The conformational musings of a medicinal chemist

Harry Finch¹,²

¹ The Coach House, Grenville Court, Britwell Road, Burnham, Slough SL1 8DF, UK
² Unit 310 Ducie House, Ducie Street, Manchester M1 2JW, UK

Structure-based drug design strategies based on X-ray crystallographic data of ligands bound to biological targets or computationally derived pharmacophore models have been introduced over the past 25 years or so. These have now matured and are deeply embedded in the drug discovery process in most pharmaceutical and biotechnology companies where they continue to play a major part in the discovery of new medicines and drug candidates. Newly developed NMR methods can now provide a full description of the conformations in which ligands exist in free solution, crucially allowing those that are dominant to be identified. Integrating experimentally determined conformational information on active and inactive molecules in drug discovery programmes, alongside the existing techniques, should have a major impact on the success of drug discovery.

Introduction
At a recent talk at the Royal Society of Chemistry in Burlington House (London, UK) I listened to Dr John Dixon describing the story about the discovery of Brilinta®, the platelet aggregation inhibitor and P2Y₁₂ adenosine diphosphate (ADP) receptor antagonist, recently introduced into clinical practice by AstraZeneca [1]. The SAR identified within the Brilinta® programme are intriguing and suggest that conformational manipulation was important (but not intentional) in its discovery [2] (Fig. 1). The 100-fold increase in potency observed when moving from a purine to a triazolopyrimidine scaffold could indicate that a different, beneficial conformational preference of the bicyclic heterocycle relative to the cyclopentane ring is adopted, whereas the range of potencies observed with the cyclopropyl diastereoisomers suggests the presence of a conformational lock. The availability of this information at an early stage in the project could have resulted in a quicker optimisation phase by eliminating the need to make and test analogues of incorrect conformation and by targeting design towards alternative ways of achieving the active or required conformation. The work performed was on molecules that would not have been handled well by in silico conformation prediction methods and no structural information about the P2Y₁₂ G-protein-coupled receptor (GPCR) target was available. I believe that new techniques such as the NMR approach described in this article would remarkably improve the efficiency of drug discovery on programmes of this nature.

Looking back... and then forward
These observations led me to reflect on structure-based approaches to drug discovery and design from the past 20 or so years, the advances made and the contribution that experimentally determined (as opposed to theoretically predicted) conformational information has to make in the years to come. As a practising medicinal chemist for over 35 years, I believe experimentally determined solution conformational information has the potential to have as great an impact as that which X-ray structure-based drug design (SBDD) has had on programmes I have been involved in since 1990 [e.g. thrombin, Factor Xa, human neutrophil elastase (HNE), hsp90 and multiple protein kinases]. I am convinced it would be highly beneficial for high-quality experimentally determined conformational and dynamic 3D data to be integrated into all medicinal chemistry programmes at all stages, complementing other sources of information such as X-ray structures of ligands and protein targets.

A recent review by Jorgensen and colleagues [3] on the significance of conformational and binding affinity changes induced by
the introduction of a single methyl group really confirms the need for medicinal chemists to understand fully the impact of simple conformational changes and how to use them in a structure-based discovery setting.

Pharma’s productivity crisis and the discovery bottleneck

The issues facing the pharmaceutical sector, such as a decline in productivity, are well documented and illustrated by the data below:

- The probability of a product in Phase I reaching market has fallen from 10% in the period 2002–2004 to 5% during 2006–2008.
- The internal rate of return on R&D has declined from 10.5% in 2010 to 7.7% in 2011, falling to 7.2% in 2012 [4].

It is clear that the drug discovery and development pathway is ripe for improvement at many stages. One major bottleneck in the drug discovery process is the discovery of high-quality hit and lead molecules to deliver improved candidate drugs because these are a fundamental requirement if the pharmaceutical and biotechnology industry is going to be successful in providing medicines that will help alleviate a wide spectrum of ineffectively treated or untreated diseases. Obviously, despite the considerable progress made over the years, there is still a lot to be done [5].

