Review

Structural Biology and the Design of New Therapeutics: From HIV and Cancer to Mycobacterial Infections A Paper Dedicated to John Kendrew

Sherine E. Thomas, Vitor Mendes, So Yeon Kim, Sony Malhotra, Bernardo Ochoa-Montaño, Michal Blaszczyk and Tom L. Blundell

Department of Biochemistry, *University of Cambridge, Tennis Court Road, Cambridge, CB2 1GA UK*

Correspondence to Tom L. Blundell: *tom@cryst.bioc.cam.ac.uk*
http://dx.doi.org/10.1016/j.jmb.2017.06.014
Edited by Georg Schulz

Abstract

Interest in applications of protein crystallography to medicine was evident, as the first high-resolution structures emerged in the 50s and 60s. In Cambridge, Max Perutz and John Kendrew sought to understand mutations in sickle cell and other genetic diseases related to hemoglobin, while in Oxford, the group of Dorothy Hodgkin became interested in long-lasting zinc-insulin crystals for treatment of diabetes and later considered insulin redesign, as synthetic insulins became possible. The use of protein crystallography in structure-guided drug discovery emerged as enzyme structures allowed the identification of potential inhibitor-binding sites and optimization of interactions of hits using the structure of the target protein. Early examples of this approach were the use of the structure of renin to design antihypertensives and the structure of HIV protease in design of AIDS antivirals. More recently, use of structure-guided design with fragment-based drug discovery, which reduces the size of screening libraries by decreasing complexity, has improved ligand efficiency in drug design and has been used to progress three oncology drugs through clinical trials to FDA approval. We exemplify current developments in structure-guided target identification and fragment-based lead discovery with efforts to develop new antimicrobials for mycobacterial infections.

Early Development of Structure-Guided Drug Discovery

When Max Perutz, John Kendrew and their colleagues in Cambridge were solving the first protein structures of myoglobin and hemoglobin in the 1950s and 60s [1–3], they were already aware of the importance of their work to medicine. Understanding the impacts of mutations on oxygen affinity and subunit cooperativity in abnormal haemoglobins that resulted in inherited single-gene disorders, such as sickle-cell disease, was recognized as a major objective. Dorothy Hodgkin's Oxford laboratory collaborated with Jørgen Schlichtkrull of Novo to understand how different crystalline forms of insulin could be exploited as slow-acting therapeutics for the treatment of diabetes [4]. This became a real possibility when the structure of insulin was solved [5,6], as many insulin sequences had been defined in Fred Sanger's laboratory in Cambridge [7]. Sequences and structures stimulated ideas not only about insulin storage and receptor binding but also about producing more effective therapeutics. These speculations became real opportunities when groups in Aachen, New York, and Shanghai completed the synthesis of insulin, encouraging ideas about the design of novel synthetic insulins.

Ideas about drug design were stimulated by the determination of the first enzyme structures—lysozyme, chymotrypsin, and trypsin—in the 60s and an emerging understanding of the interactions that led to selectivity of enzyme substrate binding [8]. In the 70s and 80s, clinically important drug targets such as the aspartic protease renin [9,10], which cleaves angiotensinogen to form angiotensin I, an essential step in regulating blood pressure, were modeled on less exciting enzymes such as fungal pepsins [11,12]. The use of protein crystallography in drug discovery accelerated in the 1980s, especially by using a combination of protein structure and interactive computer graphics, such as the Evans and Sutherland

machines [13]. The model of renin [14] was used widely in structure-guided drug design in the pharma industry in the 1980s. The high-resolution X-ray structures of apo-enzymes and complexes of renin and its close homologs followed much later [15,16].

