A new tool from Ireland.

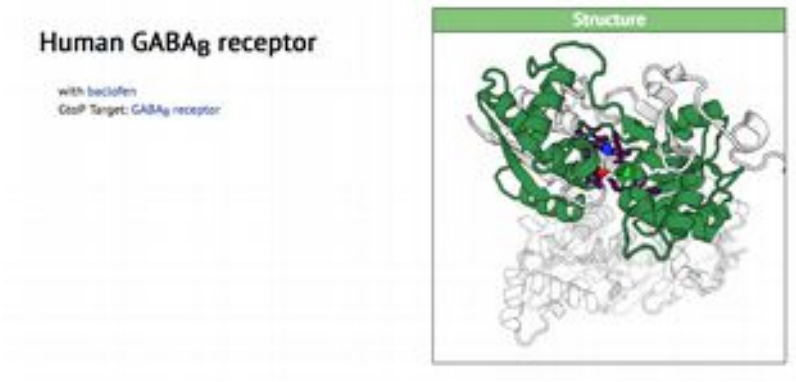
One of the somewhat unusual features of my laboratory is that we attack the frontiers of ignorance on several fronts at the same time. I know a less kind appraisal might involve the words ‘lack’ and ‘focus’, but I happen to like variety and the sense that, when we are stuck in one place, we are usually making progress at another. I do admit, though, that it can sometimes be difficult to keep everyone in the group feeling that they are one community. While the kidney folk and the synthetic biologists bond naturally over shared equipment and frustrations at the bench, they have few links with the database developers and druggies (sorry, ‘experts in pharmacoinformatics’) in the Eastern-most of the three group offices, who enter the labs only for ethanol to wipe their white-boards. I was therefore very pleased to gain funding from BBSRC for an ‘inverse pharmacology’ project that connects our pharmacology database to design problems in synthetic biology: the first of two papers emerging from the work has just been published (see Links).

Normal pharmacological research operates in a clear direction. It begins with a ‘target’ (in the early years, a nebulous physiological thing like ‘pain’, in the modern era, a molecular target known to mediate that physiological thing). Researchers then seek small molecules (or maybe big ones, nowadays) that interact with the target and either block it (in the case of pain) or perhaps stimulate it (in the case of hormone replacement, for example). Having identified them, the researchers examine the molecules for any that have no dangerous effects on anything else in the body, and which are easy to administer and basically safe: these are then developed as drugs. Several things have been done to accelerate drug discovery: the target proteins are analysed in minute detail so that their different functional domains are known, and the details of exactly how existing drugs and drug candidates bind to those proteins are generally well-studied too. All of this information is made available on large, open, international databases (the one run from this lab, the BPS/IUPHAR Guide to PHARMACOLOGY, being the usual starting point for people, but we are the first to acknowledge that it gains much of its power through links to detailed structural databases run by others). So, to recap, normal pharmacology starts with target proteins, identifies small molecules that bind to them, characterizes how these bind and how that affects the target and, in the course of other work, deposits this information in databases.

A major part of synthetic biological research consists of developing ‘designer proteins’, that never evolved in natural life, to do a specific job. Since protein design *ab initio* is an almost impossible

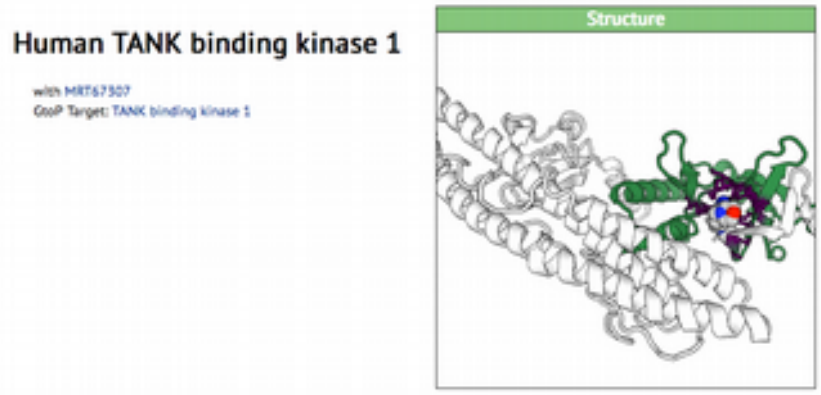
task (protein structures are really hard to predict from first principles), proteins design usually involves combining well understood elements from natural proteins, perhaps with a few minor changes. One feature that synthetic biologists usually want to include is external control, so that a protein's activity can be controlled carefully, usually in a way that requires an external 'drug' to activate it, so it becomes inactive without that drug: this is an example of fail-safe design.

