Structure-based drug design has become a key tool for the development of novel drugs. The process involves elucidating the three-dimensional structure of the potential drug molecule bound to the target protein that has been identified as playing a key role in the disease state. Using this three-dimensional information facilitates the process of making improvements to the potential drug molecule by highlighting existing and possible new interactions within the binding site. This knowledge is used to inform increases in potency and selectivity of the molecules as well as to help improve other drug-like properties. The speed and numbers of samples that can be studied, combined with the improved resolution of the structures that can be obtained using synchrotron radiation, have had a significant impact on the utilization of crystallography in the drug discovery process.

1. Introduction

Understanding the structure of a protein provides insights into its function, the mechanism by which this function occurs and, in some cases, how the function is regulated. Most drug targets are proteins that form part of catabolic or regulatory pathways, so understanding how these proteins operate is key to modulating their effect in intervention with a diseased state. Crystallography has been the cornerstone in the last 20–30 years for the discovery of small-molecule drugs in what has become known as structure-based drug design (SBDD) [1–5]. During this period, structural biology has attracted no fewer than eight Nobel Prizes in Chemistry, highlighting the significance of crystallography and related biophysical techniques (nuclear magnetic resonance (NMR) and electron microscopy) in providing structural insights into cellular processes.
Synchrotrons have been an underpinning tunable resource to produce high-intensity X-rays to help elucidate these protein structures since the first experiments at Hamburg in 1970. The UK has been a leading light in this discipline, supported by the building of the world’s first second-generation synchrotron at Daresbury in 1981, and then, more recently, with the construction of the Diamond Light Source, which had ‘first light’ in 2006 and now includes five operational state-of-the-art macromolecular crystallography beamlines (I02, I03, I04, I04-1 and I24) and more in the construction phase (a long-wavelength beamline, a sub-micrometre focus beamline and an in situ beamline); see http://www.diamond.ac.uk/Beamlines/Mx.html.

2. Structure-based drug design

SBDD began in the late 1970s with approaches pioneered in the UK at Wellcome [5] and in the USA at the University of Pittsburgh, Abbott and Agouron, which was the first SBDD biotechnology company. Larger pharmaceutical companies began to adopt the approach in the 1980s and early 1990s, embracing the approach as SBDD matured, creating multidisciplinary teams of structural biology, computational chemistry (including bioinformatics and chemoinformatics), medicinal and synthetic chemistry to help advance programmes where structural information on the target protein was available with classical targets such as HIV protease and serine proteases such as thrombin and factor Xa and then Ser/Thr kinases [6].

The process of getting a drug to market is long, often 12–15 years from initial patent filing of the chemical matter. We have now started to see the fruits of these early programmes, with inhibitors on the market for a number of protein targets, including angiotensin converting enzyme ACE (Captopril), HIV-protease (Viracept among others), neuraminidase (Relenza and Tamiflu), Abl kinase (Gleevec) and EGFR kinase (Iressa, Tarceva) [3–5] and also a number of antibacterials (for a review, see Simmons et al. [7,8]).

Several lessons have been learnt during the last few decades on the exploitation of structural information to guide drug discovery programmes. It is key that the structural information is available at an early stage of the programme, which may often involve the structural biology being initiated ‘at risk’ prior to the full drug discovery project being started. Ideally the ability to generate multiple structures should be in place at the early screening stage to help identify the mode of action of potential leads from biological screening and additionally to provide alternative hit finding through fragment screening [9] (discussed later). The ability to generate multiple structures from multiple chemotypes (compounds of different chemical classes) allows for increased understanding of key interactions and analysis of the plasticity of the ligand binding site and also the role of key water molecules. Turnaround of structures should ideally be fast (within a few weeks) to respond quickly to changes in the structure–activity relationship of the compounds in a fast design, synthesis and screening campaign. Structural biology should be as cost-effective as possible to guide design and save costs by reducing the number of cycles of synthesis and screening and ensuring that the lead selection and optimization process progresses as fast as possible.

