9

Structural Proteomics by NMR

G. Marius Clore

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland

9.1 INTRODUCTION

Nuclear magnetic resonance (NMR) is a powerful spectroscopic technique that permits the detailed study at atomic resolution of the three-dimensional structure and dynamics of macromolecules and their complexes in solution. In this brief chapter, I discuss various aspects of NMR that are pertinent to structural proteomics, that is, the high-throughput study of protein–protein complexes at the atomic level. Structural work on complexes has gained increasing importance since it is evident that the structure of a complex yields far greater functional insight than the structures of its individual component proteins.

Macromolecular structure determination by NMR is intrinsically a highly specialized, labor-intensive, and time-consuming technique. In addition, for a system of any reasonable size (say, greater than about 70 residues), isotopic labeling with ^{15}N and ^{13}C is required. Numerous reviews have been written on the subject detailing the experimental and computational methodologies involved (Wuthrich, 1986; Clore and Gronenborn, 1989, 1991, 1998; Bax and Grzsiek, 1993; van de Ven, 1995; Cavanagh et al., 1996). Determining the structure of a single protein by NMR can be broken down into essentially four steps: (i) sequential resonance assignment making use of a number of experiments to identify through-bond connectivities along the backbone and side chains (usually 3D triple resonance experiments); (ii) assignment of cross-peaks in nuclear Overhauser enhancement spectra (usually 3D and 4D) to obtain short (≤5–6 Å) interproton distance restraints, which provide

Proteomics for Biological Discovery, edited by Timothy D. Veenstra and John R. Yates.

the main source of geometrical information; (iii) measurement of additional NMR observables that provide useful conformational information (these may include three-bond scalar couplings that are related to torsion angles by simple empirical equations; $^{13}C\alpha/^{13}C\beta$ chemical shifts that are related empirically to backbone ϕ/ψ torsion angles; and long-range orientational restraints, such as residual dipolar couplings measured in dilute liquid crystalline media); and (iv) calculation of the three-dimensional structure from the experimental NMR restraints using simulated annealing. Generally, an iterative refinement strategy is employed (Clore and Gronenborn, 1989, 1991): calculations are initially carried out with a limited set of interproton distance restraints corresponding to NOE cross-peaks with unambiguous assignments; further interproton distance restraints from the remaining NOE cross-peaks are subsequently added in an iterative manner on the basis of a successively calculated series of structures. While improvements in spectrometer technology (e.g., the advent of cryoprobe technology that increases the signal-to-noise ratio three- to fourfold; higher field magnets that increase spectral resolution, thereby reducing spectral overlap) have reduced the measurement time to some extent, collecting all the data necessary to solve a NMR structure at high accuracy still requires several months. Similarly, improvements in spectral analysis software have permitted the introduction of some degree of automation (Montelione, 1991; Gerstein et al., 2003; Yee et al., 2003), but extensive human intervention is still necessary to fully and reliably interpret the data in all but the simplest of cases.

In this light, what contribution can NMR make to structural proteomics? There are two major methods for deriving high-resolution structural information at atomic resolution: NMR spectroscopy in solution and single crystal X-ray diffraction. In rare instances, electron microscopy is also capable of providing high-resolution information in the solid state. In addition, mass spectrometry in combination with crosslinking data is potentially capable of providing low-resolution structural information when combined with the computational techniques conventionally employed to derive structures from NMR data. If crystals can be obtained rapidly, there is little doubt that crystallography, particularly with the advent of synchrotron X-ray sources, offers the fastest route to high-resolution structure determination. However, complexes are generally more difficult to crystallize than isolated proteins, and it is usually the case that weak complexes (with K_D values in the 1–100 μM range) are extremely difficult to cocrystallize. In the case of NMR, complexes are amenable to structural investigation providing exchange is either fast (weak binding) or slow (tight binding) on the chemical shift time scale. If exchange, however, is intermediate on the chemical shift time scale, the signals are broadened out, precluding any detailed structural work.

A full structure determination of a protein–protein complex by NMR is extremely time consuming. For example, in the case of the 40 kDa EIN·HPr complex from the bacterial phosphotransferase system, the total NMR measurement time alone was ~3500 hours (or 4.8 months) (Garrett et al., 1999). Clearly, therefore, the conventional approach is not suitable for high throughput. Fortunately, new developments have significantly shortened the amount of time required by making full use of prior knowledge in the form of existing high-resolution crystal structures of the free proteins (Clore, 2000; Schwieters and Clore, 2001; Clore and Bewley, 2002). With this information in hand, it is possible to derive high-resolution

structures of complexes using limited intermolecular NOE data to provide translation (as well as orientational) information and residual dipolar coupling data (Prestegard et al., 2000; Bax et al., 2001) to generate very accurate orientational information. This chapter presents the underlying principles behind this approach and illustrates its application to a variety of protein–protein complexes. In addition, strategies are discussed whereby translational information from NOE data can be replaced entirely by highly ambiguous intermolecular distance restraints derived from $^{15}N/^{1}H_N$ chemical shift perturbation mapping (Clore and Schwieters, 2003).

9.2 INTERMOLECULAR DISTANCE RESTRAINTS

As noted earlier, the nuclear Overhauser effect (NOE) is the primary source of geometric information for NMR-based structure determination (Wüthrich, 1986; Clore and Gronenborn, 1989). The NOE (in the initial rate approximation) is proportional to the sixth root of the distance between two protons. Consequently, the upper limit for interproton distances that can be detected using the NOE is 5–6 Å. The key to deriving intermolecular NOE-derived interproton distance restraints lies in combining various isotope (^{15}N and ^{13}C) labeling strategies with isotope filtering experiments that permit one to detect NOEs on protons attached to specific isotopes of nitrogen and carbon (i.e., NMR active such as ^{15}N or ^{13}C, or NMR inactive such as ^{14}N and ^{12}C) (Clore and Gronenborn, 1998). For example, in a complex comprising one protein labeled uniformly with ^{13}C and the other at natural isotopic abundance (i.e., ^{12}C), one can selectively detect NOEs from protons attached to ^{13}C and protons attached to ^{12}C. Typical labeling schemes and the corresponding intermolecular NOEs observed are illustrated in Figure 9.1, and an example of the data obtained from a 3D ^{13}C-separated/^{12}C-filtered NOE experiment is shown in Figure 9.2.

