

# The use of biophysical methods increases success in obtaining liganded crystal structures

Chun-wa Chung

Structural and Biophysical Sciences,  
GlaxoSmithKline Research and Development,  
Medicines Research Centre, Gunnelswood  
Road, Stevenage SG1 2NY, England

Correspondence e-mail: cc16943@gsk.com

Received 2 October 2006  
Accepted 30 November 2006

In attempts to determine the crystal structure of small molecule–protein complexes, a common frustration is the absence of ligand binding once the protein structure has been solved. While the first structure, even with no ligand bound (apo), can be a cause for celebration, the solution of dozens of apo structures can give an unwanted sense of *déjà vu*. Much time and material is wasted on unsuccessful experiments, which can have a serious impact on productivity and morale. There are many reasons for the lack of observed binding in crystals and this paper highlights some of these. Biophysical methods may be used to confirm and optimize solution conditions to increase the success rate of crystallizing protein–ligand complexes. As there are an overwhelming number of biophysical methods available, some of the factors that need to be considered when choosing the most appropriate technique for a given system are discussed. Finally, a few illustrative examples where biophysical methods have proven helpful in real systems are given.

## 1. Introduction

### 1.1. Why do we need to know about complexes?

Crystal structures of protein–small molecule complexes often provide deeper insights into the structure and mechanism of a protein than are revealed by an apo structure alone. In some instances, nascent ligand pockets reveal themselves in the protein only when prompted to open up when an appropriate ligand is bound (Scapin *et al.*, 2003). These can show unexpected opportunities in recognition or selectivity that cannot be readily predicted by bioinformatics or computational analysis. Intricacies of enzyme mechanism only surmised and guessed at by detailed kinetic experiments can be greatly facilitated by the ability to trap complexes along the reaction pathway (Bitto *et al.*, 2006). Whilst the site of ligand binding is often known, sometimes only a crystal structure of the liganded protein can truly confirm the nature of the interactions involved and the extent of the binding site. On rare occasions, the unexpected can occur and ligands are found in completely unpredicted allosteric sites, which would have been difficult to unravel except by the ability to crystallize the complex (Horn & Shoichet, 2004). Of course, for structure-based drug design, iterative complex crystal structure determination is vital to guide the progress of synthetic chemistry efforts (Williams *et al.*, 2005). For the crystallographic data to have impact, complex structures must keep pace with the chemistry within a project. Progression of a crystallographic system from one that is able to produce single ‘one-off’ structures to one that is robust enough to cope with the demands of generating complex structures ‘on demand’ is still one of the most challenging aspects of structure-based

drug discovery. Even for established systems, the failure rate in obtaining X-ray complexes can be significant and highly compound-dependent. The scale of the problem and the wasted resources cannot be underestimated. For example, it has been stated by one major pharmaceutical company that 65% of their desired complex data sets turned out to be unliganded when the structures were finally solved. No doubt some systems demonstrated more success than others, but in our experience all crystallographic systems will generate a proportion of failed complexes within their lifetime.

### 1.2. Why can there be problems in generating structures of complexes?

When crystallography is used as a screening tool to identify ligands that bind to a protein (Hartshorn *et al.*, 2005), it is expected that many apo structures will be observed, as there will be a significant proportion of inactive molecules tested. However, when attempts are made to obtain a small molecule–protein structure, there is usually prior knowledge and expectation that the chosen ligand will bind. Sometimes, this expectation is unwarranted and the compound does not and has never bound to the protein in any form. Whilst this may seem a trivial reason for failure, it is unfortunately more common than one might hope and is worth bearing in mind before exploring other avenues. For ligands arising from a screening campaign, it is important to remember that all techniques have limitations and artefacts, and it may be wise to reconfirm the initial observation, maybe using an orthogonal method to avoid continued frustration (Chung *et al.*, submitted work). For example, a primary screen that monitors changes in fluorescence polarization on displacement of a fluorescent compound from the active site of a protein may be followed by an orthogonal activity assay that measures the inhibition of substrate turnover by a colorimetric readout. Ideally, the fluorescence and colorimetric emission and excitation wavelengths should not overlap. Activity in only one of the assays may highlight molecules with spectral properties that interfere with the assay readout or molecules that interact directly with the active-site probe (*e.g.* fluorescent compound or substrate) rather than the protein itself. Alternatively, the secondary assay may be one that monitors ligand binding directly without the need for a probe molecule, such as surface plasmon resonance (SPR), nuclear magnetic resonance (NMR) or isothermal titration calorimetry (ITC).

The most probable common sources of failure to generate a small molecule–protein complex relate to differences between the assay conditions originally used to identify the interaction and those necessary to generate high concentrations of homogeneous complex for crystallization trials. Issues can arise from the protein itself and the conditions used for complex formation. Often, the crystallographic construct is not identical to the assay protein. Sometimes, it is truncated to what is believed to be the critical and most compact form, with sites of potential heterogeneity, such as phosphorylation or other post-translational modifications, mutated away. Unfortunately, these changes may significantly alter the binding

affinity of the ligand, possibly to the extent that the ligand has little residual affinity for the crystallographic construct, making it difficult to generate a complex during the cocrystallization or soaking procedures. Success may also be influenced by the practical details of how complexation is attempted. For example, the way in which the ligand is added to the protein, the length of the incubation period, buffer conditions and the presence of other components, such as detergents and cofactors, all influence the outcome.

Once liganded, a protein complex may have markedly different solubility and self-association properties from the apo protein. The original apo crystallization conditions may no longer favour the formation of a crystalline lattice for the complex and re-screening for new crystallization conditions may be required. For weaker complexes, there can be a substantial proportion of free protein within the crystallization drop and it is possible that the unbound protein crystallizes more readily than the complex, resulting in only unliganded data sets being collected. This is less likely to occur for high-affinity ligands where little unliganded protein is present and the favourable lattice energy for apo crystal formation must be sufficiently high to compensate for complex dissociation.

Finding factors that drive the solution equilibrium towards complex formation and are compatible with crystallization will clearly enhance the pursuit of X-ray complexes. Unfortunately, while many factors can be important, those that are critical will be system-dependent. It is vital to identify the key attributes of the protein and of the conditions that are most likely to lead to successful complex formation on a case-by-case basis. This paper therefore details how biophysical methods may be used to address two fundamental questions: (i) does a compound bind to the crystallographic protein? and (ii) if so, under what conditions?

### 1.3. Which method to use?

There are a myriad of ways of monitoring ligand binding and no single ‘right’ method. For any system, it will be possible to apply a range of methods. Therefore, what are the considerations one should think about when trying to make a choice?

(i) Consider the system, the attributes of the system, the handling issues and limitations. For example, the availability of ligand and protein reagents, the robustness of the protein and whether there are any intrinsic probes that may be useful. For cytochrome P450s and other haem-containing proteins, UV methods are well established and may offer a site-specific and easy way to check complex formation.

