Biochemistry

 Volume 41, Number 1 January 8, 2002

*Current Topics*

# Mapping Protein−Protein Interactions in Solution by NMR Spectroscopy†

Erik R. P. Zuiderweg*

*Biophysics Research Division, Department of Biological Chemistry, and Department of Chemistry, The University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109*

*Received October 2, 2001; Revised Manuscript Received November 8, 2001*

ABSTRACT: NMR is very well suited to the study of especially weak protein−protein interactions, as no crystallization is required. The available NMR methods to this end are reviewed and illustrated with applications from the recent biochemical literature: intermolecular NOEs, cross-saturation, chemical shift perturbation, dynamics and exchange perturbation, paramagnetic methods, and dipolar orientation. Most of these methods are now routinely applied for complexes with total molecular mass of 60 kDa and can likely be applied to systems up to 1000 kDa. A substantial fraction of complexes studied show distinct effects of induced fit affecting structural and dynamical properties beyond the contact interface.

Until recently, NMR[1] spectroscopy was probably best known as a method to determine three-dimensional structures of biological macromolecules without the need for crystallization (for a recent review of many of the underlying methods, see ref *1*). Currently, however, the study of protein−protein interactions by NMR appears to be the widest application. This evolution became possible because of better NMR equipment, development of more efficient isotopic labeling schemes, especially perdeuteration (*2*), and the advent of TROSY (*3*, *4*).

Below, I review the literature showing that it is now *possible* to study relatively large complexes by NMR, but I start with the question if it is also *desirable* to use NMR for such studies. For this forum, it will not be necessary to review how important protein−protein interactions are for biochemical processes in living organisms. The list of known protein−protein interactions is already very long, but it will increase dramatically in the coming few years, by virtue of two-hybrid assays, functional genomics, FRET, and large-scale chip-based proteomics. There is thus an overwhelming number of complexes for which we need to know fine details of the interface to understand how life is encoded and how diseases can be cured. It is becoming clear that such detailed knowledge cannot be reliably obtained from the high-resolution structures of the individual components. This is because these structures do not disclose where the interactions take place and because many proteins adapt their conformations dramatically to improve the fit. There are even extreme cases of induced fit where one of the interaction partners is virtually unstructured before the interaction takes place (*5*).

For now, NMR cannot compete with such spectacular X-ray accomplishments as the structure determinations of supramolecular assemblies such as the ribosome, proteosome, and the transcription apparatus. However, not all important complexes may crystallize, or they may not crystallize in a biologically relevant conformation. This could especially be the case when the interactions between the proteins are weak. Many of the protein−protein interactions necessarily are of weaker affinity for reasons of reversibility. For instance, the cell recognition receptors CD42 and CD58 have a $K_d$ of only 90 $\mu$M (*6*). The ubiquitous SH3 domains that play such an important role in the MAP signal-transduction pathway have affinities of the order of 10 $\mu$M for their cogent signals (*7*). Although NMR is by no means *limited* to the study of such

† Supported by Grants GM52421 and GM63027 from the National Institutes of Health and Grant MCB 9814431 from the National Science Foundation.

* Corresponding author. Phone: (734) 936-3850. Fax: (734) 764-3323. E-mail: zuiderwe@umich.edu.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; FRET, fluorescence resonance energy transfer; HSQC, heteronuclear single-quantum correlation.

10.1021/bi011870b CCC: $22.00 
Published on Web 12/05/2001

weak protein−protein interactions, the method is very *well suited* to study weakly interacting systems, as no potentially perturbing crystallization is required. The design of pharmacological agents is aimed at obtaining molecules that bind tightly and specifically to the intended target. In the process, however, one often begins with compounds that bind much less tightly. This has led to the development of a very active area called structure−activity relationships by NMR, where lead compounds with affinities as low as 10 mM can be rapidly screened for interaction with the active site of the target protein (*8*).

*Nuclear Overhauser Effect.* An unambiguous way of mapping biomacromolecular interactions is the intermolecular nuclear Overhauser effect (NOE). The NOE measures interproton distances with the basic $r^{-6}$ distance proportionality. NOE is the NMR equivalent of the well-known FRET (*9*). Using NOEs, a full three-dimensional structure of the complex is determined, using many precise NOE-derived distance constraints between the two interacting partners, in addition to the thousands of constraints within the individual macromolecules. This method is only applicable when the interaction between the molecules is relatively tight ($K_d \leq 10\ \mu$M). Arguably the most powerful method to obtain the intermolecular NOE is the isotope-edited NOE (*10*), also named half-filter (*11*), which unambiguously discriminates between NOEs within a macromolecule and between macromolecules. This requires that the two interacting macromolecules have different isotopic labeling patterns. For instance, one partner has no labeling (i.e., $^{1}$H, $^{14}$N, $^{12}$C) while the other is labeled with stable isotopes, e.g., $^{1}$H, $^{15}$N, $^{12}$C or $^{1}$H, $^{15}$N, $^{13}$C or sometimes even $^{1}$H, $^{2}$H,$^{15}$N,$^{13}$C. The principle of distinguishing between protons residing on labeled and unlabeled macromolecules is an essential ingredient of many of the methods discussed and is explained in the Appendix.

The isotope editing methodology is naturally very well suited for the study of labeled proteins and unlabeled nucleic acids. An example is the complex of the lac repressor headpiece protein with lac operator DNA, for which the high-resolution solution structure of a 26 kDa trimolecular complex has been determined (*12*). There are no crystal structures of the lac repressor/lac operator system that show how the protein−DNA recognition takes place. The NMR NOE structure of a 38 kDa trimolecular complex of two copies of the U1A protein with PIE RNA was determined to high resolution as well (*13*). It revealed an induced fit in the proteins upon RNA binding that allowed a previously unpredicted protein−protein contact to occur, explaining the cooperativity in regulation of polyadenylation by human U1A protein.