The NOE is not the only method that can be used to derive intermolecular distance restraints. It is also possible to derive distance restraints using a combination of crosslinking, proteolytic digestion, and mass spectrometry (Bennett et al., 2000; Sinz and Wang, 2001; Schulz et al., 2003). In many cases, however, the data will not yield unique crosslinking partners but multiple possibilities. In addition, it is possible to use another NMR-based approach, which involves derivatizing a suitable surface accessible cysteine (which may have to be introduced by site-directed mutagenesis) on one protein with either a spin label or a metal binding site (such as EDTA) and measuring paramagnetic relaxation enhancement effects on the other protein to yield long-range (15–35 Å) distance restraints (Voss et al., 1995; Battiste and Wagner, 2000; Mal et al., 2002; Dvoretsky et al., 2002; Iwahara et al., 2003). In general, however, such a strategy can only be applied in a rational manner if one already has a good idea of the interaction surfaces involved in complex formation. To some exent, such information can be derived rather easily by either $^{15}N/^{1}H_N$ chemical shift perturbation mapping (Walters et al., 2001; Zuiderweg, 2002) or cross-saturation experiments (Takahashi et al., 2000; Nakanishi et al., 2002). The latter experiment is far more challenging experimentally since it necessitates that one of the proteins is not only ^{15}N labeled but fully deuterated as well.

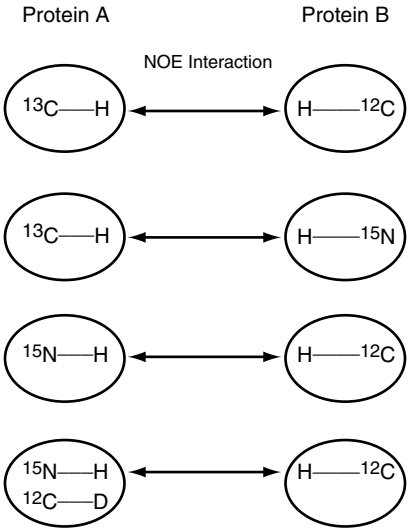


Figure 9.1. *Typical isotope labeling schemes used in the study of protein–protein complexes and corresponding intermolecular NOEs observed. If not indicated, the nitrogen or carbon isotope is ^{14}N and ^{12}C, respectively. In the fourth labeling scheme, ^{12}C attached protons in protein A are deuterated to narrow the lines of the NH resonances. This is useful for larger complexes.*

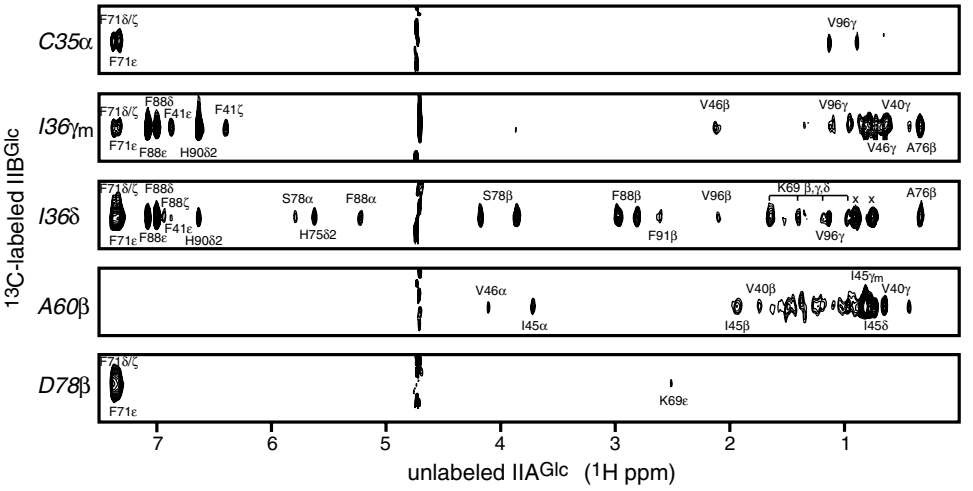


Figure 9.2. *Intermolecular NOEs in the IIA^{Glc}·IIB^{Glc} complex. Strips from a 3D ^{13}C-separated/^{12}C-filtered NOE spectrum recorded at 800 MHz on a 1 : 1 IIA^{Glc}($^{12}C/^{14}N$)·IIB^{Glc}($^{13}C/^{15}N$) complex, illustrating NOEs from protons attached to ^{13}C on IIB^{Glc} to protons attached to ^{12}C on IIA^{Glc}. Reproduced from Cai et al. (2003).*

9.3 ORIENTATIONAL RESTRAINTS

Long-range orientational restraints can be derived from the measurement of residual dipolar couplings (Tjandra et al., 1997b; Bax et al., 2001) and chemical shift anisotropy (Cornilescu et al., 1998; Wu et al., 2001) in liquid crystalline media and

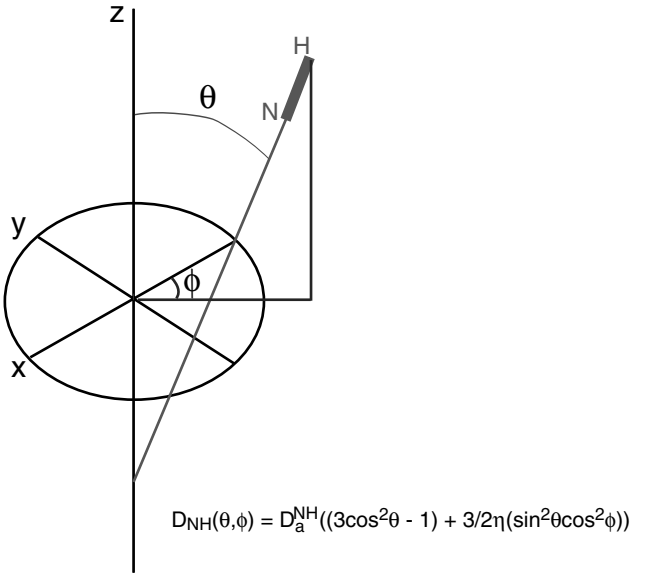


Figure 9.3. *Schematic illustration of orientational information derived from residual dipolar coupling measurements. The observed dipolar coupling, D_{NH}, is dependent on the angle θ between the N-H interatomic vector (shown as the thick line) and the z axis of the tensor, the angle ϕ, which describes the position of the projection of the interatomic vector on the xy plane of the tensor, the magnitude (D_a^{NH}) of the principal component of the tensor, and the rhombicity (η) of the tensor. (See color insert.)*

in suitable cases from heteronuclear T_1/T_2 data (Tjandra et al., 1997a). The characteristic feature of these various parameters is that they yield direct geometric information on the orientation of an interatomic vector(s) with respect to an external axis system (e.g., the alignment tensor in liquid crystalline media, the diffusion tensor for relaxation measurements) expressed in terms of two angles: θ, the angle between the interatomic vector and the *z* axis of the tensor, and θ, the angle that describes the position of the projection of the interatomic vector on the *xy* plane of the tensor (Figure 9.3).

