Pure Appl. Chem., Vol. 73, No. 9, pp. 1429–1436, 2001.

NMR spectroscopy in drug design*

Markus Heller[1] and Horst Kessler[1,2,†]

[1]Institut für Organische Chemie und Biochemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany; [2]Novaspin Biotech GmbH, Mühlfeldweg 46, 85748 Garching, Germany

Abstract: The process of preclinical drug discovery consists of two steps: finding of initial hits (binding ligands to a medicinal relevant target, usually a protein) and lead optimization. Nuclear magnetic resonance spectroscopy is a powerful tool that can provide valuable information to every step of drug development. NMR is commonly used for characterizing the structure and molecular dynamics of target or ligand molecules. During the structure-based lead optimization, NMR provides insight into the structural and dynamical properties of the target-ligand complex. Recently, the use of NMR in the lead finding process by screening technologies has been shown. For the latter use, new techniques have also been developed. Those techniques, in combination with high throughput, have lead to an efficient screening of libraries composed of small molecules. In this article, the role of NMR during the discovery of a drug candidate is described.

INTRODUCTION

Over the past years, NMR has been mainly used for the characterization of ligand molecules, regarding their three-dimensional structure and molecular dynamics. Subsequently, those studies have been extended to target molecules. More recently, ligand-receptor complexes have been studied and used for rational improvement of ligands. New applications cover the field of lead finding. Libraries of small compounds, often obtained by combinatorial chemistry, are screened for potential binding to a target molecule.

The general approach for hit finding and ligand optimization can be briefly summarized as follows (see Fig. 1): The target protein of interest is overexpressed and isotope-labeled in bacteria. Subsequently, the three-dimensional structure (and often its molecular dynamics as well) is determined via NMR spectroscopy. The next step is the identification of lead compounds in an NMR screening, preferably in a high-throughput manner. Several experiments have been developed for this purpose, for example, saturation transfer difference spectroscopy (STD) [1–3], (reverse) NOE pumping [4–5], and transferred NOE [6–8], which can be used without isotope-labeled protein. During the lead evolution, the analysis of interactions between the initial hit and the protein leads to an optimized lead structure. This article describes the contribution of NMR to lead finding and lead optimization. For a description of specific NMR experiments and assignment strategies, the reader is referred to the literature.

*Plenary lecture presented at the Hungarian–German–Italian–Polish Joint Meeting on Medicinal Chemistry, Budapest, Hungary, 2–6 September 2001. Other presentations are published in this issue, pp. 1387–1509.

†Corresponding author: E-mail: kessler@ch.tum.de

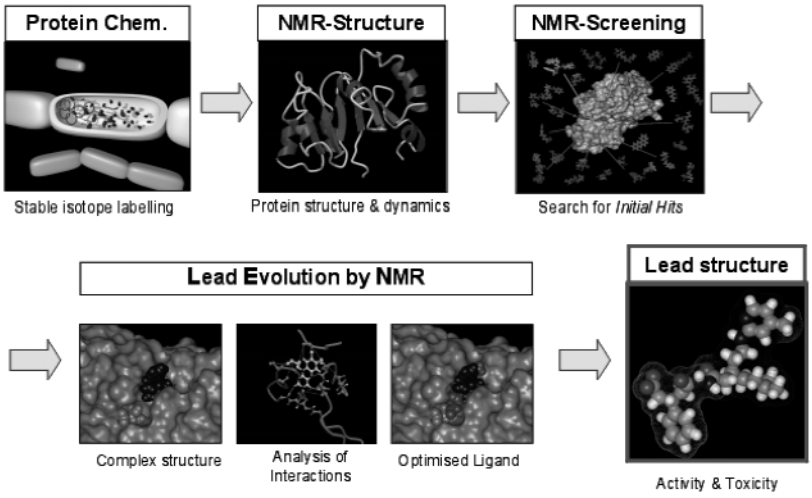

Fig. 1 General approach of an NMR-guided lead finding.

CHARACTERIZATION OF THE TARGET MOLECULE

Most of the target molecules are proteins with a molecular weight above 10 kDa. Isotope labeling with ^{15}N or ^{13}C (and ^{2}H) is normally required. For this purpose, an overexpression system with high cell density and high expression rates of the desired protein is essential. The most common systems for stable isotope labeling are bacteria like *E. coli* and *B. subtillis*, or yeast cells, e.g., *P. pastoris*.

