Studies of protein–ligand interactions by NMR

J. Clarkson and I.D. Campbell

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

Abstract

Solution-state NMR has become an accepted method for studying the structure of small proteins in solution. This has resulted in over 3000 NMR-based co-ordinate sets being deposited in the Protein Databank. It is becoming increasingly apparent, however, that NMR is also a very powerful tool for accessing interactions between macromolecules and various ligands. These interactions can be assessed at a wide variety of levels, e.g. qualitative screening of libraries of pharmaceuticals and ‘chemical shift mapping’. Dissociation constants can sometimes be obtained in such cases. Another example would be the complete three-dimensional structure determination of a protein–ligand complex. Here we briefly describe a few of the principles involved and illustrate the method with recent examples.

Introduction

Solution-state NMR has become an accepted method for studying the structure of small proteins in solution. NMR is, however, also a powerful tool for assessing interactions between macromolecules and various ligands. Information about these interactions can be obtained on various levels. One example is qualitative screening of pharmaceutical libraries to assess binding. Another is ‘chemical shift mapping’, where interacting residues can be identified and mapped on the surface of a known protein structure. Yet another example is the complete three-dimensional structure determination of a protein–ligand complex. This paper very briefly describes some of the basic principles involved in studies of protein–ligand interactions and indicates the range of possible applications.

Basic principles

The reader is reminded of some essential aspects of solution-state NMR as applied to protein–ligand interactions (see Figure 1). A key feature is the chemical shifts (resonance position) which are sensitive both to chemistry and to the environment. Chemical shifts for the free and bound states will, in general, be different because of changes in the environment. Other key NMR parameters, such as the nuclear Overhauser effect and transverse relaxation rate (related to linewidth), depend mainly on dipole–dipole interactions, generated between nuclear dipoles tumbling in solution. In an equilibrium exchange (Figure 1), the NMR parameters of the ‘free’ and ‘bound’ states will be different. The differences in NMR parameters are due to a change in the relative mass of the two states, which cause a change in the molecular tumbling rates in solution. NMR parameters are also strongly influenced by the rate of exchange between free and bound states ($k_{off}$); and the results obtained depend on whether the exchange regime is ‘fast’ or ‘slow’ – see also legend to Figure 1.

Monitoring small-molecule ligands

This is a field that has shown relatively rapid developments, especially in the area of drug screening in the pharmaceutical industry. Several reviews have been published recently [1,2]. Small molecules tumble and diffuse much more rapidly when they are free in solution compared with when they are bound to a protein. By monitoring changes in NMR properties, selective binding to a target receptor can be detected. In practice about 10 small-molecule drug leads, taken from a carefully chosen library of compounds, are added simultaneously to a protein drug target.

Mapping binding sites on the macromolecule

There are powerful two-dimensional NMR experiments that can be used to efficiently monitor specific groups in a protein. Particularly powerful are the HSQC (heteronuclear single-quantum coherence) [3,4] and TROSY (transverse relaxation-optimized spectroscopy) [5] experiments. In an $^{15}$N-labelled protein, all amino acids, except proline, give rise to single peaks in a $^{1}H$–$^{15}$N HSQC or TROSY spectrum. An example of a two-dimensional TROSY spectrum of a protein is given in Figure 2(A). Using a variety of procedures these individual $^{15}$N–$^{1}H$ cross-peaks can be assigned to specific backbone resonances in the protein (a non-trivial step that is made easier by labelling the protein with $^{13}$C as well as $^{15}$N).

When a ligand is added to an assigned spectrum, of the kind shown in Figure 2(A), the resonances shift or broaden selectively. These perturbations can be used to identify the location of binding sites. Some sort of weighted plot is usually used [e.g. $\sqrt{((\delta_N/5)^2 + \delta_H^2)}$] to allow for

Key words: NMR spectroscopy, protein–ligand interactions, solution-state NMR.

Abbreviations used: HSQC, heteronuclear single-quantum coherence; TROSY, transverse relaxation-optimized spectroscopy; NOE, nuclear Overhauser effect.