3. A changing landscape: advances in structural biology methods

This 30-year period has seen significant advances in many techniques that have revolutionized the process of producing a three-dimensional set of coordinates of a ‘target protein’ with a bound ligand or lead compound in a drug discovery programme. It is the speed with which such structures can now be elucidated that makes crystallography a key tool that is now fully embedded in the drug discovery process, particularly in the early stages of target validation, hit identification and lead optimization. Crystallography provides information on the mode of action of a ligand and affords researchers details of the interactions that the ligand makes with the protein to guide medicinal chemistry follow-up to improve potency and ligand efficiency, selectivity, removal of toxicophores [10], tuning of physio-chemical properties and, often importantly, routes to provide novel intellectual property positions in competitive drug
targets. Advances in these techniques include: molecular biology, e.g. parallel cloning, using, for example, Gateway and LIC cloning systems; the use of alternative expression systems to *Escherichia coli*, such as insect cell and mammalian cell systems; improved expression vectors and purification techniques (allowing incorporation of SeMet for phasing of structures); and genome sequencing and the advent of gene synthesis to quickly and cheaply build the target sequence of interest. Technical advances include the use of cryo-cooling of crystal samples to reduce the effects of radiation damage in the X-ray beam, resulting in full datasets routinely being collected from one sample. (Cryo-cooling involves mounting of samples in a loop with a surrounding cryo-protectant, often the crystallization buffer with added glycerol.) Cryo-cooling also affords easy storage and shipping of samples for analysis at any synchrotron globally. Structural genomics programmes that used the aforementioned advances to progress many targets in parallel to structurally cover fold and target spaces for the druggable genome [11] have also driven developments in automation, including advances in synchrotron radiation and sample handling robots such as those developed at the APS and ESRF.

4. Advances in synchrotron beamlines

Synchrotron radiation has been a critical factor for the structure determination of most crystallographic problems through techniques exploiting the nature of tunable wavelengths, such as multiple-wavelength anomalous dispersion [12] and single-wavelength anomalous dispersion, where the high brilliance and non-divergence of the X-ray beam make data collection on even the most challenging samples possible. Improvements at the machine end, insertion devices, downstream optics and automatic beam position monitoring, have delivered stable, high-quality beams that can be focused down to provide 5–10 μm beams with sufficient flux to obtain good diffraction from the smallest samples and have made many data collections routine.

Advances in detectors, moving from film to multi-wire ion chambers, then image plates to CCDs and, most recently, to large pixel array detectors, have increased sensitivity and speed of data collection, reducing a process that used to take hours or even days to a couple of minutes or even less by combining shutterless data collection with fast, accurate goniometers [13].

Beamlines such as those at Diamond have been developed with automation and throughput as key goals to ensure high levels of productivity. Continuous software development and exploitation of the advances in computational speed allow results to be produced within minutes of the dataset being collected. The use of robotics to rapidly exchange samples, with optics and image recognition for aligning samples automatically, coupled with sample tracking (e.g. ISPYB [14]) and automatic data processing (XIA2 [15]) have helped to create fast structure solution pipelines. These pipelines allow the user to solve straightforward structures (data collection, automatic processing, structure solution and initial structure refinement) in a matter of minutes with little or no user intervention. Beamlines continually evolve, balancing the need to produce a flexible experimental environment for more challenging crystallographic samples, while maintaining a robust operation with a straightforward user interface. In more recent years, this high level of automation and robustness of both the hardware and software has led to the utilization of remote access data collections. In this process, the scientists ‘mail in’ their samples (pre-frozen crystals mounted on standardized pins in pucks (figure 1)), which are then loaded into a liquid-nitrogen-filled storage dewar that is accessed by a crystal mounting robot and placed on the goniometer for data collection. The user can select the sample and change the experimental details using the normal user interface, but from a remote location over the Internet. Results are then automatically generated and the smaller sized results files transferred back in real time for further analysis.

5. Fragment-based approaches

Traditional approaches for finding lead compounds relied on screening of large collections of compounds (millions of drug-like molecules) in a suitable cell-based or enzyme assay to
identify promising ‘hits or leads’ for further enumeration by medicinal chemists to understand the structure–activity relationships. A drawback of this approach is that the screening files were often biased towards previous targets with fairly highly decorated chemical scaffolds. Additionally, access to such large files was limited to the large pharmaceutical companies, although open screening initiatives are now becoming accessible (e.g. European Lead Factory, MRC-AstraZeneca, GSK-Tres Cantos open lab foundation, Evotec-European Screening Port (Fraunhofer)).

