Biochemical Society Transactions (2017) **45** 303–311
DOI: 10.1042/BST20160422

Review Article

Genomes, structural biology and drug discovery: combating the impacts of mutations in genetic disease and antibiotic resistance

Arun Prasad Pandurangan, David B. Ascher, Sherine E. Thomas and Tom L. Blundell

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, U.K.

Correspondence: Tom L. Blundell (tlb20@cam.ac.uk)

For over four decades structural biology has been used to understand the mechanisms of disease, and structure-guided approaches have demonstrated clearly that they can contribute to many aspects of early drug discovery, both computationally and experimentally. Structure can also inform our understanding of impacts of mutations in human genetic diseases and drug resistance in cancers and infectious diseases. We discuss the ways that structural insights might be useful in both repurposing off-licence drugs and guide the design of new molecules that might be less susceptible to drug resistance in the future.

Structure-guided drug discovery

Ideas about the use of protein structure in drug discovery emerged in parallel with the first three-dimensional structures defined by X-ray crystallography. An early focus on the structure of haemoglobin in the 1950s stimulated thoughts about sickle cell disease in the Perutz laboratory in Cambridge [1,2], and the structure of lysozyme in 1966 in the Phillips' laboratory at the Royal Institution in London described a well-defined active-site cleft, interactions with substrate and early hypotheses about ligand binding and mechanism [3]. The structure of insulin in 1969 [4] was already a subject of interest in the pharma industry as insulin crystals were used to treat diabetics. The many sequences of insulin, already defined by Sanger in Cambridge, together with efforts to synthesize the protein underway in New York, Aachen and Shanghai, led to thoughts about design of more effective and longer-lasting insulins for therapeutic use [5]. The availability of further sequences in the years following led to early attempts to model human α-lactalbumin, a homologue of lysozyme found in milk [6,7], and then later relaxin [4] and insulin-like growth factor [8], both distant homologues of insulin of interest in biomedicine.

Although some laboratories, such as that of Dorothy Hodgkin, were well linked in to the pharma industry, interest elsewhere in structure-guided drug discovery increased only as experimental structures of enzymes and their complexes became available in academia, often for paralogues of current drug targets in industry. An example of this trend was renin, which had been known for many years to be involved in regulating blood pressure through its role in processing angiotensinogen to angiotensin I [9]. Aspartic proteinase structures of paralogues of renin had been defined in the 1970s, and these allowed models to be created for renin, for example, by Sibanda and colleagues [10]. These were used in some of the earlier structure-guided drug discovery campaigns for antihypertensives. The X-ray structures of complexes of renin inhibitors complexed with paralogues of renin appeared later [11] and were eventually followed by high-resolution structures of renins [12,13].

Tom L. Blundell was awarded the Biochemical Society Award in 2013; this review is based on the work for which he won the Award.

Received: 26 November 2016
Revised: 22 January 2017
Accepted: 01 February 2017

Version of Record published: 13 April 2017

PORTLAND PRESS

Biochemical Society Transactions (2017) **45** 303–311
DOI: 10.1042/BST20160422

Chemical library screening strategies exploiting structure

These developments in structure-guided drug discovery occurred in parallel with an increasing interest in target-agnostic phenotypic screening as well as roboticised screening of individual targets with chemical libraries of increasing size. However, for drug-like molecules (optimally of molecular mass ∼500 Da), screening requires libraries much larger and more diverse than the $\sim 10^6$ molecules available to most large pharma companies. The urgent need for a more innovative approach was underlined by the exponential increase in costs in drug discovery research and development in the 1990s, but the relatively low output of new drug approvals. Two innovative approaches involving protein structures that have provided complementary strategies have been virtual screening of very large numbers of compounds and experimental fragment-based drug discovery.

Virtual screening explores possible binding sites *in silico* for a large number of small molecules [14–17], assessing electrostatic, van der Waals or hydrogen-bonding interactions involved in molecular recognition of candidate molecules for binding a target of known structure. Virtual screening methods (such as GOLD [18], AUTODOCK [19] and GLIDE [20]) are used to dock large and diverse sets of drug-like molecules, in order to identify compounds that might provide useful 'hits'.

In contrast, fragment-based drug design is mainly an experimental approach to the challenge of moving from hits to leads. It depends on decreasing size and complexity by exploiting a fragment library, reduced in molecular mass (<300 Da) and in number to around a thousand molecules, which are screened against a target of interest using biochemical, biophysical and structural methods [21,22]. The fragments are developed into lead candidates by chemically growing or linking the fragments, thereby exploring the chemical space available for binding to the target protein very effectively. Although low-molecular-weight fragments have relatively lower potency than the more complex molecules found in typical high-throughput screening compound libraries, small fragments that bind, do so by making well-defined and high-quality interactions and by displacing 'unhappy' water molecules at hotspots on the protein [23,24].

At first fragment-based drug discovery was focused on 'druggable' targets with large, well-defined cavities, such as protein kinases; it was pioneered both in large companies such as Abbott, who used SAR by NMR (structure–activity relationships by nuclear magnetic resonance (NMR) [25]), as well as in small start-ups such as Astex, which has focused on high-throughput X-ray crystallography to screen fragments [21]. Increasingly now fragment-based approaches involve, first, a range of biophysical methods such as surface plasmon resonance (SPR) and thermal shift to screen a fragment library, and, second, others to provide a detailed analysis of the three-dimensional structure of the fragment complex by X-ray crystallography or NMR, the thermodynamics by isothermal calorimetry and kinetics by SPR. Molecular dynamics can also be used to explore different conformers of the protein and even reveal cryptic sites.

