Progress in Nuclear Magnetic Resonance Spectroscopy 73 (2013) 1–16

Contents lists available at SciVerse ScienceDirect

Progress in Nuclear Magnetic Resonance Spectroscopy

journal homepage: www.elsevier.com/locate/pnmrs

Using chemical shift perturbation to characterise ligand binding

Mike P. Williamson *

Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK

Edited by David Neuhaus and Gareth Morris

ARTICLE INFO

Article history:
Received 22 January 2013
Accepted 18 February 2013
Available online 21 March 2013

Keywords:
Chemical shift
Protein
Exchange rate
Dissociation constant
Docking

ABSTRACT

Chemical shift perturbation (CSP, chemical shift mapping or complexation-induced changes in chemical shift, CIS) follows changes in the chemical shifts of a protein when a ligand is added, and uses these to determine the location of the binding site, the affinity of the ligand, and/or possibly the structure of the complex. A key factor in determining the appearance of spectra during a titration is the exchange rate between free and bound, or more specifically the off-rate k_{off}. When k_{off} is greater than the chemical shift difference between free and bound, which typically equates to an affinity K_d weaker than about 3 μM, then exchange is fast on the chemical shift timescale. Under these circumstances, the observed shift is the population-weighted average of free and bound, which allows K_d to be determined from measurement of peak positions, provided the measurements are made appropriately. ^{1}H shifts are influenced to a large extent by through-space interactions, whereas $^{13}C\alpha$ and $^{13}C\beta$ shifts are influenced more by through-bond effects. ^{15}N and $^{13}C'$ shifts are influenced both by through-bond and by through-space (hydrogen bonding) interactions. For determining the location of a bound ligand on the basis of shift change, the most appropriate method is therefore usually to measure ^{15}N HSQC spectra, calculate the geometrical distance moved by the peak, weighting ^{15}N shifts by a factor of about 0.14 compared to ^{1}H shifts, and select those residues for which the weighted shift change is larger than the standard deviation of the shift for all residues. Other methods are discussed, in particular the measurement of $^{13}CH_3$ signals. Slow to intermediate exchange rates lead to line broadening, and make K_d values very difficult to obtain. There is no good way to distinguish changes in chemical shift due to direct binding of the ligand from changes in chemical shift due to allosteric change. Ligand binding at multiple sites can often be characterised, by simultaneous fitting of many measured shift changes, or more simply by adding substoichiometric amounts of ligand. The chemical shift changes can be used as restraints for docking ligand onto protein. By use of quantitative calculations of ligand-induced chemical shift changes, it is becoming possible to determine not just the position but also the orientation of ligands.

Contents

* Tel.: +44 114 222 4224; fax +44 114 272 8697.
E-mail address: m.williamson@sheffield.ac.uk

0079-6565/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.pnmrs.2013.02.001

1. Introduction

Chemical shift perturbation (CSP, also known as chemical shift mapping or complexation-induced changes in chemical shift, CIS) is a very simple experimental technique for studying binding to a protein. In the standard experiment, one needs an ^{15}N-labelled protein plus an unlabelled ligand, which can be a small molecule or another macromolecule. The ligand is titrated into the protein, monitored at each stage of the titration by acquiring a 2D HSQC spectrum [1]. Using a moderately highfield spectrometer with a cryocooled probe, one can acquire HSQC spectra in about 30 min for proteins at concentrations of 200 μM or more. This means that one can acquire a complete titration in about a day. With TROSY and perdeuterated protein [2] one can observe proteins of several hundred kDa, so that (for a well-behaved protein, and with a bit of effort) one can acquire CSP data on most targets of interest. If one is studying the binding of two proteins to each other, then each protein can be labelled in turn, providing information about both partners: indeed, by use of ^{13}C labelling on one protein but not the other, one can observe both ^{15}N-labelled proteins separately and simultaneously, in relatively small systems at least [3].

The chemical shift change is very sensitive to structural changes, and can be measured very accurately, meaning that almost any genuine binding interaction will produce CSPs. The analysis is also simple, at least in its basic form: measure the chemical shifts at each titration point, follow the movement of peaks, and measure how each peak moves throughout the titration. The peaks that move the most are very likely to map to the binding site for the ligand. Moreover, the shape of the titration curve (chemical shift vs. concentration of ligand) can often be fitted straightforwardly to obtain a value for the dissociation constant of the ligand, K_d. CSP is the only technique that can directly provide both a K_d value and a binding site from the same set of measurements [4]. The only important caveat, as with any quantitative measurement, is that during the titration, it is important to keep experimental conditions as consistent as possible. In particular, it is important to use the same buffer for protein and ligand, because small changes in pH or salt concentration can alter protein signals and confuse the analysis. Similarly, if the ligand is only soluble in an organic solvent such as DMSO, then the titration must be set up in such a way as to keep the DMSO concentration constant.

The technique can be useful even without a chemical shift assignment of the HSQC spectrum. Most usefully, if a ligand does not bind, then there will be no chemical shift changes seen. CSP is widely used in drug discovery for this reason: many other techniques such as spectrophotometry, calorimetry or enzyme assay are prone to giving false positive results, whereas CSP in general does not. It is thus a useful and moderately high-throughput method for checking whether potential ligands really do bind, and forms the basis for the 'SAR by NMR' methodology [5]. Furthermore, CSP can be used to obtain K_d values in the absence of assignments; and one can compare which signals move on addition of different ligands, and thus ascertain whether different ligands bind in the same binding site or not. CSP is however much more powerful when the assignments are known. Fortunately, triple resonance techniques mean that backbone assignments are often obtainable quickly and even automatically [6]. CSP can be used with solid-state spectra as well as solution, making it even more versatile [7].

CSP is remarkably reliable as a guide to interaction sites, both of ligand with protein and of protein with protein. Provided that the crystal structure of the protein is known and the spectrum assigned, a big advantage of this method is that it is not necessary to calculate an NMR structure; one can use the crystal structure and simply map chemical shift changes onto it. Alongside this, the increased number of assignments of proteins with known structures, and the vastly increased speed of computers, has also meant that we are now better able to understand the origins of chemical shifts in proteins. CSPs are thus entering an exciting new phase, in which we can make quantitative use of the shift changes to probe the geometry of the interactions.

We therefore start with a brief discussion of the origins of chemical shifts in proteins, and go on to consider how CSPs can be applied. Because CSPs are experimentally and conceptually simple, there has been surprisingly little analysis of their application: remarkably, this is the first article specifically on CSPs to appear in *Progress in NMR Spectroscopy*. Hiding behind the simplicity, there are a range of issues that one needs to be aware of, most importantly to do with multiple binding modes, as discussed below; and with the problems arising when the system is not in fast exchange – a situation not always easy to spot.

2. Origins of chemical shift effects in proteins

2.1. Calculation of chemical shifts

There are two main approaches towards the calculation of chemical shifts in proteins. One is to use quantum chemical methods, most commonly standard packages such as Gaussian 98 [8], which calculate the electron densities in molecules, and therefore allow calculations of the shielding of nuclei from the external magnetic field by their electrons, which is what ultimately is responsible for the observed chemical shifts. In the past, the problems with these methods have been that their accuracy (by which we mean the agreement with experimentally determined shifts) is questionable, and they are slow, implying that they can only be applied to very small molecular fragments. There is a more subtle difficulty with quantum chemical methods, in that the result is essentially just a calculated shift for a given molecular configuration, which offers little help in understanding what aspect of the structure has produced the calculated shift. The user therefore has to select his or her structural models carefully, in order to gain useful insight. The methods are improving rapidly, as are the computers that they run on, and quantum chemical calculations are now the method of choice for ^{15}N and ^{13}C′ nuclei, and probably for other carbon nuclei also. For ^{1}H, the difficulty is that chemical shift effects arising from through-space interactions are just as important as through-bond interactions; and that the shielding of the proton by its surrounding electron is weaker than that of heavier nuclei, implying that the chemical shift range for ^{1}H is smaller than it is for ^{15}N or ^{13}C, and thus calculations need to be relatively more accurate. Chemical shifts in ^{1}H therefore are harder to calculate

to useful accuracy, and often need inclusion of more atoms, to account for through-space effects properly. At present it is true to say that quantum chemical calculations have not been as helpful in understanding chemical shifts in ^{1}H in proteins as they have for other nuclei.

The second method is an empirical one. Chemical shifts are modelled as being the summation of a number of independent effects. Equations are written for each of these effects, based on theories as to possible origins of chemical shifts (or even just geometrical factors such as distances: [9]), and the parameters in these equations are fitted so as to obtain the best fit between experimental and observed shifts. Clearly there are also problems with this approach. Firstly, the method is only as good as the equations used. Second, any problems with the experimental data will introduce inaccuracies in the calculation: these could include errors in the experimental shift tables (a common problem though probably not a major source of error, although errors in referencing remain an irritation), or differences between the structures used in the calculations and the 'real' structures. The structures used in the calculations are usually crystal structures, while of course the 'real' structures giving rise to the measured chemical shifts are solution ensemble averages. There is thus an error arising from the slight differences between solution and crystal structures, and from any conformational averaging effect. Conformational averaging clearly does have an effect on shifts [10]. Related to this is the observation that in an empirical method, one takes the experimental distribution of shifts, and tries to fit some equation to it. The fitting process inherently and necessarily smoothes the data, with the result that extreme and unusual shifts are never predicted well.

Empirical methods have done surprisingly well: it does indeed appear that most chemical shift effects are simply additive. These methods are probably the most useful for calculations of ^{1}H shifts, and remain useful for ^{13}Cα and ^{13}Cβ shifts. The factors governing ^{15}N and ^{13}C′ shifts have so far proved too complex for empirical methods, which have only recently become interpretable at all as a result of quantum mechanical calculations. Thus, our understanding of chemical shifts in proteins is moving slowly from being based on empirical methods to quantum chemical methods. In particular, Oldfield and Case have shown that by choosing suitable molecular fragments, it is possible to put together a toolbox for predicting shifts in proteins (as a series of look-up tables), which complement and may eventually replace those derived from empirical methods [11–13]. The descriptions that follow are based largely on their work, and on our own empirical studies on both ^{1}H and ^{13}C shifts [14–16].

2.2. Proton shifts

Different chemical environments of ^{1}H cause very different primary chemical shifts: for example, backbone amide HN groups of proteins generally resonate in the range 7–9 ppm whereas CH_3 groups generally resonate in the range 0–2.5 ppm. These shifts are largely caused by inductive, through-bond effects, due to partial charges on the heavy atoms. Conveniently, they can normally be ignored by considering only the *secondary* shifts, which are defined as being the difference between the observed shift and the 'random coil' shift, where 'random coil' means the ensemble of structures adopted by an unstructured peptide. Clearly, this is a slightly vague definition, but has served well. Lists of random coil shifts have been compiled for various solution conditions [17–23].

Using this definition, we may define the secondary shift as being composed of a sum of effects:

$$\sigma = \sigma^{p} + \sigma^{ani} + \sigma^{ring} + \sigma^{E} \quad (1)$$

where σ^{p} arises from paramagnetic effects, σ^{ani} arises from the magnetic anisotropy around bonds, σ^{ring} arises from aromatic ring currents, and σ^{E} arises from electric fields caused by charged atoms. σ^{p} can be ignored in most cases, except in proteins containing paramagnetic centres. Of the remaining three, σ^{ring} can cause very large effects when it is present: in particular, when a proton is close to an aromatic ring. σ^{ani} mostly arises from carbonyls and amide groups, which contain the most anisotropic distributions of electrons in proteins. Many protons are close to such groups, and thus a large number of protons experience sizeable secondary chemical shift effects from this origin, over a total spread of about ±1 ppm. σ^{E} is an inductive effect, due to neighbouring charged atoms either pulling

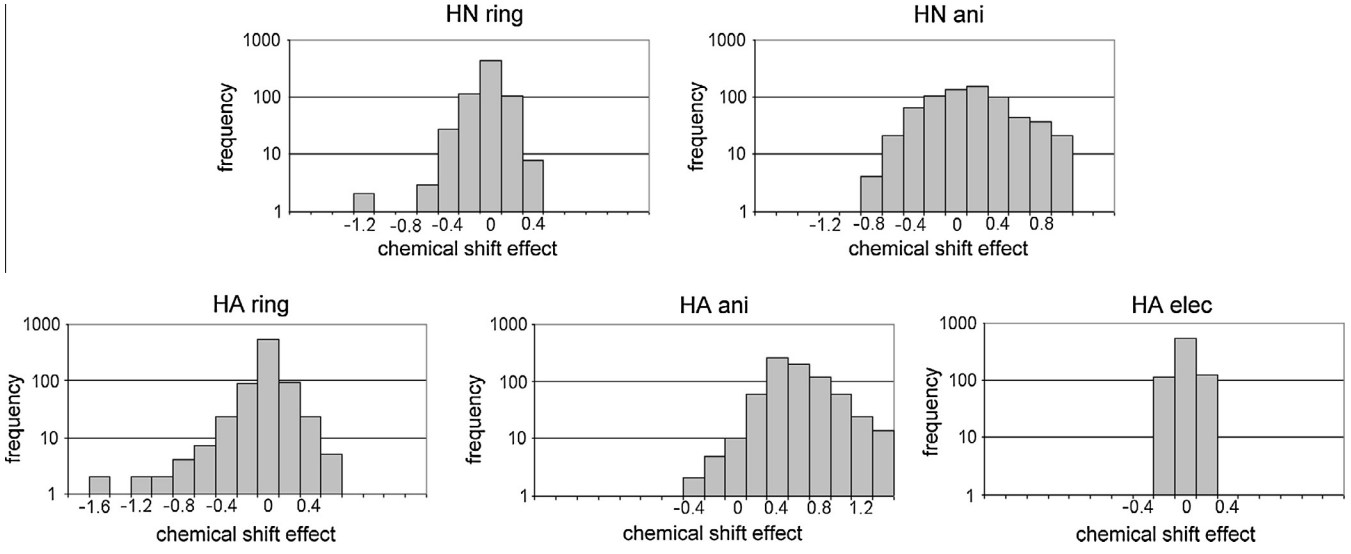

Fig. 1. Range of chemical shift effects caused by ring currents, bond magnetic anisotropy and electric field effects, calculated for Hα and HN protons in a set of representative diamagnetic proteins. The calculations were carried out with the program *total* [14], in which no electric field effects are calculated for HN. The histograms are represented using a logarithmic scale for frequencies, because of the very large number of protons with small effects, which would otherwise make the smaller numbers of protons with large effects impossible to see. Although the ring current shifts have a narrower distribution, they produce more protons with strongly shifted signals. Ring current effects tend to be skewed towards upfield shifts, because protons above or below the plane of the aromatic ring tend to have larger upfield effects than are the downfield effects to protons at the side of the ring (mainly because protons above or below the ring can approach it more closely than can those at the side). Note that these are intramolecular effects, and very few intermolecular effects will be as large. However, this demonstrates the maximum range expected.

electron density away from a proton, or pushing it towards a proton. Individual fully or partially charged atoms can cause quite large effects, but the overall sum from all neighbouring charged atoms is small, partly because of the fairly uniform distribution of local charges within a protein. It therefore causes the smallest effect of the factors discussed here. Another reason for this is that the electric field effect is attenuated by the dielectric constant, which in water is large. (Inside proteins, it is much smaller; nevertheless it is still large enough to make σ^E a small effect compared for example to the situation for small molecules in organic solvents.)

