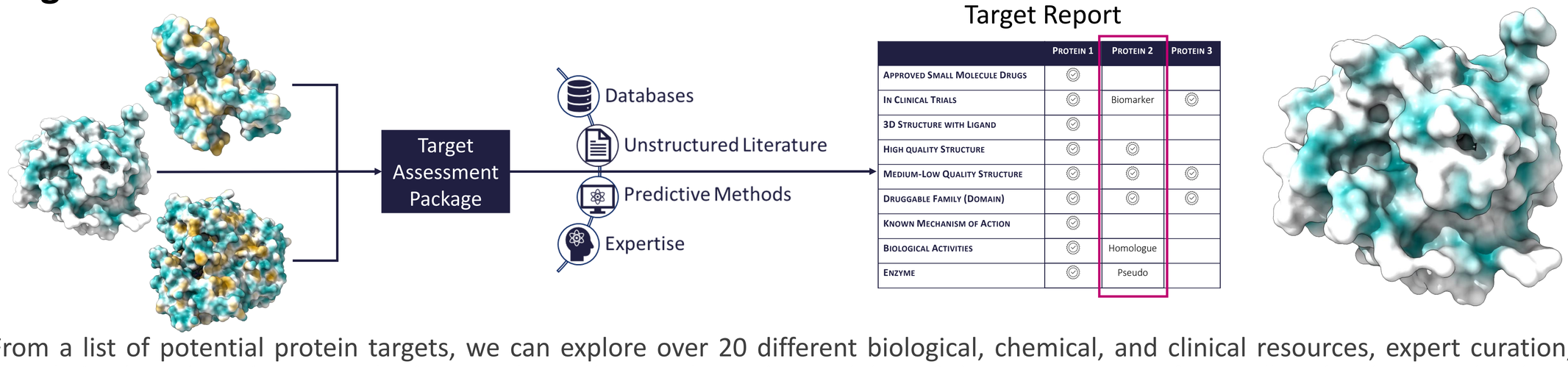
md.catapult.org.uk

The Challenge:

Identify small molecules that bind to a protein and may have some beneficial activity.

Lately, there has been an explosion of data in the public domain relating to the human genome. Such explosion comes from the sequencing of the so-called Dark Genome, a more in-depth analysis of the genome-disease association, large-scale sequencing efforts of Genomics England (and beyond), and biobanks across the world. All over the world, experts are asking the question: what other proteins ***could*** be druggable? How do we best use the data and techniques available to explore beyond the traditional? How can we repurpose existing compounds to find novel applications?

Target Assessment



	Protein 1	Protein 2	Protein 3
Approved Small Molecule Drugs	✓		
In Clinical Trials	✓	Biomarker	✓
3D Structure with Ligand	✓		
High quality Structure	✓	✓	
Medium-Low Quality Structure	✓	✓	✓
Druggable Family (Domain)	✓	✓	✓
Known Mechanism of Action	✓		
Biological Activities	✓	Homologue	
Enzyme	✓	Pseudo	

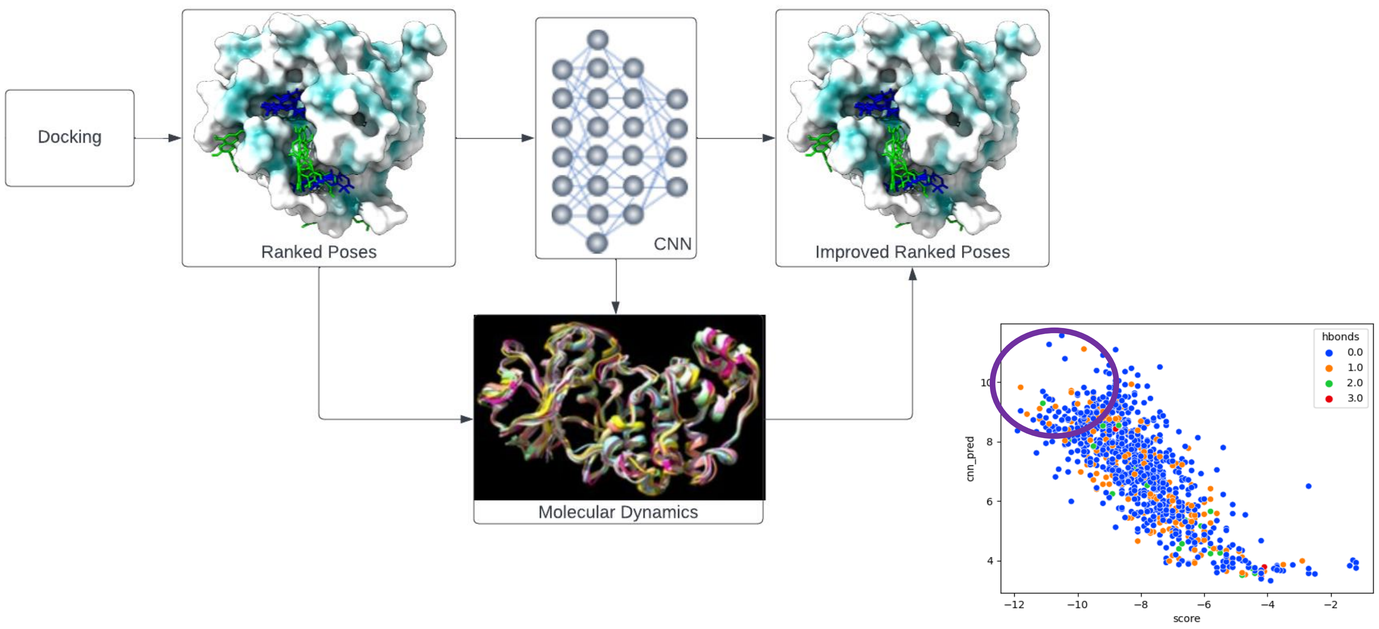
From a list of potential protein targets, we can explore over 20 different biological, chemical, and clinical resources, expert curation, and knowledge-based predictive methods to evaluate the novelty and tractability of the protein target for small molecule drug development. Based on the availability of data, or a detailed homology assessment if little or no data are available, if the target is already in clinical trials for the condition of interest, predicted druggability, and availability of high-quality 3D-structures, a recommendation is made for progression to the next stage: Deep Chemotyping.