For most practical purposes, residual dipolar couplings provide the easiest method for deriving orientational information. In an isotropic medium, the dipolar couplings average to zero. In the solid state, the maximum value of the N–H dipolar coupling is 20.7 kHz. To effectively measure dipolar couplings in solution, therefore, it is necessary to devise means of inducing only a small (ca. 10^{-3}) degree of order such that the N–H dipolar couplings lie in the ±20 Hz range. Experimentally, this is achieved by dissolving the protein or protein complex of interest in a dilute, water-soluble, liquid crystalline medium. Examples of such media include lipid bicelles (Tjandra and Bax, 1997), filamentous phages such as fd or pf1 (Clore et al., 1998; Hansen et al., 2000), rod-shaped viruses such as tobacco mosaic virus (Clore et al., 1998), and polyethylene glycol/hexanol (Rückert and Otting, 2000).

9.4 CONJOINED RIGID BODY/TORSION ANGLE DYNAMICS

In many instances, protein complex formation involves no significant changes in backbone conformation. Thus, if the structures of the individual proteins are already known at high resolution and it can be shown that the backbone conformation remains essentially unchanged upon complex formation (e.g., by comparison of dipolar coupling data measured on the complex with the X-ray structures of the free proteins), one can then make use of conjoined rigid body/torsion angle dynamics to rapidly solve the structure of the complex on the basis of intermolecular NOE data and backbone N–H dipolar couplings (Clore, 2000; Schwieters and Clore, 2001). In this procedure, only the interfacial side chains are allowed to alter their conformation. The backbone and noninterfacial side chains of one protein are held fixed, while those of the second protein are only allowed to rotate and translate as a rigid body. This has been applied with considerable success in the case of several 30–40 kDa complexes of the bacterial phosphotransferase system (Wang et al., 2000; Cornilescu et al., 2002; Cai et al., 2003). A comparison of the structure of the EIN·HPr complex obtained using the conventional full structure determination approach (Garrett et al., 1999) with that obtained by conjoined rigid body/torsion angle dynamics is shown in Figure 9.4.

It should be emphasized that conjoined rigid body/torsion angle dynamics can readily be extended to cases where significant changes in backbone conformation are localized to specific regions of the protein, such as the binding interface. In such cases, both the interfacial side chains and the relevant portions of the protein backbone would be given torsional degrees of freedom, and the experimental data would also have to include intramolecular restraints (NOE, dipolar coupling, etc.) relating to that portion of the backbone. This, for example, is the strategy that was employed to solve the structure of the IIA^{Mtl}·HPr complex (Cornilescu et al., 2002). This was necessitated because the crystal structure of IIA^{Mtl} (van Montfort et al., 1998), which contains multiple copies of IIA^{Mtl} in the unit cell, revealed alternate conformations for four loops in relatively close proximity to the putative interaction surface with HPr.

9.5 DOCKING BASED ON $^{15}N/^{1}H_N$ CHEMICAL SHIFT PERTURBATION AND N–H DIPOLAR COUPLINGS

Providing the complex under study can be aligned in a suitable liquid crystalline medium, the measurement of dipolar couplings is straightforward and permits one to determine the relative orientation of two proteins in a complex. Dipolar couplings, however, do not yield any translational information, which is essential for docking. Clearly, NOE-derived intermolecular interproton distance restraints provide the most useful and reliable source of translational information. However, intermolecular NOEs are not always easy to observe and their unambiguous assignment is still difficult and time consuming, particularly for larger complexes. Backbone $^{1}H_N$ and ^{15}N chemical shifts, on the other hand, are highly sensitive to environment and have been used extensively to rapidly map interaction surfaces on proteins (Walters et al., 2001). Not surprisingly, examination of the NMR literature reveals hundreds of examples of chemical shift mapping studies; to date,