Fig. 1 shows typical ranges for the three effects. The dominance of ring current effects is of particular interest because a large majority of drugs (estimated as >95% [4]) contain at least one aromatic ring: we may therefore expect most of the direct effects on protein chemical shift caused by titration with ligands to be due to either ring current effects or hydrogen bonding to amide protons.

These equations have been applied to a range of proteins, and give fits with experimental shifts, with root-mean-square (RMS) differences of approximately 0.25 ppm for ^{1}H–C and 0.5 ppm for ^{1}H–N. More sophisticated equations with more terms give better fits, currently up to about twice as good [24]. The poor fit for HN has a number of causes. The largest HN shifts generally arise from hydrogen bonding interactions. We do not in general know what hydrogen bonding interactions are made to solvent, and thus so far there is no good way to calculate this effect. Secondly, because hydrogen bonding interactions tend to be very short-range and highly directional, the effect on HN chemical shift of very small coordinate changes is large. Thus, small random errors in the structure used for the calculation (or small changes between the solution ensemble and the crystal) can have large effects on the calculated shift. The conclusion is therefore that any change in HN shift is likely to arise from a change in hydrogen bonding, but the exact size of the shift is difficult to calculate.

In summary, ^{1}H secondary shifts in proteins are dominated by ring current effects and (for amide protons) hydrogen bonding, and can be calculated reasonably accurately. However, the observed shift changes are often the sum of a number of competing smaller effects, implying that it is often difficult to relate a given shift change to a single structural change.

2.3. Cα and Cβ shifts

Carbon shifts, like proton shifts, arise from shielding of the nucleus by electrons. The difference between carbon and proton is that carbon has electrons in both *s* and *p* orbitals, whose hybridisation varies according to the geometry of bonds around the carbon. Thus, carbon shifts are much more dependent than proton shifts on local dihedral angles. The long-range effects described above for ^{1}H have the same effect on carbon as they do on proton (in ppm). However, the total shift range of carbon is greater than that of proton (implying that the relative importance of long-range effects is less), and in addition carbon atoms are normally less surface exposed than protons, and so are further removed from any external effects. Thus, through-space effects have only a minor effect on Cα and Cβ shifts.

The Cα and Cβ atoms are in the 'middle' of an amino acid. The consequence is that their secondary chemical shifts are affected very little by neighbouring amino acids. There is a small sequence effect, but very little effect from the conformations of neighbours. Most of the effects on Cα and Cβ shifts are from the backbone conformation of the amino acid itself, with small extra contributions from the orientation of the amino acid sidechain and hydrogen bonds to the amide bonds on either side. The backbone conformation contributes up to about 4 ppm to the secondary shift, while sidechain conformation has an effect of up to 0.6 ppm, and hydrogen bonding up to 0.9 ppm, depending on main-chain conformation. There are programs available to calculate Cα and Cβ shifts, which are simple additive calculations.

This all means that any effects of ligand binding on Cα and Cβ shifts are most likely to be due to local structural rearrangements, rather than any direct through-space interaction. The same is probably true of other sidechain carbon shifts.

2.4. ^{15}N and ^{13}C′ shifts

Nitrogen and carbonyl carbon shifts are by far the hardest to calculate, because they are affected by so many different factors. The identity and sidechain conformation of the preceding amino acid causes nitrogen secondary shifts of up to 22 ppm, of which up to 8 ppm can be contributed just by the sidechain conformation. As is the case for Cα and Cβ, there is a contribution (of up to 13 ppm) from backbone dihedral angles within the same amino acid, but for ^{15}N the ψ angle of the preceding residue also matters. Finally, and significantly, hydrogen bonding has an effect of up to 8 ppm. Interestingly, the chemical shift of an amide nitrogen is affected more by hydrogen bonding to its directly attached carbonyl group than it is by hydrogen bonding to the amide nitrogen itself. Thus there are many factors that contribute significantly to observed shifts. Unfortunately, this means that at present there is no simple way to calculate or interpret ^{15}N chemical shift changes. This is a pity, because of the importance of ^{15}N HSQC in CSP.

Carbonyl carbon shifts follow much the same pattern as ^{15}N shifts, with the differences that all the effects are smaller, and for carbonyls it is the following residue that is important, not the preceding one. Carbonyl shifts are thus at present equally difficult to calculate.

3. Rates and affinities

For a protein *P* binding reversibly to a ligand *L* at a single site, given by $P + L \leftrightarrow PL$, characterised by rate constants for forward and back reactions of k_{on} and k_{off}, the dissociation constant K_d is equal to $[P][L]/[PL]$, where $[P]$, $[L]$ and $[PL]$ represent the concentrations of free protein, free ligand and complex. K_d can be thought of as the concentration of ligand and protein required to saturate half the binding sites. The forward and back rates are given by $[P][L]k_{on}$ and $[PL]k_{off}$ respectively. At equilibrium the forward and back rates are equal, implying that the dissociation constant K_d is also equal to k_{off}/k_{on}. For a diffusion-controlled binding (i.e., binding to a sterically available site), k_{on} is typically around 10^9 M^{-1} s^{-1}[25]. There is thus an approximate relationship between K_d and k_{off}, that $k_{off} \approx 10^9$. K_d (with K_d expressed in M).

This has an important effect on the NMR spectra of an exchanging system. We consider the HSQC spectrum of a protein as ligand is titrated in. When the exchange rate is slow on the chemical shift timescale, or in other words when k_{off} is significantly slower than the difference in Hz between the chemical shifts of free and bound protein, then as ligand is titrated in, the free signal gradually disappears and the bound signal appears, the intensities of the two peaks reflecting the concentrations of free and bound protein. On the other hand, when exchange is fast, i.e. when k_{off} is much greater than the chemical shift difference, then the signals will move smoothly from their position in the free spectrum to those in the bound spectrum, with the frequency of the signal at any titration point being the weighted average of free and bound shifts (Fig. 2). When the exchange rate is similar to the shift difference, then signals broaden and shift at the same time.

Importantly, most of this review assumes that exchange is fast. On addition of a ligand, one might typically expect to see chemical shift changes of up to about 0.5 ppm on ^{1}H and 3 ppm on ^{15}N. At

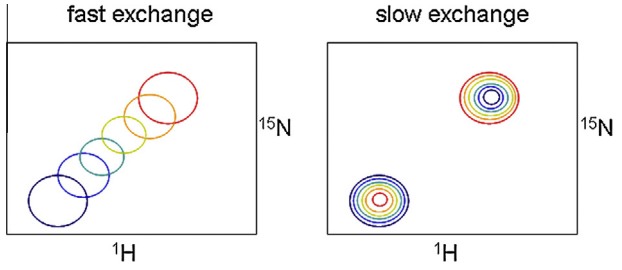

Fig. 2. The dependence of two-dimensional NMR peak shape on exchange rate. (Left) Fast exchange: peaks move smoothly from free (blue) to bound (red). In the limit of very fast exchange, peaks have the same shape throughout. As they move out of this limit, peaks may become broader when in equilibrium between free and bound, and then sharpen up again close to saturation. (Right) Slow exchange: the free peak (blue) decreases in intensity as the bound peak (red) increases.

600 MHz, these translate to 300 Hz for ^{1}H and 180 Hz for ^{15}N. 'Much greater' in the paragraph above means at least 10× faster, implying that the fast exchange limit in practice means that k_{off} must be faster than about 3000 s^{-1}. From the relationship above, this means that K_d must be greater than about 3 μM to be in the fast exchange limit, a useful rule of thumb. This value was however based on assumptions that are sometimes incorrect. For instance, in some cases, the on-rate is much slower than diffusion-controlled, leading to slow exchange even for weak binding.

In passing, we note that amide groups may sometimes have ^{1}H resonances that are in slow exchange but simultaneously have ^{15}N resonances that are in fast exchange. This leads to strange titration data, in which 2D crosspeaks in the ^{15}N HSQC spectrum grow and shrink in a slow-exchange manner, but their ^{15}N shifts change during the titration in a fast-exchange manner.

The rule of thumb described above can be relaxed a little and still be useful. In particular, if there are two competing binding events, one weak and one strong, then competition between the two binding sites can make it possible to measure affinities at the strong site even with dissociation constants below 10 nM [26,27].

During a titration, one typically knows the total concentrations of ligand and protein, which we can write as $[L]_t$ and $[P]_t$. These are the sum of free and bound forms:

$$[L]_t = [L] + [PL] \quad \text{and} \quad [P]_t = [P] + [PL] \tag{2}$$

We also know that the observed chemical shift change in fast exchange is the weighted average of the shifts in the free and bound states, or in other words

$$\delta_{obs} = \delta_f f_f + \delta_b f_b \tag{3}$$

where f_f and f_b are the fractions of free and bound. Finally, because

$$f_f + f_b = 1 \tag{4}$$

this allows us to express the fraction of ligand bound as

$$f_b = (\delta_{obs} - \delta_f)/(\delta_b - \delta_f) \tag{5}$$

A little algebra results in the equation**

$$\Delta\delta_{obs} = \Delta\delta_{max}\left\{([P]_t + [L]_t + K_d) - [([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t]^{1/2}\right\}\Big/2[P]_t \tag{6}$$

where $\Delta\delta_{obs}$ is the change in the observed shift from the free state, and $\Delta\delta_{max}$ is the maximum shift change on saturation (usually obtained as part of the fitting procedure, because it is often not measurable experimentally). This equation allows us to fit K_d from measured values of the chemical shift at different concentrations of protein and ligand. The fitting is easily set up in a spreadsheet or other fitting routine.

It is worth making two points here on fitting of dissociation constants. First, fitting many titration curves simultaneously (i.e., using the same values of K_d but different δ_b) gives a much more accurate estimate of K_d [26]. The more curves the better, at least up to 20. Secondly, it can be difficult to know the exact protein concentration. In this case, the protein concentration can be treated as another variable in the fitting, which often leads to significantly improved fits [26]. This is however something that needs watching carefully, since large changes in the protein concentration away from the expected value are a sign that something is wrong somewhere.

This equation shows that a good estimate of K_d can only be obtained if the concentration of protein and ligand is somewhere close to K_d. A more detailed analysis [28,29] shows that the optimum value for the protein concentration $[P]_t$ is $0.5K_d$. Values up to a factor of 10 less or more are still usable, as long as the range of concentrations of ligand is large enough, but the error in the fit rises dramatically away from these conditions [30]. In particular, the ligand should ideally span the range from $0.4[P]_t$ to $10[P]_t$ ($K_d/5$ to $5K_d$), with the optimal number of titration points being in the range 15–20 [28]. The lowest concentration of protein that can usefully be employed for HSQC titrations under optimum conditions of sensitivity is currently around 10 μM, implying that dissociation constants stronger than about 1 μM are too strong to be fitted from HSQC spectra. Fortuitously, this is also close to the limit for fast exchange derived above. At the other end of the scale, CSP data can be used to fit dissociation constants as weak as 10 mM, which is close to the upper limit for biologically relevant affinities. Thus, CSP data can be used to derive K_d values over a span of roughly 10^4 in affinity, covering the 'weak' end of biological affinities [31]. As it happens, a high proportion of binding events, both protein/protein and protein/ligand, come into this range.

The range of protein and ligand concentrations described above is set by the need to cover a range of fractional saturations of protein by ligand, ideally up to almost full saturation. In practice this is often difficult to achieve, due to the solubility or availability of ligand (particularly if the ligand is another protein). For an 'ideal' protein concentration of $K_d/2$, and a maximum ligand concentration of $5K_d$, the saturation still only reaches 82% (calculated using Eq. (6) above), implying that it is very difficult to achieve complete saturation of protein by ligand. If conditions close to full saturation are not attainable, the error in the fitted K_d will be larger [28]. One way to reduce the error is to use several different protein concentrations, or to deliberately dilute the protein during the titration, either by starting with a concentrated ligand/protein solution and diluting it, or using a dilute solution of ligand [30].

