Automated NMR Resonance Assignment of Large Proteins for Protein–Ligand Interaction Studies

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · DECEMBER 2010

Impact Factor: 12.11 · DOI: 10.1021/ja108383x · Source: PubMed

3 AUTHORS:

Alvar Gossert
Novartis
16 PUBLICATIONS **316** CITATIONS
SEE PROFILE

Sebastian Hiller
University of Basel
60 PUBLICATIONS **1,597** CITATIONS
SEE PROFILE

César Fernández
Novartis
36 PUBLICATIONS **1,638** CITATIONS
SEE PROFILE

J|A|C|S COMMUNICATIONS

Published on Web 12/16/2010

Automated NMR Resonance Assignment of Large Proteins for Protein−Ligand Interaction Studies

Alvar D. Gossert,[†] Sebastian Hiller,[‡] and César Fernández*,[†]

Novartis Institutes for Biomedical Research, Novartis Pharma AG, 4002 Basel, Switzerland, and Biozentrum, University of Basel, 4056 Basel, Switzerland

Received September 16, 2010; E-mail: cesar.fernandez@novartis.com

Abstract: The detection and structural characterization of protein−ligand interactions by solution NMR is central to functional biology research as well as to drug discovery. Here we present a robust and highly automated procedure for obtaining the resonance assignments necessary for studies of such interactions. The procedure relies on a combination of three automated projection spectroscopy (APSY) experiments, including the new 4D APSY-HNCACB, and the use of fractionally deuterated protein samples. This labeling pattern increases the experimental sensitivity on the one hand, but it leads to peak multiplets on the other hand. The latter complications are however overcome by the geometric APSY analysis of the projection spectra. The three APSY experiments thus provide high precision chemical shift correlations of the backbone and side chain methyl groups, allowing a reliable and robust assignment of the protein by suitable algorithms. The present approach doubles the molecular size limit of APSY-based assignments to 25 kDa, thus providing the basis for efficient characterization of protein−ligand interactions at atomic resolution by NMR, such as structure-based drug design. We show the application to two human proteins with molecular weights of 15 and 22 kDa, respectively, at concentrations of 0.4 mM and discuss the general applicability to studies of protein−protein and protein−nucleic acid complexes.

Structural characterization of protein−ligand interactions by NMR requires sequence-specific resonance assignment of the protein.[1] The assignment of backbone and side chain resonances by conventional methods typically requires two weeks of instrument time and up to several weeks of interactive spectra analysis. Efficient assignment procedures based on conventional triple-resonance experiments or reduced dimensionality spectroscopy using semi-automated analysis have enabled considerably shorter durations[2−10] but usually still require either long acquisition or analysis times. Therefore, the application of NMR in industrial settings is generally limited to ligand binding detection and no structural characterization of the protein−ligand interaction is performed.

Automated projection spectroscopy (APSY) provides a method for efficient automatic resonance assignment of proteins.[11] In an APSY experiment, chemical shift correlations for a high-dimensional NMR experiment are calculated geometrically from low-dimensional projections by the dedicated algorithm GAPRO. One of its main advantages is the robustness in delivering precise, high-dimensional peak lists, which are devoid of artifacts.[12] The high dimensionality of the APSY experiments allows the use of a minimal number of experiments and resolves overlap problems. With the APSY peak lists, assignment algorithms obtain sequence-specific resonance assignments with high reliability.[11,13] APSY-based backbone assignment has so far been reported for proteins up to 12 kDa in a fully automated way[11] and up to 19 kDa in a semi-interactive procedure.[14] The limiting factor for the amenability of APSY to larger proteins is the low sensitivity of the high-dimensional NMR experiments for high molecular weight proteins rather than spectral crowding.[12] Thus, in order to further expand the size limit for the APSY-based automated analysis, the sensitivity of the APSY experiments has to be improved. In this work, we present a new protocol for APSY-based automated assignments of the backbone and the aliphatic side chain resonances of large,

