Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Limitations and lessons in the use of X-ray structural information in drug design

Andrew M. Davis\textsuperscript{1}, Stephen A. St-Gallay\textsuperscript{1} and Gerard J. Kleywegt\textsuperscript{2}

\textsuperscript{1}Department of Medicinal Chemistry, AstraZeneca R&D Charnwood, Bakewell Road, Loughborough, Leicestershire LE11 5RH, United Kingdom
\textsuperscript{2}Department of Cell and Molecular Biology, Uppsala University, Biomedical Centre, Box 596, SE-751 24, Uppsala, Sweden

Introduction

The drug discovery industry has never been under such pressure to deliver as it is today. While R&D costs spiral higher, attrition statistics in clinical development show no signs of improvement and may actually be getting worse. For example, the FDA highlighted in March 2004 that a drug entering phase 1 clinical development in the year 2000 was less likely to reach the market than one that entered clinical development in 1985 [1]. It is a truism that much of the fate of a drug candidate in clinical development is embedded in the chemical structure and, hence, is in the control of the chemist at the point of design. Clearly, there are serious limitations on what drug designers have so far learnt about the relationship between chemical structure and attrition through clinical development.

Drug designers are ever-optimistic, and the hope is that an increasing understanding of the interactions of drug candidates with their protein targets, at the molecular level, may allow quality to be built into drug candidates at the design stage. Indeed, there are a growing number of targets that we are trying to inhibit or activate for which we do understand the relationship between inhibition or activation of the target at the molecular level and their pathophysiological effect. Advances in structural and molecular biology, as well as in biophysics, have led to the determination of high-resolution atomic structures of many of the protein targets of drug discovery projects. For instance, p38 kinase is a target that is well validated in the clinic with respect to its role in inflammation, and many high-resolution crystal structures are available. The same is true for thrombin for anti-thrombosis and renin for hypertension, iNOS for inflammation, EGF receptor tyrosine kinase for cancer, as well as many antibacterial and antiviral targets. Crystal structures of some of the most important drug-metabolising enzymes are also known [2–7], which potentially enables the structure-based rational optimisation of potency, selectivity and metabolism. However, routine control of potency, selectivity and metabolism based on the use of structural information has yet to become a reality.

While high-resolution protein structure information can be derived by both NMR spectroscopy and X-ray crystallography, most structures by far have been determined by crystallographic methods. The application of NMR spectroscopy to routine protein structure determination is limited, as it requires large quantities of soluble, multiply labelled protein, considerable time and is limited to comparatively small proteins. Solid-state NMR spectroscopy is being applied to...
protein structural work and removes the limitation of molecular size but is still in its infancy. In other words, we are reliant on X-ray crystallography as the major source of structural information, and therefore this review focuses upon the use of protein and protein–ligand structures derived by this technique.

The use of X-ray crystal structures to design-in potency and selectivity is not trivial and is fraught with difficulties and ambiguities that can mislead the unwary medicinal chemist. We have previously highlighted some of the limitations and ambiguities of the use of X-ray crystal structures in ligand and drug design [8], and provided some hints for modellers on how to assess the validity or reliability of crystal structures.

However, the literature still abounds with examples of failures, surprises and warnings that should be heeded by drug-discovery scientists. Here we highlight some recent examples, in the hope that medicinal chemists and crystallographers will be stimulated to have a more informed and fruitful dialogue, and that structural information will be used more optimally in drug discovery.

Assumptions
When protein crystal structures are used in structure-guided design, a number of fundamental assumptions are commonly made:

1. **The protein structure is correct.** It is usually assumed that the amino acid sequence is known and correct, that the structure model (including the water structure) is complete (i.e. no entities are missing) and that it is correct and known with high accuracy.

2. **The structure of the ligand and its interactions with the protein are correct.** An obvious assumption is that the chemical composition of the ligand is known and that its placement in the active site and its conformation is correct. A corollary of (1) and (2) is that the interactions between the receptor and the ligand are known, understood and correct.

3. **The protein–ligand structure is relevant for drug design.** It is usually tacitly assumed that the conditions under which the complex was crystallised are relevant, that the observed protein conformation is relevant for interaction with the ligand (i.e. no flexibility in the active-site residues) and that the structure actually contributes insights that will lead to the design of better compounds.

While these assumptions seem perfectly reasonable at first sight, they are not all necessarily true. Instead, each of the assumptions has to be carefully verified on a case-by-case basis. The following sections (and also reference [8]) describe some recent examples of cases where several of the above assumptions turned out to be invalid. They should act as a warning to any medicinal chemist embarking on a structure-guided drug-design project.