Several isotope-labeling patterns are used (Fig. 2) [9–12]. For most applications, a global labeling with ^{15}N and ^{13}C is sufficient. This labeling enables the NMR spectroscopist to add a third (or fourth) dimension to a two-dimensional $^{1}H,^{1}H$ correlation spectrum. As a result, the signal overlap of the proton spectrum is resolved in the dimension of the heteronucleus, i.e., ^{15}N or ^{13}C attached to the protons of relevance. If fast relaxation becomes a problem, all nonexchangeable protons can be substituted by deuterons. This leads to a significant decrease in relaxation rates and sharper signals. For very large targets, a domain-specific labeling scheme may be helpful where only the corresponding domain delivers NMR signals. This reduces problems with signal overlap. Spectral assignment can also be facilitated with an amino acid-specific labeling strategy or by amino acid-specific pulse sequences [13–14].

After assigning all resonances as far as possible, the secondary structure elements (such as α-helices or β-strands) are determined via scalar couplings, chemical shift indices, and sequential NOE connectivities. Subsequently, the three-dimensional structure is derived from NOE connectivities between secondary structure elements, which represent short-range information [15]. However, additional long-range restraints can be derived from residual dipolar couplings [16].

The aggregation state of the protein may be of relevance for the activity and is related to the three-dimensional structure. For example, the active form of Riboflavin synthase (RiSy), a target for antibacterial drugs, is a trimer [17]. Each monomer consists of two domains with high sequence homology. Hence, RiSy looks like a hexameric protein. When the aggregation state is known from other information, the quaternary structure can be modeled from the NMR structure of the monomeric domain. In case of RiSy, the structure of the N-terminal domain is a dimer, which has been solved by NMR as a complex with two bound ligands (riboflavin; see below and Fig. 3).

All the structural information mentioned above can be established by X-ray crystallography as well if crystals are available. But there is another unique piece of information about a macromolecule that cannot be studied by X-ray crystallography: the dynamical properties. NMR is well suited for studying molecular dynamics [18]. Mobility can be probed by NMR relaxation experiments on differ-

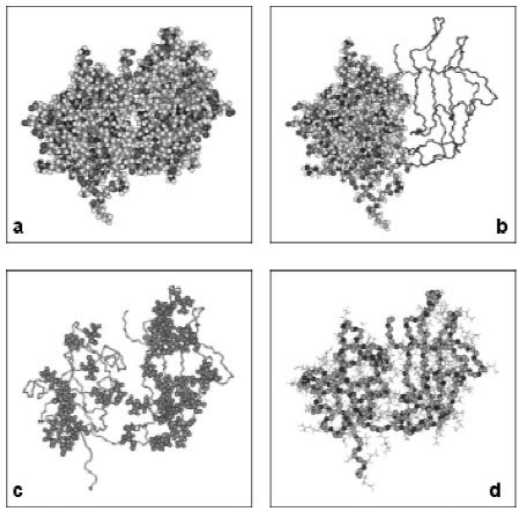

Fig. 2 Illustration of different isotope labeling strategies: a) global labeling of the whole protein (^{15}N, ^{13}C, ^{2}H); b) domain-specific labeling to reduce the number of signals in NMR spectra (this approach is useful for large targets); c) spectral assignment is facilitated with amino-acid specific labeling; d) position-specific labeling (all nonexchangeable protons are substituted by deuterons) helps to reduce detrimental relaxation effects.

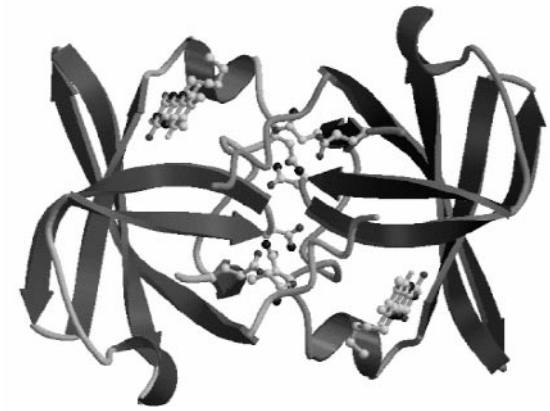

Fig. 3 The NMR-structure of the N-terminal riboflavin synthase (RiSy) homodimer. Two riboflavin ligands are shown as ball-and-stick models.