The saturation curve given by Eq. (6) is shown in Figs. 3 and 4, which allows us to make two trivial (but often overlooked) points. The first is that one needs significant curvature in order to be able to fit effectively. The black curve shown in Fig. 3 corresponds to very weak binding (K_d = 1 mM), and is not sufficiently curved to give a reliable fit (i.e., the curve reaches nowhere near saturation). By contrast, the red curve corresponds to very tight binding (K_d = 1 μM), and is essentially two straight lines; again, it cannot be fit, except possibly to give an upper limit for K_d. It is worth noting that the position of the bend in the red curve corresponds to 1:1 binding: an error in the estimated protein or ligand concentration, or binding not in a 1:1 ratio, will give a 'bend' at some concentration other than the expected equimolar point, and is a useful pointer to such problems. The optimal curve (i.e. a protein concentration of $K_d/2$, as discussed above) is close to the blue curve. The second point is that the shape of the saturation curve depends not just on the ratio of ligand to protein, or on the ligand concentration, but on *both* ligand *and* protein concentration. It is therefore not informative merely to show ligand concentration in such a plot – one must describe the protein concentration too (and how it

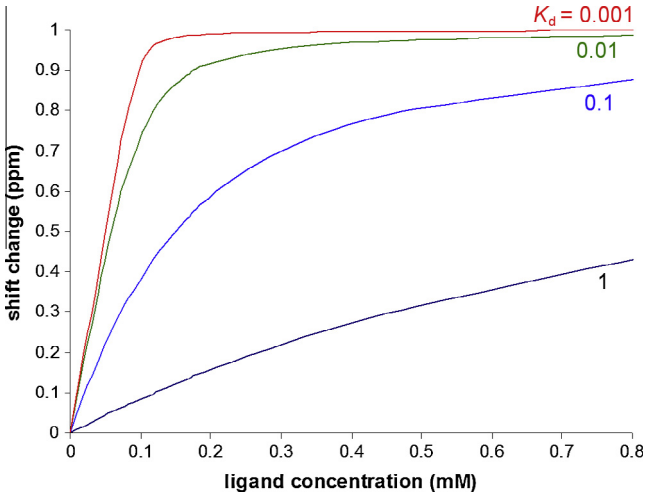

Fig. 3. Calculated saturation curves for a single-site binding equilibrium. All curves are calculated for a protein concentration of 100 μM, a maximum shift change on saturation ($\Delta\delta_{max}$) of 1 ppm, and a highly concentrated solution of ligand, such that the addition of ligand causes effectively no dilution of protein. The K_d values are given in mM, implying that for the curve of K_d = 0.1 mM, the protein concentration is the same as K_d.

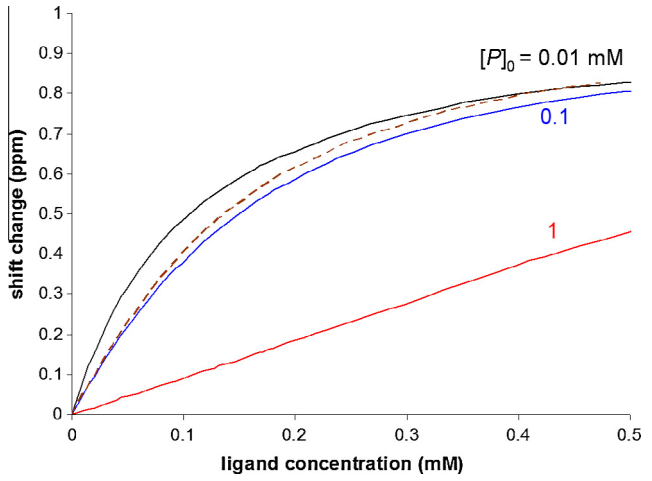

Fig. 4. Calculated saturation curves for a single-site binding equilibrium, taking into account dilution through the titration. Curves were calculated for a K_d of 0.1 mM. For the solid curves, the stock concentration of ligand was 5 mM, implying that for the highest ligand concentration shown on the graph, the protein concentration has been reduced by 17% as a result of addition of ligand. In the dashed curve, the stock ligand concentration was 0.5 mM, implying that the protein concentration is reduced during the titration from 0.1 mM to less than 10 μM by the end, explaining why it crosses over the solid curve above it.

varied during the titration). The curves in Fig 4 all correspond to the same K_d, and the same range of ligand concentrations, but have either different protein concentrations or simulate titrations done in different ways, as explained in the legend.

Eqs. (2)–(6) above assume that the ligand binds to the protein in a 1:1 ratio. This is by far the most common case. If the ligand binds to multiple sites all with the same affinity, then Eq. (6) can be modified very simply [32]:

$$\Delta\delta_{obs} = \Delta\delta_{max}\left\{(n[P]_t + [L]_t + Kd) - [(n[P]_t + [L]_t + K_d)^2 - 4n[P]_t[L]_t]^{1/2}\right\}\Big/2n[P]_t \quad (7)$$

where n is the number of equivalent sites. If there is more than one site and the affinities are different, the analysis is much more difficult and is discussed in Section 6.

4. Locating the binding site using fast exchange shifts

4.1. Choice of nuclei

In this section, we consider how chemical shift changes can be used to map ligand binding, considering the different factors involved, in order of increasing complexity of analysis.

By far the most common method of measuring chemical shift changes in a protein is to use ^{15}N HSQC spectra, which provide a rapid and simple way of locating changes. This only requires ^{15}N labelling of the protein, which is usually an easy and cheap option. The ^{15}N HSQC spectrum is the easiest to assign, it is sensitive, and it is usually well resolved, particularly in comparison with the ^{13}C/^{1}H HSQC. It detects just one signal per amino acid, from the backbone (excepting proline), plus a small selection of sidechain signals (Asn, Gln and Trp, and sometimes, depending on the solution conditions, Arg). As it happens, these are the functional groups that often tend to form specific interactions with ligands, especially when we consider that ^{15}N shifts are strongly affected by hydrogen bonding to the adjacent carbonyl (i.e., hydrogen bonding to the backbone carbonyl of residue i gives rise to large shift changes for the backbone ^{15}N of residue $i+1$), implying that direct interactions between ligand and backbone carbonyl will also be observed in the shift of the adjacent amide nitrogen. Hydrogen bonding interactions with the acidic sidechains of Glu and Asp often produce significant changes in the backbone ^{15}N, and in the ^{15}N of the following residue [33], and are thus also picked up using ^{15}N HSQC. However, we should bear in mind that by looking only at ^{15}N HSQC spectra, we may produce a biased view of the binding site, which downplays hydrophobic interactions (and of course interactions with proline). The bias against hydrophobic interactions does not in practice seem to be a problem. Partly this is possibly because the most common way of mapping changes in HSQC spectra onto the protein is to colour the surface of the protein by chemical shift change, colouring the entire amino acid residue when any NH signal in that residue changes. Thus, provided that it leads to a change in the backbone HN shift(s), an interaction with a hydrophobic sidechain exposed on the surface of a protein subjectively 'looks' inherently important because of the large coloured area depicted on the protein surface, because hydrophobic sidechains often have a larger surface area than hydrophilic sidechains [34]. It therefore appears that the different biases for and against identification of hydrophobic binding work in opposite directions if the whole sidechain is coloured: the biggest problem is probably that effects on Gly and Ala look insignificant merely because they have a small surface area.

As noted above, the other obvious spectrum to use, namely a 2D ^{13}C HSQC, is often not useful because both the ^{13}C and the ^{1}H resolution is poor, because some signals will be obscured by the water peak, and because chemical shift changes in ^{13}CH are usually smaller than those in ^{15}N. However, it does have some advantages [35]. ^{13}C shifts are less influenced than ^{15}N shifts by structural changes in the protein, implying that ^{13}C shifts may be clearer indicators of direct binding-induced effects. In addition, the smaller chemical shift changes for ^{13}C may mean that the spectra are closer to fast exchange, and therefore easier to interpret, in cases where the off-rate is slow. The other quick and sensitive 2D experiment, 2D ^{13}C HNCO, would be a good option, although more complicated to interpret because it gives the shift of the ^{1}HN in one residue and ^{13}C′ in another.

The question of which nuclei are the most informative has been studied in detail. From the discussion in Section 2, one can conclude that ^{1}HN shifts are sensitive both to hydrogen bonding and to ring currents; ^{15}N to hydrogen bonding (to the nitrogen and the preceding carbonyl) and to conformational effects; ^{13}C′ to

hydrogen bonding (to the carbonyl and the following nitrogen) and conformation; and aliphatic ^{13}C to conformation only. Since our interest is mainly in the effect of ligand binding, and explicitly not conformational change, this would imply that we should not use $^{13}C\alpha$ and $^{13}C\beta$. Schumann et al. [36] found that use of more nuclei provides better discrimination. ^{15}N, ^{1}HN and $^{13}C'$ were similar in their ability to detect the binding site, with $^{13}C\alpha$ slightly worse. Stratman et al. [37] have looked at different nuclei for their power in defining the location of the binding site, and conclude that $^{1}H\alpha$, $^{13}C\alpha$ and ^{15}N are the most useful, with ^{1}HN and $^{13}C\beta$ being worse. It therefore appears that for practical applications (acquiring the most data in the shortest time), ^{15}N HSQC spectra probably remain the best option.

It is interesting to note that ligand-induced changes in ^{1}HN and ^{15}N shift are poorly correlated. Even for well-defined interactions such as protonation of a sidechain, the changes are still uncorrelated [33]. There are probably two reasons for this. One is that ^{15}N shifts are strongly influenced by hydrogen bonding to the carbonyl of the preceding residue, which has very little effect on the proton shift. Second, electric field effects (such as hydrogen bonding) have an effect proportional to the cosine of the angle to the relevant bond (Fig. 5). These angles are quite different for ^{1}H and ^{15}N.

There are some cases where ^{13}C shifts (from ^{13}C HSQC spectra) are clearly useful. In particular, there is growing interest in using changes in $^{13}CH_3$ signals. This is especially true for larger proteins, where TROSY ^{15}N HSQC spectra are relatively weak and crowded, but methyl groups remain easily observable. Because there are three protons in the methyl group and the signals are slowly relaxing and consequently sharp, these signals tend to be of high intensity: an early study [38] showed that protein concentrations as low as 15 μM give useful spectra. It is becoming common to label methyl groups specifically, for example by feeding with ^{13}C-methyl labelled α-ketovalerate and α-ketobutyrate in an otherwise deuterated and ^{12}C background [39], which produces $^{13}C^{1}H_3$ labels just in the methyls of valine, leucine and isoleucine: there are also efficient (though still somewhat laborious) methods for assigning these methyl signals from such samples. Binding sites are frequently rich in hydrophobic sidechains, making this a powerful technique [40–42], though more challenging both in the protein production and the assignment. This makes the observation of CSPs in methyl groups a useful way of screening by detecting binding and comparing binding sites, but a laborious way of mapping them in detail [43]. The $^{13}C\varepsilon$ resonance of methione is often well resolved, and in some proteins (for example calmodulin [44]) forms a key part of the binding site, and has been used to detect binding [45]. The low occurrence of methionines means that methionine scanning (i.e., introduction of methionines by site-directed mutagenesis residue by residue as probes of binding) could turn out to be a useful technique [46], not least because the introduction of single methionines one by one makes them easy to assign.

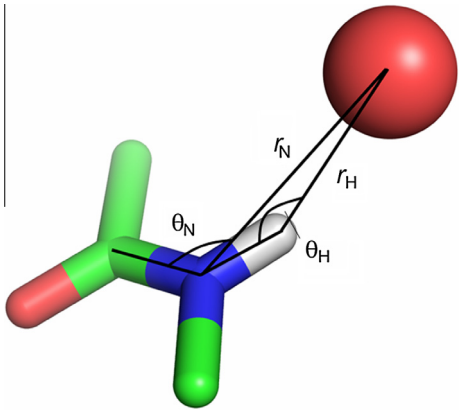

Fig. 5. The electric field effect caused by a charged atom (shown as a sphere) is proportional to $q\cos\theta/r^2$, where q is the charge on the atom [107]. The angle θ is the angle made to the bond along which electron density is pushed or pulled: where there is a choice, this will be the most polarisable bond. For $^{1}H_N$, this is the H–N bond. For N, this is the N–C′ bond. Because the relevant angles for H and N can be completely different, the effect on the chemical shifts of H and N can also be very different.

The sensitivity of cryocooled probes means that it is now practical to follow ligand titrations on methyl groups in unlabelled proteins. A recent example used methyl SOFAST experiments to measure HSQC spectra on 1 mM bovine serum albumin in 2 h each [47].

4.2. Weighting of shifts from different nuclei

In this section, we move from considering which nuclei to use, to considering how to weight the relative chemical shifts of the different nuclei. The standard practice (assuming one is using ^{15}N HSQC spectra) is to measure chemical shift changes in ^{1}H and ^{15}N shifts in ppm, multiply the ^{15}N shift changes by some scaling factor α, and then calculate the average (or summed) Euclidean distance moved:

$$d = \sqrt{\frac{1}{2}\left[\delta_H^2 + (\alpha \cdot \delta_N^2)\right]} \tag{8}$$

or more generally for several nuclei

$$d = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\alpha_i \delta_i)^2} \tag{9}$$

Is this the best method, and what value of α should one use?

There is no ideal or theoretically justifiable method. As we have seen, the origins of ^{1}H and ^{15}N shifts are different, and some interactions may cause large changes in one but not the other. There is therefore no theory that can provide a weighting. It is perfectly possible to treat ^{1}H and ^{15}N shifts separately [48], but for convenient analysis it is preferable to combine them. Values of α used have ranged from 0.1 [49,50] to 0.45 [51], with many examples in between. To provide some kind of justification for a particular value, several groups have taken the chemical shift range for ^{1}H and ^{15}N, and used this to provide a scale. For example, if the range of backbone ^{15}N shifts is 130.8–108.8 or 22.0 ppm, and the range of ^{1}HN shifts is 10.04–6.92 or 3.12 ppm, then the scale factor should be 3.12/22.0 or approximately 0.14 [52]. However, each protein gives a different value, so there is clearly no single experimentally based scale either. A simple option is to use chemical shift changes in Hz rather than ppm, which for ^{1}H–^{15}N HSQC data corresponds to using $\alpha = 0.1$ [53]. There is however no particular theoretical justification for this treatment either. One could use different values of α for each amino acid. For example, it has been noted that the spread of chemical shifts for glycine is different from other amino acids [54], justifying $\alpha = 0.14$ for most residues but $\alpha = 0.2$ for glycine [52,55].