Assumption 1—the protein structure is correct

In the past 50 years, the determination of crystal structures of proteins, DNA, RNA, viruses and all manner of complexes has yielded fantastic insights into the molecular basis of myriad biological systems and processes. However, many scientists who use structural information seem to be unaware of the fact that an X-ray crystal structure is one crystallographer’s subjective interpretation of an experimental electron density map expressed in terms of an atomic model [9,10]. Such an interpretation may contain errors, it may not be complete, and there may exist alternative interpretations for parts of it. In general, the resolution of the study and the degree of experience of the crystallographer are the most important factors determining the quality of the final model that is deposited in the Worldwide Protein Data Bank (wwPDB) [11]. Even when the resolution is very high, and the experimental data allow modelling of the structure at a great level of detail, there may still be parts of the structure for which the electron density is poor or that another crystallographer might have interpreted differently. If the resolution is low, the probability of errors and incomplete modelling of the data is much greater. Nevertheless, most chemists who undertake structure-based design treat a protein crystal structure reverently as if it was determined at very high resolution, regardless of the resolution at which the structure was actually determined (admittedly, crystallographers themselves are not immune to this practice either). Also, the fact that the crystallographer is bound to have made certain assumptions, to have had certain biases and perhaps even to have made mistakes is usually ignored. Assumptions, biases, ambiguities and mistakes may manifest themselves (even in high-resolution structures) at the level of individual atoms, of residues (e.g. sidechain conformations) and beyond. In general, the more serious such problems are, the less often they occur. However, even in this day and age it does occasionally happen that a crystal structure is published and deposited and later turns out to be seriously flawed.

The recent retraction in the journal Science of five papers and five structures (wwPDB codes 1JSQ, 1PF4, 1Z2R for MsbA and 1S7B, 2F2M for EmrE) of ATP-binding cassette (ABC) transporters [12] sent shockwaves through the crystallographic and membrane protein communities. The fact that there was a problem with these models came to light when the structure of the related protein SAV1866 [13] was determined. This showed that the structures of MsbA and EmrE were incorrect and that their shape roughly resembled the mirror image of the correct structure. In this case, unfortunately, it appears that the incorrect structures have had serious adverse effects on the development of the field and possibly also on the distribution of grant money [14]. The location, direction and connectivity of several trans-membrane helices were incorrect [13], and these errors completely invalidated any inferences made from the models. Some crystallographic statistics indicated problems, but this only caused the authors to resort to inappropriate refinement protocols. Although a ‘software error’ may have caused the initial problems, it is hard to see how the alarm bells that must inevitably have been set off could have been ignored during the five independent structure determinations (especially as some of the later models were determined to higher resolution at 3.7–3.8 Å). Moreover, the models were incompatible with much of the available biological data [17]. Unfortunate deposition practices (of the crystallographers and the journals [15]) meant that for the first model only C-α coordinates were deposited, and for only one of the five models were the experimental data ever deposited. This means that for a long time it was impossible to even reproduce the electron density maps for any ABC-transporter structure. A more detailed and technical discussion of this case can be found in [16–18].

In 2001, the structure of the *Staphylococcus aureus* transcription regulator SarA was described in *Nature* along with a complex of SarA with DNA [20] (wwPDB codes 1FZN and 1FZP, respectively). Later that year, the structure of a homologous protein, SarR, was
The lesson for modellers

Fortunately, cases where (almost) the entire protein structure is modelled incorrectly are very rare. This usually occurs only if the resolution is low and if sensible procedures for model building, refinement and validation are jettisoned. However, it is important to remember that almost every crystal structure, even at atomic resolution, can have problematic parts or aspects [8]. Inspection of electron density together with the model (and preferably in the presence of the crystallographer!) may help in identifying any such parts and in assessing the reliability of the model in areas of particular interest (binding site, catalytic residues, interaction motifs, etc.). From 1 February 2008, the Worldwide Protein Data Bank have amended their deposition practices to require the deposition of structure factors as well as coordinates for all new structures. This means that all new publicly available structure models will be able to be viewed with the experimental data from which they were derived. The interested reader can obtain further information on good practice in the validation of protein structures from references [28, 29] and a tutorial is available at http://xray.bmc.uu.se/embo2001/modval/. Furthermore, one should keep in mind that ‘high-resolution questions’ (e.g. pertaining to non-bonded distances or the precise position, orientation and conformation of sidechains and other interesting moieties) can only be answered reliably when high-resolution data are available. This is a truism of which even seasoned crystallographers need to be reminded occasionally!