ent time scales ranging from fast motions in the ns-region to slow motions in the ms-range. Molecular motion in the binding site is of outstanding importance for the kinetics of protein ligand complex formation and dissociation. It has been shown that slow motions (millisecond to microsecond range) are of greater importance for ligand binding than fast motions [19–20]. Furthermore, the mobility of the binding site determines the structural range of ligands that can be bound. Upon ligand binding, local dynamics of the protein are perturbed (see below). The solvent accessibility can be used to map solvent exposed residues of a protein. Regions with poor solvent accessibility are often buried in a hydrophobic core and are, therefore, less accessible for ligands. Solvent accessibility can be probed either with paramagnetic relaxation [21–22] or exchange rates of amide protons [23–24].

NMR SCREENING

Once the drug target is characterized, initial hits have to be found. Several NMR techniques are applicable for this purpose; some of them can be performed in a high-throughput manner. For a detailed discussion of the techniques, the reader is referred elsewhere [25–28]. In this paragraph, only the basis of NMR screening is described.

The effects resulting from the binding of a ligand to a protein can be divided into two classes: global and local effects. The former are size-dependent and therefore are suited to the observation on the ligand, while the latter are restricted to the binding region and can be monitored on both the target and the ligand (see Fig. 4).

Upon binding to a macromolecule, the apparent molecular weight and hydrodynamic radius of a small ligand change dramatically by several orders of magnitude. Those changes can be detected with several different NMR experiments. Relaxation filtering and diffusion editing reveal binding, but cannot provide any structural information [29]. In contrast, NOE-based methods (NOE pumping, reverse NOE pumping) or STD not only prove interactions, in addition they can be used to characterize the binding epitope of the ligand [2]. Ligand-based methods do not require labeled protein, since ligand signals are observed. Another advantage is the required amount of protein, which is only a fraction compared to that of protein-detected methods. Some methods use spin-labeling techniques for second site screening or covalent modification of the target with a spin label near the active site, which leads to another dramatic decrease in the amount of protein needed [30–31]. For searching initial hits, ligand-based methods can be performed in a high-throughput manner. In medicinal chemistry, possible drug candidates are often stored in libraries. With ligand-based techniques, those libraries can be directly screened as single compounds or in mixtures of about 5 to 10 compounds. Binding ligands are directly identified via their NMR spectra [32–35]. Time-consuming deconvolution steps of the mixtures are not necessary.

However, those methods also have disadvantages: In contrast to heteronuclear single quantum correlation (HSQC)-based methods (see below), observation of the ligand signals cannot distinguish

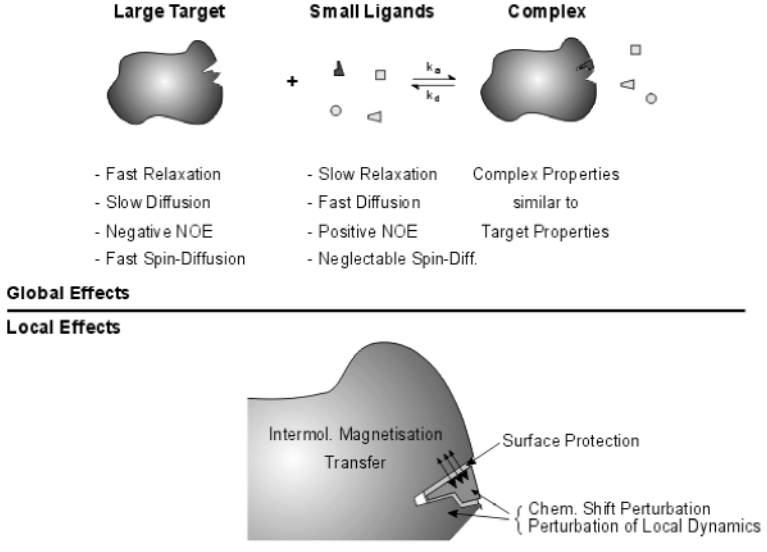

Fig. 4 The effects that arise from binding of a ligand to a target protein are illustrated. The upper part shows the global effects that are advantageously monitored on the ligand. Local effects (lower part) can be detected on both the ligand and the target.

between specific and unspecific binding. Another drawback of the techniques discussed above is the fact that they are based on the detection of the dissociated free form of the ligand. This fact implies a minimal dissociation constant (Kd) of ca. 10^{-7} M. In addition, a sufficiently large difference between the molecular weight of the small compounds and the target molecule is required, which is usually no problem in practice. The size of the protein is of no interest; the target may even be immobilized or bound in lipid vesicles [2,36].