An example of a full structure determination of a relatively large protein−protein complex is the $M_r$ 40000 phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr (*14*). Despite the availability of several NMR and crystal structures of the individual proteins, no crystal structure for the complex has been reported. This high-resolution protein−protein complex, determined by thousands of NOEs (and residual dipolar couplings; see below), is an example of almost perfect lock-and-key surface complementarity that requires virtually no changes in conformation of the partner proteins relative to that in their respective free states.

Such a lock-and-key result is not general for protein interactions. An example to the opposite is the translation initiation factor eIF4G1, which undergoes an unfolded to folded transition upon binding to cap-binding protein eIF4E (*5*). Also, the interaction between 4E binding protein 1 and the initiation factor eIF4E, which inhibits translation, is an induced fit to a completely disordered protein molecule (*15*). A recent example using the full NOE method on systems pertaining to intracellular signal transduction is the work describing the complex between fragments of SNT adapter proteins, fibroblast growth factor receptors (FGFRs), and neurotrophin receptors (TRKs) (*16*). The complex is of moderate affinity (10 $\mu$M), and no crystal structures are known. The NMR structures show that the PTB domain of the SNT adaptor interacts with FGFRs or TRKs much closer to a lock-and-key fashion than the examples above and revealed how SNTs serve as molecular switches to mediate the interplay between FGFR and TRK signaling.

*Cross-Saturation.* A new application, called cross-saturation, obtains low-resolution, but highly relevant, interface information quickly (*17*). Cross-saturation is governed by the same physical processes as the NOE experiment. The donating partner protein is not labeled, while the observed protein is perdeuterated and $^{15}$N-labeled, but its amide deuterons are exchanged back to protons. The NMR experiments starts with a steady-state saturation of, exclusively, all aliphatic proton resonances of the donating partner. Cross-relaxation carries the saturation from the donor to the acceptor protein amide protons, where it is detected using a $^{15}$N−$^{1}$H HSQC or, for larger proteins, $^{15}$N−$^{1}$H TROSY. Those acceptor $^{15}$N−$^{1}$H cross-peaks that change in intensity upon the donor saturation are very likely to be close to the intermolecular interface. When the structure of the acceptor protein is known from either X-ray or NMR and when its $^{15}$N−$^{1}$H resonances are assigned, the acceptor side of the interface is known. The labeling can be reversed to obtain the other interface. This experiment is robust in the sense that only protons close to the interface will "light up", even if long-range conformational changes occur. Because of its experimental simplicity, the method can be widely applied.

As stated above, the NOE-based methods only work when complexes are relatively tight (10 $\mu$M or tighter). One reason for this is that weaker complexes are probably best described by an ensemble of interconverting structures, for which the concentrations of the individual structures are too small to give rise to detectable NOEs. The cross-saturation experiment promises to be more sensitive for weaker interactions than the two- or three-dimensional isotope-edited NOE, because longer magnetization transfer times can be used. This was recently demonstrated for the interaction between flavodoxin and a domain of methionine synthase with a $K_d$ of ~10 mM (*18*).

*Chemical Shift Perturbation Mapping.* Chemical shift perturbation is the most widely used NMR method to map protein interfaces. In a nutshell, the $^{15}$N−$^{1}$H HSQC spectrum of one protein is monitored when the unlabeled interaction partner is titrated in, and the perturbations of the chemical shifts are recorded. The interaction causes environmental changes on the protein interfaces and, hence, affect the chemical shifts of the nuclei in this area. The methods are well reviewed in, e.g., refs *19* and *20*. In some cases, as illustrated above, the entire protein may change conformation, and all shifts may be affected; then the chemical shift

perturbation fails as a mapping device but is an excellent indicator of allosteric processes (*21*, *22*).

An early application of modern shift mapping for the delineation of protein−protein interfaces appeared in 1993, when the surface of HPr with the A-domain of eEnzyme II was mapped independently by two groups by titrating the unlabeled A-domain into a solution of $^{15}$N-labeled HPr (*23*, *24*). Another early report was the identification of the contact surface of a streptococcal protein G-domain with a human Fc fragment (*25*). The RAS−RAF system is one of the prototypes of signal transduction. In 1995, the interaction surface of c-RAF-1 with Ras-GMPPNP was mapped by NMR (*26*). This study provided important complementary information to the wealth of structural data available for the RAS side of the complex.

Chemical shift perturbations are very sensitive to subtle effects. For instance, these methods allowed the novel find that zinc fingers can, in addition to their DNA-binding function, also display protein−protein interactions between them of rather modest affinity ($10^{-5}$ M) (*27*). Shift perturbation has also identified how a fibronectin-binding protein from *Staphylococcus aureus* binds with a module pair from the N-terminal region of fibronectin (*28*). Both involved molecules are rather floppy individually but become more ordered in the complex even though the affinity of binding was very weak. Chemical shift mapping is obviously not limited to protein−protein interactions. A current example of the mapping of the interaction between proteins and nucleic acids is the RNA recognition by the human U1A protein (*29*). The interaction interface of proteins with membranes was identified for cPLA2-C2 using shift mapping with low molecular weight micelles (*30*).