Assumption 2—the structure of the ligand and its interactions with the protein are correct

In a pharmaceutical context, the structures of protein–ligand complexes are of major interest. Although one might think that,
once the structure of a protein is known, the crystallographic modelling of its ligand complexes should be straightforward, this is not always the case. In this section, we discuss some examples of problems that may arise. For instance, we have previously [8] described a case in which a ligand that had been modelled in complex with an enzyme was later found to be absent after closer inspection of the electron density [30,31]. However, the opposite situation also arises frequently, namely that there is density for some entity (either retained from the expression system or present in the crystallisation soup) that is either not recognised as such by the crystallographer, or for which it is not possible to establish the chemical identity of the compound (or mixture of compounds).

Peroxisome proliferator-activated receptors (PPARs) are nuclear receptors for fatty acid ligands. When the structure of the ligand-binding domain of PPAR-β/δ was determined in 1999 [32] (wwPDB code 2GWX), it was a major surprise to find that the apo-form of the domain displayed essentially the same conformation as the activated (ligand-loaded) form (Figure 2a). A few years ago, the apo-structure was re-determined in a different laboratory [33] and it was found that it contained a mixture of fatty acids (acquired from the bacterial expression system) in the ligand-binding site, with cis-vaccenic acid being the major component (wwPDB codes 2AWH and 2B50). This observation prompted the authors of the latter work to re-evaluate the earlier structure [34]. Using the original diffraction data of Xu et al. and their own high-resolution model as a molecular replacement probe, they solved and refined the structure anew (wwPDB code 2BAW). Initial electron-density maps clearly revealed the presence of a fatty acid in the binding site (Figure 2b). The density for this ligand (probably cis-vaccenic acid as well) had originally been modelled as a network of water molecules. Moreover, two molecules of n-heptyl-β-D-glucopyranoside, a compound that is important for crystallisation, which had previously been overlooked were identified in the density. The re-evaluation of the older structure has several important implications [33]. First, the conundrum of the activated conformation in the ‘apo’ structure has been explained. Second, previously determined binding data involving this protein domain must be treated with caution, unless it is clear that any fatty acids were removed before the analysis. Third, the fact that the receptor binds endogenous fatty acid ligands tightly means that the possibility exists that it is constitutively active.

We have previously described how careful crystallographic modelling and refinement practices can sometimes help to detect errors that had previously gone unnoticed [8].

In a high-throughput screen of compounds against dihydrofolate reductase, compound 2 was identified as a relatively potent and structurally unusual inhibitor [35], (Figure 3). When the crystal structure of the complex was determined, the electron density revealed that compound 2 could not readily be modelled. Since high-resolution data had been collected, it was possible to use the (weak) anomalous signal from the sulfur atoms to propose that compound 2 was in fact compound 1 [36] (wwPDB entry 2ANQ). This hypothesis was subsequently confirmed with 13C-NMR methods and was also supported by SAR data (derivatives of compound 2 were inactive, whereas those based on compound 1 showed significant activity).

In 2004, Holmner et al. described the structure of a hybrid between cholera toxin and heat-labile enterotoxin in complex with a pentasaccharide ligand [37] (wwPDB entry 1TL0).
ligand was assumed to contain an N-acetylgalactosamine unit and the absence of any density for the N-acetyl group was taken as an indication that the group was flexible. However, it was discovered later that the ligand in fact contained a glucose unit instead of the presumed N-acetylgalactosamine unit [38] (wwPDB entry 2NZG). Fortunately, in this case the mistake did not affect the conclusions drawn from the structure.

More generally, crystallographic modelling of oligosaccharides appears to cause particular problems [39–42]. Crispin et al. reported that about a third of the structures in the wwPDB that contain oligosaccharides have ‘significant errors in carbohydrate stereochemistry, nomenclature or even consistency with the electron density maps’ [41]. Some structures even contain previously unobserved glycosidic linkages that are incompatible with the known biosynthetic routes of N-glycan processing. Clearly, high-resolution decisions, such as the assignment of unusual structural features (especially if they are incompatible with known chemistry or biology) should not be made on the basis of low-resolution electron density.

The reasons why ligands in general tend to be modelled much less accurately by crystallographers in terms of stereochemistry and geometry than amino acids have been discussed [43,44]. Besides in the papers cited in this section, examples of ligands with poor stereochemistry in the wwPDB are also discussed in references [45–48].