In target-based experiments, the molecular weight of the protein is restricted to an upper limit of approximately 40 kDa (in case of oligomer proteins, larger molecular sizes are accessible). For larger macromolecules, relaxation rates are too fast and spectra suffer from poor resolution. Those effects can be partly overcome by special techniques, but those suffer from intensity losses and therefore require either a larger amount of protein or longer measuring time [37–39]. In contrast to ligand-based experiments, observation of the target molecule is not restricted to an affinity limit. Target-based methods usually rely on the acquisition of $^{1}H,^{13}C$ or $^{1}H,^{15}N$ HSQC spectra. HSQC-based methods allow a differentiation between specific and unspecific binding. If the spectral assignment of the protein is known, *ligand-induced changes in chemical shifts* of target resonances can be mapped on the target structure [40]. With this information, more detailed information about the binding site is provided. An alternative approach relies on the *perturbation of local dynamics* of the protein upon ligand binding. A comparison of the dynamics of the uncomplexed protein with the protein dynamics in presence of a ligand reveals an intermolecular interaction. Additionally, *exchange rates of amide protons* in a complexed protein can be measured and compared to those of the uncomplexed target molecule.

Ligands to be screened may be either acidic or basic. Upon addition of these substances to the protein sample, pH changes in the sample may occur that cannot be compensated for by the buffer. In this case, pH-sensitive protein resonances have to be known to avoid false positive hits. Such pH effects are in general of no relevance for ligand-based screenings.

LEAD EVOLUTION BY NMR

Usually, the first step of the lead evolution is the determination of the ligand structure in the protein ligand complex. In the early stages of lead evolution, ligands usually bind weakly to the target molecule. In this case, rapid exchange between the free and the bound form of the ligand occurs. For fast exchanging systems, the conformation in the bound ligand state can be determined via transferred NOE experiments [6–8]. For some applications, for example, distinguishing between *syn* and *anti* conformation of glycosidic bonds in nucleotides, a qualitative analysis of transferred NOE experiments is sufficient. For a complete description of the ligand conformation in the bound state, a more quantitative analysis is necessary. In this case, it has to be established that the observed NOE intensities are free from exchange contributions (i.e., the condition of fast exchange is fulfilled) or indirect magnetization transfer involving several protons ("spin diffusion"). The technique described permits the bound ligand conformation to be determined even if the protein is not amenable to NMR. If the ligand binds tightly to the target (i.e., slow exchange), isotope labeling is required for the differentiation between ligand and protein resonances. Ligand signals are separated from protein signals with dedicated "editing" experiments. "Editing" experiments also provide a powerful tool for the structure determination of a protein ligand complex. For this purpose, a complete spectral assignment of the protein is required and readily available for proteins up to 40 kDa.

In the study of the N-terminal domain of RiSy, a $^{15}N,^{13}C$-labeled N-terminus of RiSy and a nonlabeled ligand (Riboflavin) were examined in an isotope-filtered, ^{13}C-edited NOESY experiment [9]. This experiment connects carbon atoms of the protein with protons of the ligand via the following scheme: After frequency labeling of the carbon nuclei, magnetization from protons directly attached to the carbons (not observed) is transferred to protons of the binding ligand. The effect of magnetization transfer through the space is called "nuclear Overhauser effect" (NOE). Figure 5 shows a scheme of

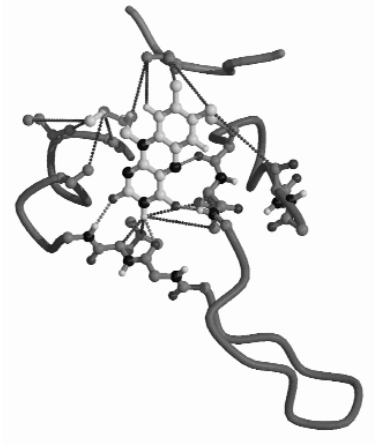

Fig. 5 Binding mode of riboflavin to RiSy. Experimental NOE contacts are shown in dark gray, H-bonds derived from molecular modeling are light gray.

riboflavin bound to the N-terminal domain of RiSy. Experimental NOE contacts obtained from such an isotope-filtered experiment and calculated H-bonds are indicated.