To whom correspondence should be addressed (e-mail iain.campbell@bioch.ox.ac.uk).

©2003 Biochemical Society
Figure 1 | Schematic representation of ligand binding

The NMR properties of free and bound states are different. These include the chemical shifts ($\delta_f$ and $\delta_b$), relaxation rates and diffusion rates. The NMR signals from free and bound states can be in the ‘fast’ (one averaged signal), ‘slow’ (two non-averaged signals) or intermediate exchange regime (broad lines). If $K_d > 10^{-5}$ M the exchange regime is normally fast while if $K_d < 10^{-6}$ M, the exchange rate is often slow. $k_{on}$ and ($\delta_f - \delta_b$) are key parameters in deciding the exchange regime. The relaxation rates in the two states are also different, especially in the case of a small, rapidly tumbling, free molecule binding to a large, slowly tumbling, protein.

Figure 2 | $^1$H-$^{15}$N TROSY spectra, obtained at a $^1$H frequency of 750 MHz, of isotope-labelled ($^{15}$N and $^2$H) SpoIIAA

(A) SpoIIAA alone. (B) SpoIIAA in a SpoIIAA–SpoIIAB complex. (C) Resonances of SpoIIAA that are significantly shifted (grey) or perturbed (black) by SpoIIAB are mapped on the known structure of SpoIIAA.

This kind of shift mapping has been used extensively in NMR studies of protein–ligand interactions (reviewed in [6,7]). Another important mapping method is ‘cross-saturation’ where a relatively large unlabelled protein is irradiated with radio frequency. Association of a smaller deuterated protein can be detected and mapped by observing intensity changes in assigned resonances of the smaller protein [8]. Site-specific changes in amide solvent exchange rates, observed in the presence of ligand, can also be mapped on a structure (see e.g. [9]).

It is worth pointing out here that NMR can also contribute to an understanding of the kinetics of complex formation. The line shapes of peaks in an HSQC spectrum, corresponding to the free and bound states, are sensitive to the on and off
rates, especially $k_{\text{off}}$. An example is a study of the binding of different peptides to an SH2 domain [10]. Peptides with different affinities were shown to bind in different ways and with different kinetics.

**Defining the structure of protein–ligand complexes**

The ultimate goal is to define the structure of the complex. An elegant early example was the determination of the calmodulin peptide complex [11]. We will complete this brief review with some comments on the definition of such structures.

**Structure of the complex**

Shift mapping is generally not sufficient to define the exact structure of the complex (see also the next section). It is important to collect additional structural restraints, such as NOEs (nuclear Overhauser effects) and coupling constants, for the complex. Observation of NOEs between protein and ligand are facilitated by differential isotope labelling (e.g. $^{15}$N on protein, $^{13}$C on ligand). A variety of elegant NMR methods have been developed to differentiate the spectra of labelled and unlabelled protein [12] and determine intermolecular NOEs. A recent example, from our own work, is the determination of the structure of two fibronectin modules in complex with a peptide from a pathogenic bacterium [13]. It is generally easier to manipulate the isotopic composition of the recombinantly expressed protein than that of the peptide ligand. In this example, a variety of labelling patterns of the protein were produced using the isotopes $^{2}$H, $^{13}$C and $^{15}$N while the peptide was unlabelled. $^{127}$NOEs were observed between peptide and protein. Important additional structural restraints were also obtained, including $^{13}$C chemical shifts and residual dipolar coupling (an explanation of the application of these methods is beyond the scope of this brief review). In this case, the uncomplexed fibronectin modules had considerable relative flexibility, both in the absence and presence of the peptide. X-ray crystallography which is, in many ways, the preferred method for determination of the structures of complexes, failed here because suitable crystals could not be formed.