(ii) Consider the information that is critical to progress the experiments. For example, whilst quantitative affinity measurements may be a bonus, it may be sufficient to merely differentiate between ligands that are able to form complexes and those that are not.

(iii) Consider the technique or maybe combination of techniques that allow access to this critical information, preferably with least effort, least ambiguity and with the type of reagents that are available.

Not all types of information can be accessed by every method, so it is very important to match the limitations of the protein system to the strengths and failings of the techniques. Failure to do this can compromise the quality of the information or lead to significantly more effort than necessary to obtain a definitive result. For example, if it is critical to understand the affinity of a number of ligands as a function of pH (*e.g.* if crystals only form at low pH) and to know their binding stoichiometry of the interaction (*e.g.* to help interpret the crystallographic electron density), then it may be sensible to consider methods that allow the generation of all these data in a straightforward fashion. SPR and ITC are two methods that can determine affinity and stoichiometry simultaneously, provided care is taken in determining the active protein concentration. In some instances, these direct methods may have advantages over activity assays, which may have a restricted operational pH range, and displacement assays, where a probe and its affinity profile over the pH range is required.

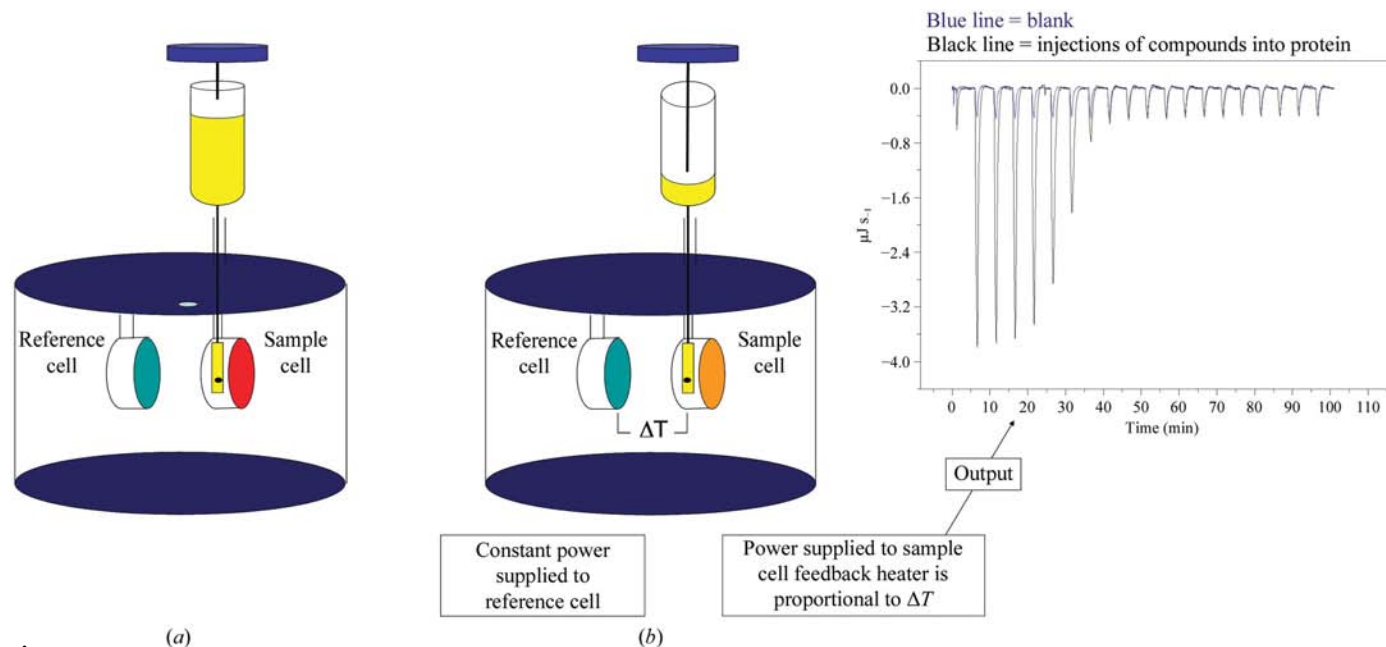
Many biophysical methods can measure ligand binding at low concentrations, with ultimate sensitivity being achieved by single-molecule techniques. None of these are discussed in this article, as it is impossible to cover the vast array of methods able to monitor protein–ligand interactions in a single review. Instead, the focus is to introduce four readily available techniques and to highlight how they may be successfully applied to guide crystallization attempts of protein–ligand complexes. Three of these, isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR) and dynamic light scattering (DLS), can study proteins at or close to crystallographic concentrations in a nondestructive manner. They have the

advantage of being able to look at the very same sample that goes on to crystallization trials. The fourth method measures shifts in thermal stability on ligand binding. This is normally run at lower protein concentrations and is a destructive technique. However, it is included because it has several unique attributes.

An understanding of the requirements for complex formation may also help in soaking experiments, although additional complications such as ligand solubility, the kinetics of binding within a protein crystal, the fragility of crystals on soaking *etc.* are also factors which are not covered in the context of this paper.

## 2. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is probably the gold standard of direct binding methods and involves measuring the heat produced (exothermic) or taken in (endothermic) during a reaction, such as the binding of a ligand to a protein (O'Brien *et al.*, 2001). The experiment takes place in an isothermal titration calorimeter at constant temperature. Typically, the ligand solution is repetitively injected in small aliquots from an automatic syringe into a thermally isolated stirred cell containing the protein (Fig. 1). The first injection of ligand produces the largest amount of protein–ligand complex and therefore the largest heat change. On subsequent injections there are fewer unoccupied sites on the protein, so less additional complex is formed and the heat change decreases. Ultimately, all the protein sites are saturated and at the end of the titration no further heat change arising from complexation is observed.



**Figure 1**  
 (a) Schematic diagram of an isothermal titration calorimetry instrument, consisting of a sample cell that contains one binding component, an automated syringe that contains the other binding partner and a reference cell. (b) Differences between the sample-cell and reference-cell temperatures induced by binding are initially translated to the power needed to bring the two samples back to the same temperature, before conversion to a binding enthalpy in molar terms.