Structure matters

Drug design using X-ray crystallography

It is now fully accepted that SBDD has an important role in drug discovery and the identification of developable molecules [6–11]. This has never been more obvious than in the now maturing area of fragment-based drug discovery [12–19] where SBDD is considered an essential requirement to deliver drug molecules from low-affinity fragments and has been illustrated by the recently launched Zelboraf® (vemurafenib, PLX4032), which originated from a fragment starting point [20].

Testimony to the importance of the availability of X-ray structural information to drug discovery programmes has been the acquisition by large pharmaceutical companies of small biotechnology groups with specialized knowledge in X-ray structure determination (Table 1) and a significant number of collaborative deals with others (such as the Ribotargets/Vernalis collaboration with Novartis on Hsp90, one in which I was involved). Most pharmaceutical companies now have internal access to X-ray crystallographic capabilities and most pharmaceutical and biotechnology companies involved in medicinal chemistry and drug discovery will now automatically incorporate structure-based methods as part of their discovery paradigm.

Industry interest in role of X-ray crystallography in drug discovery and design

In addition to the above, the Structural Genomics Consortium (SGC), a public–private partnership sponsored by numerous pharmaceutical companies, was formed with a mandate to promote the development of new medicines and to determine, on a large scale, and to make openly available, 3D structures of human proteins that represent potential drug targets [21]. The SGC is now responsible for >25% of all structures deposited into the Protein Data Bank (PDB) each year [22].

The X-factor

X-ray crystallographic data are currently the only universal experimentally defined ‘structure’ element for SBDD and are produced as three main categories:

- small molecule X-ray crystal structures;
- protein X-ray crystal structures;
- protein–ligand X-ray co-crystal structures.

Use of information from these approaches has been very successful and multiple medicines and candidate molecules have been generated and continue to be produced using these X-ray crystallography methods [23]. Despite these successes there are potential drawbacks. For example, use of only small molecule X-ray crystal structures to define molecule conformations can be problematic because crystal packing interactions can result in

### Table 1

<table>
<thead>
<tr>
<th>Year</th>
<th>Target</th>
<th>Acquirer</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>3D-Pharmaceuticals</td>
<td>J&amp;J</td>
<td>US$88 million</td>
</tr>
<tr>
<td>2005</td>
<td>Syrrx</td>
<td>Takeda</td>
<td>US$270 million</td>
</tr>
<tr>
<td>2008</td>
<td>SGX Pharmaceuticals</td>
<td>Lilly</td>
<td>US$64 million</td>
</tr>
<tr>
<td>2011</td>
<td>Astex</td>
<td>SuperGen</td>
<td>US$150 million</td>
</tr>
</tbody>
</table>

conformations that do not entirely represent the minimum energy conformation [24]. The relevance of small molecule X-ray information to drive drug design requires additional information such as protein–ligand co-crystal structure information or computational-chemistry-driven mapping of SAR data onto the solid state conformation to link the X-ray information to the biologically relevant pharmacophore.

Out of the above alternatives the current preferred and most often available option is the X-ray crystal structure of a small molecule (or series thereof) bound to its biological target as well as the apo structure of the biological target. This allows the differences between the apo target structure and that of the bound target structure to be seen and hence subsequent hypotheses on how to improve the binding interaction to be proposed and challenged in iterative rounds of informed changes to the drug template under study. Obviously, although X-ray crystal structures of ligand–target complexes have been crucial to the development of SBDD methods, the data only represent a snapshot of the real situation. For example, even though a drug molecule can be observed bound in the biological target and its key interactions identified (van der Waals, H-bonds, etc.), the bound conformation adopted by the ligand represents only one of the conformations available to that molecule in solution and the energy penalty required for the drug molecule to be restricted into the particular conformation required for binding is unknown [24].

Having referred above to X-ray structures of protein–ligand complexes, it is important to understand that really we are, of course, discussing models. Correct interpretation of the electron density in and around the protein binding site is far from a straightforward task and there is evidence that the quality of ligand models published (in the PDB) has been rather poor [25]. An additional problem is that many biological targets have not succumbed to X-ray crystallographic techniques and even those for which there are plenty of examples (e.g. protein kinases) can prove difficult, time consuming and expensive with regard to finding conditions under which a crystal form can be reliably produced. For example, some membrane-bound proteins such as ion channels have long been viewed as intractable. Advances in GPCR crystal structure determination are having an impact at various companies and academic institutions and workers at HepTares have solved structures of ligands bound in their agonist/antagonist modes and have published a conformational explanation for activation, at least from the perspective of the protein [26–28]. However, the exciting advances made in GPCR protein crystallography by these pioneers [29,30] also serve as a reminder that proteins often need to be engineered to deliver very specific constructs that are amenable to crystallisation. In overall summary, in drug discovery programmes where no biological target structural information is available, all stages – from hit identification through to the optimised candidate – are usually performed without any, or only with minimal, structural insight based on experimental data.