This is where inverse pharmacology comes in. Normal pharmacology has given us a vast amount of information about how drugs – including ones already known to be safe for clinical use – interact with proteins. Our idea for 'inverse pharmacology' was to build informatic tools that raid the database for this kind of information, 'understand' what interactions are most useful for synthetic biologists, and present their suggestions in a prioritized list. The point about interactions likely to be most useful to synthetic biologists is an important one because databases are so large these days that the problem is always too much information rather than too little, and intelligent selection is critical. In the case of drug-protein interactions, the key measure of usefulness for synthetic work is the relationship between the drug binding site and the underlying structure of the protein. Proteins are made of chains of amino acids, of which there are 20 choices: the order of amino acids in any particular protein is its primary structure. The amino acid chain folds, often with the help of cellular protein-folding systems, into 3-dimensional secondary (local) and tertiary (overall) structures: they may even associate with other protein chains to form quaternary structures. Small molecules, whether drugs or natural molecules connected with the function of the protein, tend to bind in 'pockets' of the three-dimensional structure, in which the pattern of minute electric charges caused by the behaviour of chemical bonds in the small molecule is complementary to the pattern in the pocket or, alternatively, because they make a much less bad mismatch than would the water molecules that would otherwise be in the pocket. Sometimes, these pockets are formed of amino



An example of drug (the blobby thing in the middle with the red, blue and grey spheres) in a binding pocket formed by regions of the protein (the green spirals) that are far apart in primary amino acid sequence but brought physically together in space by the folding of the protein.

acids that are close neighbours in the primary structure of the protein, and sometimes they are formed by amino acids that are far apart along the primary amino acid chain but that are brought together by the final folding up of that chain to make the mature protein. This second type of pocket is almost useless to synthetic biology because its formation will be a property of the whole natural protein and not something that can be 'borrowed' from a small part of it and used in isolation. The first type of pocket, though, is much more hopeful. This will be especially true of the pocket could be expected to form relatively independently of what the surrounding bits of protein are doing. Finding example of these modules was the aim of the first part of our "inverse pharmacology" project: the second part was providing a practical demonstration, and that will be the topic of a later blog post (when that paper, currently being revised, has been published).



A pocket made from amino acids close to one another: the drug is the brightly coloured set of blobs on the right-hand side of this molecular structure diagram, the protein is uncoloured for irrelevant residues and green for the 'drug-binding domain', which 'hangs off' the rest of the protein as a coherent unit: potentially useful as a protein module that can be borrowed from this protein and incorporated into a designer protein that will bind and may be controllable by this drug.

The first part of the project needed someone to build a computer-based tool, and we needed someone who could program computers and also understand molecular biology. Conducting interviews was interesting: we were quite disappointed with the performance of several candidates with formal qualifications in computing, even to PhD level, but who seemed very short on ideas about how to go about the task. But then there was Sam Ireland, a fresh biology graduate with a strong interest in computing born of self-education and semi-commercial app' development. Sam engaged with the problem at once, in interview, and over a half hour or so of conversation it was obvious that he was completely up to the task. He was looking for a one-year position between graduation and starting a PhD at the University of London, and we seemed a perfect fit for one

another.

Having Sam in the group was great fun – he was a rare example of a person who chose to attend both the computing and wet lab meetings, and so got to know everyone really well and contributed excellent suggestions in both areas. He wisely chose to make as much use as possible of Chris Southan's extraordinary knowledge of chemoinformatic systems, and Jo Sharman's advice on coding in our software environment, and within a few months an 'alpha version' of the inverse pharmacology tool was born. He called it SynPHARM, a name that always raises smiles at meetings when I say it out loud.

By the time Sam left for his PhD, SynPHARM was a fully functioning tool attached to the Guide to PHARMACOLOGY database (the screenshots above come from it), and Alazne Dominguez-Monedero had made great progress in producing real proteins to exploit the idea of inverse pharmacology to add drug control to gene editing systems (that's the paper-in-revision I mentioned earlier). The tool is open to anyone, for any use (see Links), though if you are not heavily into protein design it will probably not make much sense to you. Sam is now working on other things in London, but his tools are being used by a number of other groups in the world. He achieved a tremendous amount in a short time, and I suspect that the name Sam Ireland will be one to watch in the future.

Jamie Davies
Edinburgh
July 2018

Links:

The paper: https://pubs.acs.org/doi/pdf/10.1021/acsomega.8b00659

SynPHARM: http://synpharm.guidetopharmacology.org/

Guide to PHARMACOLOGY: http://www.guidetopharmacology.org/