The lesson for modellers

Some cases cited in this section are examples of serious mistakes that should hopefully not occur too often. However, in an everyday structure-based design context other problems are likely to occur, and it is good to be aware of these pitfalls. Many of these problems have been discussed previously [8], such as difficulties involved in determining the orientation of asparagine, glutamine and histidine sidechains and in the assignment of density features to water molecules. Other problems may arise when the density for a ligand is poor and the placement of the ligand by the crystallographer is questionable (but perhaps not debated!). As a matter of fact, modellers can make an important contribution themselves to the structure determination of protein–ligand complexes. First, their knowledge of organic chemistry and stereochemistry of small molecules is often better than that of a protein crystallographer, so the modellers could help formulate appropriate refinement dictionaries with proper restraints and target values for bond lengths, etc. Second, their knowledge of, and eye for, judging protein–ligand interactions could help the crystallographer, by proposing ligand poses that both fit the electron density and make good sense in terms of protein–ligand interactions.

Assumption 3—the protein–ligand structure is relevant for drug design

The widespread availability of protein structures has revolutionised the drug discovery industry, with the wwPDB now containing over 50 000 publicly available structures, with many more only available within pharmaceutical companies’ secret vaults. With recent breakthroughs, structures of pharmaceutically important membrane proteins are now also becoming accessible. Protein–ligand structures have even spawned a new drug discovery paradigm, in one of the areas most refractory to rational design, lead discovery, with fragment-based screening using crystal structures.

But our understanding of the thermodynamics of protein ligand complexation is still developing. X-ray crystal structures that question or confront our understanding are opportunities to develop our learning.

The crystallisation conditions are relevant for drug design

The conditions of crystallisation are often assumed to be absolutely relevant to the conditions of the biological assay. However, changes in buffer constituents, pH and crystallisation conditions can have a profound effect on the conformation of both ligands and proteins. For instance, the severe acute respiratory syndrome (SARS) coronavirus main protease was crystallised at different pH values and in complex with a specific inhibitor. The structures revealed substantial pH-dependent conformational changes and an unexpected mode of binding for the substrate-analogue inhibitor [49]. At a pH value of 6, the structure of the monomers in the homodimer differs (one being in the active and the other in the inactive conformation) and the inhibitor binds in a different mode to each monomer.

It is often assumed that, at least for a single composition of mother liquor and independent of whether a complex is formed by soaking or co-crystallisation, a single reproducible crystal structure can be determined. However, a recent study of aldose reductase in complex with tolrestat and zopolrestat highlights that these assumptions are not necessarily valid [50]. The authors observed a flip of a peptide bond depending on whether zopolrestat was soaked in or co-crystallised. The peptide flip resulted in the rupture of a key hydrogen bond to the ligand. With tolrestat as the ligand, complexes with two different stoichiometries were obtained, one with one and one with four inhibitor molecules bound. Accommodation of four ligands caused appreciable shifts of two helices that interact with the additional ligands.

An example of two different binding modes observed for chemically similar ligands is provided by the crystal structure of E. coli ketopantoate reductase in complex with 2’-phospho-ADP-ribose (a fragment of NADP+ that lacks the nicotinamide ring) [51]. In an attempt to crystallise the ternary complex of the enzyme with pantoate and NADPH, crystals were obtained but the density showed no traces of pantoate and only a fragment of NADPH could be located. This fragment was confirmed to be 2’-phospho-ADP-ribose by mass-spectrometry experiments. Compared with the complex with NADP+, the ligand binds in the opposite orientation. Isothermal titration calorimetry with several mutants showed that the unusual binding mode is caused by changes in the protonation state of binding groups at low pH. The implication of this observation for fragment-based approaches to ligand design is that a binding mode may be altered profoundly by variations in crystallisation conditions as well as by (unexpected or intentional) modifications of ligands.

The effects of protein flexibility are small or at least understood

A number of reviews [8,52,53] have stressed the importance of protein flexibility in protein–ligand interactions. Protein flexibility induced by ligand binding is now being addressed not only as a real problem for predictive modelling, but also as a real opportunity for drug-design purposes. An increasing number of papers are published that describe attempts to incorporate protein flexibility
into modelling and the consequences of modelling protein flexibility in docking and virtual-screening applications [54–60].

To account for protein flexibility, pharmacophore models can be based on multiple conformations of the protein. In a study of such models for HIV1 protease, a single ensemble of 28 NMR models was found to display more structural variation than a collection of 90 crystal structures [61]. Pharmacophore models based on either set of structures worked well in discriminating true and decoy inhibitors, but the model based on the NMR ensemble appeared to be the most accurate yet general representation of the active site. It is not clear whether the increased structural variation of the NMR ensemble reflects modelling of true dynamics or under-determination of the structure by the data. Nevertheless, use of a more diffuse protein model would appear to be beneficial in virtual-screening applications.