In the past decade, biophysical approaches such as NMR, surface plasmon resonance (SPR) as well as crystallography have been used to search for much weaker starting points from smaller molecules or scaffolds, which can then be more efficiently grown to produce highly ligand-efficient leads. In this process, known as fragment screening, much smaller collections (500–4000 compounds of molecular weight 250–350 daltons) are screened using high-concentration biological assays, NMR, SPR and crystallography. Crystallography often focuses on the functional catalytic domains of proteins to simplify the system and provide more chance of these regular globular domains crystallizing. This affords a system for high-throughput structure determination, which is a prerequisite for this approach, as X-ray crystallography is a key step in revealing the binding mode of the scaffolds. This then ensures that appropriate chemistry is used to grow the compounds in a direction to increase the potency of the ligands by targeting key interaction points in the ligand binding site. Synchrotron beamlines, with their high-speed data collection and automated processing of structure solution using pipelines such as DIMPLE [16], are ideally suited to this kind of screening and crystallographic readout (figure 2). A number of fragment-based new chemical entities are now in early-stage clinical trials for multiple disease indications. These routine types of data collections can now be controlled remotely and results viewed interactively; in some cases, the whole collection can be run in a fully automated mode.

Most large pharmaceutical companies have internal structural biology groups and access to large compound files; they use fragment screening to augment their approaches to hit finding. Smaller companies, biotechnology companies and some academic groups involved in drug discovery, with no access to large multi-million compound files, use fragment screening,
combined with in silico virtual screening, to identify potential hits or leads for a programme and rely on academics or an increasing number of contract research organizations to provide crystallographic support for these.

6. Current and future challenges

Progress in sample preparation methods and synchrotron data collection techniques means that we now have high-resolution, three-dimensional structural information available for all major classes of drug targets, including structurally challenging targets such as ion channels and, most recently, the G protein couple receptors (GPCRs) with the structure of human B2 adrenergic receptor [17]. Synchrotron radiation has been critical in obtaining the first structures of GPCRs and has been employed in the fast follow-up of SBDD exploiting these targets and other related members of this large target class [18–20]. Structural biology of this class of target has primarily been exploited by two biotechnology companies, Heptares Therapeutics and Receptos, both of which have SBDD of GPCRs as a key component of their business model and, as a consequence, use synchrotrons extensively (figure 3).

In addition to the use of crystallography to develop small molecules in the drug discovery area, three-dimensional structures of protein–protein complexes and multi-domain proteins are now an area of intense interest for the design of allosteric modulators, which, rather than blocking the substrate binding site of a protein, instead change the conformation of the target protein, affecting its function. Crystallography is also being used in the development of biologicals (predominantly humanized antibodies and smaller derivatives) as therapeutics through characterization of their binding to their targets [21].

Crystallography continues to evolve to work on more challenging samples and to make the collection of data more routine. While cryo-cooling of samples works in many cases, there are some systems where either the physical or osmotic shock of mounting or cryo-protection damages the sample to the point where no useful data can be obtained. In these cases, analysis of the sample in the environment in which it is grown is an approach that is being exploited by several research groups. This has led to the investigation by a number of groups into in situ data collection techniques [22–24] to screen samples for diffraction quality but also with the ultimate
aim of collecting full datasets to solve the structures of these samples, removing the bottleneck of manually mounting hundreds of samples for data collection (figure 4).

7. Summary

Synchrotrons have provided the structural biology community with high-brilliance, low-divergence, tunable monochromatic X-rays with wavelengths suitable for the determination of high-resolution crystal structures of proteins of therapeutic interest. This has been an invaluable tool in providing robust and rapid access to the highest-quality data to drive these drug design programmes forwards. This is reflected in the trend for increasing usage of Diamond beamtime by industry, particularly the pharmaceutical and biotechnology sector. Continued investments in machine architecture, beamline optics, sample environments, downstream processing and real-time feedback on experiments continue to improve the process of solving structures. This allows the study of more difficult and complex systems but also reduces the overhead for more straightforward crystallographic systems, thereby opening up the technology to a wider biological user group.

References