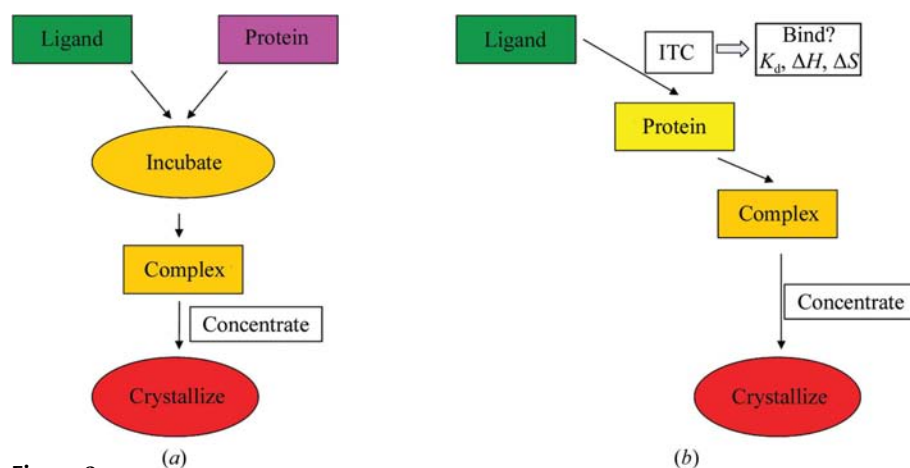
In theory, it is possible to obtain from a single experiment very precise estimates of the association and dissociation constants ( $K_a$  and  $K_d$ ), the stoichiometry of the interaction ( $n$ ) and the reaction enthalpy ( $\Delta H$ ). Thus, the entropy change for the binding can also be calculated *via* the following thermodynamic relationships

$$\Delta G = -RT \ln K_a, \quad (1a)$$

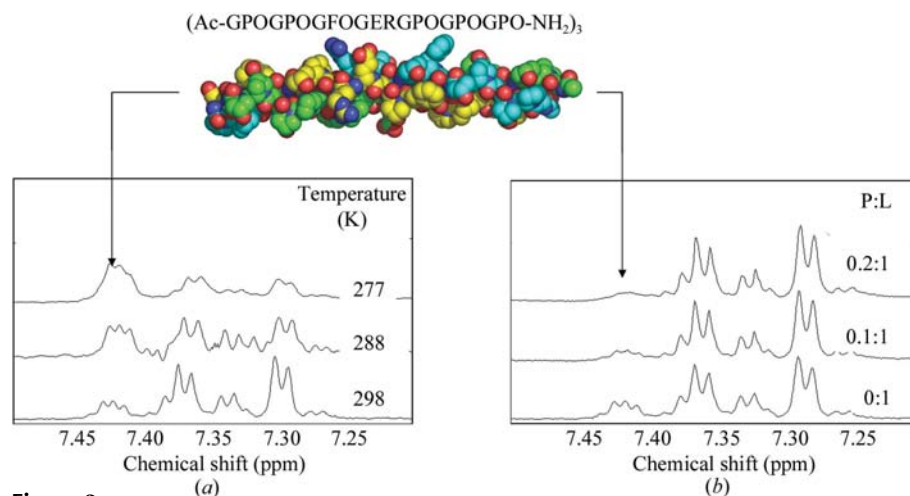
$$\Delta G = \Delta H - T\Delta S, \quad (1b)$$

$$K_d = 1/K_a. \quad (1c)$$

In practice, to accurately and simultaneously determine all these parameters ( $\Delta G$ ,  $\Delta H$ ,  $\Delta S$ ,  $n$  and  $K_a$ ) requires the initial molar concentration of protein in the cell to be defined by



**Figure 2**  
(a) Flowchart of a typical protocol to produce a solution of a complex, where protein and ligand are simply combined. (b) The simple combination of protein and ligand may be replaced by an ITC, where additional data on the binding stoichiometry, enthalpy and affinity can be gathered *en route* to crystallization trials.



**Figure 3**  
(a) The aromatic portion of the spectrum of the collagen-like peptide Ac-(GPO)<sub>2</sub>GFOGER(GPO)<sub>3</sub>-NH<sub>2</sub> is shown at 277, 288 and 298 K. The far left peak highlighted in this spectrum corresponds to the triple-helical form of the peptide and increases in abundance at lower temperatures. (b) Spectra of the same aromatic region at 298 K at varying protein:ligand (P:L) ratios is shown. The preferential broadening of the triple-helical form on protein addition is highlighted by the arrow.

$$10 < (K_a \times [\text{protein}]) < 100. \quad (2)$$

Therefore, for a binding event with  $K_a = 10^5 M$  ( $K_d = 10 \mu M$ ), the optimal cell concentration should lie between 100 and 1000  $\mu M$ . For a protein of  $M_r$  50 000 with a Microcal ITC instrument, which requires a minimum of 1.8 ml of solution to fill the cell, this equates to between 9 and 90 mg of protein for each experiment. For ligands with higher affinities, the protein concentration can be dramatically reduced, but the minimal concentration required then becomes dependent on the magnitude of  $\Delta H$ . If the concentration becomes too low, then the signal will become dominated by background contributions, such as those arising from ligand dilution and buffer mismatch. A typical ITC experiment is likely to use greater than  $\sim 0.5$ –1 mg of protein regardless of the affinity. Running ‘an ideal’ ITC experiment requires not only good experimental design and practice, but also some prior knowledge of  $\Delta H$  and  $K_d$ , which means an initial ITC experiment is often necessary to estimate these values.

It is possible to envisage how the simplicity and high information content of the ITC experiment, combined with its nondestructive nature, could be incorporated into a cocrystallization protocol where confirmation of complexation is required but protein supply is limited. This is illustrated in Fig. 2. Fig. 2(a) shows a ‘typical’ cocrystallization protocol where ligand is added to the protein and the mixture pre-incubated on ice for anything from minutes to hours. Sometimes this is performed at modest protein concentrations so that a concentration step is required and sometimes at higher concentrations. Regardless of the precise details of the process, there are often no checks to confirm the success or completeness of complex formation for small ligands. If this final ‘complex’ solution fails to generate crystals, it is unclear whether this is a complexation issue or a crystallization problem. Alternatively, if only apo crystals are formed, then it is unclear whether any complex was ever present. Fig. 2(b) shows how the ITC experiment can be used as a more controlled way of combining the protein and ligand, with the additional advantage of generating thermodynamic and binding information ‘for free’. After a successful ITC experiment, where ligand has been injected into a cell containing protein, the final cell contents contain a

confirmed saturated protein complex ready to be concentrated for cocrystallization trials.

This has been used for a number of nonphosphorylated kinase domains and for domains from a transcriptional regulator with no measureable enzyme activity. A typical starting point may use 10–20  $\mu\text{M}$  protein within the cell and 100–200  $\mu\text{M}$  ligand injections at 277 K. In the case of poorly soluble ligands, the titration can be reversed, with 10–20  $\mu\text{M}$  ligand placed into the cell and the protein contained within the syringe. In this configuration additional ligand must be added to the cell contents after the ITC experiment to ensure that the protein is saturated prior to concentration for crystallization trials.

There are situations where this simple 'ITC' modification to the complexation protocol is not appropriate or informative. For example, if the ligand is very sparingly soluble in aqueous solution then an aqueous titration is not possible even if the titration format is reversed. It is also not applicable where the binding event is entirely entropically driven. However, this method of recycling the final ITC contents for crystallization has provided an efficient use of protein in several instances, especially when the crystallographic constructs lack an alternative assay. The unique thermodynamic deconvolution of the binding energy into its enthalpy and entropy contributions by ITC should also be appreciated.