Until Otsuka purchased it in 2013, Astex brought compounds into clinical trials within the company, but also importantly developed strategic alliances with larger companies, including Jannsen, Novartis, AstraZeneca and GlaxoSmithKline. Astex has made impressive progress in clinical trials, achieving a milestone on 1 November 2016 with US FDA's filing of a new drug application for LEE011 (ribociclib), a drug that targets protein kinase CDK4, which was developed in an alliance with Novartis; this will be used in combination therapy with letrozole as a first-line treatment for advanced breast cancer.

In academia there have been attempts to use fragment-based drug discovery to target sites that have been previously defined as 'undruggable', most often interfacial or allosteric sites [26]. An example of this has been the use of fragment-based drug design to target the binding site of BRCA2-BRC repeats on Rad51, which catalyses an ATP-dependent DNA strand exchange in repair by homologous recombination of DNA double-strand breaks. This interaction involves concerted folding and binding of the BRC repeat, a foldable amino acid sequence within intrinsically disordered regions of BRCA2, onto the globular structure of Rad51 [27]. The BRCA2 interacts first through docking of a phenylalanine within a conserved FXXA sequence into a well-defined pocket of Rad51. This provides an anchor for the subsequent folding as a β-turn of the BRC repeat sequence, in order to allow the conserved alanine to bind into a smaller, more hydrophobic pocket. A further interaction is formed by the folding as an α-helix of a C-terminal region of the repeat motif into a shallow binding cleft. This was proposed as a possible site for targeting inhibitors and required very different chemistry from the 'drug-like' molecules designed to bind classical targets like protein kinases. The small pockets are well suited to bind fragments [28] and Cambridge labs in Biochemistry, Chemistry and Oncology have subsequently developed nanomolar inhibitors to bind this site [29].

Biochemical Society Transactions (2017) **45** 303–311
DOI: 10.1042/BST20160422

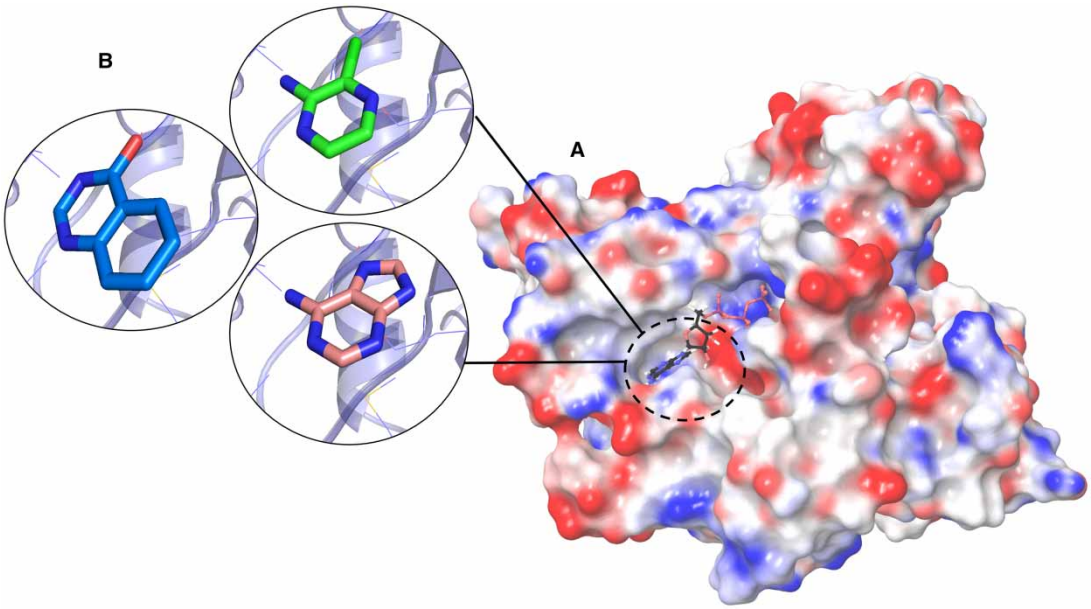

Figure 1. Targeting essential enzymes in *Mycobacterium abscessus*: Fragment hits for PurC (SAICAR Synthetase), an enzyme of *de novo* purine biosynthesis.

(**A**) Surface electrostatic representation of PurC in complex with ATP. (**B**) Fragments occupying the adenylyl pocket of PurC making strong polar contacts. Figure prepared using PyMOL and Maestro (Schrödinger, LLC).

Fragment-based drug discovery has also been used to develop new antimicrobials, particularly for tuberculosis and to target other mycobacteria. Very little work is currently developed in big pharma against pathogens that are responsible for infectious disease in developing countries or those that cause disease to small sections of the population, for example by *Mycobacterium abscessus*, which causes life-threatening lung infections in cystic fibrosis patients [30]. Here, the fragment-based approach has the advantage of working with very small molecules that can penetrate the cell walls. Figure 1 illustrates the binding of fragments to an essential enzyme in *Mycobacterium abscessus*, a non-tuberculous Mycobacterium (Thomas, S.E., Mendes, V. and Blundell, T.L., unpublished data). Our screening data have shown that fragments can bind at allosteric and interfacial sites. The allosteric sites are often not identified as functional, but occasionally even these prove to have an impact on binding at the active site, as we have recently discovered with Mycobacterium tuberculosis CoaBC enzyme (Mendes, V., Blaszczyk, M. and Blundell, T.L., unpublished results).

Mutations and disease

Many disease-related mutations, which are usually in residues that are conserved or conservatively varied during evolution, have an impact on the function of proteins. Classification and better interpretation of disease and non-disease-associated mutations is challenging but central to the understanding of genetic disease. Structural analyses of disease-causing mutations indicate that most occur at solvent-inaccessible and hydrogen-bonded residues [31].