Compound Libraries

Repurposing libraries are built from approved drugs (ATC, EMA, & FDA), compounds in late-stage clinical development. Other sources of compounds are those known to be active in the protein already, or in close homologues. Known biological entities, such as molecules in the PDB, and vendor libraries, such as mcule, Enamine, and Zinc.

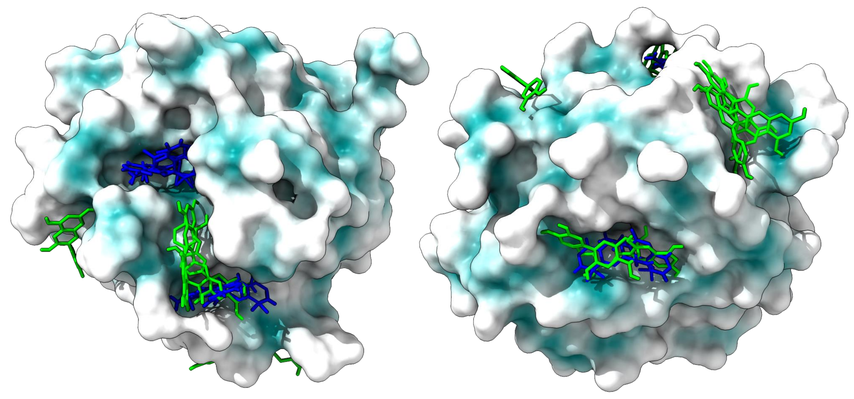
We work closely with the client to build the most appropriate compound library

Deep Chemotyping



One Protein, Many (Potential) Sites

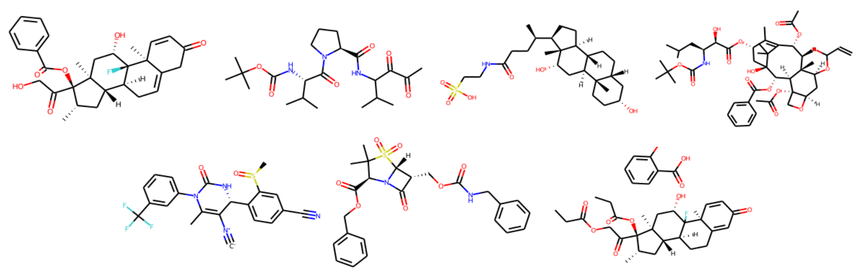
Using full-protein as well as site specific docking methods, we can explore more of the protein than before, potentially identifying allosteric sites, protein-protein binding grooves, as well as traditional binding sites.



Deep Chemotyping utilises traditional docking methods in which the ligand is allowed to be flexible, but the protein is held rigidly. By adding molecular dynamics into the pipeline, re-incorporating protein flexibility. The machine learning method of convolutional neural networks (CNN) are used to find the most promising poses to return an improved ranking of the small molecule results.

Compound Assessment

Brings together many different resources, expert curation, and knowledge based predictive methods to evaluate the compounds identified. The data gathered include clinically relevant information such as dosing, side effects, delivery route, and therapeutic classification. Calculated data include models such as blood-brain barrier or cell membrane penetrability, solubility, activity, and other physico-chemical properties that may affect drug-likeness. We also gather information on clinical trials, patents, and literature occurrence to help drive novelty and suitability assessment.



Acknowledgements: Original Ares Project work by Charles Bury, Kepa Burusco-Goni, Anna Pallo, John P. Overington, and Andrew Pannifer.