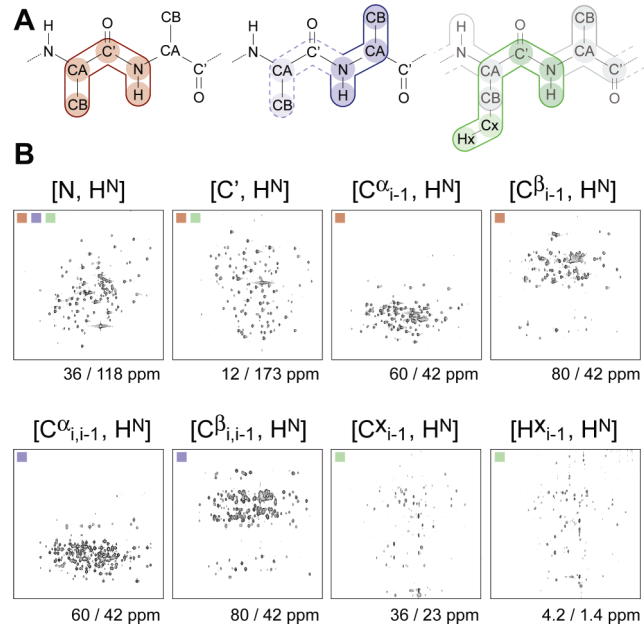

Figure 1. The APSY CB-CA-CM sequential resonance assignment strategy. (A) Chemical shift correlations obtained by the 5D APSY-HNCOCACB[15] (red), the 4D APSY-HNCACB (blue), and the 5D APSY-HC(CC-TOCSY)CONH[16] (green). Each correlation contains the nuclei marked by colored circles. The 4D APSY-HNCACB gives rise to an intra- and interresidual correlation (solid and dashed line, respectively). For the 5D APSY-HC(CC-TOCSY)CONH experiment, H^x-C^x denotes an aliphatic moiety. (B) Subset of 2D APSY projection spectra of the 22 kDa kRas. The small colored squares denote the APSY experiment type in the same color code as in (A). The first two projections are recorded in more than one experiment, indicated by multiple squares. The projections correspond to correlations of a single nucleus in the indirect dimension with the amide proton in the direct dimension, as indicated above each spectrum. The sweep widths/carrier frequencies in the indirect dimensions are given below the spectra. The direct dimension shows $\delta(^1H^N)$ of the range of 6−10 ppm. The individual projection spectra of the 5D and 4D experiments were recorded in 80 and 153 min, respectively, at a temperature of 23 °C on a 600 MHz spectrometer equipped with a cryogenic probe head.

[†] Novartis Pharma AG.
[‡] University of Basel.

10.1021/ja108383x

partially deuterated proteins, which enables the characterization of protein−ligand complexes at atomic resolution in a fast and efficient way (Figure 1).

For the backbone assignment, we selected the "CA-CB" assignment strategy (Figure 1A).[17] The CA-CB strategy requires the recording of two APSY NMR experiments, the 5D APSY-HNCOCACB[15] and the previously undescribed 4D APSY-HNCACB (Figure 2). In this experiment, each backbone amide

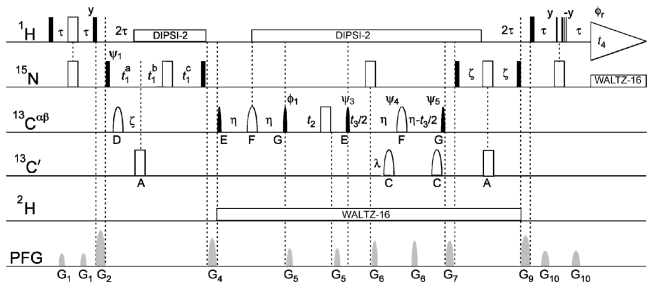

Figure 2. Pulse sequence used for the 4D APSY-HNCACB experiment. The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts are resolved in two separate dimensions,[18] t_3 and t_2, respectively, using two constant-time periods on C^{α} ($2\eta = 13.6$ ms). All experimental details are included in the Supporting Information.

moiety gives rise to a pair of 4D chemical shift correlations, combining the ^1H_i, $^{15}N_i$, $^{13}C\alpha_i$, $^{13}C\beta_i$ and the ^1H_i, $^{15}N_i$, $^{13}C\alpha_{i-1}$, $^{13}C\beta_{i-1}$ chemical shifts, respectively. For glycine residues the C^{α} chemical shift is recorded in both the C^{α} and C^{β} dimensions. The main advantage of the APSY 4D-HNCACB compared to a conventional 3D HNCACB experiment is that the direct result from the APSY experiment is a precise, complete, and artifact-free 4D peak list, in which the C^{α} and C^{β} chemical shifts are readily correlated. Backbone assignment is achieved by combining the APSY 5D and 4D peak lists to fragments and then by mapping these fragments onto the sequence based on their C^{α} and C^{β} chemical shifts. The availability of two chemical shifts leads to minimal ambiguity when building the spin system fragments, and the characteristic C^{β} chemical shift information allows reliable mapping of even short fragments on the amino acid sequence.[17] Furthermore, for large proteins, the CA-CB strategy has the highest possible sensitivity, and it is compatible with backbone deuteration, which is also a requirement for larger target molecules (Figure 1B, Table S1). These features make the CA-CB strategy the method of choice for automated backbone assignment of proteins above 20 kDa.