Based on these observations, we proposed that the effect of a mutation on protein stability could be calculated by considering the statistical potential of a residue to change across evolution. The method site-directed mutator (SDM) uses a set of conformationally constrained, environment-specific substitution tables, which consider a residue's conformation, solvent accessibility and side-chain hydrogen bonding, to calculate the difference in stability between the wild-type and mutant protein structures [32,33]. More recently, we have used graph-based signatures to represent the wild-type structure environment in order to predict the effect of mutations on stability [34,35], and interactions with other proteins [35–39], nucleic acids [35], small molecules [40–42] and metal ions [43].

These programs have provided insights into mutations that lead to a range of human genetic diseases, including cancers. For example, von Hippel–Lindau Syndrome leads to the development of clear cell renal

Biochemical Society Transactions (2017) **45** 303–311
DOI: 10.1042/BST20160422

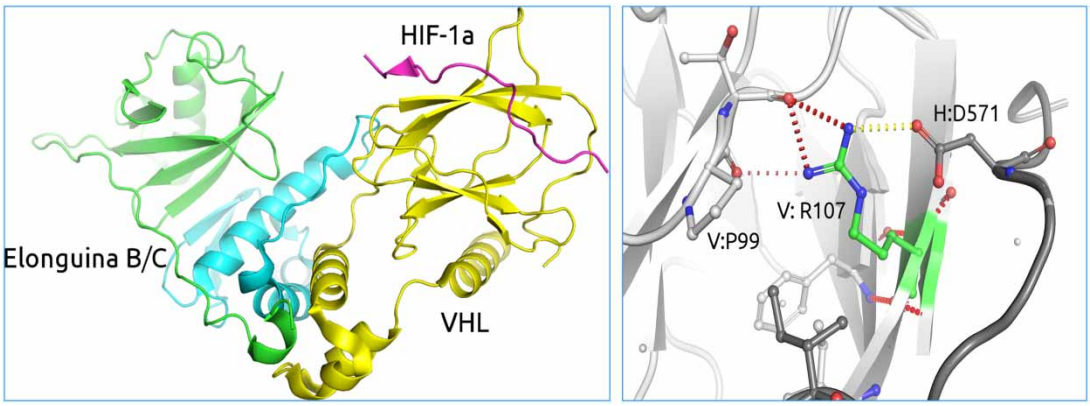

Figure 2. Genetic diseases mutations: clear cell renal carcinoma in von Hippel–Lindau disease.
Left-hand side panel: the structure of the ternary complex of pVHL with Elongin C and Elongin B, critical for pVHL stability and function [50]. Right-hand side panel: Inter-subunit interactions mediated by arginine (R107). In VHL disease some mutations, for example at R107, alter the charge complementarity of the subunits and destabilise protein–protein interactions, essential to function.

carcinoma and is caused by mutations in the VHL gene (Figure 2). By considering the effect of the mutations upon the stability of the VHL protein and how they alter the affinity for its various binding partners, HIF-1α and Elongin B and C, we could accurately identify a patient's risk of developing clear cell renal carcinoma [43,44]. In subsequent studies looking at other human Mendelian diseases, we have consistently observed that mutations affecting protein stability represent the majority of disease mutations, but those that affect key protein–protein interactions have proved also to be common and important [45–48]. Interestingly, changes in protein stability have even been linked to population phenotypic variability, including drug responses [49].

The same molecular consequences also drive the development of drug resistance. Looking at single-point coding mutations in *Mycobacterium tuberculosis*, we observed strong correlations between the structural features of mutations and their link to antibiotic sensitivity [51]. This suggested the possibility that we could predict resistance mutations before they arise, allowing us to suggest optimal therapeutic interventions based on genomic sequences and guide drug development.

We have applied this theory to our recent efforts to develop small-molecule inhibitors of GuaB2 to treat *Mycobacterium tuberculosis*. Analysis of the crystal structure of GuaB2 with VCC234718 revealed the most likely resistance mutation would occur at Y487, altering interactions with the inhibitor but not NAD, while not disrupting protein stability or the interactions between the homo-tetramer units. Y487C was the main resistance mutation observed [52]. By modifying the inhibitor, altering the interactions being made, the resistance profile could be altered. Other molecules were identified that did not make these interactions and were active against the Y487C mutant [53] (Figure 3). While these inhibitors were shown to be active against Mtb *in vitro* through the inhibition of GuaB2, further work is needed to explore the potential therapeutic benefit of GuaB2 inhibition *in vivo*.

These analyses highlight the power of considering the structural environment of a mutation in order to understand the molecular and biological consequences. While many disease mutations are buried, simple classification of residues by solvent accessibility does not provide all of the information encompassed within the graph-based signatures, necessary to consider longer-range allosteric effects and to distinguish disease-causing mutations. This led us to consider other structural features to locate a residue within the interior of the structure.

Refining our understanding of the impacts of mutations on protein function

The interiors of protein structures can be further classified using the conservation of residues in an orthologous family, as well as the packing density and the depth of the residue. These can provide more detailed information regarding the structural environment used in our substitution tables in SDM, allowing more accurate prediction of the effects of mutation and better identification of disease-related mutations.

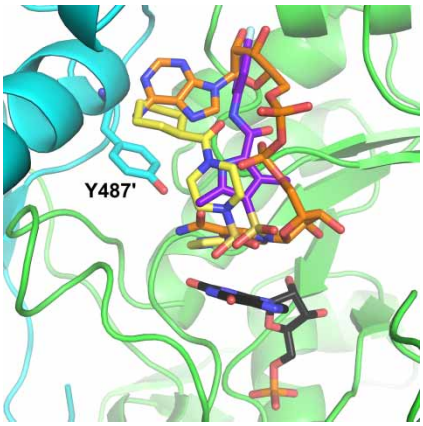

Figure 3. Avoiding resistance hotspots.

The GuaB2 inhibitor VCC234718 (represented as yellow sticks [52]) stacks on top of IMP (represented as black sticks) in a similar manner to NAD (represented as orange sticks), but makes different interactions to the neighbouring protomer unit through Y487, which is mutated leading to resistance. Optimisation of a separate scaffold (Compound 6, represented as purple sticks [53]) more closely mimicked the interactions made by NAD, and was active against the Y487C mutant.