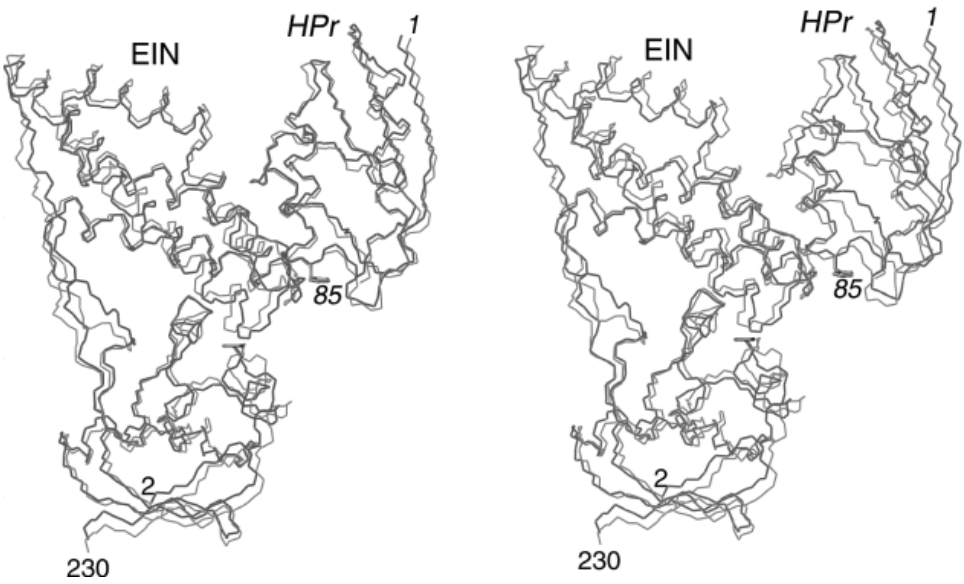


Figure 9.4. *Comparison of the structure of the EIN·HPr complex obtained using the conventional full structure determination approach (red) with that obtained by conjoined rigid body/torsion angle dynamics on the basis of 231 backbone N–H dipolar coupling data and either a full complement (blue) or partial complement (green) of NOE-derived intermolecular interproton distance restraints. The full complement of intermolecular NOEs comprises 109 interproton distance restraints; the partial complement consists of only eight intermolecular methyl proton–NH interproton distance restraints. The relative orientation of the proteins in all three calculated structures is identical. The backbone rms difference between the conventional NMR structure (red) and the structure calculated by docking the X-ray coordinates of the free proteins using the full complement of intermolecular NOEs (blue) only reflects the differences in the NMR and X-ray coordinates of the individual proteins, and these differences are within the uncertainty of the NMR coordinates. Adapted from Clore (2000). (See color insert.)*

however, only a handful of structures of macromolecular complexes have been determined by NMR.

Recently, we have shown that it is possible to convert chemical shift perturbation maps into highly ambiguous intermolecular distance restraints, which, in combination with orientational restraints from dipolar couplings, can reliably and accurately dock the partner proteins in a complex by means of rigid body/torsion angle dynamics calculations (Clore and Schwieters, 2003). Clearly, this methodology provides a powerful tool for high-throughput structural proteomics and, moreover, can greatly accelerate the determination of higher accuracy NMR structures of complexes (including the detailed placement of interfacial side chains) by providing a good starting point for the assignment of intermolecular NOE data.

The chemical shift maps are represented by a set of "r^{-6}-summed" distance restraints as follows (Clore and Schwieters, 2003). Given N_a residues on protein A and N_b residues on protein B that have been localized to the protein–protein interface by chemical shift mapping, a set of $N_a + N_b$ ambiguous distance restraints (d_{aB} and d_{bA}) is derived between all hydrogen, nitrogen, and oxygen atoms (i) of each

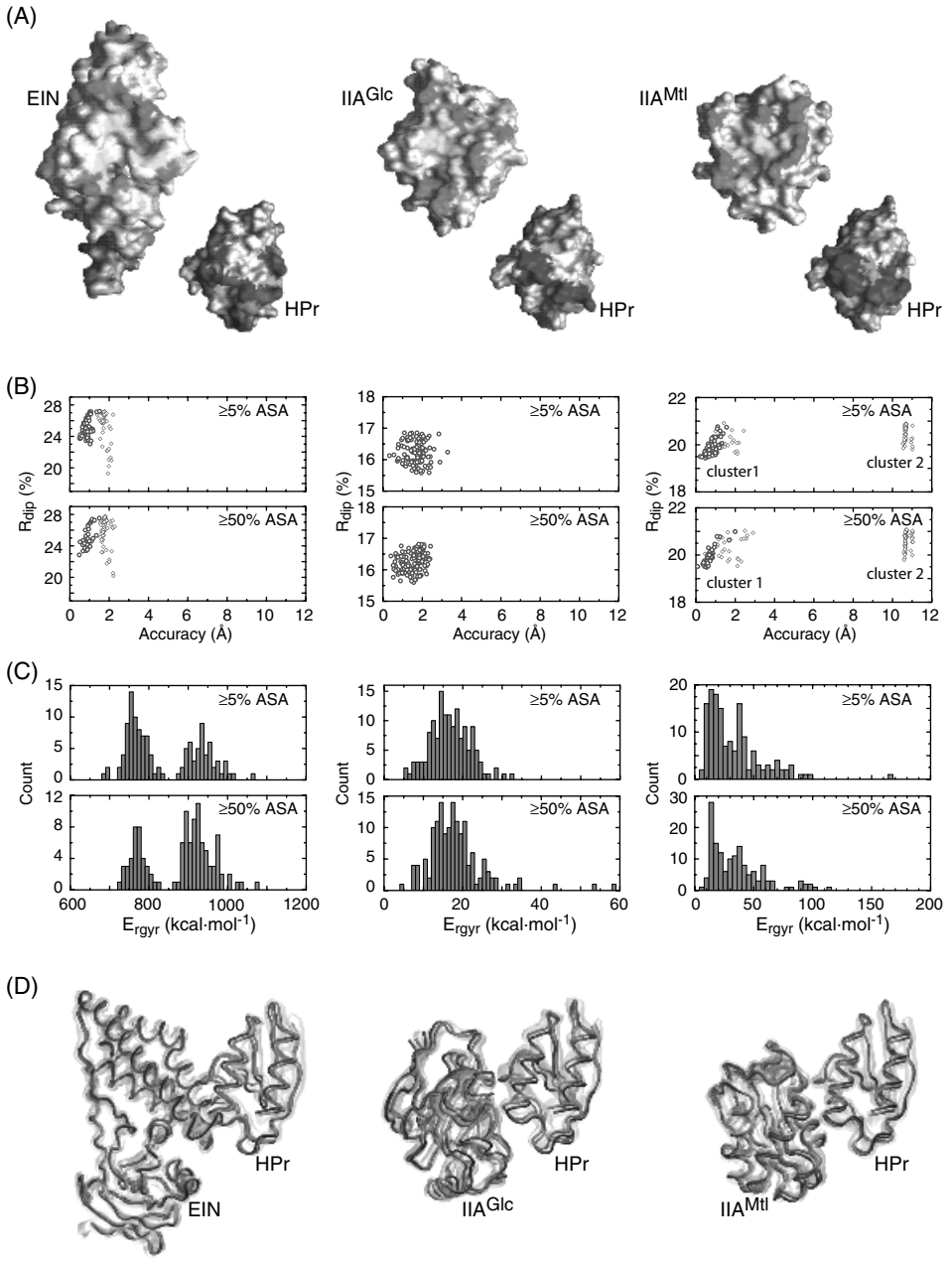
Figure 9.5. *Results of docking calculations for the EIN·HPr (left), IIA^{Glc}·HPr (middle) and IIA^{Mtl}·HPr (right) complexes on the basis of highly ambiguous distance restraints derived from $^{15}N/^{1}H_N$ chemical shift perturbation maps and backbone N–H dipolar couplings. (A) Interfacial residues (blue/cyan for HPr, red/orange for the three enzymes, and purple for active site histidines) identified by $^{1}H_N/^{15}N$ chemical shift perturbation are displayed on a molecular surface representation of the proteins. (The blue and red colored interfacial residues indicate residues with an accessible surface area (ASA) in the free proteins ≥50% of that in an extended Gly-X-Gly peptide; the cyan and orange colored residues indicate interfacial residues in the free proteins with 5% ≤ ASA < 50%. (B) Plots of the dipolar coupling R-factor (R_{dip}) versus accuracy for the converged structures characterized by no violations >0.5 Å in the highly ambiguous intermolecular distance and $R_{dip} \leq R_{dip}^{median}$. In the case of the EIN·HPr (left panel) and IIA^{Mtl}·HPr complexes (right panel), the circles and diamonds indicate structures in the lower and higher energy populations, respectively, of the radius of gyration energy function (E_{rgyr}) distribution. (C) Histograms of the E_{rgyr} distributions for the converged structures. The E_{rgyr} distribution is unimodal for the IIA^{Glc}·HPr complex (middle), but bimodal for the EIN·HPr (left) and IIA^{Mtl}·HPr (right) complexes. For the bimodal distributions, the lower and higher energy E_{rgyr} populations are colored red and blue, respectively. Note that in the case of the IIA^{Mtl}·HPr complex, all the structures in lower energy E_{rgyr} population reside in the correct cluster 1 ensemble; all the structures in the incorrect cluster 2 ensemble reside in the higher energy E_{rgyr} population. (D) Backbone (depicted as tubes) best-fit superpositions of the average coordinates (red) of the converged structures on the previously determined NMR structures (blue) solved on the basis of intermolecular NOEs and residual dipolar couplings. In the case of the IIA^{Mtl}·HPr complex, the mean coordinates are derived from the cluster 1 ensemble. The ensemble distributions of the docked structures are depicted by isosurfaces of the reweighted atomic density maps. Reproduced from Clore and Schwieters (2003). (See color insert.)*