Virtual screening ... virtual reality
The above paradigm for X-ray crystallography SBDD relies on an already identified ‘fragment’, ‘hit’ or ‘lead’ molecule and structural data concerning how the ligand binds to its biological target to proceed. The identification of these starting points for SBDD programmes is varied but one strategy adopted universally is ligand-based virtual screening [31–39]. This computational process identifies several key features (a pharmacophore) in a set of similar or varied molecules and uses these key features to search huge databases of molecules, which are proprietary and commercially available, that will have similar features. This filtering process allows a much smaller subset of the available molecules to be screened in a biological assay, saving time and money.

Publications appear incredibly frequently describing the successes achieved with virtual screening as a hit and lead identification strategy and, in fact, for academic institutions or small companies that cannot access or afford an expensive and time-consuming HTS campaign this is often the only realistic approach they can adopt to find starting points for drug discovery programmes [40]. One of the problems with this technique is that unless the generated pharmacophore map is based on X-ray data of ligands bound in the appropriate or a related drug target then a very coarse 3D pharmacophore is produced. This is because the 3D arrangement of the key pharmacophoric moieties when predicted via computational methods can be unreliable. Additionally, during the virtual screening process the conformation of molecules to be ‘Docked’ into the protein structure is generated ‘on the fly’ using computationally predicted methods. It is accepted in the computational community that it is necessary to take into account protein flexibility when conducting virtual screening campaigns and, hence, it would seem logical to introduce the experimentally determined flexibility of a known ligand when investigating protein–ligand interactions [41–43]. The generalised application of sophisticated, dynamic conformational NMR data, which is now available, would allow this to be accomplished with a much greater degree of accuracy.

Conformation by design
Despite this wealth of structural information, medicinal chemists and computational chemists, whether involved in hit identification, lead identification or lead optimisation, have never had the opportunity to use fully experimentally determined structures of a ligand in solution, the most relevant phase to that which the ligand is subjected to when interacting with its biological target, whether in vitro or in vivo. Although molecular dynamics simulations boot-strapped with sparse experimental data have been used as a proxy for a fully experimental ligand solution structure, the predictive usefulness of such simulations is fundamentally dependent upon the force-fields used, which are well-known to be often sorely lacking. A major step forward would be the availability of reliable experimentally determined conformational information for the ligand in solution and for this to be used in an iterative sense during drug discovery programmes. Where available, this ligand solution conformational information could be combined with X-ray ligand–target data and, altogether, this would provide a profoundly powerful dataset for ligand design and/or optimisation studies.

In my experience, there have been occasions during drug discovery programmes when aspects of the conformation of molecules in solution have been studied in conjunction with NMR spectroscopy groups but the information provided has been rudimentary at best and I cannot remember it especially impacting a discovery programme [44–46]. This is not caused by a lack of
understanding in the industry about what NMR spectroscopy can provide – I actually selected an NMR spectroscopy book by Jackman and Sternhell [47] as the prize I received for results during my studies for a Higher National Certificate in Chemistry in 1970. Therefore, I would argue that currently a synthetic organic or medicinal chemist in the pharmaceutical or biotechnology business is incredibly familiar with NMR techniques when it comes to using them to determine and confirm the 2D structure of a synthesised molecule. It has struck me that gathering and, importantly, extracting from existing data additional conformational information on active and inactive compounds would be a major step forward. In fact, having the experimentally determined set of preferred conformations of drug molecules of interest delivered to the desktop of a practising medicinal chemist should be an aspiration and one that, in years to come, should be routinely fulfilled.