We have recently analysed the residue side-chain relative solvent accessibility (RSA) [54], packing density, [55] and depth below the surface (http://cospi.iiserpune.ac.in/depth) [56,57] of residues across ∼15 000 high-resolution (≤2.5 Å), representative protein structures. This demonstrated that solvent-inaccessible residues (RSA <7%) exhibit varying levels of packing, with those closer to the core of the protein being more tightly packed. This suggests that further categorisation<UK ise of the environment-specific substitution tables in SDM with respect to depth and packing density might be useful in the prediction of the impact of mutations on protein stability [32].

We have found that residue conservation progressively increases as a function of residue packing, but the dependence on depth is more residue-specific (Figure 4; Pandurangan, A.P., work in progress). We are now modifying the environment-specific substitution tables to reflect these observations. Both these structural features could also add useful information while training models using machine-learning approaches to predict the impact of mutations on protein stability and affinity.

We have also asked whether we could better identify disease-causing mutations based on residue depth, packing density and conservation scores of the protein orthologues. To test this idea, we have investigated mutations occurring in alkaptonuria (AKU), which was the first characterised human Mendelian disease and is also known as black bone disease. AKU is caused by mutations in homogentisate-1,2-dioxygenase (HGD; PDB ID: 1EY2) that lead to the accumulation of reactive metabolites. There are over 80 known single-point mutations in HGD that have been linked to the development of AKU. We have previously shown that these mutations affect the activity of the protein complex primarily through destabilisation of the individual protomer structure but also through disruption of protomer–protomer interactions within the hexamer [45,46].

To visualise the relationship between residue conservation, depth and packing, we mapped the respective properties onto the structure of HGD (Figure 5). The residue conservation score was calculated using ConSurf [59]. In AKU, about 70% of the disease mutants are highly conserved (conservation scores >7), whereas for non-disease residues only 36% are highly conserved. Disease mutants proximal to the interface and the iron-binding site are well conserved for the purpose of maintaining interaction and function, respectively. In general, residue conservation could act as a potential indication to distinguish between disease mutant and non-disease residues.

To better dissect the properties unique to disease mutants, we looked at the correlation between depth and conservation for disease mutants and non-disease residues. As expected, depth was strongly correlated with conservation for both disease mutants and non-disease residues. For disease-causing mutants, the correlation between packing and conservation was significantly higher compared to the non-disease residues in AKU. Residue conservations have been previously shown to be greater than expected from analysis of amino acid, environment-dependent substitution tables in residues that are closer to functional binding sites [60], and have been shown to be dependent on the distance of the residue from the catalytic site [60,61].