This question has been examined systematically [36]. The authors considered two other ways of combining shifts: either to use a simple sum of absolute shift changes, $d = |\delta_H| + \alpha|\delta_N|$, or to consider also the sign of the chemical shift change. Their conclusions were that consideration of the sign of the chemical shift change gives a worse result (that is, a poorer definition of which residues form part of the interaction site); and that simple summation and Euclidean distance are roughly equally effective. Similarly, there is no obvious choice as to the best value of α. Our own results [56], using a protein which undergoes very little conformational change on ligand binding, so that almost all the shift changes derive from direct binding interactions, suggest that a value of α of 0.1 is too small (Fig. 6). We therefore suggest that basing the value

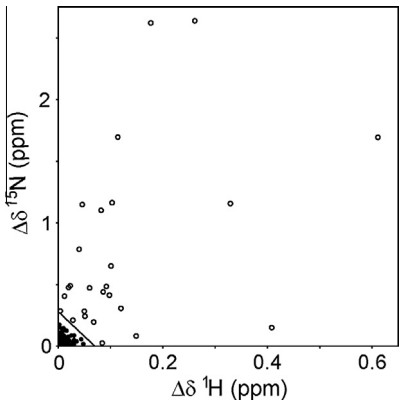

Fig. 6. Chemical shift changes observed in the ^{15}N HSQC spectrum of the family IIb xylan binding module from *Cellulomonas fimi* xylanase D on addition of xylohexaose [56]. Changes for sidechain and backbone HN are shown. Changes adjudged large are shown as open circles. The line dividing large and small shifts is near the bottom left of the plot: in this case it was calculated using a value of $\alpha = 0.25$ (Eq. (8)).

of α on the shift range, and thus the use of $\alpha = 0.14$ for most residues, and $\alpha = 0.2$ for glycine, is a reasonable compromise. We therefore conclude that standard practice (Euclidean weighting, with $\alpha \approx 0.14$) seems to be roughly optimal.

The optimum value of α for carbon resonances is again similar to the value one would obtain based on shift ranges, and is around 0.3 [57,58], being conveniently similar for Cα and C′.

4.3. Threshold value

This section addresses the question of how to decide which shifts are large enough to be considered indicators of the binding site. Standard practice is to calculate the standard deviation σ of the Euclidean chemical shift change, and to identify residues for which the shift change is greater than σ; or sometimes to use 2σ as the cutoff. Clearly, increasing the size of the cutoff reduces the number of residues considered, and in general would be expected to improve the specificity (residues included are genuinely in the binding site) but decrease the sensitivity (some residues in the binding site are missed).

The analysis of Schumann et al. [36] concludes that the best method is to calculate the standard deviation σ of the shift changes; exclude any residues for which the shift change is greater than 3σ (to avoid biasing the distribution by including the small number of residues with very large shift changes); recalculate σ; and iterate these calculations until no further residues are excluded. The corrected standard deviation was expressed as σ_0. The cutoff to be used is then σ_0 itself. This will necessarily be a smaller number than the σ calculated by simply using all residues together: it will therefore generate more false positives, but conversely include a greater number of genuinely interacting residues. The authors also point out that removing buried residues from the analysis and only considering surface-exposed residues helps.

The magnitude of the cutoff is always a balance between specificity and sensitivity, and the question of the best cutoff value really depends on which factor one wants to emphasise. Having said this, the benefit of a defined cutoff such as $1\sigma_0$ is that it removes the temptation to set one's own level so as to produce the desired result.

It is worth noting that the absolute shift measured varies widely for different systems. Partly this is because of the degree of saturation reached. If the ligand binds in a single well-defined site with a single orientation, then there should be several chemical shift changes (d) in excess of 0.2 ppm. In a fairly typical example of protein–protein interactions [59], the largest shift changes were of 0.8 ppm in ^{1}HN, 1.5 ppm in ^{15}N, and 1.0 ppm in ^{13}C (leucine and valine methyls). However, if the ligand binds in a less well defined orientation, then the changes will be smaller. One study of a phenylalanine-containing peptide binding to a protein found the largest change in d to be only 0.08 ppm (although this study did not use a saturating concentration of ligand as binding was very weak) [60]. Nevertheless, this represents genuine binding: it is small because the ligand can adopt a range of bound conformations, and therefore the shift is averaged, as discussed below.

5. Slow and intermediate exchange, and broadening

Slow exchange leads to a change in the appearance of the spectra during a titration (Fig. 2) but does not affect the value of $\Delta\delta_{max}$. Therefore in order to identify ligand binding sites, the same analysis of shift changes can be made as for fast exchange, the main difference being that the assignment of the bound spectrum has to be done all over again, as there is no straightforward way to assign the bound spectrum from the free. This is often not a practical proposition, and a number of authors have used a 'minimum chemical shift procedure' [58,61], in which each free resonance is linked to the signal in the bound spectrum that has moved the least from the position of the free. In this way, a chemical shift change can be ascribed to each resonance. The true shifts may be larger, but they cannot be smaller. Therefore the only error introduced by the procedure is to miss some signals that have in fact moved further than assumed. It is of course possible that the bound signal may have been missed because it is broadened beyond detection. However, this would at least imply a change in the vicinity of the amide group affected, and so is fairly unlikely to lead to false conclusions.

Although, if in slow or intermediate exchange, the ligand binding site can be identified from large changes in a similar way as for fast exchange, the same is not true for the measurement of K_d. As soon as the system starts to deviate from fast exchange, the observed shift is no longer simply the weighted average of free and bound shifts. In the limit of slow exchange, the observed shift does not move at all, it merely disappears. As the exchange regime moves from fast to intermediate, the observed shift starts to move less than predicted based on the simple weighted average. This leads to a very strange-looking relationship between the observed shift and the population-weighted average (Fig. 7a) [62,63]. The consequence is that the protein titration curve has an odd sigmoidal appearance (Fig. 7b). As noted recently, this shape looks deceptively like cooperative binding of two ligand molecules [64]. Significant deviations from the expected population-weighted average like that shown in Fig. 7b would normally be accompanied by severe line broadening and so should be easy to spot.

If the titration is done with observation of the ligand signal rather than the protein signal, then at low concentrations of ligand, and under conditions of intermediate exchange, the chemical shift of the ligand remains close to fully bound for the initial part of the titration: the titration curve for the ligand at low concentration therefore lies above the magenta curve of Fig. 7b. In other words, the affinity of the ligand appears to be greater than it actually is. This can lead to apparent affinities up to 100 times stronger than they really are [65].

When the system is genuinely in slow exchange, the affinity can be determined by plotting the intensity of the signal as a function of ligand concentration and fitting it to Eq. (6), because (as for chemical shift in the fast exchange case) the intensity of the signal is simply proportional to the concentration of free/bound protein. It is preferable to use the intensity of the bound signal rather than the free [66,67].

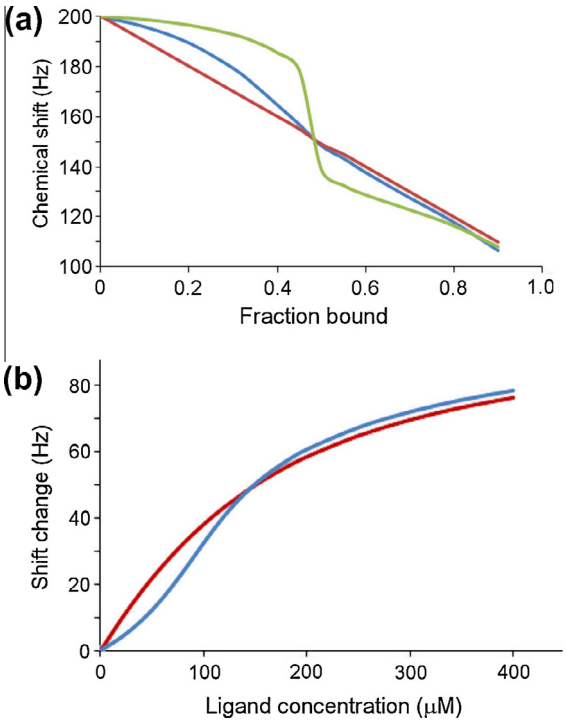

Fig. 7. (a) The apparent chemical shift of a protein signal as ligand is added in a case where the binding equilibrium is in moderately fast exchange. Magenta: simple population-weighted shift (fast exchange). Blue: apparent chemical shift for an off-rate of 444 s^{-1}. Green: apparent chemical shift for an off-rate of 222 s^{-1}. Calculated using the equations of [62] with free shift 200 Hz, bound shift 100 Hz. The green line corresponds to an off-rate where free and bound shifts just coalesce, and the blue line has an off-rate twice as large (corresponding respectively to Fig. 1D and Fig. 1C of [62]). Thus in practice anything as extreme as the green line is unlikely to be observed because the signals will be too broad. However, situations similar to the blue line are feasible. (b) Simulated titration curve for a free shift of 200 Hz, a bound shift of 100 Hz, K_d 100 μM, protein concentration 100 μM, and an off-rate of 444 s^{-1} (blue), compared to the standard titration curve for fast exchange (magenta).

From the discussion in Section 3, the exchange regime is likely to be leaving the fast exchange limit when the fitted K_d becomes stronger than about 10 μM, though if the off-rate is exceptionally fast, it could remain in the fast-exchange limit up to considerably stronger values of K_d. In the (not uncommon) situation where the binding of the ligand is sterically hindered, so that on- and off-rates are slower than those governed by simple diffusion limits, then the signals can be affected by slow exchange at much weaker affinities. Slow exchange usually leads to signal broadening, implying that signal broadening during a titration is an indication of possible slow exchange. However, signals broaden even in the moderately fast-exchange regime [63]: interestingly, the maximum broadening occurs not at the midpoint (when half the protein is bound and half is free) but when 1/3 is bound [63,65].

This argument has two implications: first, that it is not always obvious from the spectra whether the system is in fast exchange or not, and second, that if the system is not in fast exchange, then the K_d values obtained by fitting to the standard Eq. (6) will be inaccurate. In practice, if the fitted K_d is weaker than 10 μM and there is no evidence of extensive line broadening, the fitted values should be close (within 50%) to the true value; but one should always be aware of the possible errors. More accurate determinations of affinities can be obtained by fitting the lineshapes. There are reviews of this field [68–70]. Fitting of lineshapes is by no means easy, especially in view of other complications discussed below [64]. Probably the simplest way to derive reliable values of K_d in intermediate exchange conditions is to use competition experiments, in which high concentrations of a weak ligand are used to displace a stronger ligand: this puts the exchange into the fast regime [29].

It is a very common observation that signals broaden or even disappear during a titration. Broadening is not necessarily a consequence of intermediate-to-slow exchange, so in the rest of this section, we consider some of the circumstances that can give rise to broadening.

Broadening can be caused mainly by one of two effects: slow tumbling or exchange. If the unlabelled ligand is of very high molecular weight, or if addition of ligand leads to oligomerisation or aggregation, then signals will broaden during the titration due to a slowing down of the tumbling rate. In such cases, usually *all* the signals of the labelled protein will broaden (except possibly some from termini or other highly mobile regions), and very often they will never recover, even at high ligand concentrations. This behaviour can be a good indicator of binding (though not necessarily specific binding), especially if peak intensity is restored on addition of a competitive ligand [71]. In the case of exchange broadening, a range of phenomena (to be discussed next) can lead to peak broadening during a titration, and it is important to remember that broadening is not necessarily any indication that the affinity is strong. In particular, the degree of broadening therefore should never be used as a measure of affinity. In such circumstances, it would be advisable to use other biophysical techniques such as light scattering, analytical ultracentrifugation, or fluorescence to provide more information on the system.

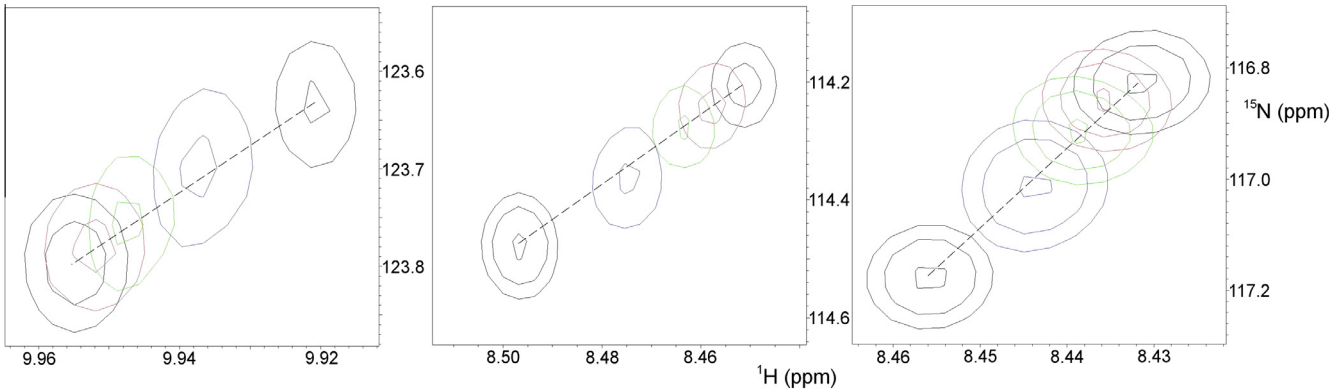

Fig. 8. ^{15}N HSQC signals from a titration between a carbohydrate binding domain and a sugar. In this titration, signals located close to the binding site (not shown here) broaden markedly during the titration and start to sharpen again with a large excess of ligand. The figure shows three cross peaks that are further from the binding site and thus have only small chemical shift changes. The K_d is 65 μM: protein concentration 50 μM, with ligand:protein ratios of 0, 1.5, 4, 8 and 24. One would therefore not expect the ligand binding to be outside the fast exchange regime. In each case, the shifts are not quite in a straight line, indicative of an intermediate formed during the titration, showing a conformational change in the bound protein. Therefore the broadening is very likely due to the slow conformational change rather than slow off-rate.