With this backdrop in mind, researchers at C4X Discovery set about trying to maximise the conformational information for flexible ligand molecules that can be collected using NMR spectroscopy and combined these data with new mathematical models for describing conformational behaviour in solution [48]. This new approach allows the preferred solution conformations of biologically relevant molecules to be accurately measured directly from experimental NMR data. This is achieved by analysing the molecule in solution, freeing rational 3D drug discovery from its previous dependence on protein structure data and computational simulations. Molecules can be analysed under physiologically relevant conditions with standard NMR magnets (300–800 MHz), and the resulting data are interpreted according to the published methods [48]. The methodology is independent of computational chemistry and eliminates the problems of virtual conformations and under-constrained ensembles that have previously plagued the field by using multiple kinds of NMR data, an order of magnitude more experimental measurements than typical and its novel representation of conformational behaviour. The output is an accurate description of the complete dynamic conformational behaviour of a ligand: essentially the ensemble of 3D conformations the molecule naturally oscillates through in solution. An analysis as complex and thorough as this is not possible with conventional tools for NMR analysis [49]. Thus the methodology combines data from a range of independent experimental types [nuclear Overhauser effect (NOE), residual dipolar coupling (RDC), scalar couplings, etc.] into one consistent picture while, moreover, accounting for all the data contained in those spectra. By contrast, conventional NMR tends to use only a fraction of the available cross-peak data from NOE spectroscopy (NOESY). This extensive set of experimental data provides the opportunity to determine the full range of conformations a molecule adopts.

Each experimentally derived ensemble describes not just the bond torsions and atomic distances within each 3D conformation but also quantifies the occupancy of each conformational mode, thereby identifying preferred conformations. The method has reportedly been successfully applied across a range of targets and molecular types, including enzyme co-factors, endogenous peptides, candidate drugs, oligonucleotides, complex natural products, macrocyclic compounds and carbohydrates [50]. Specific examples include lisinopril, angiotensin-(1–7) (the heptapeptide natural inhibitor of angiotensin-converting enzyme), hyaluronan, carazolol, ivermectin and the corticotropin-releasing factor (CRF)1R antagonist CP376395 [50]. The quantification of free ligand conformational preferences and their relationship to the bioactive conformation have recently been described for streptomycin (Fig. 2) and illustrates in detail the data generation and structural outputs feasible with this NMR technology [48].

For small molecular weight ‘drug-like’ molecules, the latest implementation of this NMR method is reportedly routinely able to solve structures in one week and, for the second or third compound in a hit or lead series, one or two days [50]. These latest improvements address a potential weakness of the method and it is therefore anticipated that, where used, this approach should be a relatively low cost enterprise compared with the

![FIGURE 2](image-url)

**FIGURE 2**

Relationship between free (a) and crystal (b–e) conformations of streptomycin. (a) In solution, streptomycin oscillates (light grey) about two main conformational families (black) in a ~60:40 ratio (left:right) [48]. These two families differ principally in the conformation of the lower glycosidic linkage between the ribose and glucosamine residues. (b) The known ribosome-bound conformations (blue) are the same as the predominant conformation in solution (grey), within the resolution of the X-ray data (3–3.5 Å) – PDB codes 1FJG, 4DR3, 4DRS-7 [54]. (c) By contrast, the free crystal conformation (green) – oxime derivative CSD code STOSEH10 [55] is the same as the second dominant conformation in solution (grey). (d,e) Interestingly, two micromolar affinity RNA-aptamer-bound conformations (yellow) – 1NTA and 1NTB [56,57] – that were discovered by artificial selection correspond most closely to one of the two families each. The obviously less close shape correspondence of these aptamers is likely to account for their lower binding affinity. Insets kindly provided by Dr Charles Blundell (C4X Discovery) from the original data [48].
infrastructure required to solve new protein and protein–ligand X-ray structures repeatedly. Other possible weaknesses could be those typically expected of an NMR-based approach [i.e. potentially high sample concentrations (>0.2 mM) or lack of resolution in repetitive molecules] but there are also tried and tested solutions to such problems.

