NMR studies of protein-small molecule and protein-peptide interactions

Guan, J.

6 Concluding remarks

NMR in protein-ligand interactions

NMR is a powerful tool at various steps of drug discovery and development. From ligand screening, hit validation to structural characterization, NMR is commonly used to characterize the structure and dynamics of the protein target and the ligand.

Throughout this work, NMR spectroscopy has been extensively used to study molecular interactions. 1D ^{1}H-NMR was used in Chapters 2 and 4 in various ways, such as TINS (target-immobilized NMR screening) fragment screening, characterizing ligand binding with a T_2 relaxation filter, as well as paramagnetic pseudocontact shifts (PCSs). Isotope filtered/edited NOESY type experiments were used in Chapter 3. Chemical shift perturbation (CSP) analysis using [^{1}H,^{15}N]-HSQC titrations were used in Chapters 3 and 5 to determine binding constants of protein-small molecule and protein-peptide complexes. In addition, paramagnetic NMR spectroscopy was used to obtain PCS restraints in Chapter 4 (1D ^{1}H-NMR for ligand PCSs and [^{1}H,^{15}N]-HSQC for protein PCSs) and paramagnetic relaxation enhancement (PRE) restraints in Chapter 5 ([^{1}H,^{15}N]-HSQC). This demonstrated the robustness of classical NMR approaches and the never-ending evolution of new NMR techniques that complement the conventional NMR approaches in structural studies.

Similarity and difference between CPMG and PRE: relaxation, and relaxation

A T_2 relaxation filter can be used to filter out the sharp NMR signals of ligands from those of macromolecules, such as proteins. The filter can be applied using a Carr-Purcell-Meiboom-Gill (CPMG) or a spin-lock pulse sequence. When the ligand is bound to the protein, it becomes slow-tumbling, and consequently the signal intensity of the ligand is reduced. With a reasonable fraction of bound ligand, changes in chemical shifts induced by ligand binding can also be observed, if no extensive line-broadening is used in spectra processing. By observing the differences in chemical shifts as well as the intensities between the free and the bound state, the binding ligands can be distinguished from the non-binding ones. It

should be noted that, in contrast to PRE (discussed below), the broadening effect is global on the bound ligand, i.e. there is no distance dependence between the bound ligand nuclei and the protein nuclei. For molecules with labile protons, the exchange process can also have significant broadening effects on the labile protons, but not on the non-labile ones.

PRE NMR spectroscopy has proved to be a powerful technique for studying transient complexes.[120,121] PRE arises from the unpaired electrons, which have very strong magnetic moments. These unpaired electrons may enhance relaxation rates of nearby nuclei because of the dipolar interaction between the unpaired electron and the nucleus as well as the fast longitudinal relaxation of the electron spin. The paramagnetic center can be an intrinsic metal binding site, or an engineered external tag on the protein surface. For synthetic peptides, the paramagnetic amino acid TOAC can be incorporated in peptides via a peptide bond, making it very convenient in automated peptide synthesis. Like for the NOE, the magnitude of PRE is strongly dependent on the distance between the paramagnetic center and the observed nucleus (r^{-6}). Even lowly populated states which only exist shortly in the vicinity of the spin label can be observed. Therefore, similar to NOE in a certain extent, PRE is very sensitive to changes in distance. In contrast to NOE (effective range ~5 Å), PRE detects long range (10-35 Å) interaction with high sensitivity. Compared with the T_2 relaxation filter which simply removes broad signals based on the molecular size, PRE induces resonance broadening depending on the distance.

Comparison of protein-ligand structures determined by NOE and by PCS

The experimental data from NOESY experiments can be used as proton-proton distances between protein-ligand (intermolecular NOE) and within the protein residues (intramolecular NOE). With the use of filtered NOESY experiments, one can be used to distinguish protein-protein and protein-ligand NOE. In Chapter 3, this strategy was successfully applied to determine the ligand binding mode using 43 intermolecular NOE restraints. Given the large number of restraints with a nature of short-range interaction, the ligand binding mode can be determined with high resolution. Although robust, however, the requirement of nearly complete backbone and side chain assignments for the bound protein is time-consuming and sometimes error prone. Besides, $^{13}C/^{15}N$-labeled proteins must be produced in relatively large quantity compared with the small quantity used in other NMR techniques. This can be costly if many ligand bound complexes need to be studied.

Pseudocontact shifts (PCSs) are one of the most well-known paramagnetic effects. PCSs are distance and orientation dependent, and the effect is predictable. Significant PCSs can be