As discussed above, intermediate or slow exchange between free and bound states leads to broadening. However, many other exchange phenomena can also occur, which may also lead to broadening. For example, there may be a pre-equilibrium conformational change of the protein before binding ligand; or a structural rearrangement of the protein/ligand complex after binding; or protein dimerisation before or after binding. For all of these, even if the ligand-binding step itself is fast, broadening is observed at the specific sites involved, which can look a lot like simple free/bound exchange broadening. The clearest way to tell the difference is that the equilibria described here all involve at least one additional state in addition to the free and bound states. This is likely to lead to 2D spectra during the titration that move in a non-linear manner from free to bound (Fig. 8) [72]. The deviations from linearity will depend on the populations of the intermediate states, and can therefore be small and difficult to see. Further unusual behaviour can also occur, such as linewidths that do not simply broaden and then sharpen again during the titration, or splitting of peaks during a titration. Thus, unusual changes in lineshape during a titration indicate that there is something more complicated happening, but it is not a simple matter to work out the origin of the effect. In particular, it is in general not possible to derive affinities from any simple measure of linewidth. One cannot even be confident that the signals that are most broadened represent the binding site: they could easily arise from allosteric change or dimerisation.

6. Multiple binding modes

6.1. Ligand binding or conformational change?

When a ligand binds to a protein, it can produce chemical shift changes either by direct interactions at the binding site, or by inducing a conformational change in the protein, which we can consider as some kind of allosteric change. In either case, the chemical shift change is similarly dependent on the fraction of protein bound, implying that the two types of interaction cannot be separated on the basis of shift changes (see Section 6.3).

Provided that a ligand binds in a single location, it is likely that any shift changes far from that location must be due to conformational effects. This would include of course chemical shift changes

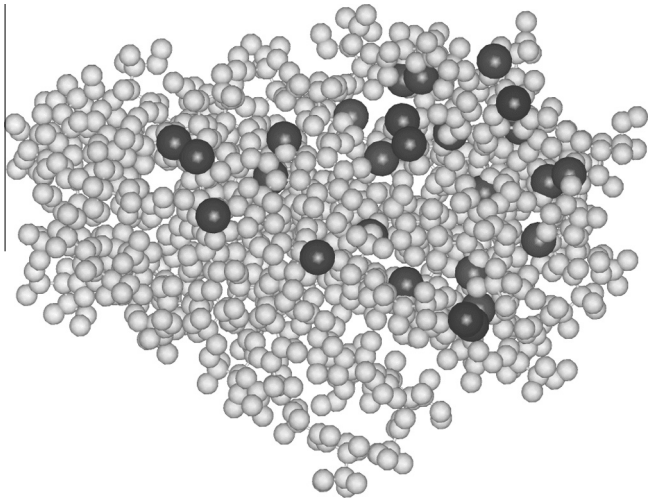

Fig. 9. Location of the large chemical shift changes shown in Fig. 6. Only those atoms having large shifts in Fig. 6 are identified, as large black spheres on the amide proton. All other atoms are depicted with 0.6× van der Waals radius, to facilitate observation of buried atoms. The viewing angle is approximately directly above the binding site, which in this case is an extended site running from top centre to right centre of the diagram.

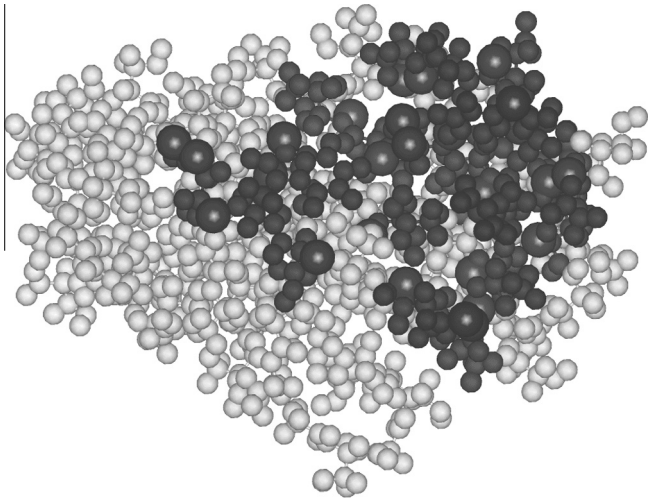

Fig. 10. Location of the large chemical shift changes shown in Fig. 6. The difference from Fig. 9 is that in this figure all atoms of any amino acid whose backbone or sidechain amide group has a large shift change are picked out in black. Even in cases where the amide proton was buried in Fig. 9, some atoms of the corresponding residue are surface-exposed and consequently show up in this representation.

inside the protein (that is, in residues with zero surface area), which are almost invariably a result of conformational change. Indeed, it has been suggested that chemical shift changes that do not map to a single area can be a good indicator of conformational change [1]. A good example is provided by carbohydrate binding proteins, since they usually have a well-defined binding site, they are relatively rigid, and the ligand has no aromatic rings, and so would not be expected to have any direct effects on chemical shift at interior residues. The chemical shift changes seen on binding of xylan to a xylan binding module are shown in Fig. 6. Of the large chemical shift changes in Fig. 6, almost all are from HN groups very close to or at the surface (Fig. 9). Interestingly, when one adopts the normal practice and colours in all the atoms from any amino acid that has large HN shift changes, virtually all the changes are from residues on the surface: most of the shifted HN that are below the surface have exposed sidechains (Fig. 10). Thus in this protein, essentially all the large effects seen, including all the large effects seen on the surface, can be ascribed to residues in direct contact with the ligand. Because the ligand has no aromatic rings, the largest effects (of up to 0.6 ppm in ^{1}H and 2.5 ppm in ^{15}N) are direct hydrogen bonds to the amide group. This is an unusually clean example, because conformational change within the protein is very limited, and the direct ligand-induced chemical shift changes are so large. Nevertheless it is a surprisingly common observation that the largest shift changes on addition of a ligand are local and due to direct binding: in other words that the chemical shift changes arising from large-scale conformational changes on binding often seem to be small.

There are of course many counter-examples of large shift changes seen on ligand binding that can only be due to conformational rearrangement within the protein. An early example comes from work on dihydrofolate reductase (DHFR) [73]. NADPH, a cofactor in the reaction of DHFR, was added to a complex of DHFR with its inhibitor methotrexate. Large chemical shift changes were seen near the NADPH binding site, but a number of additional chemical shift changes were seen extending well away from the binding site, which were suggested to be due to a rigid-body motion of a helix abutting the NADPH binding site. This is an example of allosteric changes resulting from an enzyme binding to its specific substrate/cofactor, and similar large-scale effects have been seen in protein-DNA interactions [74], where many of the large chemical shift changes are a long way from the binding site.

One can argue that the proteins that are more likely to have large chemical shift changes arising from conformational change are enzymes and large multi-domain systems. It is therefore perfectly possible that such effects are still rare in the literature because most NMR studies are still on systems that are too small to have big conformational change. In a genuinely allosteric system, a small fraction of the shift changes were over 0.5 ppm in ^{1}H [48]. There is thus no law that says that allosterically derived shift changes must be small. We therefore have to conclude on the same point that we made at the start of this section, that there is no way from the data alone to distinguish between shift changes that are a direct result of binding, and shift changes that result from an allosteric change in the protein structure.

6.2. Does the ligand bind at several sites?

Multiple binding modes can be readily detected by conducting HSQC titrations and plotting the results as ^{15}N shift vs. ^{1}H shift (for example, by superimposing the HSQC spectra). If there is only a single binding mode, then the titrations should produce a straight line in such a plot, whereas multiple binding will in general produce curved plots, because the secondary interactions will almost always have different effects on the chemical shifts than the primary interaction [75] (Fig. 11).

The above result raises an interesting point. On examination of Fig. 11, it can be seen that the NH signals move in a straight line (indicating binding at a single site) up to roughly 0.5 equivalents of ligand bound (black to cyan). After this point, secondary binding effects start to be seen, due to weaker binding at a second site. In this case [26], the two dissociation constants were fitted to be 24 and 164 μM. From a consideration of binding equilibria, this is what we would expect: if the primary binding event has an affinity at least 10 times greater than any secondary event, then added ligand will initially bind almost exclusively to the primary site, but once most of the primary binding sites are full, then the ligand will start to bind in other places. This implies that the simplest way to reduce the effect of secondary binding is to titrate in ligand only as far as approximately 0.5 equivalents. For typical affinities, this will reduce the fraction of ligand bound at secondary sites to less than 10% (Fig. 12). If the aim is to determine a ligand binding site, and not to measure K_d, there is no need to titrate the ligand in to obtain complete saturation of the binding site!

Nonlinear plots do not necessarily imply multiple interactions: they can also be caused by something more complicated than a two-state binding interaction. A conformational change, either as

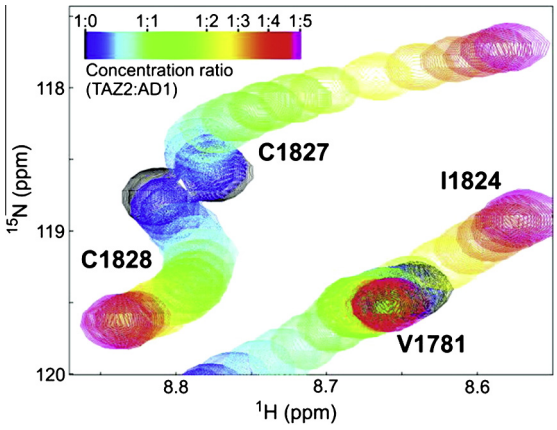

Fig. 11. A small region of the ^{15}N HSQC spectrum of ^{15}N-labelled TAZ2 titrated with unlabelled p53 AD1 domain (residues 13–37). The colour of the cross-peaks changes from black (free protein) to magenta (1:5 ratio). Adapted with permission from M. Arai et al. J. Am. Chem. Soc. 2012 **134**:3792–3803. Copyright (2012) American Chemical Society.

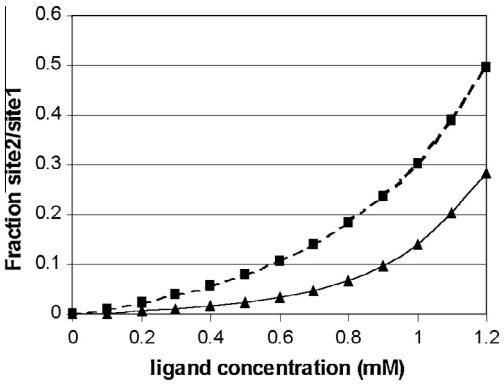

Fig. 12. Fraction of ligand bound at a weaker secondary site, compared to that bound at the primary site. Calculations were done for a protein concentration of 1 mM, and a dissociation constant at the primary binding site of 5 μM. The dissociation constant at the secondary binding site was 50 μM (dashed line) or 200 μM (solid line). For the 50 μM secondary binding, a 1:1 ligand:protein ratio already has 30% of the ligand binding at the secondary site – the concentration of [Protein.Site 1] is 610 μM, while the concentration of [Protein.Site 2] is 190 μM. However at a 0.5:1 ligand:protein ratio, only 8% of the ligand is bound at the secondary site (bound concentrations of 430 μM and 35 μM respectively). With a secondary binding dissociation constant of 200 μM, a 0.5:1 ligand:protein ratio produces only 2% of ligand bound at the secondary site.

$$P + L \leftrightarrow P^{*} + L \leftrightarrow PL,$$

or

$$P + L \leftrightarrow PL \leftrightarrow P^{*}\ L$$

will also give rise to curved titrations (and very often broadened signals), and can produce some remarkably messy and intractable results.

Multiple binding turns out not to be as problematic as might be expected. The most common type of multiple binding is strong binding at a specific site, accompanied by weak non-specific binding, often at multiple sites. Non-specific binding usually causes little conformational change in the protein, and the interactions are not well localised. Hence the chemical shift changes caused are generally small. For example, a study of a phenylalanine-rich peptide binding to its specific target protein reported specific chemical shift changes at the binding site of only 0.06 ppm [60]: small though these were, they were larger than almost all non-specific effects.

Protein–protein interactions are also subject to multiple binding modes, and here again the non-specific chemical shift changes are small, even for reasonably localised binding [76]. A good example can be found in interactions between cytochrome *c* and partner proteins [77]. Cytochrome *c* has a positively charged patch near the surface-exposed part of the haem group, and interacts with partner proteins in a relatively undefined geometry dominated by charge–charge interactions. In the complex with its physiological partner cytochrome *c* peroxidase, the largest $\Delta\delta_{av}$ values are only 0.1 ppm or less (with the exception of a single value of 0.3 ppm). In the complex with adrenodoxin, a ferredoxin that can transfer electrons to cytochrome *c* but is not a physiological ligand, the largest $\Delta\delta_{av}$ is 0.05 ppm. In both these cases, the largest chemical shift effects are consistent with a single binding surface close to the haem group. By contrast, in the complex between cytochrome b_5 and myoglobin, which is weak and very dynamic, the largest $\Delta\delta_{av}$ are less than 0.02 ppm, and no specific binding surface is evident.

In all these cases, titrations with substoichiometric amounts of ligand are usually enough to distinguish specific from nonspecific binding [78]. A similar strategy was used for the binding of cytidyl triphosphate (CTP) to the dimeric enzyme CTP:glycerol-3-phosphate cytidyltransferase (GCT) [48], which binds one CTP with a

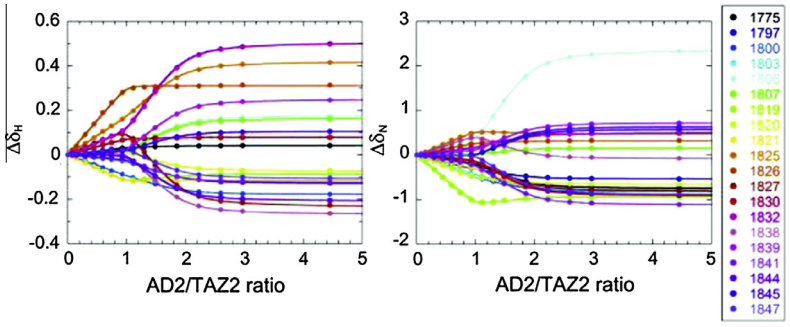

Fig. 13. ^{1}H and ^{15}N titration curves for ^{15}N-labelled TAZ2 titrated with unlabelled p53 AD2 domain (residues 38–61). Some signals show tight binding, saturating at a 1:1 ratio, indicated by the sharp bends at this point (compare Fig. 3, red curve). Others show a weaker binding, the curves running from 1:1 and bending more gradually at a 1:2 ratio. Adapted with permission from M. Arai et al. J. Am. Chem. Soc. 2012 **134**:3792–3803. Copyright (2012) American Chemical Society.

dissociation constant K_d of $\leqslant$5 μM, and a second with a much weaker affinity of 400–800 μM. On addition of one equivalent of CTP, chemical shift changes can be mapped to a contiguous region of surface, consistent with localised binding of one molecule of CTP to the dimer. On addition of further CTP, almost all of the shift changes are due to the weaker allosteric binding.