Surprising ligand-induced conformational changes continue to be reported. Even in protein active sites that have been studied extensively and for which many crystal structures are available, surprises can still occur that provide further opportunities for medicinal chemistry exploitation. For instance, it is well known that human aldose reductase possesses two main binding pockets: a rigid anion-binding pocket, and a very flexible hydrophobic pocket, which can be open or closed depending on the bound ligand. Hence, it was surprising to find that a potent ligand from a naphthol[1,2-d]isothiazole acetic acid series extended the anion-binding site by opening a new subpocket [62].

Modelling of two series of adenosine kinase inhibitors proved confusing when the expected binding mode predicted from modelling studies failed to predict the observed structure-activity relationships [63]. 5-Iodotubercidin is chemically similar to adenosine and binds in a similar fashion to the enzyme. A new series of alkynylpyrimidine inhibitors was assumed to bind with the pyrimidine ring oriented in the same way as that of 5-iodotubercidin. However, when the crystal structure of a complex was determined it was found that the alkynylpyrimidine compound displayed a distinctly different binding mode. The adenosine-kinase structure accommodates the inhibitor by a 30° rotation of the small domain relative to the large domain, and the inhibitor binds in the opposite manner to what was expected.

It has previously been suggested that hydrophobic residues might play an important role in ligand-induced conformational changes in a protein active site. This hypothesis has recently been confirmed by an analysis of 98 high-resolution apo/holo structure pairs; 41 pairs that displayed little conformational change upon ligand binding, 35 with a moderate degree of induced fit, and 22 with substantial changes on ligand binding [64]. Pockets that did not undergo significant conformational changes tended to be dominated by polar active site residues and hydrogen-bond interactions. Binding pockets that did undergo ligand-induced conformational alterations, on the other hand, were found to be hydrophobic in nature, and tryptophan had a high propensity to occur in these flexible active sites.

The thermodynamics of drug–receptor interactions are understood

The free energy of binding is the result of large changes in enthalpy and entropy to the system, of which the observed protein–ligand complex is only one part, the whole system comprising the unbound ligand and protein, the bound ligand–protein complex, bulk solvent and its re-organisation on ligand binding and any displaced solvent molecules from the apo-protein active site. So, while the observation of an interaction in a crystal structure provides no information on the overall thermodynamic changes the system has undergone, the analysis of structural information is helping to at least unravel some unusual thermodynamic observations. Several recent studies have demonstrated the complexity and subtleties of the thermodynamics of drug–ligand/receptor interactions.

Hydrogen bonds between the ligand and protein are often seen as very important enthalpy-driven interactions, facilitating both specificity and selectivity. Hydrophobic interactions can yield significant increases in affinity as well, but are often viewed as non-specific and entropy-driven. However, detailed calorimetric studies have found many examples of enthalpy-driven hydrophobic interactions [67], as well as of hydrogen bonds formed between the ligand and protein that are entropy-driven [68].

Homans has analysed the hydrophobic binding of ligands to mouse urinary protein-1 and concluded that suboptimal hydration of the protein binding site causes the exchange of hydrophobic ligands with solvent to become enthalpy-driven, owing to favourable solute-solute dispersion forces. This leads to significant gains in binding affinity when shape complementarity is optimised [69].

Klebe has shown that even the seemingly most trivial structural homologations can highlight surprising complexity in structure-activity relationships [70]. A pair of thrombin inhibitors where the S3/S4 hydrophobic binding group is homologated from cyclopentyl to cyclohexyl gave identical binding affinities, instead of the expected increase in affinity of 3–4 kJ/mol for the incorporation of an additional methylene unit. X-ray crystallography showed that Silverman and co-workers designed hydroxyl and amino-substituted arginine analogues as peptidomimetic inhibitors of neuronal nitric oxide synthase (nNOS) [65]. Crystal structure analysis revealed the presence of an important structural water molecule that was hydrogen-bonded between the two propionate moieties of the NOS haem group (Figure 4a and c). Molecular modelling led to the design of N-hydroxy and N-amino analogues, which were shown by crystallography to successfully displace the active-site water molecule, but failed to improve in vitro potency significantly (Figure 4b and c). The hydrogen bonds to the propionate groups caused small displacements of the ligand that weakened other hydrogen-bonding interactions with the protein.