measured up to 60 Å away from the paramagnetic center with strong lanthanides, thus being an important source of long-range restraints, particularly for large proteins or complexes. In Chapter 4, we demonstrated that it is possible to determine the ligand binding site and obtain a low-resolution structure using only restraints from simple 1D-^{1}H NMR spectra. This requires attachment of two-armed paramagnetic lanthanide tags (CLaNP-5) on three positions of the protein via disulfide bond linkage, one tag at a time. Besides, the ligand has to be in fast exchange on the NMR time scale. The ligand position is determined relative to the different paramagnetic tensor frames. Lanthanide positions can be modeled and tensor magnitudes (axial and rhombic components) estimated with sufficient accuracy if the protein structure is available. In Chapter 4, a low-resolution structure of the ligand binding mode has been determined using 21 ligand PCSs from three sets of para- and diamagnetic 1D-^{1}H NMR spectra. The final structure is similar to the structure derived from NOE restraints (Chapter 3). The quality of the structure can be improved if the tensor positions are optimized using experimental protein PCSs. Nevertheless, the structure information obtained with the predicted tensors is sufficient for the early stage of drug discovery. The advantage of this approach is that protein isotope labeling and protein backbone assignments are not compulsory. Most important are the ligand assignments, which are typically much easier to obtain. The resolution can further be improved by increasing the number of restraints. This can be achieved by using more than three tagging sites, using a tag with different tensor frames (such as CLaNP-7[188]), or measuring additional PCSs of the bound ligand in ^{13}C-, ^{19}F-, or ^{31}P-NMR spectra (if the ligand has fluorine or phosphorous atoms). Sometimes introducing two additional cysteines on the protein can affect the stability and structure of the protein. Therefore, it is advised to prepare more mutants than actually needed, and select the ones which are least affected by the extra cysteines and by the lanthanide tags.

Chemical shift perturbation analysis in protein-ligand complexes

CSP analysis has been frequently used to study protein-protein and protein-ligand interactions. The interaction is accompanied by changes in the chemical environment of the observed nuclei at the binding interface as well as at a remote region. This information is subsequently converted into binding affinities (K_D), and used to qualitatively estimate the degree of dynamics. CSPs are averaged over all orientations. The absolute size of CSPs has also been reported to correlate with the degree of dynamics: smaller chemical shift changes indicates higher mobility.[165,222,225,226] Observation of distant conformational changes indirectly caused by ligand binding has been shown in the example of FKBP12 bound to a

small molecule (Chapter 3). In Chapter 5, the overall small size of chemical shift perturbations strongly suggests that the Pc-tetralysine peptide complexes are highly dynamic. Therefore CSP analysis can be used as a tool to study protein-small molecule and protein-peptide dynamics. Further developments in the quantitative analysis of CSP may be needed to draw a clear line between the encounter complex and the specific complex.

Dynamics in transient protein-peptide complexes

To study interactions in transient complexes, the traditional NOE method is difficult to apply. The sensitivity of PRE to lowly populated states makes it a versatile technique to study dynamics in complex formation.[120] On the spectroscopy side, the interactions can be detected by simply acquiring standard HSQC spectra. On the sample preparation side, if the molecule (a protein or a ligand) is diamagnetic, an additional attachment step is required to introduce the paramagnetic source. It is also possible to replace the metal ion in metalloproteins with a suitable ion, to suppress or introduce paramagnetism to the system. In chapter 5, the electron transfer protein, plastocyanin (Pc), naturally harbours to a copper ion. In our *in vitro* study, the copper was replaced by a zinc ion to suppress PRE caused by the copper (II). The paramagnetic amino acid, TOAC, was attached on the peptide as the source of PRE. PRE data represent all orientations of the partners in the complex, including the sparse states. For encounter complexes that are mainly driven by long range electrostatic interactions, multiple orientations can exist with the same energy, like 'velcro'.[197] By comparing CSP and PRE maps of the Pc-tetralysine peptide complexes with the results from Monte Carlo (MC) simulations, it was found that the results of CSP maps and those of MC simulations were in high agreement for Pc from *Populus nigra* and *Dryopteris crassirhizoma*, which have many negative charges on the protein surface. For complexes with weaker charge complementarity, such as tetralysine binding to Pc from *Phormidium laminosum*, the binding is weak as observed in CSP and PRE data. The PRE data provide indisputable evidence that the peptide binds in multiple orientations.

Computational tools for studying molecular interactions

Modeling the positions of the lanthanide binding tag CLaNP-5 was first introduced in pseudoazurin and showed high accuracy.[144] The modeling approach is simple and can be valuable for proteins that cannot be produced with isotope labels. If the secondary structure of the tagging site is flexible, fluctuations due to the mobility of the residues should be considered. In Chapter 4, the binding mode of the small-molecule ligand bound to FKBP12

was calculated relative to the positions of the lanthanides, which were derived from modeling and later optimized using experimental data.

PCSdock (Chapter 4) provides a fast, rough estimate of the ligand binding site based on ligand PCSs. This simple procedure uses ligand PCSs as the input restraints, provided the tensor parameters and the protein structure are available. The ligand structure and assignments are not required. PCSdock simply shows the grid points that satisfy the input PCSs, which leads to a grid in which the ligand is likely to bind.

An ensemble docking approach (Chapter 5) was used to visualize the PRE data of the complexes of Pc-charged peptides. For specific complexes, PRE should represent a single orientation of the binding partners. For encounter complexes that can exist in multiple orientations, interpretation of PRE data using an ensemble of orientations provides better visualization than using a single orientation. However, the ensemble docking approach still provides limited solutions. As MC simulations consider only electrostatic interactions, the results of experimental PRE of complexes that are purely driven by electrostatics are expected to correlate well with those of MC simulations. In our study, the results of theoretical MC simulations with experimental CSP and PRE analysis on the charged peptide and the three Pcs with different surface charge properties provided evidence for dominant non-specific electrostatic interactions as well as other transient short-range interactions, such as hydrogen bonds and hydrophobic contacts.