### 3. Nuclear magnetic resonance

NMR measures the resonance characteristics of magnetically active nuclei with nuclear spin  $>1/2$ , such as naturally abundant  $^1\text{H}$  and  $^{19}\text{F}$  or less abundant isotopes such as  $^{15}\text{N}$  and  $^{13}\text{C}$  if the sample has been suitably enriched. It is a versatile and information-rich technique which can be tailored to emphasize the most informative NMR attribute for any particular purpose. Consequently, there are a huge range of experiments available. For example, combining experiments that highlight molecular bond connectivities with those that give through-space interactions can allow *de novo* three-dimensional structures to be determined (Wishart, 2005; Nietlispach *et al.*, 2004). Here, discussions are restricted to experiments designed for monitoring ligand binding (Guenther *et al.*, 2004; Peng *et al.*, 2004).

It is useful to divide these NMR methods into two distinct categories, ligand-based methods and protein-based methods, as these have different properties.

#### 3.1. Ligand-based NMR methods

Ligand-based methods, such as STD (Krishnan, 2005) and waterLOGSY (Dalvit *et al.*, 2001), scrutinize only the ligand signals as a function of protein binding. Firstly, the position, or chemical shift, of the resonances can change. The direction and magnitude of the change depend on differences in the ligand's chemical environment between the bound and free states and have been used to differentiate between several postulated binding modes. Secondly, the shape of the signals may change on binding, indicating that the NMR relaxation

properties of the ligand have been altered. A common observation is peak broadening, as illustrated in the next example.

One difficulty in using ligand resonances is spotting them amongst potentially thousands of protein resonances. One solution is to use NMR-active nuclei present in the ligand but absent in the protein, such as  $^{31}\text{P}$  or  $^{19}\text{F}$ . Another solution is to have a large excess of ligand over protein and then monitor changes in ligand parameters as a function of sub-stoichiometric protein additions. For the excess ligand signals to contain information about the bound state, the ligand-exchange rate needs to be rapid compared with the chemical shift difference between the bound and free signals. Typically, this means modest (micromolar) to low-affinity (millimolar) interactions are best suited to this method.

Many biologically interesting protein–ligand interactions do not have a simple 1:1 stoichiometry and finding appropriate conditions to generate a complicated multicomponent complex can be immensely challenging. When one component is in an oligomeric equilibrium with oligomerization constants comparable to crystallographic concentrations, the situation becomes even more complex.

The synthetic collagen-like peptide Ac-(GPO)<sub>2</sub>GFO-GER(GPO)3-NH<sub>2</sub> (Emsley *et al.*, 2004) is a useful tool for exploring protein–collagen interactions as the peptide is able to self-associate into the triple-helical structure typical of collagen. This peptide has a single aromatic phenylalanine residue, so the large number of signals in the aromatic region of the NMR spectrum of this peptide indicates the presence of several species in solution (Fig. 3*a*). The concentration (not shown) and temperature-dependence (Fig. 3*a*) of these NMR signals suggest that these species correspond to different oligomeric states of the peptide in dynamic and reversible equilibrium, the triple-helical form being more stable and abundant at lower temperatures.

When a sub-stoichiometric amount (0.1:1) of a collagen-binding domain from an integrin is added to the peptide at 298 K, selective broadening of the triple-helical resonances are observed, indicating that the protein preferentially interacts with this oligomer (Fig. 3*b*). No changes in the other resonances are seen, an observation that is consistent with a relatively weak interaction that competes with the oligomeric equilibrium but is unable to significantly perturb it under these conditions.

Generating a homogeneous 1:1 protein:triple-helical collagen peptide complex for crystallization may therefore be difficult. However, for any given protein:collagen concentration, lowering the temperature is likely to be beneficial, as the NMR results show that this increases the proportion of the triple-helical form of the peptide in solution. This suggestion was borne out in crystallization trials, where extensive screening produced complex crystals only at temperatures  $<283\text{ K}$  and these crystals were found to be highly temperature-sensitive.

This example illustrates how the high molecular interpretability of NMR spectra (*e.g.* peaks corresponding to individual atoms in molecules) and richness of information can give

insights into complex multicomponent systems. As a ubiquitous technique, able to work with as little as 5  $\mu\text{l}$  of sample in a nondestructive manner, it is one of the most versatile tools available for biophysical analysis in a crystallographic context.

One major disadvantage of ligand-based experiments is that no information can be inferred about the site of interaction and specific and nonspecific interactions cannot be differentiated. To ensure that a ligand binds in the desired site, an additional experiment must be performed where the desired site is made unavailable for binding. This may be achieved by using a competitor for the active site or by using a mutant protein no longer competent to bind ligand. Site-specific information can also be gained using spin-labelled ligands (Jahnke *et al.*, 2000; Jahnke, 2002). Another way is to use protein-based NMR experiments, where a binding site may be determined without prior knowledge or tool compounds in favourable circumstances.

### 3.2. Protein-based NMR methods

Protein-based NMR experiments concentrate on changes in the protein spectrum upon ligand binding. As proteins have a much larger number of atoms than their ligand counterparts, many protein-based experiments focus on signals from a selected subset of these atoms. To make the information even clearer, these signals are often dispersed into two or more dimensions to give multidimensional plots, as shown schematically in Fig. 4.

Frequently, proteins enriched in rarer NMR nuclei are used to increase the sensitivity and range of chemical subsets accessible.  $^{15}\text{N}$  enrichment is the most popular and cost-effective option and subsequent discussions will be limited to use of the ubiquitous two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  correlation spectrum shown schematically in Fig. 4(a). In this spectrum each peak corresponds to a directly connected proton-nitrogen pair, such as the NH of an amide, which means that every residue in the protein backbone, except for proline, gives rise to a peak in this spectrum. Provided there are no global changes in structure upon ligand binding, the addition of a ligand will cause only a subset of peaks proximal to the binding site to be perturbed in a  $^1\text{H}$ - $^{15}\text{N}$  spectrum, as illustrated in Fig. 4(b). This enables binding to be monitored and the site of binding to be identified.

The power of this experiment lies with the easy access to residue-resolved information and the fact that at crystallographic concentrations this spectrum can be acquired in a matter of minutes with the sample returned intact. The HSQC spectrum also contains peaks arising from side-chain N-H pairs from residues such as asparagine, glutamine *etc.*, so sometimes it is possible to gain information about the interactions made by these side chains.