Biochemical Society Transactions (2017) **45** 303–311
DOI: 10.1042/BST20160422

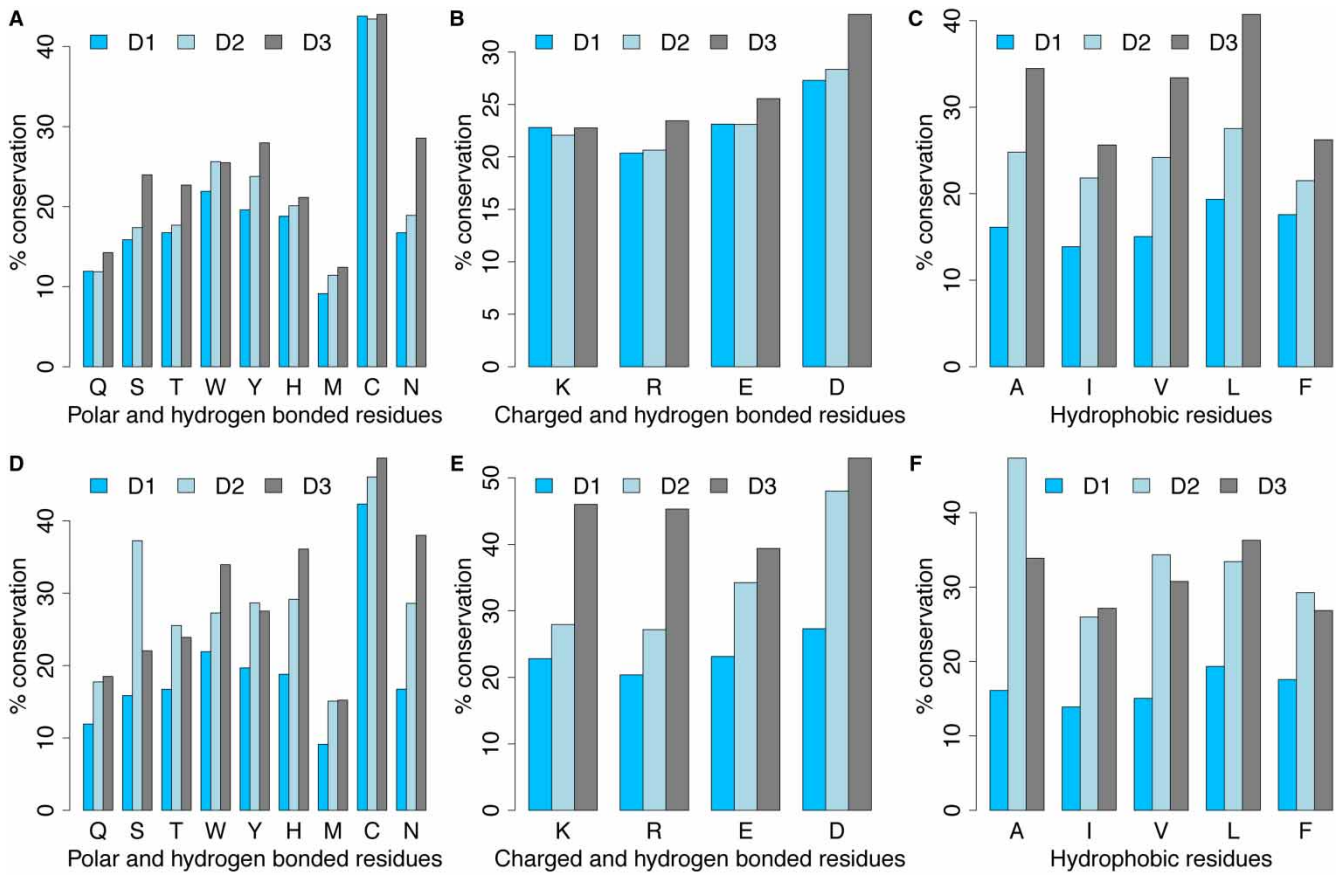

Figure 4. Residue conservation as a function of residue packing (A–C) and depth (D–F).

D1 corresponds to all residues with RSA > 7%. The remaining inaccessible residues were categorised into the D2 and D3. For the case of packing (**A–C**), D2 and D3 corresponds to residue packing values ≤0.56 and >0.56, respectively. For depth (**D–F**), D2 and D3 corresponds to residue depth values ≤6 and >6 Å, respectively. Residue conservation with packing and depth were analysed for residues grouped on the basis of polar, charged and hydrophobic properties. Residue conservations were obtained from the substitution tables calculated using the ULLA program [58]. Ulla uses structural alignment of protein families annotated by the program JOY, indicating the local conformation and side-chain environment [54] for the purpose of calculating substitution tables.

This suggests that residue packing, depth and conservation are crucial parameters for the understanding of evolutionary constraints in protein structure and have potential applications in predicting the impact of mutations on protein stability and interactions. While residue depth is well correlated with residue conservation, the residue packing density offers a greater potential to differentiate between disease and non-disease mutations. We are now investigating the impact of water molecules that are bound in the interior and are completely solvent-inaccessible in order to further refine the calculation of packing densities.

In many cases multiple mutations have impacts on resistance, some increasing the resistance and others compensating for the original mutation. One approach is to use SDM or mCSM (see above) to estimate whether a single mutation is stabilising or destabilising. We model this mutation into the protein and treat it as the wild-type. We can then explore whether further mutations are likely to be stabilising, compensating for the first or lead to further instability. In future, we intend to look at the impacts of multiple mutations using machine-learning methods.

Future challenges and applications of structure-guided drug discovery

Structure-guided approaches have demonstrated clearly that they can contribute to many aspects of early drug discovery, both computational and experimental. Structure can also inform our understanding of impacts of mutations in human genetic diseases and drug resistance in cancers and infectious disease. The Open Source

Biochemical Society Transactions (2017) **45** 303–311
DOI: 10.1042/BST20160422

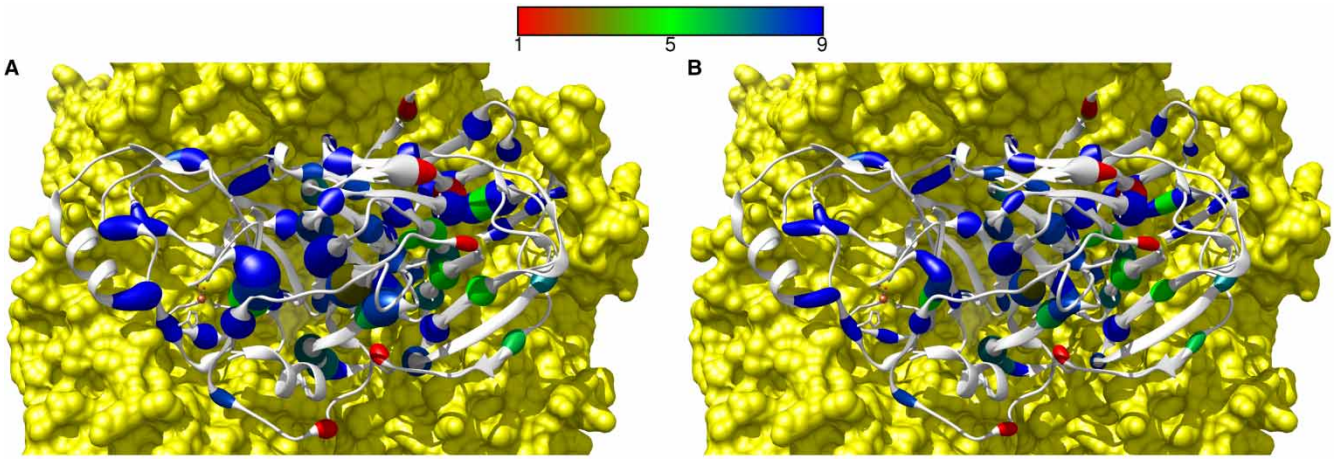

Figure 5. Mapping of residue conservation along with (A) residue packing and (B) residue depth for disease-related mutations in AKU.
In (**A**) and (**B**), one of the protomers is shown as a ribbon diagram and the remaining five protomers are shown as a surface representation coloured in yellow. The disease mutants are coloured based on residue conservation. The conservation score range between 1 and 9 with low (in red) and high (in blue) score values corresponding to low and high conservation, respectively. In (**A**), the increasing thickness of the ribbon corresponds to the increasing residue packing in the protomer. Similarly, in (**B**), the increasing thickness of the ribbon corresponds to the increasing residue depth in the protomer. The colour gradient bar representing the residue conservation score is shown on top of the figure.

Drug Discovery programme in India (http://www.osdd.net/about-us) has developed approaches using structural information to identify current, off-licence drugs that might be repurposed for use against infectious diseases in developing countries.