Side chain chemical shift correlations are recorded with the 5D APSY-HC(CC-TOCSY)CONH experiment, and they are added to the assigned backbone by matching $^{13}C'$, ^{15}N, and $^1H^N$ chemical shifts with the ALASCA algorithm.[16] For the purpose of ligand binding characterization, among all amino acid side chains we focus on the methyl groups of Ile, Val, Leu, Thr, and Ala residues. These moieties feature high intrinsic sensitivity and are often located in positions structurally close to the bound ligand. Therefore, a fractionally deuterated protein sample is used, produced by *E. coli* cells in ^{15}N-labeled minimal medium with 99.8% D_2O and [*U*-^1H,^{13}C]-glucose.[19,20] With this protocol, methyl groups are protonated to 50−100%, but the α position is protonated to 5−8%.[21] The backbone deuteration increases the sensitivity of the experiments substantially, and a single sample can thus be used for both backbone and side chain assignment. For example, the 5D APSY-HNCOCACB experiment with a fractionally deuterated/protonated sample of the 22 kDa protein kRas has a sensitivity gain of 10 over a fully protonated sample (Figure 3A and B).[18]

A drawback of the fractional labeling scheme is a resulting heterogeneous deuteration, leading to multiple NMR signals for

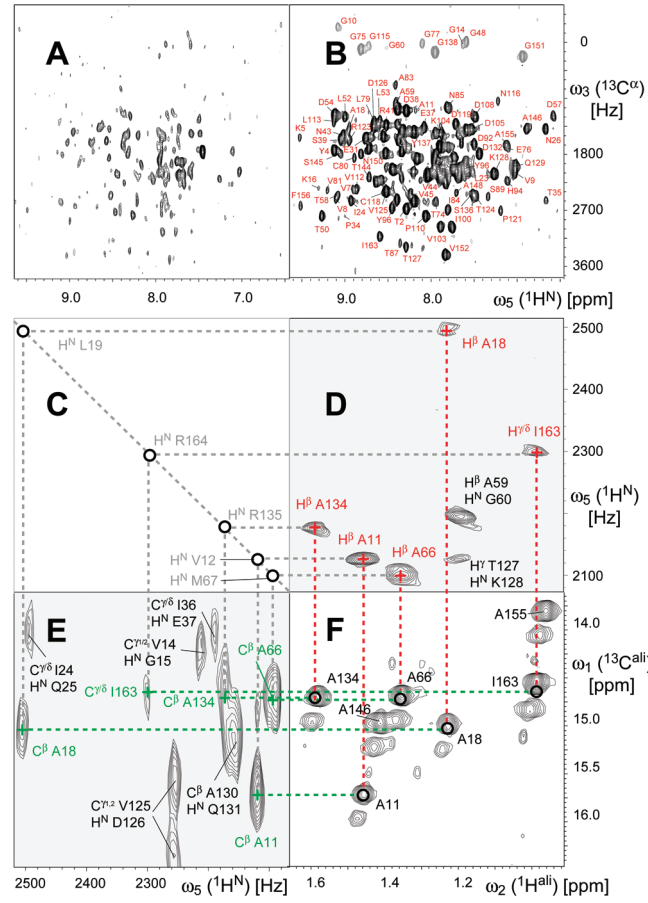

Figure 3. Effects of fractional deuteration on backbone and side chain spectra. (A) The (0°, 0°, 0°) 2D-projection of the 5D APSY-HNCOCACB experiment[15] of protonated kRas and (B) same for 80% fractionally deuterated kRas.[19] Differences in intensity of a factor of 10 are achieved for several resonances. Selected C^{α} assignments are indicated in red. (C−F) Disentangling isotope shift multiplets of methyl resonances with APSY geometric analysis. 2D projection spectra (gray background) from the 5D APSY-HC(CC-TOSCY)CONH experiment, where the $^{13}C^{ali}_{i-1}$ (green, E) and the $^1H^{ali}_{i-1}$ (red, D) chemical shifts are correlated with the $^1H^N_i$ chemical shift respectively. (F) Part of the methyl region of a [^{13}C,^1H] ct-HSQC spectrum of kRas (1−180), where the triplets due to the different deuterium isotope shifts of CH_3, CH_2D, and CHD_2 isotopomers are resolved. Note that in the APSY projections D and E the multiplet structures are not resolved, but precise peak positions are nevertheless calculated (dashed lines). Black circles mark the final APSY peak positions, which are shifted from the CH_3 peak positions according to the admixture of the other methyl isotopomers. In C, H^N diagonal peak positions are illustrated. All spectra were recorded with same experimental conditions as those in Figure 1B.