Unlike ligand-based experiments, the  $^1\text{H}$ - $^{15}\text{N}$  correlation experiment can monitor interactions with a wide range of affinities (nanomolar to millimolar) without the need to alter the experimental conditions. For fragment-based drug-discovery approaches that attempt to start from weak but chemically attractive molecules, confirmation of binding and

the likelihood of specificity (*i.e.* site of binding) is easily achieved using this experiment and may be a critical filter for compound progression. In some systems, NMR may be ideal for use as the initial screen to identify these chemical starting points (Shuker *et al.*, 1996; Jhoti, 2005).

SH2 domains are commonly found in intracellular proteins. Their role is to sense the phosphorylation of specific tyrosine residues within their partner proteins. They often lie at control points in intracellular pathways and the ability to differentiate between tyrosine and phosphotyrosine (pY) is critical to their role in regulation.

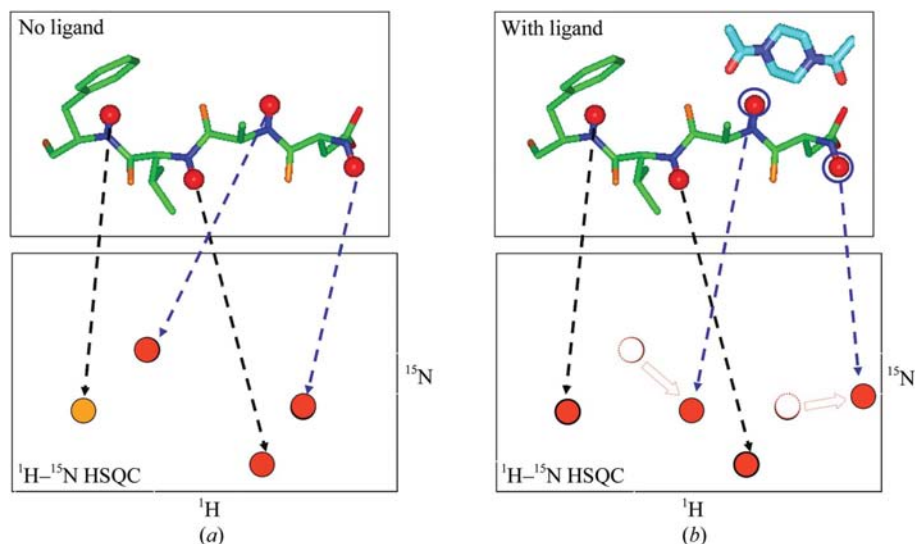
Inhibitors of this regulation event that act by binding to the SH2 domain require a suitable pY mimetic, which is probably anionic. Ideally, these mimetics need to make good hydrogen-bonding interactions in the pY pocket and yet have a low enough charge to allow the inhibitor to enter cells and reach the required site of action. One way such mimetics have been identified is to directly screen small acid mimetics, with good potential permeability properties, for their ability to bind at the pY pocket. These mimetics can then be elaborated into larger compounds to give the required potency and selectivity (Lesuisse *et al.*, 2002).

Despite the critical nature of the phosphotyrosine residue, an isolated pY moiety only has an affinity of the order of millimolar for the SH2 domain. Successful mimetics have similar potencies, making it challenging to identify and confirm the site of action of these fragments. A multitude of biophysical methods have been successfully used to screen pY mimetics. These include competition assays using pY peptides and fluorescence (Cousins-Wasti *et al.*, 1996), radioactive detection methods, SPR (Mandine *et al.*, 2001) and noncovalent MS (Bligh *et al.*, 2003). All these techniques have been able to detect the binding of low-affinity compounds to SH2 domains. However, they do not allow the site of action to be precisely localized to the pY-binding pocket, as pY itself cannot be used as a probe owing to its low intrinsic affinity.

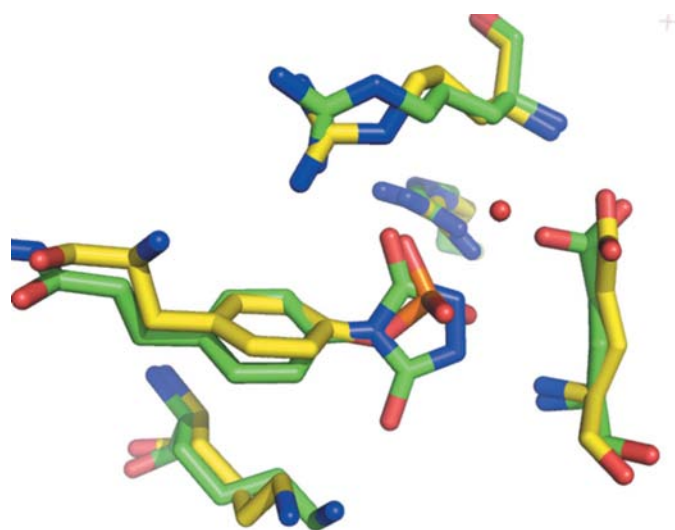
Crystallography of the pY mimetics would confirm binding at this site, but the problems associated with obtaining liganded crystal structures are amplified when very weak ligands are used and success rates can be very low. The HSQC experiment can help both to focus efforts prior to crystallization trials and to demonstrate specificity even in the absence of successful liganded structures. This demonstration can provide the confidence needed to begin chemical efforts to make inhibitors with greater potency that may be more successful in subsequent crystallization attempts.

This was indeed the case for our discovery of the urazole moiety as a novel phosphotyrosine mimetic. Fragment-based screening using a variety of biophysical methods [*e.g.* noncovalent mass spectroscopy (Bligh *et al.*, 2003), fluorescence polarization, scintillation proximity assay and NMR] on the Src SH2 domain identified a number of urazole-containing fragments which competed with phosphotyrosine peptide binding with affinities in the 1–5 mM range. These fragments satisfied many of the criteria required for an ideal pY mimetic, so there was great interest in understanding their binding mode. Their potency was so low that all crystallographic





**Figure 4**  
 (a) The bottom panel shows a schematic representation of a two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  correlation spectrum. Peaks in this spectrum can be assigned to amide NH pairs in the protein backbone. (b) On addition of a ligand, NH pairs close to the binding molecule will be selectively perturbed, allowing the binding site to be localized.



**Figure 5**  
 Overlay of the crystal structures of the Src SH2 domain with acetylated pYEEI peptide (PDB code 1a1b; in yellow) and the urazole derivative of the YEEI peptide (in green; Chung *et al.*, in preparation). The extensive hydrogen-bonding interactions made by the urazole are shown by dotted green lines. The excellent phosphate mimicry of the urazole heterocycle within this recognition pocket is evident.

attempts with the isolated fragments proved unsuccessful. However,  $^1\text{H}$ - $^{15}\text{N}$  NMR confirmed that all the urazole moieties bound within the pY pocket. This provided the evidence needed to make a modified recognition peptide with the phosphate group replaced by the urazole heterocycle. This had greater potency ( $30\ \mu\text{M}$ ) and the cocrystal structure of Src SH2 with this urazole peptide was successful. An overlay of the original pYEEI peptide complex with this variant (Fig. 5) clearly demonstrates the phosphate mimicry of this fragment within the pY pocket (Charifson *et al.*, 1997; Chung, in preparation).