Our and others' research on understanding the impacts of mutations in genetic disease has demonstrated that mechanisms often involve interfacial disruptions and allosteric impacts on multiprotein assemblies and more simple oligomeric protein functions. Here, the design of interfacial stabilisers offers an attractive way forward that depends on a detailed knowledge of protein structure. Other approaches to combatting resistance include the use of peptidic inhibitors, which have more intrinsic flexibility and can often adopt other conformations to make alternative interactions with the mutant protein.

Our analysis of mechanisms of emergence of resistance in cancer and infectious disease and the methods for predicting it have been equally dependent on structural insights on protein targets. In combating resistance, an initial objective might be to design the candidate drug to interact only with residues used by the natural ligands; this would increase the chance that mutations affect the viability of the tumour or infectious agent as well as the drug. However, if it is necessary to find drug interactions beyond the interaction site of the natural substrates and cofactors, *in silico* saturation mutagenesis can be used to inform the choice of ways to grow fragments or redesign drugs, in such a way that they would less likely give rise to resistance mutations. In all these ways structural biology appears to be a central component of future efforts in drug discovery!

Abbreviations

AKU, alkaptonuria; HGD, homogentisate-1,2-dioxygenase; NMR, nuclear magnetic resonance; RSA, relative solvent accessibility; SDM, Site-Directed Mutator; SPR, surface plasmon resonance.

Funding

D.B.A. and T.L.B. are supported by the Wellcome Trust Programme Grant [093167/Z/10/Z]. A.P.P. and T.L.B. have been supported by the Gates HIT-TB and the EU MM4TB [Project ID: 260872] programmes. S.E.T. and T.L.B. are supported by Cystic Fibrosis Trust (Registered as a charity in England and Wales (1079049) and in Scotland (SC040196).

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

Biochemical Society Transactions (2017) **45** 303–311
DOI: 10.1042/BST20160422

References

1 Perutz, R.R., Liquori, A.M. and Eirich, F. (1951) X-ray and solubility studies of the haemoglobin of sickle-cell anaemia patients. *Nature* **167**, 929–931 PMID:14843112

2 Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G. and North, A.C.T. (1960) Structure of hæmoglobin: a three-dimensional Fourier synthesis at 5.5-Å. Resolution, obtained by X-ray analysis. *Nature* **185**, 416–422 doi:10.1038/185416a0

3 Phillips, D.C. (1966) The three-dimensional structure of an enzyme molecule. *Sci. Am.* **215**, 78–90 doi:10.1038/scientificamerican1166-78

4 Adams, M.J., Blundell, T.L., Dodson, E.J., Dodson, G.G., Vijayan, M., Baker, E.N. et al. (1969) Structure of rhombohedral 2 zinc insulin crystals. *Nature* **224**, 491–495 doi:10.1038/224491a0

5 Pullen, R.A., Lindsay, D.G., Wood, S.P., Tickle, I.J., Blundell, T.L., Wollmer, A. et al. (1976) Receptor-binding region of insulin. *Nature* **259**, 369–373 doi:10.1038/259369a0

6 Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C. and Hill, R.L. (1969) A possible three-dimensional structure of bovine α-lactalbumin based on that of hen's egg-white lysozyme. *J. Mol. Biol.* **42**, 65–86 doi:10.1016/0022-2836(69)90487-2

7 Bedarkar, S., Turnell, W.G., Blundell, T.L. and Schwabe, C. (1977) Relaxin has conformational homology with insulin. *Nature* **270**, 449–451 doi:10.1038/270449a0

8 Blundell, T.L., Bedarkar, S., Rinderknecht, E. and Humbel, R.E. (1978) Insulin-like growth factor: a model for tertiary structure accounting for immunoreactivity and receptor binding. *Proc. Natl Acad. Sci. U.S.A.* **75**, 180–184 doi:10.1073/pnas.75.1.180

9 Fyhrquist, F. and Saijonmaa, O. (2008) Renin-angiotensin system revisited. *J. Intern. Med.* **264**, 224–236 doi:10.1111/j.1365-2796.2008.01981.x

10 Blundell, T., Sibanda, B.L. and Pearl, L. (1983) Three-dimensional structure, specificity and catalytic mechanism of renin. *Nature* **304**, 273–275 doi:10.1038/304273a0

11 Foundling, S.I., Cooper, J., Watson, F.E., Cleasby, A., Pearl, L.H., Sibanda, B.L. et al. (1987) High resolution X-ray analyses of renin inhibitor-aspartic proteinase complexes. *Nature* **327**, 349–352 doi:10.1038/327349a0

12 Rahuel, J., Priestle, J.P. and Grütter, M.G. (1991) The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. *J. Struct. Biol.* **107**, 227–236 doi:10.1016/1047-8477(91)90048-2

13 Dhanaraj, V., Dealwis, C.G., Frazao, C., Badasso, M., Sibanda, B.L., Tickle, I.J. et al. (1992) X-ray analyses of peptide–inhibitor complexes define the structural basis of specificity for human and mouse renins. *Nature* **357**, 466–472 doi:10.1038/357466a0

14 Abagyan, R. and Totrov, M. (2001) High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **5**, 375–382 doi:10.1016/S1367-5931(00)00217-9

15 Albiston, A.L., Morton, C.J., Ng, H.L., Pham, V., Yeatman, H.R., Ye, S. et al. (2008) Identification and characterization of a new cognitive enhancer based on inhibition of insulin-regulated aminopeptidase. *FASEB J.* **22**, 4209–4217 doi:10.1096/fj.08-112227

16 Chai, S.Y., Yeatman, H.R., Parker, M.W., Ascher, D.B., Thompson, P.E., Mulvey, H.T. et al. (2008) Development of cognitive enhancers based on inhibition of insulin-regulated aminopeptidase. *BMC Neurosci.* **9**(Suppl 2), S14 doi:10.1186/1471-2202-9-S2-S14