6.3. Disentangling multiple binding

If a ligand binds at two sites with different affinities, then a graph of shift against added ligand will not be a simple saturation curve but will have a more complicated shape. Curve fitting against a model can be used to extract not only the affinities for the two sites, but also the shift differences for each site, and therefore the locations of the two sites. An elegant example comes from Arai et al. [26]. Individual titration plots (Fig. 11), and a comparison of saturation curves (Fig. 13), clearly show multiple binding. Their analysis required two key elements. The first is that saturation curves from all titrating sites were fitted simultaneously. This is important, because it produces a much better and more reliable fit than fitting each curve independently (see also [79]). Second, they used a statistical method called singular value decomposition (SVD) to analyse the data. SVD is a multivariate technique (similar to the more familiar Principal Component Analysis or PCA) which achieves two things: it can estimate the number of meaningful components contributing to the shift change, and it obtains the best values for the contributions to these components. The number of meaningful components limits the complexity of the binding model. A simple two-state binding has only two components, free and bound. Independent binding at two sites has three components (free, $Bound_1$ and $Bound_2$), while binding at two sites that can interact has four components contributing to the shift changes (binding at both sites gives rise to a chemical shift change that is not simply a sum of the two separate effects). Thus SVD provides a statistically validated assessment of the least complicated model that satisfies the data. The importance of the second feature, the best values for the contributions, is that these are consensus (i.e., much less noisy) estimates of the shift changes for each binding event. Armed with the outcome of the SVD analysis, one can then fit the (less noisy) titration curves to obtain affinities. This permits a more accurate fitting of the affinity; alternatively, it allows one to estimate values for extremely weak or extremely strong affinities, where one is working at the edge of the range of reliable data. In this way, they were able to measure an affinity of 32 nM for the stronger binding event. The paper [26] devotes considerable effort to demonstrating that such a strong affinity can be measured reliably, not least because in this case it is effectively a competition experiment [29].

Other groups have used other multivariate techniques. Konuma et al. [80] used Principal Component Analysis (PCA) to analyse the binding of 1-anilinonaphthalene-8-sulphonate (ANS) to β-lactoglobulin, and were able to obtain affinities and sites for two binding events. For one of these, two molecules of ANS bind in a concerted manner. Selvaratnam et al. [81] also used SVD, to analyse the binding of a series of related ligands to a multidomain protein named EPAC. Having obtained the consensus shift maps, as described above, they then applied a covariance analysis in conjunction with a clustering method, to analyse patterns of shift changes from different ligands. In this way they separated out chemical shift effects due just to binding (which vary from ligand to ligand) from effects due to allosteric change (which are the same for all ligands), along the way identifying chemical shift changes due to allostery, of which few were larger than 0.2 ppm (calculated as the sum $d = |\delta_H| + 0.2|\delta_N|$).

7. Use of CSPs to guide docking

7.1. Docking using CSPs to define the binding site

There are many programs that have been developed for docking a ligand onto a protein. It is generally agreed that docking calculations are not yet sufficiently fast, and do not reproduce energies sufficiently well, for docking algorithms to work well purely based on the separated structures: they generally need additional information to guide them [82]. Calculations have shown that as few as three distance restraints can be enough to guide docking successfully [83]. Similarly, CSPs alone are not enough to dock two molecules together [84]. The simplest way in which shift changes can be used is to say that any nucleus with a shift change larger than a defined cutoff (as defined in Section 4.3 for example) is in the interface. The most popular program using this method is HADDOCK [85]. In this method, *active* residues are ones that are found in the interface as defined above, and in addition are on the surface. HADDOCK also defines *passive* residues, which are surface residues that either have a smaller shift change than the active residues or are close to the active residues. It then defines Ambiguous Intermolecular Restraints or AIRs, which are restraints between any atom on an active residue on one partner and any atom on all active plus passive residues on the other. They are ambiguous in the sense that for each active residue, they restrain one atom on that residue to be close to one atom on *any* active or passive residue on the target. (Strictly, the restraint is that the sum of $(r^{-6})^{-1/6}$ should be less than 3 Å. This is the same function as used for $1/r^{-6}$-summed NOE restraints, and means that the restraint can be satisfied either by a single short distance or alternatively by several fairly short ones.) These restraints are expressed as energies, and are often enough to direct a docking algorithm to the correct solution. One can of course add extra restraints from other experiments, such as residual dipolar couplings or paramagnetic relaxation. These two experiments in particular provide complementary

information to chemical shifts (orientation information and longer range distances respectively), and often help to improve the docking markedly. One can also add information on which residues are in the interface derived from analysing mutational data.

Similar programs have been described by others but have not achieved the popularity of HADDOCK. The program BiGGER [86] carries out a docking, and then uses chemical shift changes as a filter: any nucleus with a shift change larger than the cutoff must be within 5 Å of an atom on the partner. If not, this solution is rejected [87,88]. AutoDockFilter [89] works in a similar way. Clore and Schwieters described a very similar method to HADDOCK [90]. The program LIGDOCK [84] also uses AIRs to provide the experimental restraints, and combines them with van der Waals energies from the docked models to identify the best solutions.

For AutoDockFilter [89], a better result was obtained by doing a manual filter of the residues with shifts greater than the cutoff, by deleting residues that were spatially isolated or a long way from the major cluster of residues. Bonvin and colleagues have refined this to an automated and unbiased method called SIMPLEX, which weights the selection of a residue as perturbed or not by the weights of its neighbours [91]. This method thus downgrades the effect of outlying single residues with large shift changes, often observed as a result of conformational change in the protein, which it is claimed significantly improves the search efficiency.

7.2. Definition of ligand orientation from differential shifts

As discussed so far, the magnitude of chemical shift changes has been used just to impose a cutoff on which changes are significant and which are not. The rest of this section looks at methods that have attempted a more quantitative approach, using chemical shifts to derive structural information more directly. In particular, a more quantitative use of shifts might allow one to make deductions about the orientation ('pose') of the ligand.

An early attempt to do this came from the drug discovery group at Abbott Labs [92]. They pointed out that if one has a series of closely related ligands that all bind to a target protein in the same way, then they should produce different chemical shift changes in the target, the differences being localised to the regions of the ligands that differ. One can thus pinpoint where different parts of the ligand are binding to the target. The method was illustrated by binding of ascomycin and derivatives to FK506 binding protein. The authors note that this idea can be readily extended to protein/protein complexes: comparison of a protein/protein complex with a complex in which one residue in one of the proteins has been mutated should give good information on the near neighbours of the mutated residue. When this method works, it is quick and effective, allowing it to be applied in a high-throughput manner [93].

The problem comes when different ligands do not bind in the same way: unfortunately, this is a common problem, and one that bedevils structure/activity relationships. Even small changes to the ligand can often change its orientation. Riedinger et al. [94] had this problem with the binding of a series of related ligands, and plotted ^{1}H and ^{15}N chemical shift changes residue by residue, for the whole series of ligands. They found that some ligands clustered together in their effects, and concluded that each cluster binds in the same orientation. By mapping these clusters onto the binding site in the protein, they were able to propose how the different ligands bind, and rationalise why.

7.3. More quantitative approaches

Over the last 10–15 years there has been a major increase in our ability to calculate chemical shifts in proteins: that is, to go from the structure of a protein to its chemical shifts. We are now beginning to work out ways of going the other way, from chemical shifts to structure. Two significant papers [95,96] have shown that it is possible to use chemical shifts to score potential structures generated by computer prediction, and therefore effectively to begin to use shifts to calculate structures. If we can do this for proteins,

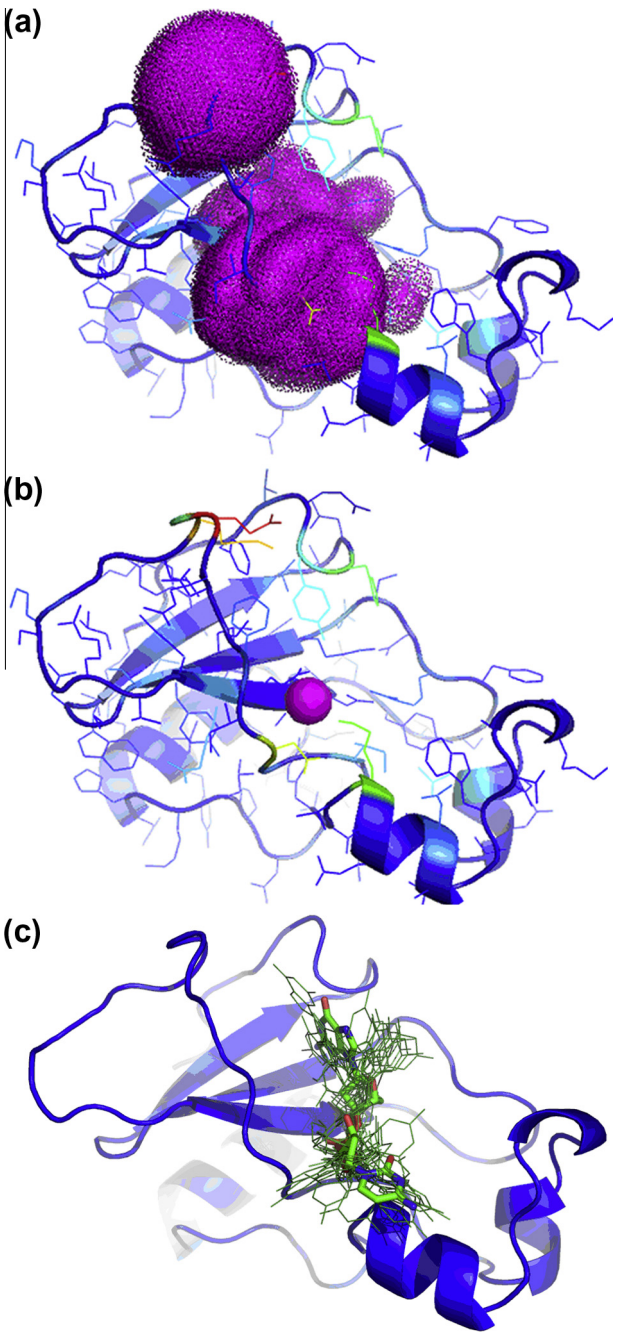

Fig. 14. Docking of d(GC) onto the ribonuclease enzyme barnase. Ligand-induced chemical shift changes are shown in panel b, colour code red > yellow > green > cyan. The largest single change is to Glu60, whose sidechain makes two hydrogen bonds to the guanosine: this chemical shift change is thus not a ring current effect at all. The *j*-surface constructed from these changes is shown in panel a. There is a large sphere at the top of the figure due to Glu60, but the densest region of the surface (formed by the overlap of spheres from several residues) is in the middle. The most likely location for the ligand is at the centre of this densest region, and is shown by the magenta sphere in panel b. This location was then used to dock d(GC) onto the protein, using the detailed shift changes to optimise the position and orientation of the ligand (c). The structure shown using sticks in panel (c) is a typical structure close to the average of the ensemble. Adapted with permission from M. Cioffi et al. J. Biomol. NMR 2009 **43**:11–19. Copyright (2009) Springer Science + Business Media B. V.

why not do something similar for what should be a simpler problem, namely the docking of a ligand into a rigid or semirigid protein?

A simple application of such a more quantitative use of shifts was proposed by McCoy and Wyss [34]. They pointed out that most protein ligands (roughly 95% of leads in one major drug design database [4]) contain aromatic rings; and that the ring current shifts generated by these rings should be the largest single effect on the shift changes induced by ligand binding. Interestingly, in their first implementation, the authors used ^{1}Hα shift changes rather than the more conventional ^{1}HN changes, because ^{1}Hα shift changes are more likely to be due exclusively to ring current shifts, whereas ^{1}HN changes can be due to hydrogen bond and solvent rearrangement, and are in general less well predicted than ^{1}Hα shifts [34]. Although this is generally true, it is not as easy to measure ^{1}Hα shifts, as discussed above.

McCoy and Wyss [97] noted that the aromatic ring current shift can be approximated by the equation

$$\Delta\delta \approx (B/r^3)(1 - 3\cos^2\theta) \quad (10)$$

where B is a proportionality constant, r is the distance to the centre of the ring, and θ is the angle between the proton-ring vector and the ring normal. This can be simplified even further, such that the maximum possible chemical shift change due to an aromatic ring (when $\theta = 0$, that is when the proton is directly under the ring) is proportional to r^{-3}. Thus, for any given shift change, one could draw a sphere around the proton, within which the ring centre should lie. For example, for a benzene ring the sphere has a 4.8 Å radius for a shift change of +0.2 ppm, and a 6.0 Å radius for a shift change of −0.2 ppm.

The method then consists of taking each proton with a binding-induced shift, and drawing an appropriately sized sphere around it. If there is a single location for the ligand, then all the spheres from all the shifted protons should intersect, the common volume defining the allowed ring locations. In practice, this is done by calculating dot densities within each sphere and then calculating the highest resultant dot density on the protein surface. The calculation can be improved by removing dots located within the protein, which are not possible locations for the ligand. The resultant distribution is described as a *j*-surface (Fig. 14). As one might expect, best results are obtained when there are many experimental data points available. In this case, calculations were made using ^{1}HN shifts. Interestingly, the standard chemical shift mapping procedure provides a rather imprecise location for the ligand, mainly because the single largest shift change (large sphere at the top of Fig. 14a and red sidechain at the top of Fig. 14b) is a hydrogen bonding interaction from a sidechain, whereas the more quantitative method identifies a very clear location, which is convincingly placed just off the van der Waals surface of the protein. The fact that a single well-defined site is identified suggests strongly that the protein structure is not perturbed by ligand binding. The calculation is extremely fast, and is a very useful way of checking for simple single-site binding, as a precursor to a more complete quantitative search, as described below.