One of the largest systems approached with shift mapping to date is the interaction of a 10 kDa fragment of $^{15}$N-labeled Hsp40 (DnaJ) with a 44 kDa fragment of Hsp70 (DnaK) and also with the complete Hsp70 (i.e., 80 kDa total) (*31*). Only the beginning of the titration could be monitored as too much broadening occurred when the full complex became abundant, as no TROSY was used. However, this information was completely sufficient to document the interaction surface on DnaJ. The authors found that addition of MgATP largely reversed the effects, a control experiment showing that the monitored interaction is the biologically relevant one. Currently, the largest complex for which full chemical shift mapping is carried out at stoichiometric ratio is the 51 kDa complex of FimC−adhesin and FimH (*32*). Using TROSY, the surface of FimC was mapped for the interactions with the partner. As predicted by TROSY theory (see below), the line widths of the FimC NMR peaks were very similar in the 23 kDa FimC and in the 51 kDa complex.

To quantify the shift perturbation, most workers quote the length of the vector that connects the two-dimensional end points, normalized by typical shift ranges (*25*). Because of the complicated relationship between structure and chemical shift (*33*), the precise value of these normalization factors is rather immaterial. In addition, one should view the quantitation as a statistical measure only: the results are most significant when a contiguous surface patch of shifting resonances is obtained.

Generally, one cannot use chemical shift changes to predict what exactly happens at the interface. However, ring current shifts caused by proximity of aromatic residues are relatively well understood (*33*) and can be used to filter proposed complex structures for consistency. Especially the shift of protein resonances caused by single aromatic residues on small ligands can be used to locate the aromatic residue in the binding pocket of enzymes (*34*).

Shift perturbation measurements just yield the locations of the interfaces on the individual binding partners. It is then still unknown how the partners interact on an atom-to-atom basis. One approach may be to combine the NMR shift mapping experiments with a few distance constraints obtained from cross-linking or other experimental techniques such as FRET or with one of the other NMR methods described below. Another possibility is to use the NMR-defined interface areas as an selection filter on computational "soft-docking" programs (*35*).

*Titrations with NMR.* NMR spectroscopists carry out titrations because this allows, in addition to the mapping of the interface, a good estimation of the affinity, stoichiometry, and specificity of binding as well as the kinetics of binding. How the chemical shifts of the labeled protein change during the titration is determined by the kinetics of the interaction.

If the complex dissociation is very fast, there is, even during the titration, only a single set of resonances whose chemical shifts are the fractionally weighted average of the free and bound chemical shifts. Here the resonances of nuclei at the interface move in a continuous fashion during the titration. This regime is referred to as fast chemical exchange and is often observed for weaker interactions. One of the many examples is given by ref *18*. It is a simple matter to follow the resonances to their "bound" position and to extract the binding constant by fitting the fractional shift against a quadratic equation depending on total protein and partner concentrations. The trajectories of the shifting resonances in fast exchange are informative. If all two-dimensional trajectories are linear and occur at the same rate, a single binding event is indicated. If the trajectories for different, resonances occur at a different rate, and/or if they are curved, more than one binding site is implicated.

If the complex dissociation is very slow, one observes one set of resonances for the free protein and one set for the bound protein. During the titration, the "free set" will disappear and will be replaced by the bound set. Most of the resonances of the two sets will overlap with each other, but the differences will mark the interaction interface. This regime is referred to as slow chemical exchange. In slow exchange one does not automatically know to which new location the resonance has moved, unless one carries out an independent assignment procedure for the bound state. Consequently, one cannot easily quantitate the degree of change. Some workers have approached this problem by assuming that the new resonance which appears closest to the "free" resonance corresponds to its bound state (*36*, *37*). In the slow exchange case, the binding constant can still be quantitated by measuring the intensities of the disappearing and/or appearing peaks as a function of the titration progression (*21*). Slow exchange processes are also accompanied by kinetic line broadening effects which allow estimations, or limits, on kinetic parameters to be made but which complicate the extraction of affinities (*21*, *38*).

In the intermediate chemical exchange case, the frequencies of the changing resonances become poorly defined, and

extensive kinetic broadening sets in (*38*). If the lines become broad enough, the resonances may disappear from the NMR spectrum. Here, the interaction interface becomes delineated by progressively disappearing resonances. An example of a titration in this regime is found for the binding of CD42 and CD58 (*6*).

A rule of thumb is that interactions with $K_d$ <10 $\mu$M are slow exchange and intermediate/fast exchange otherwise. However, exceptions can be very dramatic: slow exchange for the interaction of a peptide with a 500 $\mu$M affinity was measured for the Hsp70 chaperones (because of a slow $k_{on}$) (*21*), and fast exchange was encountered for the binding of a phosphate compound to hemoglobin with an affinity of around 1 nM (because of a multisite binding mechanism) (*39*).

The range of binding constants that can be determined in a quantitative or semiquantitative way by NMR is set by the concentration of the interaction partners that can be observed by NMR. As a rule, quantitative measures can be obtained for $K_d$ values within an order of magnitude of the concentration of the studied species. Ultraweak affinities of 10 mM have been reported from NMR data of very concentrated samples (*18*). The high-affinity limit for quantitative determination of the kDa by NMR lies currently around 3 $\mu$M. New, cryogenically cooled coil probes improve NMR sensitivity by a factor of 3. Accordingly, NMR can (soon) be used to measure dissociation constants from $10^{-6}$ to $10^{-2}$ M.

A good example of quantitative mapping is work on troponin (*40*), which is a Ca-binding protein that interacts with troponin I and is involved in muscle regulation. The NMR titration with $Ca^{2+}$ revealed that 2 equiv of $Ca^{2+}$ binds to troponin C with strong positive cooperativity and high affinity. In this process, troponin C folds from a largely unstructured state to a domain capable of interacting with troponin I. Titration of troponin C in the presence of $2Ca^{2+}$ with a troponin I fragment occurs with a 1:1 stoichiometry and with a $K_d$ of $2 \pm 1$ $\mu$M.