residue a on protein A and all hydrogen, nitrogen, and oxygen atoms (j) of all residues b on protein B, and vice versa:

$$d_{aB} = \left(\sum_{b} \sum_{ij} r_{ai,bj}^{-6} \right)^{-1/6} \quad \text{and} \quad d_{bA} = \left(\sum_{a} \sum_{ij} r_{ai,bj}^{-6} \right)^{-1/6} \tag{9.1}$$

where $r_{ai,bj}$ is the distance between atom i of residue a of protein A and atom j of residue b of protein B. The number of atoms per residue ranges from 5 for Gly to 18 for Arg. Each d_{aB} restraint therefore comprises a set of $r_{ai,bj}$ distances involving 5–18 atoms of residue a, depending on the nature of residue a, and anywhere between 50 and 250 atoms from protein B, depending on the number and type of selected interfacial residues b on protein B. The number of $r_{ai,bj}$ distances encompassed in a single ambiguous distance restraint may range anywhere from 400 to 3000. Each d_{aB} and d_{bA} ambiguous distance restraint is given an upper bound of 5 Å. This does not imply that any individual $r_{ai,bj}$ distance is 5 Å or less since d_{aB} is always smaller than the shortest $r_{ai,bj}$ distance. Moreover, a cutoff of 5 Å is actually quite generous owing to the nature of the ambiguous distance restraints. Thus, for example, if a given d_{aB} ambiguous distance restraint is made up of 20 individual $r_{ai,bj}$ distances, each 10 Å in length, the value of d_{aB} is 6 Å.

The results of such calculations are shown in Figure 9.5 for three complexes: EIN·HPr, IIA^{Glc}·HPr, and IIA^{Mtl}·HPr (Clore and Schwieters, 2003). In the case of the first two complexes, a unique orientation is obtained that is within 1–2 Å of the structure derived from intermolecular NOE data and dipolar couplings. For the IIA^{Mtl}·HPr complex, however, two alternative solutions are obtained, the first is ~1 Å from the correct solution and the second ~11 Å away.



For an asymmetric alignment tensor, N–H dipolar couplings measured in a single alignment medium are consistent with four possible relative protein–protein orientations, two of which differ by a 180° rotation about the *z* axis of the alignment tensor, and the other two by a 180° rotation about the *y* axis of the alignment tensor. In most instances, exemplified by the EIN·HPr and IIA^{Glc}·HPr complexes, the ambiguous intermolecular distance restraints derived from $^{1}H_N/^{15}N$ chemical shift

mapping resolve the fourfold degeneracy such that only a single orientation is consistent with both the ambiguous intermolecular distance restraints and the $^1D_{NH}$ dipolar couplings (Clore and Schwieters, 2003).

In unfavorable cases, such as the IIA^{Mtl}·HPr complex, the ambiguous intermolecular distance restraints only reduce the number of solutions to two. The twofold reduction in degeneracy is achieved because the ambiguous intermolecular distance restraints ensure that the two binding surfaces are apposed and interpenetration of the two molecules is prohibited by the van der Waals repulsion term. In the case of the IIA^{Mtl}·HPr complex, the persistence of twofold degeneracy arises from the fact that the x and y axes of the alignment tensor lie in the plane of the protein–protein interface, such that a 180° rotation about the z axis can occur without interpenetration of the two molecules (Figure 9.2). Fortunately, it is usually easy to distinguish the correct solution from the incorrect one using a variety of independent approaches (Clore and Schwieters, 2003). The simplest involves reexamination of the $^{15}N/^1H_N$ chemical shift perturbation maps in the light of the structures to assess whether these maps allow one to distinguish between the two alternate solution. In this instance, the incorrect solution predicts several interfacial residues that do not exhibit any $^{15}N/^1H_N$ perturbations and are therefore unlikely to be part of the binding site. Consistency with prior biochemical data can also be employed. Phosphoryl transfer occurs from His15 of HPr to His65 of IIA^{Mtl}, which places an upper limit of ~14 Å on the Cα–Cα separation between these two histidine residues. In the correct solution, this Cα–Cα distance is ~12 Å, whereas in the incorrect one it is ~17 Å. Experimentally, the twofold degeneracy can be resolved by measuring a second set of dipolar couplings in a liquid crystalline medium possessing a different alignment tensor (e.g., charged versus uncharged media). In addition, an empirical method, based on the radius of gyration, for assessing the overall packing density and surface complementarity can also be employed.