17 Sigurdardottir, A.G., Winter, A., Sobkowicz, A., Fragai, M., Chirgadze, D., Ascher, D.B. et al. (2015) Exploring the chemical space of the lysine-binding pocket of the first kringle domain of hepatocyte growth factor/scatter factor (HGF/SF) yields a new class of inhibitors of HGF/SF-MET binding. *Chem. Sci.* **6**, 6147–6157 doi:10.1039/C5SC02155C

18 Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. and Taylor, R.D. (2003) Improved protein-ligand docking using GOLD. *Proteins* **52**, 609–623 doi:10.1002/prot.10465

19 Goodsell, D.S. and Olson, A.J. (1990) Automated docking of substrates to proteins by simulated annealing. *Proteins* **8**, 195–202 doi:10.1002/prot.340080302

20 Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T. et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**, 1750–1759 doi:10.1021/jm030644s

21 Blundell, T.L., Jhoti, H. and Abell, C. (2002) High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov.* **1**, 45–54 doi:10.1038/nrd706

22 Murray, C.W., Verdonk, M.L. and Rees, D.C. (2012) Experiences in fragment-based drug discovery. *Trends Pharmacol. Sci.* **33**, 224–232 doi:10.1016/j.tips.2012.02.006

23 Snyder, P.W., Lockett, M.R., Moustakas, D.T. and Whitesides, G.M. (2014) Is it the shape of the cavity, or the shape of the water in the cavity? *Eur. Phys. J. Spec. Top.* **223**, 853–891 doi:10.1140/epjst/e2013-01818-y

24 Kulp, III, J.L., Kulp, Jr, J.L., Pompliano, D.L. and Guarnieri, F. (2011) Diverse fragment clustering and water exclusion identify protein hot spots. *J. Am. Chem. Soc.* **133**, 10740–10743 doi:10.1021/ja203929x

25 Hajduk, P.J., Meadows, R.P. and Fesik, S.W. (1997) Discovering high-affinity ligands for proteins. *Science* **278**, 497–499 doi:10.1126/science.278.5337.497

26 Winter, A., Higueruelo, A.P., Marsh, M., Sigurdardottir, A., Pitt, W.R. and Blundell, T.L. (2012) Biophysical and computational fragment-based approaches to targeting protein–protein interactions: applications in structure-guided drug discovery. *Q. Rev. Biophys.* **45**, 383–426 doi:10.1017/S0033583512000108

27 Pellegrini, L., Yu, D.S., Lo, T., Anand, S., Lee, M., Blundell, T.L. et al. (2002) Insights into DNA recombination from the structure of a RAD51–BRCA2 complex. *Nature* **420**, 287–293 doi:10.1038/nature01230

28 Blundell, T.L., Sibanda, B.L., Montalvao, R.W., Brewerton, S., Chelliah, V., Worth, C.L. et al. (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 413–423 doi:10.1098/rstb.2005.1800

29 Scott, D.E., Marsh, M., Blundell, T.L., Abell, C. and Hyvonen, M. (2016) Structure-activity relationship of the peptide binding-motif mediating the BRCA2: RAD51 protein–protein interaction. *FEBS Lett.* **590**, 1094–1102 doi:10.1002/1873-3468.12139

30 Bar-On, O., Mussaffi, H., Mei-Zahav, M., Prais, D., Steuer, G., Stafler, P. et al. (2015) Increasing nontuberculous mycobacteria infection in cystic fibrosis. *J. Cyst. Fibrosis* **14**, 53–62 doi:10.1016/j.jcf.2014.05.008

31 Gong, S. and Blundell, T.L. (2010) Structural and functional restraints on the occurrence of single amino acid variations in human proteins. *PLoS ONE* **5**, e9186 doi:10.1371/journal.pone.0009186