*Mapping with Dynamics.* In recent years $^{15}N$ relaxation has been extensively used to describe the dynamics of the protein backbone. Information is obtained on rapid fluctuations at the pico- to nanosecond time scale (quantified in order parameters), as well as information on motions at the milli- to microsecond time scale (quantified in exchange broadening). For a recent review, see ref *41*. Dynamics by NMR is a large field by itself and may provide a link between structure, function, and thermodynamics. Here, we are interested in whether changes in dynamics upon intermolecular interactions can be used to map intermolecular interfaces. Currently, the message is mixed: for some cases, dynamics becomes quenched in the interface; in others, not. An example of the former is the interaction of the NSyp SH2 domain, where residues involved in peptide binding have order parameter values slightly below average in the free state (describing enhanced picosecond to nanosecond time scale dynamics) and which increase significantly upon complexation. This indicates rigidification of the interface (*42*). Similarly, comparisons of the $^{15}N$-transverse relaxation rates of the uncomplexed lipoyl domain with a complex involving the E1 component of the pyruvate dehydrogenase multienzyme complex showed a good correspondence between sites where a loss of fast motion occurred on binding and the chemical shift perturbation map (*43*). In stark contrast, and an example of the latter, is the interaction of hydrophobic peptides with the hydrophobic contact area on the PLCC SH2 domains, where no change in the dynamics of the methyl groups could be detected using deuterium relaxation experiments (*44*). It is argued that this represents a retention of configurational entropy upon binding and hence a contribution to high-affinity binding while maintaining binding promiscuity.

An interesting discussion states that binding/active sites show milli- to microsecond dynamics as a rule, which is quenched upon intermolecular interaction (*45*−*48*). In this vision, such sites are dynamically primed for induced fit conformational changes, and a range of conformations is dynamically sampled even without the actual interaction taking place. Upon interaction, one subconformation is stabilized, corresponding to a change in population levels of the different conformations available to the protein. This is a likely scenario, and the results with the PLCC SH2 domains are perhaps an exception. Nevertheless, it appears that protein dynamics is too much an intrinsic part of the binding process to make it a reliable tool for interface mapping.

*Mapping with Amide*−*Proton Exchange.* A similar variety in observations has been made for the mapping of protein interfaces using the protection of amide−proton exchange. Amide−proton exchange in proteins occurs upon transient unfolding of structure, ranging from local to global (*49*−*51*). Solvent accessibility to the transiently unfolded areas allows the exchange of the amide protons for other protons, or deuterons, to take place on time scales ranging from minutes to months. It is to be expected that a protein−protein interaction should shield the interface area from exchange. For very tight binding interactions such as antibody−antigene interactions, this is indeed the case. The epitopes of Der p 2 defined by monoclonal antibodies could be identified by comparing amide−proton exchange rates in the absence and presence of the antibodies (*52*). However, other observations are that the amide−proton exchange rates become slower for the entire protein upon interaction with a partner system (*53*, *54*) or that exchange changes occur far beyond the interaction site (*55*). As for the dynamical studies, these added complexities likely derive from the fact that proteins are adaptable molecules, which change some of their fundamental properties upon interaction.

*Mapping with Paramagnetics.* Certain forms of paramagnetic mapping appear to be reliable tools for the delineation of protein−protein interactions. Paramagnetic ions in solution broaden the NMR resonances of the solute. The effect is caused by transient dipolar interaction between the unpaired electron spins and the nuclei in the solute. When the electronic spins behave as free magnetic dipoles, no shifts of NMR resonances occur, and the NMR broadening follows a simple $r^{-6}$ law. If a protein is placed in a solution containing paramagnetic species, only its surface residues will be affected by the paramagnetic broadening (*56*). This was utilized to map the interaction between the catalytic domain of matrix metalloproteinase 3 and its protein inhibitor TIMP by comparing the amide−proton line broadening of the metalloproteinase by paramagnetic gadolinium−EDTA in the presence and absence of TIMP (*55*).

*Mapping with Site-Specific Spin-Labeling.* The free stable radical nitroxide (spin-label) TEMPO can interact specifically with cysteine groups engineered on the protein surface.

TEMPO is extensively used in EPR site-specific spin-labeling studies (*57*). For NMR, line broadening of the resonances occurs proportional to the $r^{-6}$ distance between label and NMR nucleus.

This method was used to determine the antibody interface with a spin-labeled peptide antigen (*58*). NOESY difference spectra of the binding site region of a Fab fragment were measured in the absence and presence of the spin-labeled peptide. The line broadening in the spectra was measured and converted to distance restraints to be used for the calculation of a model for the TE33−peptide complex. These methods were also used for the mapping of intermolecular contacts of the chaperone protein DnaK with peptides (*59*). A single Cys residue on the peptide was modified with an iodoacetoamide TEMPO. The effect of peptide titrations on the $^{15}N-^{1}H$ HSQC spectrum of the protein was monitored, and a distinct and contiguous surface patch of disappearing resonances on the surface of the DnaK chaperone protein was detected.

*Mapping with Pseudocontact Shifts.* In many compounds, the electronic magnetic dipole does not stay exactly aligned with the external magnetic field when a molecule tumbles in solution. The electron−nuclear dipolar interactions in such molecules do not average to zero, and a residual electron−nuclear coupling remains. Because of further averaging processes this results in shifts of the NMR resonance positions, an effect that is referred to as the pseudocontact shift.