**Combining soft-docking calculations and NMR data**

In many cases, NMR does not provide sufficient information to produce a well-defined structure of a protein–protein complex. In such cases soft-docking calculations can be useful, provided that structures are available for the separate proteins. These calculations assume that the backbone of the protein structure remains unchanged in the complex, but conformational flexibility of the protein side chains is allowed. Algorithms developed to carry out soft-docking calculations [14] contain two main components. The first is a search of all the possible geometries, to create a large data set of possible protein complex structures. The second is to filter these geometries by evaluating their binding energies, thus predicting the ‘correct’ structure among the top ranking geometries. Computer programs, for example BiGGER [15,16] and HADDOCK [17], are now available that can incorporate NMR chemical-shift perturbation analysis, mutagenesis data and other known attributes such as the proximity of certain residues to a catalytic site. Other forms of experimental data that can be used are restraints from residual dipolar coupling and the level of symmetry in oligomers [15,18].

**A specific example**

SpoIIAA and SpoIIB are proteins involved in the regulation of sporulation in *Bacillus*. In the presence of ATP, SpoIIB acts as a kinase to phosphorylate SpoIIAA, whereas in the presence of ADP, SpoIIB forms a long-lived complex with SpoIIIA. Once phosphorylated, SpoIIAA forms a dimer that can be detected and defined by NMR [19]. The SpoIIAA–SpoIIB–ADP complex was modelled here, using both monomers from the known dimer structure of *Bacillus stearothermophilus* SpoIIAA–ADP [20]. Structures of SpoIIAA were taken from the free forms of the *B. subtilis* [21] and *B. stearothermophilus* [22] proteins. The experimental data used were NMR chemical-shift and line-shape perturbations. Fairly broad criteria were applied to discover which of the geometries broadly reflected the experimental data. These criteria were applied only to the geometries ranked in the top 50 by the soft-docking calculations, since test cases show that 80% of correct structures predicted by this program are ranked in the top 50 geometries [15]. The criteria from the experimental data were that 60% of the residues, seen to be significantly perturbed in the NMR chemical shift analysis [23], should be within 6 Å of the partner protein and that the phosphorylatable group should be within 8 Å of the terminal phosphate of the ADP moiety. This analysis showed that three geometries fitted all the criteria in the docking of the SpoIIAA structure from *B. subtilis* with monomer ‘B’ from SpoIIB and two fitted all the criteria with the *B. stearothermophilus* structure. No solutions fitted all the criteria with monomer ‘A’ from SpoIIB. The geometries that fitted all the criteria were found to be similar, using both the *B. subtilis* and *B. stearothermophilus* SpoIIAA structures, and had root-mean-square-deviation values of less than 4 Å.

A representative structure of the predicted complex is shown in Figure 3. In this structure the SpoIIAA molecule has a similar orientation with respect to the SpoIIB molecule, as in an unpublished crystallographic structure of the *B. stearothermophilus* SpoIIAA–SpoIIB complex (S. Darst, personal communication). However, the position of SpoIIAA on SpoIIB is slightly different and in the crystallographic structure completely covers the ‘ATP lid’. These differences may be due to the complex in the crystal having undergone backbone conformational changes in either the SpoIIB or SpoIIAA molecule.

In this example, other geometries fitting the experimental data could be found if lower-ranking solutions were also
considered. However, these geometries less accurately reflected the crystallographic structures. This presents a general problem with this technique in that there is no clear cut-off in the ranking of solutions within which a realistic geometry should lie. Accurate determination of this cut-off will require more analysis of experimentally derived structures to determine realistic binding energies. A second problem is that soft-docking calculations do not, in general, allow for any movement of the protein backbone. This problem is beginning to be addressed by programs such as HADDOCK that allow for movement of the backbone at the protein interface. A final refinement that will aid the accurate prediction of complex structures is improved use of experimental data. In this example, fairly crude measurements were taken to examine if particular geometries fitted experimental data. The measurements could be refined, for example by taking account of whether a particular residue has access to the solvent.

**Conclusions**

In this review we have given an indication of the power and range of solution-state NMR studies of protein–ligand complexes. Targets for NMR solution studies extend beyond those mentioned here, and include protein–nucleic acid and protein–membrane interactions. The applicability of this technique is continually being extended, yielding new insights into macromolecular interactions.

We thank BBSRC and the Wellcome Trust for financial support. We also thank Professor M.D. Yudkin for valuable discussions.

**References**


Received 20 June 2003