32 Topham, C.M., Srinivasan, N. and Blundell, T.L. (1997) Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. *Protein Eng. Des. Sel.* **10**, 7–21 doi:10.1093/protein/10.1.7
33 Worth, C.L., Preissner, R. and Blundell, T.L. (2011) SDM — a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* **39**, W215–W222 doi:10.1093/nar/gkr363
34 Pires, D.E.V., Ascher, D.B. and Blundell, T.L. (2014) DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* **42**, W314–W319 doi:10.1093/nar/gku411
35 Pires, D.E.V., Ascher, D.B. and Blundell, T.L. (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **30**, 335–342 doi:10.1093/bioinformatics/btt691
36 Ascher, D.B., Jubb, H.C., Pires, D.E.V., Ochi, T., Higueruelo, A. and Blundell, T.L. (2015) Protein–protein interactions: structures and druggability. In: *Multifaceted Roles of Crystallography in Modern Drug Discovery* (Scapin, G., Patel, D. and Arnold, E., eds), pp. 141–163, Springer Netherlands, Dordrecht
37 Jubb, H., Blundell, T.L. and Ascher, D.B. (2015) Flexibility and small pockets at protein–protein interfaces: new insights into druggability. *Prog. Biophys. Mol. Biol.* **119**, 2–9 doi:10.1016/j.pbiomolbio.2015.01.009
38 Pires, D.E.V. and Ascher, D.B. (2016) mCSM-AB: a web server for predicting antibody–antigen affinity changes upon mutation with graph-based signatures. *Nucleic Acids Res.* **44**, W469–W473 doi:10.1093/nar/gkw458
39 Coelho, M.B., Ascher, D.B., Gooding, C., Lang, E., Maude, H., Turner, D. et al. (2016) Functional interactions between polypyrimidine tract binding protein and PRI peptide ligand containing proteins. *Biochem. Soc. Trans.* **44**, 1058–1065 doi:10.1042/BST20160080
40 Pires, D.E.V. and Ascher, D.B. (2016) CSM-lig: a web server for assessing and comparing protein–small molecule affinities. *Nucleic Acids Res.* **44**, W557–W561 doi:10.1093/nar/gkw390
41 Pires, D.E.V., Blundell, T.L. and Ascher, D.B. (2015) Platinum: a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. *Nucleic Acids Res.* **43**, D387–D391 doi:10.1093/nar/gku966
42 Pires, D.E.V., Blundell, T.L. and Ascher, D.B. (2016) mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Sci. Rep.* **6**, 29575 doi:10.1038/srep29575
43 Pires, D.E.V., Chen, J., Blundell, T.L. and Ascher, D.B. (2016) In silico functional dissection of saturation mutagenesis: interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. *Sci. Rep.* **6**, 19848 doi:10.1038/srep19848
44 Forman, J.R., Worth, C.L., Bickerton, G.R.J., Eisen, T.G. and Blundell, T.L. (2009) Structural bioinformatics mutation analysis reveals genotype–phenotype correlations in von Hippel–Lindau disease and suggests molecular mechanisms of tumorigenesis. *Proteins* **77**, 84–96 doi:10.1002/prot.22419
45 Nemethova, M., Radvanszky, J., Kadasi, L., Ascher, D.B., Pires, D.E.V., Blundell, T.L. et al. (2016) Twelve novel HGD gene variants identified in 99 alkaptonuria patients: focus on 'black bone disease' in Italy. *Eur. J. Hum. Genet.* **24**, 66–72 doi:10.1038/ejhg.2015.60
46 Usher, J.L., Ascher, D.B., Pires, D.E.V., Milan, A.M., Blundell, T.L. and Ranganath, L.R. (2015) Analysis of HGD gene mutations in patients with alkaptonuria from the United Kingdom: identification of novel mutations. *JIMD Rep.* **24**, 3–11 doi:10.1007/8904_2014_380
47 Jubb, H.C., Pandurangan, A.P., Turner, M.A., Ochoa-Montaño, B., Blundell, T.L. and Ascher, D.B. (2016) Mutations at protein–protein interfaces: small changes over big surfaces have large impacts on human health. *Prog. Biophys. Mol. Biol.* doi:10.1016/j.pbiomolbio.2016.10.002
48 Jafri, M., Wake, N.C., Ascher, D.B., Pires, D.E.V., Gentle, D., Morris, M.R. et al. (2015) Germline mutations in the CDKN2B tumor suppressor gene predispose to renal cell carcinoma. *Cancer Discov.* **5**, 723–729 doi:10.1158/2159-8290.CD-14-1096
49 Silvino, A.C.R., Costa, G.L., de Araujo, F.C.F., Ascher, D.B., Pires, D.E.V., Fontes, C.J.F. et al. (2016) Variation in human cytochrome P-450 drug-metabolism genes: a gateway to the understanding of plasmodium vivax relapses. *PLoS ONE* **11**, e0160172 doi:10.1371/journal.pone.0160172
50 Stebbins, C.E., Kaelin, Jr, W.G. and Pavletich, N.P. (1999) Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 doi:10.1126/science.284.5413.455
51 Phelan, J., Coll, F., McNerney, R., Ascher, D.B., Pires, D.E.V., Furnham, N. et al. (2016) Mycobacterium tuberculosis whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. *BMC Med.* **14**, 31 doi:10.1186/s12916-016-0575-9
52 Singh, V., Donini, S., Pacitto, A., Sala, C., Hartkoorn, R.C., Dhar, N. et al. (2017) The inosine monophosphate dehydrogenase, GuaB2, Is a vulnerable new bactericidal drug target for tuberculosis. *ACS Infect. Dis.* **3**, 5–17 doi:10.1021/acsinfecdis.6b00102
53 Park, Y., Pacitto, A., Bayliss, T., Cleghorn, L.A., Wang, Z., Hartman, T. et al. (2017) Essential but not vulnerable: indazole sulfonamides targeting inosine monophosphate dehydrogenase as potential leads against Mycobacterium tuberculosis. *ACS Infect. Dis.* **3**, 18–33 doi:10.1021/acsinfecdis.6b00103
54 Mizuguchi, K., Deane, C.M., Blundell, T.L., Johnson, M.S. and Overington, J.P. (1998) JOY: protein sequence-structure representation and analysis. *Bioinformatics* **14**, 617–623 doi:10.1093/bioinformatics/14.7.617
55 Pattabiraman, N., Ward, K.B. and Fleming, P.J. (1995) Occluded molecular surface: analysis of protein packing. *J. Mol. Recognit.* **8**, 334–344 doi:10.1002/jmr.300080603
56 Chakravarty, S. and Varadarajan, R. (1999) Residue depth: a novel parameter for the analysis of protein structure and stability. *Structure* **7**, 723–732 doi:10.1016/S0969-2126(99)80097-5
57 Tan, K.P., Nguyen, T.B., Patel, S., Varadarajan, R. and Madhusudhan, M.S. (2013) Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Res.* **41**, W314–W321 doi:10.1093/nar/gkt503
58 Lee, S. and Blundell, T.L. (2009) Ulla: a program for calculating environment-specific amino acid substitution tables. *Bioinformatics* **25**, 1976–1977 doi:10.1093/bioinformatics/btp300
59 Goldenberg, O., Erez, E., Nimrod, G. and Ben-Tal, N. (2009) The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.* **37**, D323–D327 doi:10.1093/nar/gkn822
60 Chelliah, V., Chen, L., Blundell, T.L. and Lovell, S.C. (2004) Distinguishing structural and functional restraints in evolution in order to identify interaction sites. *J. Mol. Biol.* **342**, 1487–1504 doi:10.1016/j.jmb.2004.08.022
61 Jack, B.R., Meyer, A.G., Echave, J. and Wilke, C.O. (2016) Functional sites induce long-range evolutionary constraints in enzymes. *PLoS Biol.* **14**, e1002452 doi:10.1371/journal.pbio.1002452