The pseudocontact shift $\delta$ is given by

$$\delta = \frac{A}{r^3}[\chi_{ax}(3\cos^2\theta - 1) + \chi_{rhom}(\sin^2\theta \cos 2\phi)] \quad (1)$$

The parameter $A$ contains the basic magnetic properties of the nuclei and electrons and the temperature, and the terms $\chi_{ax}$ and $\chi_{rhom}$ give magnitude of the electronic magnetic susceptibility tensor in two orthogonal directions. The angles $\theta$ and $\phi$ are the polar angles between the electron−nuclear vector and the long axis of the susceptibility tensor. There are five unknowns in this equation; three of them, $r$, $\theta$, and $\phi$, yield structural information.

Most studies to determine the structure of complexes using the pseudocontact shift approach were carried out with cytochrome heme proteins. Examples are the interaction of plastocyanin with paramagnetic cytochrome *f* (*60*), the interaction of cytochrome $b_5$ with paramagnetic cytochrome *c* (*61*), and the interaction of yeast iso-1 cytochrome *c* with paramagnetic cytochrome *c* peroxidase (*62*). In all of these studies, the three-dimensional structures of the interacting partners were known. The cytochrome $b_5$/cytochrome *c* interaction is extremely weak (around 1 mM), and no three-dimensional structure was known prior to the NMR study.

Typically, of the order of 50 resonances become perturbed by the paramagnetic interaction in these complexes. Since the structure of the monitor protein is known, this amounts to 50 pieces of data relating to eq 1. The equation is thus overdetermined and can be solved by a multiparameter fit, and one obtains the distance and *orientation* of the paramagnetic center on the other protein with respect to the reporting protein. The process can be carried out to a very high precision. For example, the location of plastocyanin relative to cytochrome *f* was obtained to a precision of 1 Å (*60*). Similar methodology has been used to obtain the position of an intercalating drug into the DNA minor groove (*63*). The drug requires a divalent metal ion for complex formation, which for this study was taken as the paramagnetic ion $Co^{2+}$. Pseudocontact shifts were quantitated, and the structure of the complex was determined to a precision of 0.7 Å, even without the use of NOEs.

*Relative Positioning of Proteins Using Dipolar Couplings.* When a biomolecule tumbles isotropically in solution, the dipolar coupling between the nuclei in a molecule fluctuates but is on average zero. Recently, several workers have realized that the tumbling can be made to be spatially anisotropic. This happens when the protein aligns in very high magnetic fields (*64*) or when the solution is made to be anisotropic by dilute solutions of bicelles (*65*) or filamentous phage (*66*), which form liquid crystals. The result is that the dipolar interactions are not totally averaged to zero and that small dipolar splittings are retained in the solution NMR spectra (of the order of a few to a few tens of hertz, representing a spatial anisotropy of $10^{-5}-10^{-4}$.)

The resulting residual dipolar interactions are again given by eq 1, but now $\delta$ represents the residual dipolar coupling. The factor $A$ now describes nuclear properties (no temperature), and $\theta$ and $\phi$ are the polar angles of the internuclear vector with respect to the orientation of the protein in the magnetic field. $\chi_{ax}$ and $\chi_{rhom}$ are now the magnitudes of the residual dipolar interaction for vectors parallel with the longest and second longest axis of the alignment tensor, respectively. As above, the angles $\theta$ and $\phi$ represent the structural information. The distance $r$ is generally a constant, as it is common to measure the residual dipolar couplings for NH vectors from differences in coupled $^{15}N-^{1}H$ HSQC spectra in the absence or presence of alignment. This method is now used to obtain relative domain orientations in multidomain proteins (*67, 68*) or relative protein orientations in multiprotein complexes (*69*). When the structures of the individual proteins/domains are known, the relative orientations of all dipolar vectors within the individual proteins/domains are given. The measured dipolar couplings, in turn, reflect the orientation of the dipoles with respect to the magnetic field. It is then possible to carry out a computer procedure rotating these individual proteins/domains to fit the measured dipolar couplings. If the $\chi_{ax}$ and $\chi_{rhom}$ are found to be equal for the different proteins/domains, it strongly indicates that the individual proteins/domains move together as a single entity. Then, the orientations of the ordering tensors (the axes) for the different molecules must be collinear, which defines the orientation of the individual molecules with respect to each other. The method was for the first time demonstrated for a true two-protein complex with the 40 kDa complex of enzyme I and the histidine phosphocarrier protein (*69*). While this complex had been previously determined from thousands of NOEs (see above), it was much more efficient to dock the molecules together using only a few intermolecular NOEs and a few hundred dipolar coupling orientational restraints for each of the proteins. The methods have been implemented in a new computer program (*70*).

Orientational constraints can also be obtained from anisotropic $^{15}N$ relaxation in proteins. The measurements to obtain the relative orientation of domains are much more involved but measure similar parameters. The relative orientation of

loosely connected subdomains in individual proteins, such as the orientation of multiple zinc fingers in a larger transcription factor (*71*) or the more rigid relative orientation of SH2 and SH3 domains in a combined SH2−SH3 construct, have been determined with this method (*72*). The methods hold promise to determine the relative orientation of the components in larger molecular machines as well. Computer programs to carry out the necessary math are available (*73*).