Reflecting on the use of this new ligand conformational information, it is clear that for structure-based discovery programmes the strategy adopted when the conformational preferences are determined for the available ligands depends upon whether they are potent or weakly binding ligands. For potent ligands these data should provide a novel route to predicting the bioactive conformation without X-ray co-crystallography. Further template optimisation would look to reduce the conformational freedom and, when optimal geometry has been achieved, the SAR of local interactions between the ligand and its target can be investigated. The situation with a weakly binding ligand would be based upon designed structural changes that were able to favour differing subsets of the conformations determined for the parent molecule. In this way it would be possible to ‘home in’ on the probable bioactive conformation and then focus further optimisation efforts towards that shape. Iterative rounds of molecular changes followed by further NMR-derived 3D solution structures would provide the information on which to formulate and reappraise SAR hypotheses around potency and or selectivity (e.g. a related target, an unrelated target or a CYP450 enzyme), as is the normal practice with SBDD.

Using such an approach, determination of multiple ligand structures (selective and nonselective compounds) in a C4X Discovery internal orexin antagonist discovery programme has reportedly identified the geometry favouring OX1 antagonists [50]. These data have been used to design low nanomolar inhibitors with 1000-fold selectivity for OX1 over OX2, with no detrimental impact on ligand lipophilic efficiency (LLE) or ligand efficiency (LE) indices, in just a few months.

**Potential ways of using this enhanced conformational information**

Apart from the obvious uses of experimentally determined conformational information that are described above, there are other areas where it could be expected to have major implications. A few of my personal thoughts are detailed here:

- A database of experimentally derived solution conformations of drug-like molecules and fragments that can be searched and used to predict the conformations of new molecules. Such a database should provide a vastly more accurate basis for conformation generation, compared with those generated theoretically, and thereby impact hit identification strategies such as virtual screening and scaffold hopping. It could be imagined that the availability of such a database of experimentally derived conformations would, in combination with computational prediction methods to fill the gaps, provide a major enhancement in success rates because of the higher quality of the input data.
- The preferred solution conformations of potent peptide ligands, whereby the dominant conformations of the experimentally determined solution structure can be assumed to be, or be close to, the bioactive shape, could be measured and this would enable the identification of key pharmacophore points and facilitate ‘peptide to small molecule’ discovery.
- Subtle differences in preferred conformations could be experimentally derived and subsequently used to define the molecular requirements for agonism, partial agonism, inverse agonism, antagonism (including slow off-rate kinetics) and bias signalling in related ligands in a range of biological targets (e.g. GPCRs, nuclear receptors).
- It is known that some compounds in the MW range 700–1500 Da do have acceptable bioavailability and/or permeability properties—for example cyclosporine, 1203 Da, %F = up to 89% [51]—despite not conforming to Lipinski’s Rule-of-Five [52], but the reasons for this are currently unknown. It could be that such molecules adopt different conformations in aqueous and lipid environments and these changes facilitate their permeability or recognition by active transporters. It is interesting to speculate that experimentally derived conformational information measured in the different environments that molecules reside in *in vivo* (e.g. lipid vs aqueous) could potentially elucidate the ‘rules’ that could allow larger ‘drug-like’ molecules for large surface area protein–protein interactions to be discovered.
- The generation of and use of macrocyclic structures in drug discovery is undergoing a renaissance and a macrocyclic summit recently highlighted the need to create an understanding of the rules that govern the behaviour of macrocycles and thus enable developers to identify drug-like compounds more readily [53]. Knowledge of the experimentally determined conformational structure of compounds of this nature would be available via the new NMR method and could have a major effect on the knowledge base required to understand and more fully exploit this area of drug discovery.

**Concluding remarks**

It is my firm belief that the ability to measure accurate, experimental conformational information in the solution state on the whole range of ligands (e.g. fragments, hits, leads, peptides, carbohydrates, co-factors, macrocycles) that medicinal and computational chemists utilise in their current efforts to discover novel medicines will have a major effect on success rates and on the understanding of the interactions between a ligand and its biological target, particularly when used in combination with current structure-based methods. I anticipate that the use of this powerful NMR-based methodology will be widespread in a few years’ time.

**Conflicts of interest**

The author is a non-executive director of C4X Discovery Ltd.

**Acknowledgements**

The author would like to acknowledge helpful discussions with Emma Palmer-Foster, Charles Blundell, Thorsten Nowak and Barrie Martin that were held during the production of this manuscript.