The *j*-surface is a good way of defining the binding site for an aromatic ligand, but is poor at generating orientations. Cioffi et al. therefore combined the *j*-surface with a more detailed calculation of ligand-induced chemical shift changes [98–100]. The *j*-surface was used to locate the binding site; this location was then fed into a docking program, which produced 100–1000 ligand/protein 'poses'. For each of these poses, the ligand structure was used to calculate the expected ^{1}H shift changes in the protein, as described in Section 2.2; and the position of the ligand was varied to optimise the match between calculated and experimental shift changes [98]. The resultant poses (location and orientation) match structures derived from crystallography well, and in some cases are actually more accurate [100]. As one might expect, the method is less good if the protein structure alters on binding ligand [99]. A similar method has been described using quantum mechanical (MNDO) shift calculations [101,102]. A more recent attempt along these lines determines the match between calculation and experiment using a correlation coefficient rather than an RMSD; this avoids the need to estimate the degree of saturation of the protein binding site, and reduces the impact of outliers, for example caused by allosteric change [103]. It also applies the shift calculation at the stage of initial docking, thus giving a high hit rate of useful poses. A recent version of HADDOCK also uses this approach [37]: it uses shifts in the standard qualitative way first, to locate the binding site, but then uses a more complete calculation (using the empirical calculation carried out by ShiftX [104]) to optimise the poses. The authors note that this two-stage approach works much better than a straightforward chemical shift-based approach. They also note that the best shifts to use for the detailed calculation are ^{1}Hα, ^{13}Cα and ^{15}N, but not ^{1}H$_N$.

None of these methods works well if the protein structure changes substantially on binding to ligand. In this case, it has been shown that docking based only on the shift changes can work [105]. Crystal structures of two proteins were used to dock them together, using a procedure that allows flexible parts of the proteins to move during the docking. For each complex structure generated, chemical shift calculations (again based on ShiftX [104]) were carried out and used to score the structure. This chemical shift-based 'biasing' of the energies calculated during the docking is enough to produce the correctly docked structures as the lowest energies. The authors comment that although the chemical shift restraints do not provide detailed information on the complex, the simultaneous use of a large number of shifts is sufficient to produce a well-docked structure. They have even suggested that chemical shifts can be used to generate the structure of a complex, even if the structures of the free proteins are not known [106].

7.4. The future

The examples presented here show that a rigorous statistical treatment of CSPs not only gives better estimates of binding affinities, but also has the potential to separate strong binding from weak secondary binding, and possibly even to separate shifts caused by binding at the binding site from allosteric effects. This would be a major improvement in the reliability of CSPs for mapping ligand binding. The use of methyl labelling, particularly in larger proteins, is likely to grow considerably, and probably gives a much more reliable indicator of binding because the chemical shifts of ^{13}C and ^{1}H$_C$ are likely to be more directly affected than those of ^{15}N by ligand binding rather than allosteric effects. Methods for the calculation of chemical shifts in proteins continue to improve, and are likely to lead to much more accurate determination of the binding orientation of ligands.

Acknowledgements

I thank Mark McCoy (Schering-Plough, NJ) and the editors for helpful comments.

References

[1] E.R.P. Zuiderweg, Mapping protein-protein interactions in solution by NMR spectroscopy, Biochemistry 41 (2002) 1–7.
[2] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, Proc. Natl. Acad. Sci. USA 94 (1997) 12366–12371.

[3] A.P. Golovanov, R.T. Blankley, J.M. Avis, W. Bermel, Isotopically discriminated NMR spectroscopy: a tool for investigating complex protein interactions *in vitro*, J. Am. Chem. Soc. 129 (2007) 6528–6535.

[4] M.A. McCoy, D.F. Wyss, Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations, J. Am. Chem. Soc. 124 (2002) 11758–11763.

[5] S.B. Shuker, P.J. Hajduk, R.P. Meadows, S.W. Fesik, Discovering high-affinity ligands for proteins: SAR by NMR, Science 274 (1996) 1531–1534.

[6] M.P. Williamson, C.J. Craven, Automated protein structure calculation from NMR data, J. Biomol. NMR 43 (2009) 131–143.

[7] S.G. Zech, E. Olejniczak, P. Hajduk, J. Mack, A.E. McDermot, Characterization of protein-ligand interactions by high-resolution solid-state NMR spectroscopy, J. Am. Chem. Soc. 126 (2004) 13948–13953.

[8] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A.J. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian 98, Pittsburgh, PA, 1998.

[9] K.J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, M. Vendruscolo, Fast and accurate predictions of protein NMR chemical shifts from interatomic distances, J. Am. Chem. Soc. 131 (2009) 13894–13895.

[10] P. Robustelli, K.A. Stafford, A.G. Palmer, Interpreting protein structural dynamics from NMR chemical shifts, J. Am. Chem. Soc. 134 (2012) 6365–6374.

[11] X.P. Xu, D.A. Case, Probing multiple effects on ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts in peptides using density functional theory, Biopolymers 65 (2002) 408–423.

[12] A.C. de Dios, J.G. Pearson, E. Oldfield, Secondary and tertiary structural effects on protein NMR chemical shifts: an *ab initio* approach, Science 260 (1993) 1491–1496.

[13] D. Sitkoff, D.A. Case, Density functional calculations of proton chemical shifts in model peptides, J. Am. Chem. Soc. 119 (1997) 12262–12273.

[14] M.P. Williamson, T. Asakura, Empirical comparisons of models for chemical shift calculation in proteins, J. Magn. Reson. Ser. B 101 (1993) 63–71.

[15] M.P. Williamson, T. Asakura, in: D.G. Reid (Ed.), Methods in Molecular Biology, Humana Press, New Jersey, 1997, pp. 53–69.

[16] M. Iwadate, T. Asakura, M.P. Williamson, Cα and Cβ carbon-13 chemical shifts in proteins from an empirical database, J. Biomol. NMR 13 (1999) 199–211.

[17] A. Bundi, K. Wüthrich, ^1H NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH, Biopolymers 18 (1979) 285–297.

[18] G. Merutka, H.J. Dyson, P.E. Wright, Random coil ^1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG, J. Biomol. NMR 5 (1995) 14–24.

[19] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, ^1H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. 1. Investigations of nearest-neighbor effects, J. Biomol. NMR 5 (1995) 67–81.

[20] S. Schwarzinger, G.J.A. Kroon, T.R. Foss, J. Chung, P.E. Wright, H.J. Dyson, Sequence-dependent correction of random coil NMR chemical shifts, J. Am. Chem. Soc. 123 (2001) 2970–2978.

[21] A. De Simone, A. Cavalli, S.-T.D. Hsu, W. Vranken, M. Vendruscolo, Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins, J. Am. Chem. Soc. 131 (2009) 16332–16333.

[22] K. Tamiola, B. Acar, F.A.A. Mulder, Sequence-specific random coil chemical shifts of intrinsically disordered proteins, J. Am. Chem. Soc. 132 (2010) 18000–18003.

[23] M. Kjaergaard, F.M. Poulsen, Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution, J. Biomol. NMR 50 (2011) 157–165.

[24] B. Han, Y. Liu, S.W. Ginzinger, D.S. Wishart, SHIFTX2: significantly improved protein chemical shift prediction, J. Biomol. NMR 50 (2011) 43–57.

[25] A.R. Fersht, Structure and Mechanism in Protein Science, W. H. Freeman, New York, 1999.

[26] M. Arai, J.C. Ferreon, P.E. Wright, Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP, J. Am. Chem. Soc. 134 (2012) 3792–3803.

[27] A. Carrington, A. McLachlan, Introduction to Magnetic Resonance with Applications to Chemistry and Chemical Physics, Harper & Row, New York, 1967.

[28] J. Granot, Determination of dissociation constants of 1:1 complexes from NMR data: optimization of the experimental setup by statistical analysis of simulated experiments, J. Magn. Reson. 55 (1983) 216–224.

[29] L. Fielding, NMR methods for the determination of protein-ligand dissociation constants, Prog. Nucl. Magn. Reson. Spectrosc. 51 (2007) 219–242.

[30] C.J. Markin, L. Spyracopoulos, Increased precision for analysis of protein-ligand dissociation constants determined from chemical shift titrations, J. Biomol. NMR 53 (2012) 125–138.

[31] J. Vaynberg, J. Qin, Weak protein-protein interactions as probed by NMR spectroscopy, Trends Biotechnol. 24 (2006) 22–27.

[32] N.J. Baxter, T.H. Lilley, E. Haslam, M.P. Williamson, Multiple interactions between polyphenols and a salivary proline-rich protein repeat result in complexation and precipitation, Biochemistry 36 (1997) 5566–5577.

[33] J.H. Tomlinson, V.L. Green, P.J. Baker, M.P. Williamson, Structural origins of pH-dependent chemical shifts in the B1 domain of protein G, Proteins, Proteins: Struct. Funct. Bioinf. 78 (2010) 3000–3016.

[34] M.A. McCoy, D.F. Wyss, Alignment of weakly interacting molecules to protein surfaces using simulations of chemical shift perturbations, J. Biomol. NMR 18 (2000) 189–198.

[35] R.J. Carbajo, F.A. Kellas, M.J. Runswick, M.G. Montgomery, J.E. Walker, D. Neuhaus, Structure of the F_1-binding domain of the stator of bovine F_1F_o-ATPase and how it binds an α-subunit, J. Mol. Biol. 351 (2005) 824–838.

[36] F.H. Schumann, H. Riepl, T. Maurer, W. Gronwald, K.-P. Neidig, H.R. Kalbitzer, Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions, J. Biomol. NMR 39 (2007) 275–289.

[37] D. Stratmann, R. Boelens, A.M.J.J. Bonvin, Quantitative use of chemical shifts for the modeling of protein complexes, Proteins: Struct. Funct. Bioinf. 79 (2011) 2662–2670.

[38] P.J. Hajduk, D.J. Augeri, J. Mack, R. Mendoza, J.G. Yang, S.F. Betz, S.W. Fesik, NMR-based screening of proteins containing ^{13}C-labeled methyl groups, J. Am. Chem. Soc. 122 (2000) 7898–7904.

[39] N.K. Goto, K.H. Gardner, G.A. Mueller, R.C. Willis, L.E. Kay, A robust and cost-effective method for the production of Val, Leu, Ile$^{\delta 1}$ methyl-protonated ^{15}N-, ^{13}C-, ^{2}H-labeled proteins, J. Biomol. NMR 13 (1999) 369–374.

[40] I. Gelis, A.M.J.J. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, A. Economou, C.G. Kalodimos, Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR, Cell 131 (2007) 756–769.

[41] D.J. Hamel, F.W. Dahlquist, The contact interface of a 120 kD CheA-CheW complex by methyl TROSY interaction spectroscopy, J. Am. Chem. Soc. 127 (2005) 9676–9677.

[42] S. Wiesner, A.A. Ogunjimi, H.-R. Wang, D. Rotin, F. Sicheri, J.L. Wrana, J.D. Forman-Kay, Autoinhibition of the HECT-Type ubiquitin ligase smurf2 through its C2 domain, Cell 130 (2007) 651–662.

[43] J.R. Huth, C. Park, A.M. Petros, A.R. Kunzer, M.D. Wendt, X. Wang, C.L. Lynch, J.C. Mack, K.M. Swift, R.A. Judge, J. Chen, P.L. Richardson, S. Jin, S.K. Tahir, E.D. Matayoshi, S.A. Dorwin, U.S. Ladror, J.M. Severin, K.A. Walter, D.M. Bartley, S.W. Fesik, S.W. Elmore, P.J. Hajduk, Discovery and design of novel HSP90 inhibitors using multiple fragment-based design strategies, Chem. Biol. Drug Des. 70 (2007) 1–12.

[44] T. Yuan, S. Tencza, T.A. Meitzner, R.C. Montelaro, H.J. Vogel, Calmodulin binding properties of peptide analogues and fragments of the calmodulin-binding domain of simian immunodeficiency virus transmembrane glycoprotein 41, Biopolymers 58 (2001) 50–62.

[45] B.G. Szczepankiewicz, G. Liu, P.J. Hajduk, C. Abad-Zapatero, Z.H. Pei, Z.L. Xin, T.H. Lubben, J.M. Trevillyan, M.A. Stashko, S.J. Ballaron, H. Liang, F. Huang, C.W. Hutchins, S.W. Fesik, M.R. Jirousek, Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy, J. Am. Chem. Soc. 125 (2003) 4087–4096.

[46] M.C. Stoffregen, M.M. Schwer, F.A. Renschler, S. Wiesner, Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces, Structure 20 (2012) 573–581.

[47] M. Quinternet, J.-P. Starck, M.-A. Delsuc, B. Kieffer, Unraveling complex small-molecule binding mechanisms by using simple NMR spectroscopy, Chem. Eur. J. 18 (2012) 3969–3974.

[48] S.Y. Stevens, S. Sanker, C. Kent, E.R.P. Zuiderweg, Delineation of the allosteric mechanism of a cytidylyltransferase exhibiting negative cooperativity, Nature Struct. Biol. 8 (2001) 947–952.

[49] C.H. Hardman, R.W. Broadhurst, A.R.C. Raine, K.D. Grasser, J.O. Thomas, E.D. Laue, Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy, Biochemistry 34 (1995) 16596–16607.

[50] F.H.T. Allain, Y.M. Yen, J.E. Masse, P. Schultze, T. Dieckmann, R.C. Johnson, J. Feigon, Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding, EMBO J. 18 (1999) 2563–2579.