9.6 STRUCTURAL PROTEOMICS OF THE GLUCOSE ARM OF THE BACTERIAL PHOSPHOTRANSFERASE SYSTEM

In bacteria, carbohydrate transport across the membrane is mediated by the phosphoenolpyruvate:sugar phosphotransferase system (PTS), which provides tight coupling of translocation and phosphorylation (Kundig et al., 1964). The PTS is a classical example of a signal transduction pathway involving phosphoryl transfer whereby a phosphoryl group originating on phosphoenolpyruvate is transferred to the translocated carbohydrate via a series of three bimolecular protein–protein complexes (Postma et al., 1996). The first two steps of the PTS are common to all sugars: enzyme I (EI) is autophosphorylated by phosphoenolpyruvate and subsequently donates the phosphoryl group to the histidine phosphocarrier protein HPr. The proteins downstream from HPr are sugar specific, comprising four distinct families of IIA permeases. In the case of the glucose branch of the PTS, the phosphoryl group is transferred from HPr to IIA^{Glc} and thence from IIA^{Glc} to the C-terminal cytoplasmic domain (IIB^{Glc}) of the glucose transporter $IICB^{Glc}$. The complexes in this pathway are rather weak with K_D values in the 10 μM range. Although high-resolution crystal structures have been determined for three of the

four proteins, namely, EIN (Liao et al., 1996), HPr (Jia et al., 1993), and IIA^{Glc} (Worthylake et al., 1991), crystallization of these protein–protein complexes has proved to be refractory, despite many years of trying. Thus, this system provides a showcase for the impact of NMR in structural proteomics.

Figure 9.6 shows the cascade of three protein–protein complexes, EIN·HPr (Garrett et al., 199), IIA^{Glc}·HPr (Wang et al., 2000), and IIA^{Glc}·$IICB^{Glc}$ (Cai et al., 2003), involved in phosphoryl transfer in the glucose-specific arm of the PTS. These complexes shed light on understanding fundamental aspects of protein–protein recognition, mechanisms for phosphoryl transfer between proteins, and the diversity of structural elements recognized by a single protein. Specificity of the protein–protein interaction surfaces is characterized by geometric and chemical complementarity, coupled with extensive redundancy to permit the effective recognition of multiple partners. There is little or no conformational change in the protein backbone before and after association. Some interfacial side chains, however, adopt different conformations (side chain conformational plasticity) depending on the interacting partner so as to achieve optimal intermolecular interactions. A consequence of these properties is increased velocity in signal transduction by eliminating any unnecessary time delay required for significant conformational change.

The interaction surfaces for HPr on EIN and IIA^{Glc} are very similar despite the fact that their underlying structures are completely different in terms of linear sequence, secondary structure (helices for EI, β-strands for IIA^{Glc}), and topological arrangement of structural elements. HPr makes use of essentially the same surface to interact with both its upstream partner EI and its downstream partner IIA^{Glc}. Concomitantly, the binding sites for IIB^{Glc} and HPr on IIA^{Glc} overlap extensively (~85% of the binding site for IIB^{Glc} constitutes part of the binding site for HPr). One might therefore anticipate that IIB^{Glc} could also interact with EIN. However, NMR data indicate that there is absolutely no interaction between EIN and IIB^{Glc} at millimolar concentrations. From a functional perspective this is important since it ensures that the PTS cascade is not bypassed. In addition, prevention of the potential shortcut between enzyme I and $IICB^{Glc}$ for glucose phosphorylation is necessary since these proteins also regulate the functions of proteins in other pathways (Postma et al., 1996). The structural basis for specificity and discrimination lies in the different charge distributions on the interaction surfaces of HPr and IIB^{Glc} such that binding of IIB^{Glc} to EIN is precluded by electrostatic repulsion (Cai et al., 2003).

9.7 CONCLUSION

NMR is the only solution technique capable of providing high-resolution structural information on protein–protein complexes at atomic resolution. While NMR is not a high-throughput technique, recent advances have considerably enhanced the speed of NMR structure determinations and the size and complexity of protein–protein complexes that can be studied. Indeed, NMR spectroscopy combined with prior knowledge on the structures of individual proteins from high-resolution X-ray crystallography promises to provide a very powerful approach for the efficient determination of three-dimensional structures of protein–