Modelling studies suggested that hydrogen-bonding groups at the 7-position of 1,2,3,4-tetrahydrobenz[h]isoquinoline inhibitors of phenethanolamine N-methyltransferase would confer increased potency. However, all compounds synthesised were actually less active than the parent H-substituted analogue. The authors suggest that this illustrates that there are limitations to the extent to which predictions based on docking studies can be employed to guide chemistry [66].
while the cyclopentyl group gave good density in the S3/S4 pocket, the cyclohexyl compound showed ill-defined density in the difference maps. This was interpreted as either static or dynamic disorder of the cyclohexyl group. ITC revealed that the equality of the Gibbs free energy between the two inhibitors was factored into very different enthalpic and entropic contributions. While the free energy of binding of the cyclopentyl compound was factored equally between entropy and enthalpy, the cyclohexyl compound showed a relatively larger entropic advantage in binding, compensated for by a weaker enthalpic contribution to binding. This is consistent with the observations from the crystallography. The high residual mobility of the cyclohexyl group results in the loss of good enthalpic contacts but is compensated by the smaller entropy penalty paid. While enthalpy–entropy compensation is a generally observed phenomenon, reflecting the interplay between the enthalpic benefit from ‘tight’ interactions, to the entropic price paid to form them, it is still surprising that such subtle medicinal chemistry changes can result in such large unpredictable changes in the biophysics of the system, and such perfect compensation.

The lesson for modellers

While an X-ray crystal structure of a ligand bound to its target protein is seductive in its clarity, the examples given in this section highlight that the thermodynamics of the system may confound a simple and straightforward interpretation based on the X-ray crystal model. Surprises highlighted by discrepancies between expected structure–activity relationships and observations from X-ray crystal structures are important opportunities to gain a more detailed understanding of the underlying physics of the system. The combination of X-ray crystal structural information, molecular dynamics simulations and calorimetric investigations is starting to unravel these subtle complexities. These surprises also provide the potential to gain chemical novelty in drug discovery programmes, and hence one should embrace these surprises as opportunities rather than interesting but academic distractions.

Case study—development of iNOS inhibitors

An AstraZeneca project to discover inhibitors of inducible nitric oxide synthase (iNOS), as a potential treatment for inflammatory diseases, demonstrated some limitations of X-ray crystallography and modelling for compound design highlighted in this review. A series of compounds existed with promising potency and selectivity [71] and several protein–ligand structures existed with the ligand complexed to the oxygenase domain of mouse iNOS. Selectivity for iNOS over the two other constitutively active isoforms, neuronal NOS (important for gut motility and long-term memory) and epithelial NOS (important in maintaining vascular tone), was a key requirement. The potential to design-in selectivity for iNOS over nNOS and eNOS inhibition, appeared quite challenging, as the only amino acid side chain difference within the active site was an Asp in iNOS that was an Asn in eNOS (Asp 376 in mouse iNOS). Our existing ligands did not make an interaction with this iNOS Asp residue. Hence, the compound design strategy was to build-in an interaction with this residue, in the hope this would engender increased iNOS potency and selectivity over eNOS.

The programme GRID [72] was used to analyse the active sites and suggested favourable positions for a basic nitrogen atom. One feature of GRID is that it considers the benefit of making an interaction over the cost of displacing active-site water molecules. A compound bearing a benzylamine substituent was designed to place the basic nitrogen within the favourable GRID density for a
basic nitrogen atom. The binding mode of the compound was predicted using the programme GOLD\[73,74\] and indicated it was possible to build an interaction between the basic nitrogen and the Asp376 side chain, while maintaining the binding mode demonstrated by existing compounds in the same chemical series (Figure 5).

However, once synthesised and tested, the activity of this compound was found to be much lower than predicted, in fact, the worst in the series. The structure–activity relationships showed that the interaction between the base and the acid side chain dramatically reduces potency (Table 1).

An X-ray structure was determined for the designed compound complexed to iNOS, and although it was at low resolution (only 3.3 Å), inspection of the structure model seemed to confirm the binding mode predicted by GOLD with the interaction between the base and Asp376 (Figure 5).

This was puzzling and undermined our confidence in our computational chemistry tools. Various hypotheses were developed to explain the drop in potency for the benzylamine compound, while apparently binding exactly as predicted.

In other crystal structures, a water molecule was often observed hydrogen bonded to Asp376, but it could be displaced. Interestingly, it is displaced by the carboxylate group of the native ligand arginine, suggesting that either the Asp or the ligand acid group is protonated. So if Asp376 is protonated, this might account for the poor potency of the designed compound, as a strong salt bridge will not be formed.

When the hydrogen atoms were added to the water molecule by the ligand resulted in the loss of more hydrogen bonds than were gained, causing the loss in potency.

A final hypothesis was that there was a solvation energy penalty, for removing the benzyl amine compound from solution, that was not balanced by the interaction energies of the amine with the protein.

Any combination of these three considerations might have been responsible for the poor potency of this compound, while binding in the conformation we predicted.