When the structure of the protein-ligand complex is known, the intermolecular interactions can be analyzed by using computational chemistry. Molecular modeling of binding pockets or H-bond donors/acceptors in the interaction region provides further insight in the interactions; virtual docking of a ligand into the binding pocket suggests structural modifications of the lead. It is evident that this procedure fails in cases of induced fits with larger conformational changes in the target, the ligand, or both.

If two binding sites on a target molecule in spatial proximity are occupied by two different ligands, the normally weakly binding ligands can be linked together to optimize for highest binding affinity (SAR by NMR) [40].

RECENT DEVELOPMENTS IN THE FIELD OF NMR

In the past few years, much progress was made for improving NMR as a powerful tool for industrial drug research. For the determination of the three-dimensional structure of a target protein, the spectral assignment has to be as complete as possible. Since manual assignment of all peaks is very time-consuming, software packages have been developed to facilitate this task [41–42]. The introduction of cryogenic probe heads offers two different options. The required amount of substances can be lowered while keeping the experimental time the same. Alternatively, the experimental time can be shortened by a factor of about 8, which leads to a significant increase in throughput. For the data acquisition, hardware setups have been developed that allow a "just-in-time" preparation of the samples including the transfer to the magnet, locking and shimming. However, the "rate-determining step" is not the acquisition of the NMR data but its analysis. As far as the evaluation of screening data is concerned, methods have been described that calculate similarities between spectra and therefore allow for the detection of binding ligands [43]. In addition to these hard- and software developments, numerous new pulse sequences have been deduced over the years that improve previous technology in terms of sensitivity, resolution, selectivity, speed, and efficiency. Several "new" NMR parameters, such as residual dipolar coupling (RDC) [16], cross-correlated relaxation [44], "unusual" non-Karplus-type coupling constants, or scalar couplings across hydrogen bonds [45–48] provide more and more information about the interacting molecules. The next years will show if they can be used in routine work and substitute conventional structural parameters (NOE, J-couplings) or if they are only applied in special cases where more detailed information is needed.