New Paradigm of Structure-Guided Drug Discovery: Targeting HIV Protease

Probably the most influential development was the design of AIDS antivirals, based on the structure of human immunodeficiency virus (HIV) protease; these moved onto the market very quickly in the 1990s. Some basic science influenced this! In 1978, Jordan Tang who had sequenced pepsin together with crystallographic collaborators suggested that proteases had evolved from an ancestral dimer by gene duplication, fusion, and divergence to give more effective enzymes [17]. A close relative of this dimeric ancestral aspartic protease was found in the retroviral proteases, first in Rous Sarcoma Virus and then in HIV soon after the AIDS epidemic was recognized in the US and Europe. The genome of HIV, which encodes a polyprotein, was shown to include a protease [18], which was quickly recognized as a dimeric viral protease essential for the generation of infectious viral particles, and a model was produced based on the aspartic proteinase evolutionary relationship [19]. In 1989, structures followed for Rous Sarcoma Virus [20,21] and HIV protease [22,23], the structure of which was improved by further experimental structures determined independently by two labs [24,25]. The subsequent development of new AIDS antivirals by 1997, including four very successful drugs (Roche Pharmaceuticals' saquinavir, Abbot's ritonavir, Merck's indinavir, and Agouron's nelfinavir), demonstrated the importance of understanding the genome in terms of not only the functions of gene products but also their architectures for use in structure-guided drug discovery, recorded recently in an excellent history of macromolecular crystallography and its fruits [26].

The development of AIDS antivirals provided a new paradigm in drug discovery. It demonstrated that there was value in computational analysis of genomes in order to identify targets. This was an "exploration of biological space", an exciting challenge in the early 1990s, as the sequence determination of human infectious agents such *Mycobacteriaceae* that give rise to tuberculosis (TB) and leprosy, and the very much larger genome of *H. sapiens*, became real prospects. The HIV protease inhibitor story also illustrated the importance of "exploring chemical space" using protein structure to estimate the druggability of potential targets, followed by exploration of possible binding using screening libraries of chemical compounds. This idea of drug discovery can be summarized, as in Fig. 1a.

Over the subsequent 25 years, several new approaches have been introduced that exploit knowledge of the architecture of the target and screening of chemical libraries. One of the most influential has been the development of structure-guided fragment-based drug discovery (FBDD).

Protein Crystallography, FBDD, and Oncology

In the 1980s and 90s, meeting the challenge of the size and diversity of "chemical space" became a focus in the pharmaceutical industry with the realization that chemical libraries of several thousand drug-like compounds explored only a tiny area of the chemical space. In order to estimate the number, Lipinski rules assuming a molecular weight limit of 500 Da, the presence of carbon, hydrogen, oxygen, nitrogen, and sulfur, and a maximum of 4 rings lead to an estimate of 10^{63} [27]. Big pharma searched the world for new chemical diversity, often using the products of our natural environment from underdeveloped, forested areas. The chemical libraries grew to hundreds of thousands of compounds, and screening was roboticized to cope with the challenge, but a solution was also found in a different approach in which complexity of the chemicals screened was reduced by decreasing their molecular weights, which at the same time increased their promiscuity in binding targets. The innovation that allowed the decrease of the size of the chemical screening library was FBDD.

In FBDD, a fragment library often of ~1000 compounds of <300 Da is screened against the target of interest, resulting in identification of initial hits. These are then moved to lead candidates by chemically growing or linking the fragments followed by optimization of interactions, thereby exploring the chemical space available for binding to the target protein very effectively. A high-affinity lead molecule thus developed from a fragment hit retains the key binding interactions of the original fragment with the "hotspot" on the target protein. Most of the fragments have lower potency than the more complex molecules found in typical high-throughput screening (HTS) compound libraries; however, small fragments that bind do so by making well-defined and directional high-quality interactions and by displacing unhappy water molecules at the hotspots, giving rise to high ligand efficiency [28].

Early experiments used ligand-based NMR (Steve Fesik and his colleagues at Abbott) [29] and X-ray crystal screening [30,31], developed at Astex initially by exploiting high-throughput analysis of cocktails of 6 to 10 fragments soaked into apo-protein crystals. Knowledge of the structure of the complex of the fragment with target protein allowed the initial use of small, often non-chiral compounds, which were optimized using structure-guided approaches to