*Conclusions and Future Outlook.* The easiest methods, such as shift mapping and cross-saturation, give answers with the lowest resolution. This is very valuable information if the interaction interface was completely unknown before. If detailed residue-specific knowledge is required, full NOE structure determination, pseudocontact shift, and dipolar methods need to be used. It is likely that especially the latter method will play a major role in the determination of the relative orientation of individual proteins in larger complexes. All methods described can be combined with TROSY, a method that overcomes the line broadening effects that have stood in the way of the application of NMR spectroscopy to really large systems. With TROSY, NMR lines remain narrow for large molecules, and there is theoretically no size limit to a $^{1}H-^{15}N$ shift correlation-type experiment anymore (for reviews, see refs *3* and *4*). Of course, the spectra of larger molecules may become too crowded to analyze, but this can be overcome with selective labeling of individual molecules in a large complex (as above). If the NMR spectra of the individual proteins in the complex are still too complicated to analyze, they can be selectively labeled either at the residue level (*74*) or as subdomains (*75*). Currently, it is difficult to predict what practical upper limit there may be on the determination of protein interfaces by NMR, but the future looks bright: it was recently shown (*76*) that well-resolved $^{1}H-^{15}N$ spectra can be obtained for the labeled GroES (10 kDa) bound to the 60 kDa GroEL in an intermolecular 14-mer. This is a complex of total molecular mass 1000000 Da.

*Appendix.* Proteins can be easily and inexpensively labeled with $^{15}N$ by growing the *Escherichia coli* production strain in a minimal medium, using $^{15}NH_4Cl$ as the sole nitrogen source. Isotopically $^{15}N$-enriched rich media are available as well from several vendors. The most common NMR experiment carried out with a $^{15}N$-labeled protein is the $^{15}N-^{1}H$ heteronuclear single-quantum correlation experiment, or HSQC. The HSQC spectrum is a two-dimensional map that correlates the chemical shifts of the amide protons with the chemical shifts of the attached $^{15}N$ amide nitrogens. The map contains roughly as many cross-peaks as there are $^{15}N$-labeled amino acid residues in the protein or complex and thus serves as an excellent "fingerprint" of the protein.

The fundamentals of the experiment can be described as follows (see Figure 1). The first radio-frequency pulse excites the $^{1}H$ NMR transitions and leaves them in a state of resonance (coherence). For $^{15}N$-coupled protons, this coherence is subsequently transferred to the $^{15}N$ resonance though a combination of the $^{15}N-^{1}H$ scalar coupling and several $^{15}N$ and $^{1}H$ pulses during the time $T_{tr}$ (drawn line). This process of coherence transfer may be likened to a resonance energy transfer, here mediated through the particular combination of scalar coupling (provided by nature) and pulses (provided by the spectroscopist). The coherence is kept on

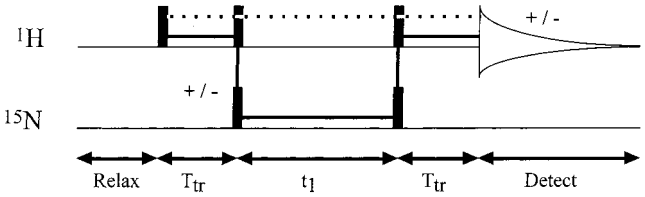


FIGURE 1: Simplified NMR pulse sequence for the $^{1}H-^{15}N$ HSQC experiment, discussed in the Appendix. Black boxes are radio-frequency pulses that excite and transfer NMR coherence. The different time intervals of the experiment are shown at the bottom of the figure. The drawn pathway is followed by coherence deriving from protons coupled to labeled amide nitrogens and the dashed pathway by all other protons. The +/− symbols indicate that the phase (sense) of the first $^{15}N$ pulse and detector are inverted simultaneously at every other experiment.

the $^{15}N$ spins for a variable amount of time (see below) and is subsequently transferred back to the protons through the reverse process, after which it is finally detected and stored. Essential is that the experiment is carried out twice using opposite phases of the $^{15}N$ pulses. This inverts all subsequent NMR signals, including the final $^{1}H$ signal. The results of the two experiments are subtracted, and only those $^{1}H$ resonances that have "experienced" a coherence transfer through the $^{15}N$ resonance (drawn path) have changed sign and are thus added in the difference spectrum. Those are of course those that have a scalar coupling to a $^{15}N$ nucleus, i.e., the amide protons that reside in a $^{15}N$-labeled protein. The other $^{1}H$ signals not coupled to $^{15}N$ do not experience the $^{15}N$ pulses and are subtracted out (dashed pathway). Thus, in a complex of a labeled and unlabeled protein, only the amide resonances of the labeled protein are observed. The one-dimensional version of this experiment is called isotope filter.

By varying the length of the $^{15}N$ resonance time, $t_1$, in a systematic manner, the final $^{1}H$ signal is modulated by the $^{15}N$ chemical shift, and a two-dimensional $^{15}N-^{1}H$ correlation spectrum is obtained after Fourier transformation. For a more complete, but still semiclassical (vector and level diagram) explanation the reader is referred to ref *77* and for a quantum mechanical derivation, to ref *78*.

## REFERENCES

1. Ferentz, A. E., and Wagner, G. (2000) *Q. Rev. Biophys. 33*, 29−36.
2. Venters, R. A., Farmer, B. T., II, Fierke, C. A., and Spicer, L. D. (1996) *J. Mol. Biol. 264*, 1101−1116.
3. Wider, G., and Wüthrich, K. (1999) *Cur. Opin. Struct. Biol. 9*, 537−651.
4. Riek, R., Pervushin, K., and Wüthrich, K. (2000) *Trends Biochem. Sci. 25*, 462−468.
5. Hershey, P. E., McWhirter, S. M., Gross, J. D., Wagner, G., Alber, T., and Sachs, A. B (1999) *J. Biol. Chem. 274*, 21297−21304.
6. McAlister, M. S. B., Mott, H. R., van der Merwe, P. A., Campbell, I. D., Davis, S. J., and Driscoll, P. C. (1996) *Biochemistry 35*, 5982−5991.
7. Mayer, B. J. (2001) *J. Cell Sci. 114*, 1253−1263.
8. Hajduk, P. J., Meadows, R. P., and Fesik, S. W.(1997) *Science 278*, 497−499.
9. Kenworthy, A. K. (2001) *Methods 24*, 289−296.
10. Fesik, S. W., Luly, J. R., Erickson, J. W., and Abad-Zapatero, C. (1988) *Biochemistry 27*, 8297−8301.
11. Otting, G., and Wüthrich, K. (1990) *Q. Rev. Biophys. 23*, 39−96.
12. Spronk, C. A., Bonvin, A. M., Radha, P. K., Melacini, G., Boelens, R., and Kaptein R. (1999) *Struct. Folding Des. 7*, 1483−1492.