#### 4. Dynamic light scattering

Most sizing methods do not have the resolution able to differentiate between a protein and ligand-bound protein and so are not generically useful for monitoring small-molecule binding. Noncovalent mass spectroscopy (Bolbach, 2005; Benesch & Robinson, 2006) is an exception, but tends to be a rather specialist activity needing considerable effort to ensure a representative result. For systems where ligand binding induces protein oligomerization or gross conformational changes, however, monitoring these significant size/shape changes can be a convenient surrogate for monitoring the ligand binding directly.

For cocrystallization trials, DLS (Brown, 1993; Schmitz, 1990) is a particularly attractive option, as it enables data to be gathered on the same

solutions as used for crystallization without dilution. This removes any concerns regarding batch-to-batch variations and the need to extrapolate between concentrations. The latter consideration is especially important in multi-component oligomeric systems, as the governing equilibria and kinetic parameters are often unknown, causing any concentration extrapolation to be unreliable.

Analytical ultracentrifugation (AUC) may also be used for such studies and is preferable if very detailed characterization of the oligomerization phenomenon is required (Gilbert, 2005). However, DLS is a technique available in many crystallography laboratories and often provides rapid and convenient access to sufficient information to guide crystallization experiments.

Dynamic light scattering monitors changes in the intensity of scattered light from a sample as a function of time. These fluctuations are caused by the Brownian motion of the molecules within a solution and can be correlated to the particles' diffusion coefficient and size *via* the Stokes–Einstein equation. It has been suggested that samples that are monodispersed by DLS (that is, uniform and consisting of only one particle size) are more likely to crystallize (Ferre-D'Amare & Burley, 1997; Winzor, 2003). As protein samples go on to experience a diverse range of conditions that may change their behaviour during crystallization screening, this criteria is rarely used to abandon crystallization trials (Stura *et al.*, 2002), but may be useful as a way of monitoring improvements in protein behaviour (Jancarik *et al.*, 2004).

A common obstacle to forming protein–small molecule complexes is the limited solubility of the ligands. At equilibrium, for the simple 1:1 interaction between a protein P and ligand L described by



the degree of complexation is dictated by the equilibrium association constant  $K_a$ ,

$$K_a = [\text{PL}]/[\text{P}][\text{L}_{\text{aq}}], \quad (4)$$

where  $[\text{P}]$  is the free protein concentration,  $[\text{L}_{\text{aq}}]$  the free ligand concentration in solution and  $[\text{PL}]$  is the complex concentration. For sparingly soluble ligands, this is complicated by the fact that the free ligand concentration  $[\text{L}_{\text{aq}}]$  cannot exceed its solubility limit. It takes a limiting value of  $[\text{L}_{\text{sat}}]$  regardless of the amount of excess ligand that is often added, as this excess is precipitated as solid. In this case, the concentration of complex achievable is determined by the maximum solubility of the ligand  $[\text{L}_{\text{sat}}]$  and  $K_a$ . For a protein–ligand solution where excess ligand (*e.g.* as solid) is present, as described by



the free ligand concentration is equal to its saturation value, *i.e.*  $[\text{L}_{\text{aq}}] = [\text{L}_{\text{sat}}]$ , and (4) becomes

$$K_a^*[\text{L}_{\text{sat}}] = [\text{PL}]/[\text{P}]. \quad (6)$$

This means that the ratio of complex to free protein is equal to the product of the affinity constant multiplied by the saturated ligand concentration and is no longer dependent on the protein concentration. For a compound with a solubility nine times its dissociation constant ( $[\text{L}_{\text{sat}}] = 9K_d = 9/K_a$ ), a 90% complex solution will be generated regardless of the protein concentration. In contrast, for a compound with a solubility comparable to its dissociation constant ( $[\text{L}_{\text{sat}}] = 1/K_d = K_a$ ), only a 50% complex solution will be possible.

Whilst the practicalities of how the ligand is added to the protein can have no bearing on the final equilibrium complex concentration, they can have a dramatic effect on the rate that the equilibrium is attained and it is useful to have tools to check this progress, as the next example illustrates.

DLS experiments were carried out on a protein known to dimerize upon ligand binding, where a large number of cocrystallizations had been attempted resulting in apo structures. Ligands for this system were identified in an assay using the intact transmembrane oligomeric receptor. However, for crystallographic studies only the excised extracellular ligand-binding domain was used. There was thus no convenient assay to confirm ligand binding and no biological interest in the crystallographic domain beyond its use to provide the molecular details of the interactions.

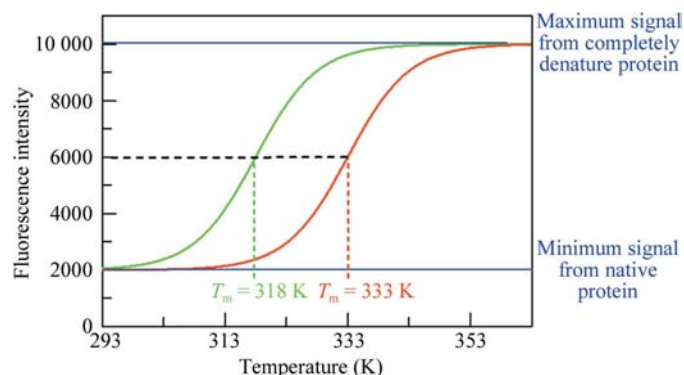
For this system, DLS provided rapid answers to a number of questions. Firstly, DLS was used at low protein concentrations in the 0.3–1 mg ml<sup>-1</sup> range to confirm that active ligands were able to enhance protein dimerization and by implication bind to the truncated protein. Moreover, the level of enhancement paralleled the potency ranking of the compounds in the intact receptor assay, providing confidence that the crystallographic protein would provide biologically relevant information.

Secondly, DLS was used to study the level of protein dimerization of samples destined for crystallization trials in order to try and improve the success rate of complex structures. A number of buffer parameters such as salt concentra-

tion and pH were explored, but the most critical determinant in this system was the manner in which the ligand and protein were combined. Initially, crystallization samples had been produced by incubating protein at ~10 mg ml<sup>-1</sup> with a vast excess of solid compound in order to avoid the use of DMSO, which was thought to be detrimental to cocrystallization. DLS of a sample incubated overnight with a low-solubility compound showed that the solution stayed monomeric after this treatment; in fact, the protein remained monomeric even after several days incubation at 277 K. In contrast, when added as a concentrated DMSO solution (100 mM), the same compound instantaneously induced complete protein dimerization, even when only a modest ligand excess was present. Clearly, in this simplistic case the kinetics of compound dissolution was the rate-limiting step and was primarily responsible for the lack of success. An efficient way of forming and checking complexes was therefore to add excess compound from a concentrated DMSO stock with no need for pre-incubation and to use this solution for cocrystallization trials. If necessary, this same solution could be spun down and used in the DLS to confirm the complexation. In this case the original DMSO sensitivity of the crystallization could be overcome, but when this is not possible dialysis of the final complex or use of a more tolerated solvent may be necessary. The important factor was being able to identify the key limitation.