[51] B.E. Coggins, X.C. Li, A.L. McClerren, O. Hindsgaul, C.R.H. Raetz, P. Zhou, Structure of the LpxC deacetylase with a bound substrate-analog inhibitor, Nature Struct. Biol. 10 (2003) 645–651.

[52] R.A. Williamson, M.D. Carr, T.A. Frenkiel, J. Feeney, R.B. Freedman, Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation, Biochemistry 36 (1997) 13882–13889.

[53] T. Terada, Y. Ito, M. Shirouzu, M. Tateno, K. Hashimoto, T. Kigawa, T. Ebisuzaki, K. Takio, T. Shibata, S. Yokoyama, B.O. Smith, E.D. Laue, J.A. Cooper, Nuclear magnetic resonance and molecular dynamics studies on the interactions of the Ras-binding domain of Raf-1 with wild-type and mutant Ras proteins, J. Mol. Biol. 286 (1999) 219–232.

[54] D.S. Wishart, B.D. Sykes, F.M. Richards, Relationship between nuclear magnetic resonance chemical shift and protein secondary structure, J. Mol. Biol. 222 (1991) 311–333.

[55] M.D. Urbaniak, F.W. Muskett, M.D. Finucane, S. Caddick, D.N. Woolfson, Solution structure of a novel chromoprotein derived from apo-Neocarzinostatin and a synthetic chromophore, Biochemistry 41 (2002) 11731–11739.

[56] P.J. Simpson, D.N. Bolam, A. Cooper, A. Ciruela, G.P. Hazlewood, H.J. Gilbert, M.P. Williamson, A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity, Struct. Fold. Des. 7 (1999) 853–864.
[57] M. Pellecchia, D.L. Montgomery, S.Y. Stevens, C.W. van der Kooi, H.P. Feng, L.M. Gierasch, E.R.P. Zuiderweg, Structural insights into substrate binding by the molecular chaperone DnaK, Nature Struct. Biol. 7 (2000) 298–303.
[58] B.T. Farmer, K.L. Constantine, V. Goldfarb, M.S. Friedrichs, M. Wittekind, J. Yanchunas, J.G. Robertson, L. Mueller, Localizing the NADP$^+$ binding site on the MurB enzyme by NMR, Nature Struct. Biol. 3 (1996) 995–997.
[59] E.F. DeRose, T. Darden, S. Harvey, S. Gabel, F.W. Perrino, R.M. Schaaper, R.E. London, Elucidation of the ε-θ subunit interface of *Escherichia coli* DNA polymerase III by NMR spectroscopy, Biochemistry 42 (2003) 3635–3644.
[60] J. Morrison, J.C. Yang, M. Stewart, D. Neuhaus, Solution NMR study of the interaction between NTF2 and nucleoporin FxFG repeats, J. Mol. Biol. 333 (2003) 587–603.
[61] H. Lüttgen, R. Robelek, R. Mühlberger, T. Diercks, S.C. Schuster, P. Köhler, H. Kessler, A. Bacher, G. Richter, Transcriptional regulation by antitermination. Interaction of RNA with NusB protein and NusB/NusE protein complex of *Escherichia coli*, J. Mol. Biol. 316 (2002) 875–885.
[62] R.E. London, Chemical shift and linewidth characteristics of reversibly bound ligands, J. Magn. Reson. Ser. A 104 (1993) 190–196.
[63] J.L. Sudmeier, J.L. Evelhoch, N.B.H. Jonsson, Dependence of NMR lineshape analysis upon chemical rates and mechanisms: implications for enzyme histidine titrations, J. Magn. Reson. 40 (1980) 377–390.
[64] E.L. Kovrigin, NMR line shapes and multi-state binding equilibria, J. Biomol. MR 53 (2012) 257–270.
[65] J. Feeney, J.G. Batchelor, J.P. Albrand, G.C.K. Roberts, Effects of intermediate exchange processes on the estimation of equilibrium constants by NMR, J. Magn. Reson. 33 (1979) 519–529.
[66] L.M.I. Koharudin, A.R. Viscomi, B. Montanini, M.J. Kershaw, N.J. Talbot, S. Ottonello, A.M. Gronenborn, Structure-function analysis of a CVNH-LysM lectin expressed during plant infection by the rice blast fungus *Magnaporthe oryzae*, Structure 19 (2011) 662–674.
[67] C. McInnes, S. Grothe, M. O'Connor-McCourt, B.D. Sykes, NMR study of the differential contributions of residues of transforming growth factor alpha to association with its receptor, Prot. Eng. 13 (2000) 143–147.
[68] B.D.N. Rao, Nuclear magnetic resonance line-shape analysis and determination of exchange rates, Meth. Enzymol. 176 (1989) 279–311.
[69] J. Sandström, Dynamic NMR Spectroscopy, Academic Press, New York, 1982.
[70] G.C.K. Roberts (Ed.), NMR of Macromolecules: A Practical Approach, IRL Press, Oxford, UK, 1993.
[71] M. Krajewski, P. Ozdowy, L. D'Silva, U. Rothweiler, T.A. Holak, NMR indicates that the small molecule RITA does not block p53-MDM2 binding *in vitro*, Nature Med. 11 (2005) 1135–1136.
[72] T. Mittag, B. Schaffhausen, U.L. Günther, Tracing kinetic intermediates during ligand binding, J. Am. Chem. Soc. 126 (2004) 9017–9023.
[73] S.J. Hammond, B. Birdsall, M.S. Searle, G.C.K. Roberts, J. Feeney, Dihydrofolate reductase ^{1}H resonance assignments and coenzyme-induced conformational changes, J. Mol. Biol. 188 (1986) 81–97.
[74] M.P. Foster, D.S. Wuttke, K.R. Clemens, W. Jahnke, I. Radhakrishnan, L. Tennant, M. Reymond, J. Chung, P.E. Wright, Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA, J. Biomol. NMR 12 (1998) 51–71.
[75] C.J. Craven, B. Whitehead, S.K.A. Jones, E. Thulin, G.M. Blackburn, J.P. Waltho, Complexes formed between calmodulin and the antagonists J-8 and TFP in solution, Biochemistry 35 (1996) 10287–10299.
[76] P. Rajagopal, E.B. Waygood, J. Reizer, M.H. Saier, R.E. Klevit, Demonstration of protein-protein interaction specificity by NMR chemical shift mapping, Protein Sci. 6 (1997) 2624–2627.
[77] J.A.R. Worrall, W. Reinle, R. Bernhardt, M. Ubbink, Transient protein interactions studied by NMR spectroscopy: the case of cytochrome *c* and adrenodoxin, Biochemistry 42 (2003) 7068–7076.
[78] J. Krishnamoorthy, V.C.K. Yu, Y.-K. Mok, Auto-FACE: an NMR based binding site mapping program for fast chemical exchange protein-ligand systems, PLoS ONE 5 (2010) e8943.
[79] A.J. Lowe, F.M. Pfeffer, P. Thordarson, Determining binding constants from ^{1}H NMR titration data using global and local methods: a case study using [*n*]polynorbornane-based anion hosts, Supramol. Chem. 24 (2012) 585–594.
[80] T. Konuma, Y.-H. Lee, Y. Goto, K. Sakurai, Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism, Proteins: Struct. Funct. Bioinf. 81 (2013) 107–118.
[81] R. Selvaratnam, S. Chowdhury, B. VanSchouwen, G. Melacini, Mapping allostery through the covariance analysis of NMR chemical shifts, Proc. Natl Acad. Sci. USA 108 (2011) 6133–6138.
[82] S. Barelier, D. Linard, J. Pons, A. Clippe, B. Knoops, J.-M. Lancelin, I. Krimm, Discovery of fragment molecules that bind the human peroxiredoxin 5 active site, PLoS ONE 5 (2010) e9744.
[83] E.S.C. Shih, M.-J. Hwang, On the use of distance constraints in protein-protein docking computations, Proteins: Struct. Funct. Bioinf. 80 (2012) 194–205.
[84] U. Schieborr, M. Vogtherr, B. Elshorst, M. Betz, S. Grimme, B. Pescatore, T. Langer, K. Saxena, H. Schwalbe, How much NMR data is required to determine a protein-ligand complex structure?, Chembiochem 6 (2005) 1891–1898
[85] C. Dominguez, R. Boelens, A.M.J.J. Bonvin, HADDOCK: A protein-protein docking approach based on biochemical or biophysical information, J. Am. Chem. Soc. 125 (2003) 1731–1737.
[86] P.N. Palma, L. Krippahl, J.E. Wampler, J.J.G. Moura, BiGGER: A new (soft) docking algorithm for predicting protein interactions, Proteins: Struct. Funct. Genet. 39 (2000) 372–384.
[87] X. Morelli, A. Dolla, M. Czjzek, P.N. Palma, F. Blasco, L. Krippahl, J.J.G. Moura, F. Guerlesquin, Heteronuclear NMR and soft docking: an experimental approach for a structural model of the cytochrome c_{553}-ferredoxin complex, Biochemistry 39 (2000) 2530–2537.
[88] X.J. Morelli, P.N. Palma, F. Guerlesquin, A.C. Rigby, A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data, Protein Sci. 10 (2001) 2131–2137.
[89] J. Stark, R. Powers, Rapid protein-ligand costructures using chemical shift perturbations, J. Am. Chem. Soc. 130 (2008) 535–545.
[90] G.M. Clore, C.D. Schwieters, Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from ^{1}HN/^{15}N chemical shift mapping and backbone ^{15}N–^{1}H residual dipolar couplings using conjoined rigid body/torsion angle dynamics, J. Am. Chem. Soc. 125 (2003) 2902–2912.
[91] M. Krzeminski, K. Loth, R. Boelens, A.M.J.J. Bonvin, SAMPLEX: automatic mapping of perturbed and unperturbed regions of proteins and complexes, BMC Bioinf. 11 (2010) 51.
[92] A. Medek, P.J. Hajduk, J. Mack, S.W. Fesik, The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands, J. Am. Chem. Soc. 122 (2000) 1241–1242.
[93] A.A. Lugovskoy, A.I. Degterev, A.F. Fahmy, P. Zhou, J.D. Gross, J.Y. Yuan, G. Wagner, A novel approach for characterizing protein ligand complexes: molecular basis for specificity of small-molecule Bcl-2 inhibitors, J. Am. Chem. Soc. 124 (2002) 1234–1240.
[94] C. Riedinger, J.A. Endicott, S.J. Kemp, L.A. Smyth, A. Watson, E. Valeur, B.T. Golding, R.J. Griffin, I.R. Hardcastle, M.E. Noble, J.M. McDonnell, Analysis of chemical shift changes reveals the binding modes of isoindolinone inhibitors of the MDM2-p53 interaction, J. Am. Chem. Soc. 130 (2008) 16038–16044.
[95] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J.M. Aramini, G.H. Liu, A. Eletsky, Y.B. Wu, K.K. Singarapu, A. Lemak, A. Ignatchenko, C.H. Arrowsmith, T. Szyperski, G.T. Montelione, D. Baker, A. Bax, Consistent blind protein structure generation from NMR chemical shift data, Proc. Natl Acad. Sci. USA 105 (2008) 4685–4690.
[96] A. Cavalli, X. Salvatella, C.M. Dobson, M. Vendruscolo, Protein structure determination from NMR chemical shifts, Proc. Natl Acad. Sci. USA 104 (2007) 9615–9620.
[97] M.A. McCoy, D.F. Wyss, Structures of protein-protein complexes are docked using only NMR restraints from residual dipolar coupling and chemical shift perturbations, J. Am. Chem. Soc. 124 (2002) 2104–2105.
[98] M. Cioffi, C.A. Hunter, M.J. Packer, A. Spitaleri, Determination of protein-ligand binding modes using complexation-induced changes in ^{1}H NMR chemical shift, J. Med. Chem. 51 (2008) 2512–2517.
[99] M. Cioffi, C.A. Hunter, M.J. Packer, Influence of conformational flexibility on complexation-induced changes in chemical shift in a neocarzinostatin protein – ligand complex, J. Med. Chem. 51 (2008) 4488–4495.
[100] M. Cioffi, C.A. Hunter, M. Pandya, M.J. Packer, M.P. Williamson, Use of quantitative ^{1}H NMR chemical shift changes for ligand docking into barnase, J. Biomol. NMR 43 (2009) 11–19.
[101] B. Wang, K. Raha, K.M. Merz, Pose scoring by NMR, J. Am. Chem. Soc. 126 (2004) 11430–11431.
[102] B. Wang, K.M. Merz, Validation of the binding site structure of the cellular retinol-binding protein (CRBP) by ligand NMR chemical shift perturbations, J. Am. Chem. Soc. 127 (2005) 5310–5311.
[103] D. González-Ruiz, H. Gohlke, Steering protein-ligand docking with quantitative NMR chemical shift perturbations, J. Chem. Inf. Model. 49 (2009) 2260–2271.
[104] S. Neal, A.M. Nip, H. Zhang, D.S. Wishart, Rapid and accurate calculation of protein ^{1}H, ^{13}C and ^{15}N chemical shifts, J. Biomol. NMR 26 (2003) 215–240.
[105] R.W. Montalvao, A. Cavalli, X. Salvatella, T.L. Blundell, M. Vendruscolo, Structure determination of protein-protein complexes using NMR chemical shifts: case of an endonuclease colicin-immunity protein complex, J. Am. Chem. Soc. 130 (2008) 15990–15996.
[106] A. Cavalli, R.W. Montalvao, M. Vendruscolo, Using chemical shifts to determine structural changes in proteins upon complex formation, J. Phys. Chem. B 115 (2011) 9491–9494.
[107] M.A.S. Hass, M.R. Jensen, J.J. Led, Probing electric fields in proteins in solution by NMR spectroscopy, Proteins: Struct. Funct. Bioinf. 72 (2008) 333–343.

Glossary of abbreviations

CIS: Complexation-induced change in chemical shift
CSP: Chemical shift perturbation
DHFR: Dihydrofolatereductase
NADPH: Nicotinamide adenine dinucleotide phosphate, reduced form
PCA: Principal component analysis
RIS: Root mean square
SVD: Singular value decomposition