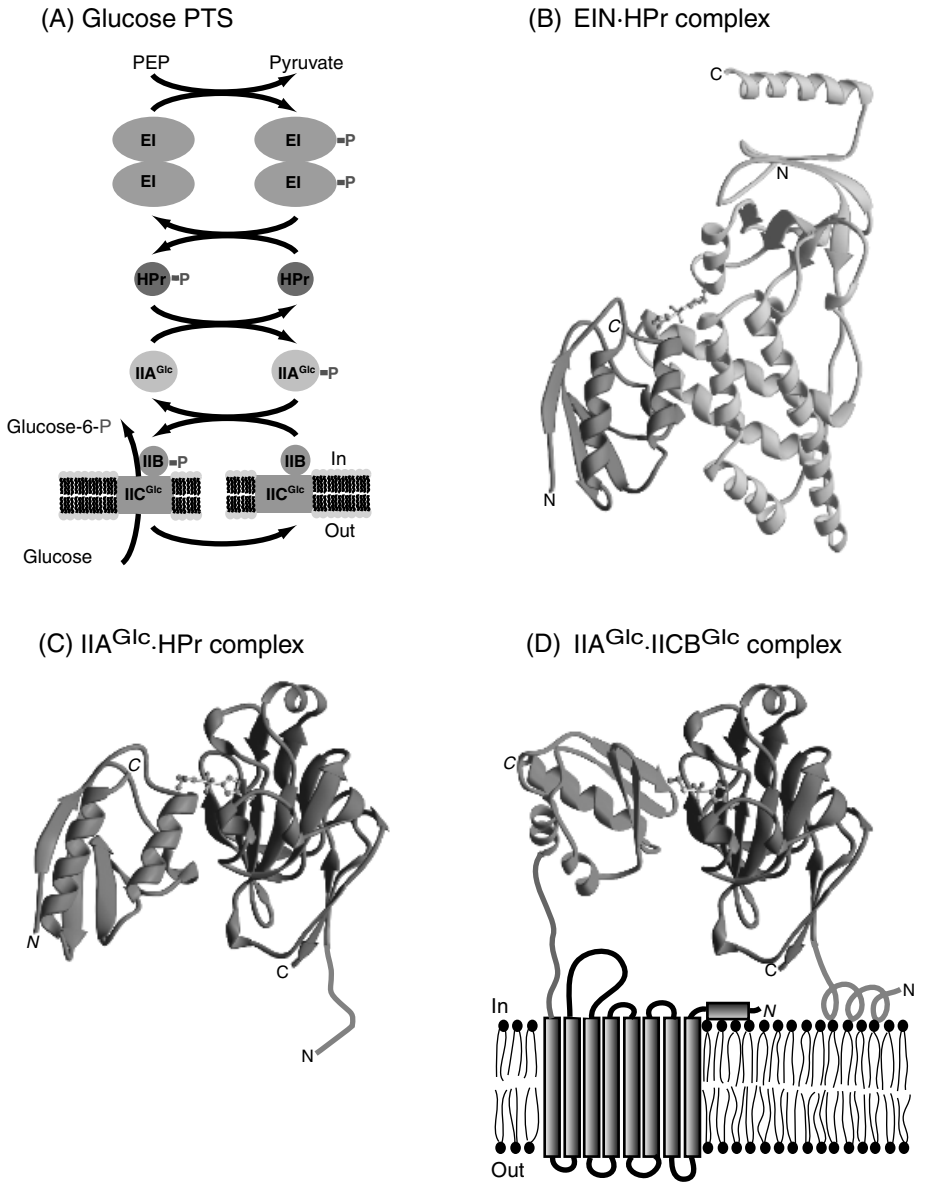


Figure 9.6. *Summary of the glucose arm of the E. coli PTS. (A) Diagrammatic illustration of the PTS cascade illustrating the transfer of phosphorus originating from phospho*enol*pyruvate and ending up on glucose through a series of bimolecular protein–protein complexes between phosphoryl donor and acceptor molecules. Ribbon diagrams of the (B) first (EIN·HPr), (C) second (HPr·IIA^{Glc}), and (D) third (IIA^{Glc}·$IICB^{Glc}$) complexes of the glucose PTS. The N-terminal domain of EI (EIN) is shown in gold, HPr in red, IIA^{Glc} in blue, and the IIB^{Glc} domain of $IICB^{Glc}$ in green. Also shown in yellow are the active site histidine residues of EIN (His189), HPr (His15) and IIA^{Glc} (His90) and the active site cysteine (Cys35) of IIB^{Glc}, together with the pentacoordinate phosphoryl group (red atoms) in the transition states of the complexes. IIB^{Glc} constitutes the C-terminal cytoplasmic domain of $IICB^{Glc}$. The transmembrane IIC^{Glc} domain of $IICB^{Glc}$ is thought to comprise eight transmembrane helices (shown diagrammatically in black). Note that the N-terminal end of IIA^{Glc} (residues 1–18) is disordered in free solution (C), but upon interaction with a lipid bilayer, residues 2–10 adopt a helical conformation (D), thereby further stabilizing the IIA^{Glc}·IIB^{Glc} complex, by partially anchoring IIA^{Glc} to the lipid membrane. Reproduced from Cai et al. (2003). (See color insert.)*

protein complexes up to ~100 kDa, particularly in cases that are refractory to cocrystallization.

Focus in this chapter was specifically on three-dimensional structure determination of protein–protein complexes. However, NMR also has a role in rapid screening and providing absolute proof for the formation of protein–protein complexes. While this does require substantially more material than mass spectrometry, typically a minimum of about 50 μM solution in a volume of 250 μL, the detection of protein–protein complex formation is straightforward and extremely rapid and can be accomplished by comparing the ^{1}H–^{15}N correlation spectrum of one of the partners, uniformly labeled with ^{15}N, in the absence and presence of the unlabeled second partner. Chemical shift perturbations and line broadening provide unambiguous proof of complex formation.

REFERENCES

Battiste JL, Wagner G. 2000. Utilization of site-directed spin labeling and high resolution heteronuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* 39:5355–5365.

Bax A, Grzsiek S. 1993. Methodological advances in protein NMR. *Acc Chem Res* 26: 131–138.

Bax A, Kontaxis G, Tjandra N. 2001. Dipolar couplings in macromolecular structure determination. *Methods Enzymol* 339:127–174.

Bennett KL, Kussmann M, Bjork P, Godzwon M, Mikkelsen M, Sorensen P. 2000. Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping: a novel approach to assess intermolecular protein contacts. *Protein Sci* 9:1503–1518.

Cai M, Williams DC, Wang G, Lee BR, Peterkofsky A, Clore GM. 2003. Solution structure of the phosphoryl transfer complex between the signal-transducing protein $IIA^{Glucose}$ and the cytoplasmic domain of the glucose transporter $IICB^{Glucose}$ of the *Escherichia coli* glucose phosphotransferase system. *J Biol Chem* 278:25191–25206.

Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ. 1996. *Protein NMR Spectroscopy: Principles and Practice*. Academic Press, New York.

Clore GM. 2000. Accurate and rapid docking of protein–protein complexes on the basis of intermolecular nuclear Overhauser enhancement data and dipolar couplings by rigid body minimization. *Proc Natl Acad Sci USA* 97:9021–9025.

Clore GM, Gronenborn AM. 1989. Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy *CRC Crit Rev Biochem Mol Biol* 24:479–564.

Clore GM, Gronenborn AM. 1991. Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* 252:1390–1399.

Clore GM, Gronenborn AM. 1998. Determining structures of larger proteins and protein complexes by NMR. *Trends Biotechnol* 16:22–34.

Clore GM, Bewley CA. 2002. Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. *J Magn Reson* 143:329–335.

Clore GM, Schwieters CD. 2003. Docking of protein–protein complexes on the basis of highly ambiguous distance restraints derived from $^{1}H_{N}/^{15}N$ chemical shift mapping and backbone ^{15}N-^{1}H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. *J Am Chem Soc* 125:2902–2912.

Clore GM, Starich MR, Gronenborn AM. 1998. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J Am Chem Soc* 120:10571–10572.

Cornilesci G, Marquardt JL, Ottiger M, Bax A. 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J Am Chem Soc* 120:6836–5837.