## 5. Thermal stability enhancements on ligand binding

The ability of ligand binding to enhance protein stability is a well recognized phenomenon. The degree of stabilization can be systematically probed by observing the increased resistance of a protein to chemically or physically denaturing conditions, such as urea or temperature. Many methods are able to monitor the extent of denaturation, *e.g.* specific enzyme activity, NMR, circular dichroism (CD) and, most recently, extrinsic fluorescence using a probe that binds selectively to unfolded protein (*e.g.* ThermoFluor; Cummings *et al.*, 2006; Koblisch *et al.*, 2006).



**Figure 6** An idealized plot of the fluorescence changes that occur during the thermal denaturation of two proteins when a fluorescent dye such as ANS or Sypro Orange is used. The left curve shows the trace from a protein that denatures with  $T_m = 318$  K; the right curve shows one with  $T_m = 333$  K.



Dyes such as ANS and Sypro Orange show enhanced fluorescence when bound to hydrophobic patches on proteins. Protein denaturation tends to expose more of these areas, so the observed fluorescence increases during the transition from native to denatured states. Fig. 6 shows the idealized transition curves of two samples with different thermal stability, where the melting temperature  $T_m$  is defined by the mid-point of the denaturation curves. Fluorescence intensity can easily be measured in a plate-based format and this method of visualizing the temperature stability allows many samples to be quickly characterized with relatively small amounts of material, typically 20–50  $\mu\text{l}$  of  $\sim 0.1 \text{ mg ml}^{-1}$  protein (Vedadi *et al.*, 2006; Ericsson *et al.*, 2006).

Proteins with very low thermal stability (<293 K) may highlight a cause for concern; however, beyond this the absolute value of  $T_m$  does not provide any information about how likely a protein is to crystallize. Many factors that govern the ability to form ordered crystals are independent of those that determine  $T_m$ ; there is no reason why a protein with a lower  $T_m$  should not crystallize more readily than one with higher  $T_m$ . The utility of  $T_m$  lies in its application in a comparative context, where changes in  $T_m$  can identify those factors that have the greatest influence on protein behaviour, at least from a stability perspective.

There are numerous examples of proteins that do not crystallize unless a ligand is present. In some instances, the mere presence of a ligand is not truly sufficient and high diffraction quality requires high ligand potencies. This is the situation for the ligand-binding domain (LBD) of many nuclear receptors (NR), where the use of CD-detected  $T_m$  experiments are well established (Watkins *et al.*, 2003). One advantage of using CD over a dye-based method is the ability to interpret the CD spectrum in terms of the secondary and tertiary structural elements within the protein, thus providing a higher information-content assay.

Nuclear receptors are involved in transcriptional regulation. Ligand binding can result in activation, deactivation or modulation of biological activity depending on the conformational change induced upon complexation and its effect on the subsequent recruitment of partner proteins. In such a complex system, NR ligands are often identified within assays that maintain significant biological context. These assays do use the isolated LBD used for crystallographic studies and  $T_m$  measurements provide a generic way of triaging compounds for direct LBD interactors, regardless of their site of binding or the conformational changes elicited. Binders should enhance the protein stability, so

$$\Delta T_m = T_m(\text{ligand}) - T_m(\text{apo}) > 0.$$

Unfortunately, the affinity of a ligand cannot be simply determined from the saturation value of  $\Delta T_m$  (Matulis *et al.*, 2005). However, it may be possible to use the variation of  $\Delta T_m$  as a function of the ligand concentration to extract an affinity constant. For a given protein, there is a tendency for  $\Delta T_m$  to increase with potency.

When faced with a selection of many ligands for cocrystallization trials, one way of prioritizing ligands may be to

profile them using a fluorescence thermal denaturation,  $T_m$ , experiment. Those that produce no enhanced protein stability do not interact strongly with the protein, at least under the chosen conditions, so may be comparatively more difficult to cocrystallize than those that show strong enhancement or binding. Those that produce large stability shifts are strong binders and may be more ideal initial starting points. For some systems both weak and strong binders will result in successful X-ray complexes, while for others only the most potent inhibitors can be observed. This again highlights why  $\Delta T_m$  cannot be used in an absolute fashion, because the absolute requirements for success are system-dependent.

When highly potent ligands known to bind to the crystallographic protein repeatedly fail to produce complex structures, the effect of the extraordinary contents of the crystallization solution are often questioned. Most biophysical methods find that the presence of high precipitant concentrations (*e.g.* salts, PEGs) interferes with binding measurements. However, we have found that the fluorescence  $T_m$  assay is remarkably tolerant of a wide variety of conditions. It is therefore possible to systematically probe the effect of every component of the crystallization solution, *e.g.* pH, salt *etc.* to pinpoint the factor that has the greatest effect on the protein stability and, by inference, complex formation. It is important to remember that a suitable control must be run for each experiment. For example, to find out whether 1 M  $\text{NH}_4\text{SO}_4$  has an effect on ligand binding, the appropriate reference  $T_m$  must be from the apo protein also in 1 M  $\text{NH}_4\text{SO}_4$ , so only the presence and absence of ligand distinguishes the two samples. This is because the solution conditions themselves can affect the intrinsic stability of a protein.

In the past,  $T_m$  measurements have been used to identify stabilizing conditions for protein storage, often using differential scanning calorimetry (DSC). The thermal denaturation experiment may be used in the same way to scan a variety of buffer conditions to optimize protein handling and storage. These conditions could include the effect of solubilizing detergents. It may be less appropriate to use  $T_m$  as a parameter to choose additives designed to effect crystal growth, as factors that modulate the growth habit may have little effect on the intrinsic stability of the protein.

Whilst in principle a generic tool, there are instances where the denaturation method fails. The denaturation process cannot be visualized for all proteins. Some proteins precipitate before dye binding occurs. Others do not fully unfold on thermal denaturation and no significant increase in dye binding takes place (*e.g.* for some disulfide-bonded proteins). More fundamentally, there are rare but documented examples of nonspecific binding resulting in protein stabilization and specific binding giving rise to unexpected destabilization as measured by  $T_m$  (Horn & Shoichet, 2004). These caveats are especially pertinent for low-affinity ligands, where relatively large compound concentrations may produce significant nonspecific effects similar to those seen with additives such as arginine, glycerol *etc.* If only large  $\Delta T_m$  (>3 K) at low ligand concentrations (<100  $\mu\text{M}$ ) are considered, the dangers of misinterpreting nonspecific interactions are reduced.