13. Varani, L., Gunderson, S. I., Mattaj, I. W., Kay, L. E., Neuhaus, D., and Varani, G. (2000) *Nat. Struct. Biol. 7*, 329−335.
14. Garrett, D. S., Seok, Y. J., Peterkofsky, A., Gronenborn, A. M., and Clore, G. M. (1999) *Nat. Struct. Biol. 6*, 166−173.
15. Fletcher, C. M., and Wagner, G.(1998) *Protein Sci. 7*, 1639−1642.
16. Dhalluin, C., Yan, K. S., Plotnikova, O., Lee, K. W., Zeng, L., Kuti, M., Mujtaba, S., Goldfarb, M. P., and Zhou, M. M. (2000) *Mol. Cell 6*, 921−929.
17. Takahashi, H., Nakanishi, T., Kami, K., Arata, Y., and Shimada, I. (2000) *Nat. Struct. Biol. 3*, 220−223.
18. Hall, D. A., Vander Kooi, C. W., Stasik, C. N., Stevens, S. Y., Zuiderweg, E. R. P., and Matthews, R. G., (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 9521−9526.
19. Pellecchia, M., Stevens, S. Y., Vander Kooi, C. W., Montgomery, D. H., Feng, E. H., Gierasch, L. M., and Zuiderweg, E. R. P. (2000) *Nat. Struct. Biol. 7*, 298−303.
20. Stevens, S. Y., Sanker, S., Kent, C., and Zuiderweg, E. R. P. (2001) *Nat. Struct. Biol. 8*, 947−952.
21. van Nuland, N. A., Kroon, G. J., Dijkstra, K., Wolters, G. K., Scheek, R. M., and Robillard, G. T. (1993) *FEBS Lett. 315*, 11−15.
22. Chen, Y., Reizer, J., Saier, M. H., Jr., Fairbrother, W. J., and Wright, P. E.. (1993) *Biochemistry 32*, 32−37.
23. Garrett, D. S., Seok, Y. J., Peterkofsky, A., Clore, G. M., and Gronenborn, A. M. (1997) *Biochemistry 36*, 4393−4398.
24. Emerson, S. D., Madison, V. S., Palermo, R. E., Waugh, D. S., Scheffler, J. E., Tsao, K. L., Kiefer, S. E., Liu, S. P., and Fry, D. C. (1995) *Biochemistry 34*, 6911−6918.
25. Otting, G. (1993) *Curr. Opin. Struct. Biol. 3*, 760−768.
26. Foster, M. P., Wuttke, D. S., Clemens, K. R., Jahnke, W., Radhakrishnan, I, Tennant, L., Reymond, M., Chung, J., and Wright, P. E. (1998) *J. Biomol. NMR 12*, 51−71.
27. Liew, C. K., Kowalski, K., Fox, A. H., Newton, A. Sharpe, B. K., Crossley, M., and Mackay, J. P. (2000) *Struct. Folding Des. 8*, 1157−1166.
28. Penkett, C. J., Dobson, C. M., Smith, L. J., Bright, J. R., Pickford, A. R., Campbell, I. D., and Potts, J. R. (2000) *Biochemistry 39*, 2887−2893.
29. Kranz, J. K., and Hall K. B. (1999) *J. Mol. Biol. 285*, 215−231.
30. Xu, G. Y., McDonagh, T., Yu, H. A., Nalefski, E. A., Clark, J. D., and Cumming, D. A. (1998) *J. Mol. Biol. 280*, 485−500.
31. Greene, M. K., Maskos, K., and Landry, S. J. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 6108−6113.
32. Pellecchia, M., Sebbel, P., Hermanns, U., Wüthrich, K., and Glockshuber, R. (1999) *Nat. Struct. Biol. 6*, 336−339.
33. Sitkoff, D., and Case, D. A. (1998) *Prog. Nucl. Magn. Reson. Spectrosc. 32*, 165−190.
34. McCoy, M. A., and Wyss, D. F. (2000) *J. Biomol. NMR 18*, 189−198.
35. Morelli, X., Dolla, A., Czjzek, M., Palma, P. N., Blasco, F., Krippahl, L., Moura, J. J., and Guerlesquin, F. (2000) *Biochemistry 39*, 2530−2537.
36. Williamson, R. A., Carr, M. D., Frenkiel, T. A., Feeney, J., and Freedman, R. B. (1997) *Biochemistry 36*, 13882−13889.
37. Muskett, F. W., Frenkiel, T. A., Feeney, J., Freedman, R. B., Carr, M. D., and Williamson, R. A (1998) *J. Biol. Chem. 273*, 21736−21743.
38. Zuiderweg, E. R. P., Hamers, L. F., Rollema, H. S., De Bruin, S. H., and Hilbers, C. W. (1981) *Eur. J. Biochem. 118*, 95−104.
39. Carrington, A., and McLachlan, A. (1967) *Introduction to magnetic resonance with applications to chemistry and chemical physics*, Harper & Row, New York.
40. Mercier, P., Li, M. X., and Sykes, B. D. (2000) *Biochemistry 39*, 2902−2911.
41. Ishima, R., and Torchia, D. A. (2000) *Nat. Struct. Biol. 7*, 740−743.