REFERENCES

1. M. Mayer and B. Meyer. *J. Am. Chem. Soc*. **123**, 6108 (2001).
2. R. Meinecke and B. Meyer. *J. Med. Chem*. **44**, 3059 (2001).
3. M. Mayer and B. Meyer. *Angew. Chem*. **111**, 1902 (1999).
4. A. Chen and M. J. Shapiro. *J. Am. Chem. Soc*. **120**, 10258 (1998).
5. A. Chen and M. J. Shapiro. *J. Am. Chem. Soc*. **122**, 414 (2000).
6. L.-Y. Lian, I. L. Barsukov, M. J. Sutcliffe, K. H. Sze, G. C. K. Roberts. *Methods Enzymol*. **239**, 657 (1994).
7. J. Feeney, B. Birdsall, G. C. K. Roberts, A. S. Burgen. *Biochemistry* **22**, 628 (1983).
8. I. L. Barsukov, L.-Y. Lian, J. Ellis, K. H. Sze, W. V. Shaw, G. C. K. Roberts. *J. Mol. Biol*. **262**, 543 (1996).
9. F. C. L. Almeida, G. C. Amorim, V. H. Moreau, V. O. Sousa, A. T. Creazola, T. A. Américo, A. P. N. Pais, A. Leite, L. E. S. Netto, R. J. Giordano, A. P. Valente. *J. Magn. Reson*. **148**, 142 (2001).
10. N. K. Goto and L. E. Kay. *Curr. Opin. Struct. Biol*. **10**, 585 (2000).
11. T. Otomo, N. Ito, Y. Kyogoku, T. Yamazaki. *Biochemistry* **38**, 16040 (1999).
12. T. Otomo, K. Teruya, K. Uegaki, T. Yamazaki, Y. Kyogoku. *J. Biomol. NMR* **14**, 105 (1999).
13. M. Schubert, H. Oschkinat, P. Schmieder. *J. Magn. Reson*. **148**, 61 (2001).
14. M. Schubert, M. Smalla, P. Schmieder, H. Oschkinat. *J. Magn. Reson*. **141**, 34 (1999).
15. K. Wüthrich. *NMR of Proteins and Nucleic Acids*, Wiley, New York (1986).
16. J. H. Prestegard, H. M. al-Hashimi, J. R. Tolman. *Q. Rev. Biophys.* **33**, 371 (2000).
17. V. Truffault, M. Coles, T. Diercks, K. Abelmann, S. Eberhardt, H. Lüttgen, A. Bacher, H. Kessler. *J. Mol. Biol*. **309**, 949 (2001).
18. L. E. Kay, *Biochem. Cell. Biol*. **76**, 145 (1998).
19. O. Jardetzky, *Prog. Biophys. Mol. Biol*. **65**, 171 (1996).
20. M. Akke, J. Lin, J. Cavanagh, H. P. Erickson, A. G. Palmer III. *Nat. Struct. Biol*. **5**, 55 (1998).
21. S. W. Fesik, G. Gemmecker, E. T. Olejniczak, A. M. Petros. *J. Am. Chem. Soc*. **113**, 7080 (1991).
22. A. M. Petros, L. Mueller, K. D. Kopple. *Biochemistry* **29**, 10041 (1990).
23. S. Koide, W. Jahnke, P. E. Wright. *J. Biomol. NMR* **6**, 306 (1995).
24. G. Gemmecker, W. Jahnke, H. Kessler. *J. Am. Chem. Soc*. **115**, 11620 (1993).
25. T. Diercks, M. Coles, H. Kessler. *Curr. Opin. Chem. Biol*. **5**, 285 (2001).
26. G. C. K. Roberts. *Drug Discov. Today* **5**, 230 (2000).
27. J. M. Moore. *Curr. Opin. Biotechnol*. **115**, 11620 (1999).
28. J. M. Moore. *Biopolymers* **51**, 221 (1999).
29. P. J. Hajduk, E. T. Olejniczak, S. W. Fesik. *J. Am. Chem. Soc*. **119**, 12257 (1997).
30. W. Jahnke, S. Rüdisser, M. Zurini. *J. Am. Chem. Soc*. **123**, 3149 (2001).
31. W. Jahnke, L. B. Perez, C. G. Paris, A. Strauss, G. Fendrich, C. M. Nalin. *J. Am. Chem. Soc*. **122**, 7394 (2000).
32. H. Barjat, G. A. Morris, S. Smart, A. G. Swanson, S. C. R. Williams. *J. Magn. Reson. B* **108**, 170 (1995).
33. K. Bleicher, M. Lin, M. J. Shapiro, J. R. Wareing. *J. Org. Chem*. **63**, 8486 (1998).
34. M. Lin, M. J. Shapiro, J. R. Wareing. *J. Org. Chem*. **62**, 8930 (1997).
35. M. Lin and M. J. Shapiro. *J. Org. Chem*. **61**, 7617 (1996).
36. J. Klein, R. Meinecke, M. Mayer, B. Meyer. *J. Am. Chem Soc*. **121**, 5336 (1999).
37. R. Riek, G. Wider, K. Pervushin, K. Wüthrich. *Proc. Natl. Acad. Sci. USA* **96**, 4918 (1999).
38. M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wüthrich. *Proc. Natl. Acad. Sci. USA* **95**, 13585 (1998).
39. K. Pervushin, R. Riek, G. Wider, K. Wüthrich. *Proc. Natl. Acad. Sci. USA* **94**, 12366 (1997).
40. S. B. Shuker, P. J. Hadjuk, R. P. Meadows, S. W. Fesik. *Science* **274**, 1531 (1996).

41. M. Leutner, R. M. Gschwind, J. Liermann, C. Schwarz, G. Gemmecker, H. Kessler. *J. Biomol. NMR* **11**, 31 (1998).
42. H. N. Moseley and G. T. Montelione. *Curr. Opin. Struct. Biol*. **9**, 635 (1999).
43. A. Ross and H. Senn. *Drug Discov. Today* **6**, 583 (2001).
44. H. Schwalbe, T. Carlomagno, M. Hennig, J. Junker, B. Reif, C. Richter, C. Griesinger. *Methods Enzymol*. **338**, 35 (2001).
45. A. J. Dingley and S. Grzesiek. *J. Am. Chem Soc*. **120**, 8293 (1998).
46. K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho, K. Wüthrich. *Proc. Natl. Acad. Sci. USA* **95**, 14147 (1998).
47. G. Cornilescu, J.-S. Hu, A. Bax. *J. Am. Chem Soc*. **121**, 2949 (1999).
48. F. Cordier and S. Grzesiek. *J. Am. Chem Soc*. **121**, 1601 (1999).