## 6. Summary

Biophysical methods are an armoury of tools that can be fashioned to provide key guiding information to enhance the success of cocrystallization experiments. This review has highlighted a few examples of their application in the focused pursuit of crystal structures of protein–ligand complexes. Even within this narrow arena, their use often provides unexpected insights into the nature of the complexes formed. In our attempts to gain a better understanding of the biological and physical world, these tools provide unique information that is unobtainable from and complementary to that provided by a molecular structure.

## References

- Benesch, J. L. P. & Robinson, C. V. (2006). *Curr. Opin. Struct. Biol.* **16**, 245–251.
- Bitto, E., Bingman, C. A., Wesenberg, G. E., McCoy, J. G. & Phillips, G. N. Jr (2006). *J. Biol. Chem.* **281**, 20521–20529.
- Bligh, S. W. A., Haley, T. & Lowe, P. N. (2003). *J. Mol. Recognit.* **16**, 139–147.
- Bolbach, G. (2005). *Curr. Pharm. Des.* **11**, 2535–2557.
- Brown, W. (1993). *Dynamic Light Scattering: The Method and Some Applications*. Oxford: Clarendon Press.
- Charifson, P. S., Shewchuk, L. M., Rocque, W., Hummel, C. W., Jordan, S. R., Mohr, C., Pacofsky, G. J., Peel, M. R., Rodriguez, M., Sternbach, D. D. & Consler, T. G. (1997). *Biochemistry*, **36**, 6283–6293.
- Cousins-Wasti, R. C., Ingraham, R. H., Morelock, M. M. & Grygon, C. A. (1996). *Biochemistry*, **35**, 16746–16752.
- Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M. & Stockman, B. (2001). *J. Biomol. NMR*, **21**, 349–359.
- Cummings, M. D., Farnum, M. A. & Nelen, M. I. (2006). *J. Biomol. Screening*, **11**, 854–863.
- Emsley, J., Knight, C. G., Farndale, R. W. & Barnes, M. J. (2004). *J. Mol. Biol.* **335**, 1019–1028.
- Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.
- Ferre-D'Amare, A. R. & Burley, S. K. (1997). *Methods Enzymol.* **276**, 157–166.
- Gilbert, R. J. C. (2005). *Analytical Ultracentrifugation*, edited by D. J. Scott, S. E. Harding & A. J. Rowe, pp. 253–269. Cambridge: Royal Society of Chemistry.
- Guenther, U. L., Fischer, C. & Rueterjans, H. (2004). *Methods and Principles in Medicinal Chemistry*, Vol. 21, edited by T. Dinger-mann, D. Steinhilber & G. Folkers, pp. 242–268. Weinheim: Wiley-VCH.
- Hartshorn, M. J., Murray, C. W., Cleasby, A., Frederickson, M., Tickle, I. J. & Jhoti, H. (2005). *J. Med. Chem.* **48**, 403–413.
- Horn, J. R. & Shoichet, B. K. (2004). *J. Mol. Biol.* **336**, 1283–1291.
- Jahnke, W. (2002). *ChemBioChem*, **3**, 167–173.
- Jahnke, W., Perez, L. B., Paris, C. G., Strauss, A., Fendrich, G. & Nalin, C. M. (2000). *J. Am. Chem. Soc.* **122**, 7394–7395.
- Jancarik, J., Pufan, R., Hong, C., Kim, S.-H. & Kim, R. (2004). *Acta Cryst.* **D60**, 1670–1673.
- Jhoti, H. (2005). *Nature Biotechnol.* **23**, 184–186.
- Koblisch, H. K., Zhao, S., Franks, C. F., Donatelli, R. R., Tominovich, R. M., LaFrance, L. V., Leonard, K. A., Gushue, J. M., Parks, D. J., Calvo, R. R., Milkiewicz, K. L., Marugan, J. J., Raboisson, P., Cummings, M. D., Grasberger, B. L., Johnson, D. L., Lu, T., Molloy, C. J. & Maroney, A. C. (2006). *Mol. Cancer Ther.* **5**, 160–169.
- Krishnan, V. V. (2005). *Curr. Anal. Chem.* **1**, 307–320.
- Lesuisse, D., Lange, G., Deprez, P., Benard, D., Schoot, B., Delettre, G., Marquette, J., Broto, P., Jean-Baptiste, V., Bichet, P., Sarubbi, E. & Mandine, E. (2002). *J. Med. Chem.* **45**, 2379–2387.
- Mandine, E., Gofflo, D., Jean-Baptiste, V., Sarubbi, E., Touyer, G., Deprez, P. & Lesuisse, D. (2001). *J. Mol. Recognit.* **14**, 254–260.
- Matulis, D., Kranz, J. K., Salemme, F. R. & Todd, M. J. (2005). *Biochemistry*, **44**, 5258–5266.
- Nietlispach, D., Mott, H. R., Stott, K. M., Nielsen, P. R., Thiru, A. & Laue, E. D. (2004). *Methods Mol. Biol.* **278**, 255–288.
- O'Brien, R., Chowdhry, B. Z. & Ladbury, J. E. (2001). *Protein–Ligand Interactions: Hydrodynamics and Calorimetry*, edited by S. E. Harding & B. Z. Chowdry, pp. 263–286. Oxford University Press.
- Peng, J. W., Moore, J. & Abdul-Manan, N. (2004). *Prog. Nucl. Magn. Reson. Spectrosc.* **44**, 225–256.
- Scapin, G., Patel, S. B., Lisnock, J., Becker, J. W. & Lograsso, P. V. (2003). *Chem. Biol.* **10**, 705–712.
- Schmitz, K. (1990). *An Introduction to Dynamic Light Scattering by Macromolecules*. San Diego: Academic Press.
- Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1996). *Science*, **274**, 1531–1534.
- Stura, E. A., Loic, M., Lortat-Jacob, H., Vives, R. & Vita, C. (2002). *Acta Cryst.* **D58**, 1670–1673.
- Vedadi, M., Niesen, F. H., Allali-Hassani, A., Fedorov, O. Y., Finerty, P. J. Jr, Wasney, G. A., Yeung, R., Arrowsmith, C., Ball, L. J., Berglund, H., Hui, R., Marsden, B. D., Nordlund, P., Sundstrom, M., Weigelt, J. & Edwards, A. M. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 15835–15840.
- Watkins, R. E., Davis-Searles, P. R., Lambert, M. H. & Redinbo, M. R. (2003). *J. Mol. Biol.* **331**, 815–828.
- Williams, S. P., Kuyper, L. F. & Pearce, K. H. (2005). *Curr. Opin. Chem. Biol.* **9**, 371–380.
- Winzor, D. J. (2003). *J. Biochem. Biophys. Methods*, **56**, 15–52.
- Wishart, D. (2005). *Curr. Pharm. Biotechnol.* **6**, 105–120.