42. Kay, L. E., Muhandiram, D. R., Wolf, G., Shoelson, S. E., and Forman-Kay, J. D. (1998) *Nat. Struct. Biol. 5*, 156−163.
43. Howard, M. J., Chauhan, H. J., Domingo, G. J., Fuller, C., and Perham, R. N. (2000) *J. Mol. Biol. 295*, 1023−103746.
44. Kay, L. E., Muhandiram, D. R., Farrow, N. A., Aubin, Y., and Forman-Kay, J. D. (1996) *Biochemistry 135*, 361−368.
45. Abbott, M. B., Dvoretsky, A., Gaponenko, V., and Rosevear, P. R. (2000) *FEBS Lett. 469*, 168−172.
46. Wyss, D. F., Dayie, K. T., and Wagner, G. (1997) *Protein Sci. 6*, 534−542.
47. Feher, V. A., and Cavanagh, J. (1999) *Nature 400*, 289−293.
48. Volkman, B. F., Lipson, D., Wemmer, D. E., and Kern, D. (2001) *Science 291*, 2429−2433.
49. Englander, S. W., Mayne, L., Bai, Y., and Sosnick, T. R. (1997) *Protein Sc*i. *6*, 1101−1109.
50. Li, R., and Woodward, C. (1999) *Protein Sci. 8*, 1571−1590.
51. Szewczuk, Z., Konishi, Y., and Goto, Y. (2001) *Biochemistry 40*, 9623−9630.
52. Mueller, G. A., Smith, A. M., Chapman, M. D., Rule, G. S., and Benjamin, D. C. (2001) *J. Biol. Chem. 276*, 9359−9365.
53. Ekiel, I., Banville, D., Shen, S. H., and Gehring, K. (1998) *Biochem. Cell Biol. 76*, 334−340.
54. Engen, J. R., Gmeiner, W. H., Smithgall, T. E., and Smith, D. L. (1999) *Biochemistry 38*, 8926−8935.
55. Arumugam, S., Hemme, C. L., Yoshida, N., Suzuki, K., Nagase, H., Berjanskii, M., Wu, B., and Van Doren, S. R. (1998) *Biochemistry 37*, 9650−9657.
56. Petros, A. M., Mueller, L., and Kopple, K. D. (1990) *Biochemistry 29*, 10041−10048.
57. Hubbell, W. L., Mchaourab, H. S., Altenbach, C., and Lietzow, M. A. (1996) *Structure 4*, 779−783.
58. Scherf, T., Hiller, R., and Anglister, J. (1995) *FASEB J. 9*, 120−126.
59. Wang, H., Pang, Y., Kurochkin, A. V., Hu, W., Flynn, G. C., and Zuiderweg, E. R. P. (1998) *Biochemistry 37*, 7929−7940.
60. Ubbink, M., Ejdeback, M., Karlsson, B. G., and Bendall, D. S. (1998) *Structure 6*, 323−335.
61. Guiles, R. D., Sarma, S., DiGate, R. J., Banville, D., Basus, V. J., Kuntz, I. D., and Waskell, L. (1996) *Nat. Struct. Biol. 3*, 333−339.
62. Worrall, J. A. R., Kolczak, U., Canters, G. W., and Ubbink, M. (2001) *Biochemistry 40*, 7069−7076.
63. Gochin, M. (2000) *Struct. Folding Des. 8*, 441−452.
64. Tolman, J. R., Flanagan J. M., Kennedy, M. A., and Prestegard, J. H. (1997) *Nat. Struct. Biol. 4*, 292−297.
65. Tjandra, N., and Bax, A. (1997) *Science 278*, 1111−1114.
66. Hansen, M. R., Mueller, L., and Pardi, A. (1998) *Nat. Struct. Biol. 5*, 1065−1074.
67. Fischer, M. W. F., Losonczi, J. A., Weaver, J. L., and Prestegard, J. H. (1999) *Biochemistry 38*, 9013−9022.
68. Mueller, G. A., Choy, W. Y., Yang, D., Forman-Kay, J. D., Venters, R. A., and Kay, L. E. (2000) *J. Mol. Biol. 300*, 197−212.
69. Clore, G. M. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 9021−9025.
70. Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001) *J. Biomol. NMR 20*, 223−231.
71. Bruschweiler, R., Liao, X., and Wright, P. E. (1995) *Science 268*, 886−889.
72. Fushman, D., Xu, R., and Cowburn, D. (1999) *Biochemistry 38*, 10225−10230.
73. Dosset, P., Hus, J.-C., Blackledge, M., and Marion, D. (2000) *J. Biomol. NMR 16*, 23−28.
74. McIntosh, L. P., and Dahlquist, F. W. (1990) *Q. Rev. Biophys. 23*, 1−38.
75. Cowburn D., and Muir, T. W. (2001) *Methods Enzymol. 339*, 41−54.
76. Wüthrich, K., Bonjour, S., Ferbndez, C., Fiaux, J., Riek, R., and Pervushin, K. (2001) *Frontiers of NMR in Molecular Biology*, Vol. VII, Big Sky, MT.
77. Zuiderweg, E. R. P., and Van Doren, S. R. (1994) *Trends Anal. Chem. 13*, 24−36.
78. Cavanaugh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy*, Academic Press, New York.

BI011870B