Cornilescu G, Lee BR, Cornilescu CC, Wang G, Peterkofsky A, Clore GM. 2002. Solution structure of the phosphoryl transfer complex between the cytoplasmic A domain of the mannitol transporter $II^{Mannitol}$ and HPr of the *Escherichia coli* phosphotransferase system. *J Biol Chem* 277:42289–42298.

Dvoretsky A, Gaponenko K, Rosevera PR. 2002. Derivation of structural restraints using a thiol-reactive chelator. *FEBS Lett* 528:189–192.

Garrett DS, Seok UJ, Peterkofsky A, Gronenborn AM, Clore GM. 1999. Solution structure of the 40,000 M_r phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nature Stuct Biol* 6:166–173.

Gerstein M, Edwards A, Arrowsmith CH, Montelione GT. 2003. Structural genomics: current progress *Science* 299:1663.

Hansen MR, Hanson P, Pardi A. 2000. Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol* 317:220–240.

Iwahara J, Anderson DE, Murphy EC, Clore GM. 2003. EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement. *J Am Chem Soc* 125:6634–6635.

Jia Z, Quail JW, Waygood EB, Delbaere LT. 1993. The 2.0 Å resolution structure of the *Escherichia coli* histidine-containing phosphocarrier protein HPr: a redetermination. *J Biol Chem* 268:22940–22501.

Kundig W, Ghosh S, Roseman S. 1964. Phosphate bound to histidine in a protein as an intermediate in a novel phosphotransferase system. *Proc Natl Acad Sci USA* 52: 1067–1074.

Liao DI, Silverton E, Seok YJ, Lee BR, Peterkofsky A, Davies DR. 1996. The first step in sugar transport: crystal structure of the amino terminal domain of enzyme I of the *E. coli* PEP:sugar phosphotransferase system and a model of the phosphotransfer complex with HPr. *Structure* 4:861–872.

Mal K, Ikura M, Kay LE. 2002. The ATCUN domain as a probe of intermolecular interactions: application to calmodulin–peptide complexes. *J Am Chem Soc* 124:14002–14003.

Montelione GT. 2001. Structural genomics: an approach to the protein folding problem. *Proc Natl Acad Sci USA* 98:13488–13489.

Nakanishi T, Miyazawa M, Sakakura M, Terasawa H, Takahashi H, Shimada I. 2002. Determination of the interface of a large protein complex by transferred cross-relaxation measurements. *J Mol Biol* 318:245–249.

Postma PW, Lengeler JW, Jacobson GR. 1996. Phospho*enol*pyruvate:carbohydrate phosphotransferase systems. In: Neidhardt FC (Ed), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington DC, pp 1149–1174.

Prestegard JH, al-Hashimi HM, Tolman JR. 2000. NMR structures of biomolecules using field-oriented media and residual dipolar couplings. *Q Rev Biophys* 33:371–424.

Rückert M, Otting G. 2000. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J Am Chem Soc* 122:7793–7797.

Schulz DM, Ihling C, Clore GM, Sinz A. 2003. Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin·melittin complex by chemical cross-linking and high resolution FTICR mass spectrometry: direct demonstration of multiple binding modes. *Biochemistry* 43:4703–4715.

Schwieters CD, Clore GM. 2001. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J Magn Reson* 152:288–302.

Sinz A, Wang K. 2001. Mapping protein interfaces with a fluorogenic cross-linker and mass spectrometry: application to nebulin–calmodulin complexes. *Biochemistry* 40:8903–7913.

Takahashi H, Nakanishi T, Kami K, Arata Y, Shimada Y. 2000. A novel NMR method for determining the interfaces of large protein–protein complexes. *Nature Struct Biol* 7:220–223.

Tjandra N, Bax A. 1997. Direct measurement of distances and angles in biomolecules by NMR in a liquid crystalline medium. *Science* 278:1111–1114.

Tjandra N, Garrett DS, Gronenborn AM, Bax A, Clore GM. 1997a. Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nature Struct Biol* 4:443–449.

Tjandra N, Omichinski J, Gronenborn AM, Clore GM, Bax A 1997b. Use of dipolar ^{15}N–^{1}H and ^{13}C–^{1}H couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Struct Biol* 4:732–738.

Van de Ven FJM. 1995. *Multidimensional NMR in Liquids: Basic Principles and Experimental Methods*. VCH Publishers, New York.

Van Montfort RL, Pijning T, Kalk KH, Hangyi I, Kouwijzer ML, Robillard GT, Dijkstra BW. 1998. The structure of the *Escherichia coli* phosphotransferase $IIA^{Mannitol}$ reveals two conformations of the active site. *Structure* 6:377–388.

Voss J, Hubbell WL, Kaback HR. 1995. Distance determination in proteins using designed metal ion binding sites and site-directed labeling: application to the lactose permease of *Escherichia coli*. *Proc Natl Acad Sci USA* 92:12300–12303.

Walters KJ, Ferentz AE, Hare NJ, Hidalgo P, Jasanoff A, Matsuo H, Wagner G. 2001. Characterizing protein–protein complexes and oligomers by nuclear magnetic resonance spectroscopy. *Methods Enzymol* 339:238–258.

Wang G, Louis JM, Sondej M, Seok YJ, Peterkofski A, Clore GM. 2000. Solution structure of the phosphoryl transfer complex between the signal transducing proteins HPr and $IIA^{Glucose}$ of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system. *EMBO J* 19:5635–5649.

Worthylake D, Meadow ND, Roseman S, Liao DI, Herzberg O, Remington SJ. 1991. Three-dimensional structure of the *Escherichia coli* phosphocarrier protein III^{Glc}. *Proc Natl Acad Sci USA* 88:10382–10386.

Wu Z, Tjandra N, Bax A. 2001. ^{31}P chemical shift anisotropy as an aid in determining nucleic acid structure in liquid crystals. *J Am Chem Soc* 123:3617–3618.

Wüthrich K. 1986. *NMR of Proteins and Nucleic Acids*. Wiley, Hoboken, NJ.

Yee A, Pardee K, Christendat D, Savchenko A, Edwards AM, Arrowsmith CH. 2003. Structural proteomics: toward high-throughput structural biology as a tool in functional genomics. *Acc Chem Res* 36:183–189.

Zuiderweg ER. 2002. Mapping protein–protein interactions in solution by NMR spectroscopy. *Biochemistry* 41:1–7.