A final chapter to this story occurred when it was being prepared for publication, as an example of the weaknesses in current computational chemistry tools. The manuscript was fully prepared and ready for submission when a second crystallographer analysed the electron densities, since the original crystallographer had left the company. It was the opinion of the second crystallographer that the data were not good enough to support the position of the amine substituent of the designed compound interacting with the Asp376 side chain, as originally suggested. The electron density for the ethylamine chain was weak, and owing to the resolution (3.3 Å) not detailed enough to be distinguished from a water molecule (Figure 6). Efforts were made to re-collect data for the complex, but without success.

Lessons for modellers
This case study reinforces many of the lessons highlighted earlier, chief among them the importance of good communication between medicinal chemistry, computational chemistry and structural chemistry when interpreting the results from

TABLE 1
NOS isoform inhibition potency

| Compound number | R1   | IC50 (µM) | iNOS | eNOS | nNOS |
|-----------------|------|-----------|------|------|------|
| Designed compound | 1EtNH2Ph | 24 | Inactive @ 100 µM | 39 |
| Un-substituted analogue | Ph | 0.52 | >100a | 15 |
| Potent analogue | 4CNPh | 0.071 | >100a | 6.6 |

*a<20% Inhibition at 100 µM.
crystallography, and the inspection of both the crystal model and the electron density maps before detailed compound design decisions are made. A number of software packages are freely available on the Internet for the visualisation of crystal structure models together with the electron density, including ‘O’ [75] and Astex Viewer [76].

Correct interpretation of the original data could have helped to draw the conclusion that the ligand/Asp376 interaction was very weak. This would certainly have saved a lot of futile deep thought, hypothesis generation and indeed medicinal chemistry design and synthesis.

Can X-ray crystal structures really aid drug design?
Structure-based design has now delivered drugs to the market for a number of important diseases, including cancer, HIV, glaucoma and hypertension [77]. Recently, the protein renin has finally succumbed to rational drug design, as aliskiren is the first renin inhibitor to reach the market after almost two decades of research by the global pharmaceutical industry to target this mechanism in the control of hypertension [78].

The pharmaceutical industry continues to invest in the collection of protein structural information to aid drug design. In ideal cases, the use of protein structure models has been used to design-in potency and selectivity, but drug design requires much more than this. Many other properties need to be built into the chemical structure that are not directly aided and may even be hindered, by the availability of protein structure models.

For instance, the pharmaceutical industry recognised a number of years ago the importance of physicochemical properties in gaining oral bioavailability [79]. But recent work suggests physicochemical property control has an even wider importance. Inflation of lipophilicity, molecular weight and hydrogen bonding interactions, and reduction in solubility, has been linked, not only to poor oral bioavailability, but also to unwanted metabolism, the number and severity of off-target activities [80], various toxicological endpoints and even to overall attrition through clinical development [81]. Drug design requires careful control of these properties, while maintaining high target affinity. We have previously suggested that the use of protein structural information has enticed medicinal chemists to inflate lipophilicity, hydrogen bonding interactions and molecular weight to fill protein pockets and maximise the interactions at the ligand protein interface [52]. Protein structural information may be used in a different way. It may allow the rational replacement of undesirable functional groups to be made, positioning of physical property-controlling groups into solvent, capitalising on protein movement to capture different protein pharmacophores and the identification of highly ligand efficient motifs [82].

Conclusions
The use of X-ray crystal structure models continues to prove valuable in drug discovery. These models provide a strong stimulus to chemical creativity, through the direct visualisation of the ligand–receptor interactions. Such interactions are otherwise rather abstract to the medicinal chemist, whose only other alternative is to blindly feel his way forward using the tools of structure–activity analysis. For users of X-ray crystal structure information, however, it is important to realise that a crystal structure is a model, a crystallographer’s partly subjective interpretation of experimental data. This interpretation may be flawed, ambiguous or inaccurate in its details (and in rare cases in its entirety). The best way to assess to reliability of a model is to discuss it with the crystallographer who produced it, to examine the model alongside the experimental electron density and to put the model to the test through iterations of structure–activity work.

Acknowledgements
The authors would like to thank the magnificent seven (sic!) referees for providing many useful comments and suggestions. AMD and SStG would like to thank Lisa-Lotte Olsson (AstraZeneca Structural Chemistry) for her patience, advice and understanding. GK would like to thank Alwyn Jones (Uppsala) for many discussions about quality control, model building errors and validation. GK was supported by Uppsala University and the Royal Swedish Academy of Sciences during part of this work.

References
1 Anon., . http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.pdf
2 Rowland, P. et al. (2006) Crystal structure of human cytochrome P450 2D6. J. Biol. Chem. 281, 7614–7622
3 Yano, J.K. et al. (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-A resolution. J. Biol. Chem. 279, 38091–38094
5 Williams, P.A. et al. (2004) Crystal structure of human cytochrome P450 3A4 bound to metyrapone and progesterone. Science 305, 683–686

8 Liu, Y. et al. (2006) Retraction. Science 314, 1856–1857

9 Nissink, J.W. et al. (2002) A new test set for validating predictions of protein–ligand interaction. Protein Struct. Funct. Genet. 49, 457–471

11 Bohmström, J. (2001) Reproducing the conformations of protein-bound ligands: a critical evaluation of several popular conformational searching tools. J. Comput. Aided Mol. Des. 10, 255–262

12 Meller, G. (2006) Scientific publishing. A scientist’s nightmare: software problem leads to five retractions. Science 314, 1856–1857

19 Liu, Y. et al. (2009) Exploring experimental sources of multiple protein structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions. J. Mol. Biol. 363, 174–187

23 Rupp, B. et al. (2001) Use of an induced fit receptor structure in virtual screening. J. Med. Chem. 44, 2570–2579

25 Mizutani, M.Y. et al. (2006) Effective handling of induced-fit motion in flexible docking. Proteins 63, 878–891

26 Rabol, O. et al. (2007) Structure-based virtual screening of FGF inhibitors: crossdecys and induced-fit effect. BioDrugs 21, 31–45

28 Kleywegt, G.J. (2000) Validation of protein crystal structures (‘Topical review’). Acta Crystallogr. D56, 249–265

33 Fytche, S.A. et al. (2006) Recombinant human PPAR-beta/delta ligand-binding domain is locked in an activated conformation by endogenous fatty acids. J. Mol. Biol. 356, 1005–1013

34 Fytche, S.A. et al. (2006) Reevaluation of the PPAR-beta/delta ligand binding domain model reveals why it exhibits the activated form. Mol. Cell 21, 1–2

35 Zoli-Juran, M. et al. (2003) High throughput screening identifies novel inhibitors of Escherichia coli dihydrofolate reductase that are competitive with dihydrofolate. Bioorg. Med. Chem. Lett. 13, 2493–2496

36 Summerfield, R.L. et al. (2006) A 2.13 A structure of E. coli dihydrofolate reductase bound to a novel competitive inhibitor reveals a new binding surface involving the M20 loop region. J. Med. Chem. 49, 6977–6986

37 Holmmer, Å. et al. (2004) Novel binding site identified in a hybrid between chlorella toxin and heat-labile enterotoxin: 1.9 A crystal structure reveals the details. Structure 12, 1655–1667

38 Holmmer, Å. et al. (2007) Novel binding site identified in a hybrid between chlorella toxin and heat-labile enterotoxin: 1.9 A crystal structure reveals the details [Erratum to document cited in CA1141:390244]. Structure 15, 253
69 Homans, S.W. (2007) Water, water everywhere—except where it matters? Drug Discov. Today 12, 534–539
70 Christof, G. et al. (2007) Thermodynamic inhibition profile of a cyclopentyl and a cyclohexyl derivative towards thrombin: the same but for different reasons. Angew. Chem. Int. Ed. 46, 8511–8514
71 Connolly, S. et al. (2004) 2-Aminopyridines as highly selective inducible nitric oxide synthase inhibitors. Differential binding modes dependent on nitrogen substitution. J. Med. Chem. 47, 3320–3323
72 Goodford, P.J. (1985) A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. J. Med. Chem. 28, 849–857
73 Jones, G. et al. (1997) Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 267, 727–748
74 Jones, G. et al. (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J. Mol. Biol. 245, 43–53
75 Anon., http://xray.bmc.uu.se/~alwyn/
76 Anon., http://www.astex-therapeutics.com/AstexViewer/index.php
77 Clark, D.E. (2007) What has computer-aided molecular design ever done for drug discovery? Exp. Opin. Drug Discov. 1, 103–110
78 Wood, J.M. et al. (2003) Structure-based design of aliskiren, a novel orally effective renin inhibitor. Biochem. Biophys. Res. Commun. 308, 698–705
79 Lipinski, C. et al. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 23, 3–25
80 Leeson, P. et al. (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. Nat. Rev. Drug Discov. 6, 881–890
81 Wenlock, M. et al. (2003) A comparison of physiochemical property profiles of development and marketed oral drugs. J. Med. Chem. 46, 1250–1256
82 Carr, R.A. et al. (2005) Fragment-based lead discovery: leads by design. Drug Discov. Today 10, 987–992