CH_3/CH_2D/CHD_2 isotopomer moieties due to isotope shifts (Figure 3F). This potential setback is however overcome during the APSY analysis by recording the individual projection spectra of the 5D APSY-HNCOCACB at reduced digital resolution, thereby collapsing the triplets into a single, broader resonance peak (Figure 3E). Since the five-dimensional APSY peak position is obtained from all projections by geometrical averaging, the resulting precision is still very high (Figure 3C−F). We term the resulting approach for automated backbone and side chain methyl assignment APSY CA-CB-CM.

We demonstrate the feasibility of the APSY CA-CB-CM approach with two human proteins that are typical targets for NMR drug discovery projects, the 22 kDa protein kRas at 0.4 mM concentration and a 15 kDa drug target protein (protein A) at 0.3 mM concentration. These data sets were recorded using standardized sets of projection angles (Tables S2−S7), following the generally

valid sensitivity considerations for projection spectra given in ref 11, which include the use of sweep-width matching angles to optimize the spectral resolution.[9,11] For each of the two proteins, 76 h of instrument time were used for the two backbone experiments (13 projections of the 4D APSY-HNCACB in 33 h and 32 projections of the 5D APSY-HNCOCACB in 43 h) and an additional 44 h for the side chain experiment (34 projections of the 5D APSY-HC(CC-TOCSY)CONH), summing up to a total of 5 days. The backbone resonances and side chain C−H moieties were analyzed and assigned to the protein spin systems in a matter of minutes by routines in CARA and MATCH.[13] Data handling, including processing of spectra, brief visual inspection of results of the different algorithms, and format changes of data amounted to half a day of total human interaction.

The overall completeness of the backbone resonance assignment was 95% for the 22 kDa protein kRas and 98% for the 15 kDa protein A (Figures 4 and S2). The missing assignments comprise

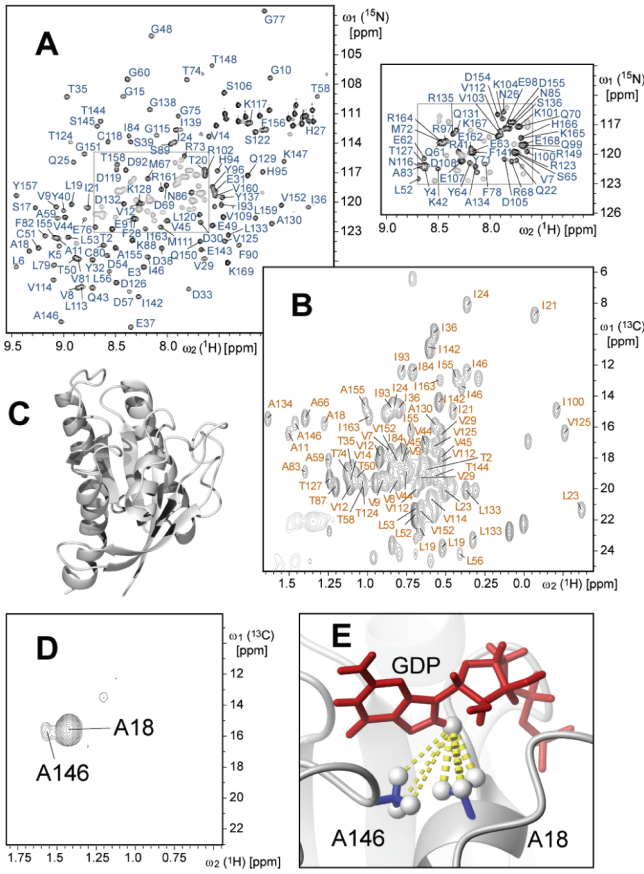

Figure 4. Structural characterization of the kRas−GDP protein−ligand interaction based on APSY experiments. (A) APSY-based backbone and (B) APSY-based side chain resonance assignments were used to assign intermolecular NOE signals (D) between [*U*-^{13}C,^{15}N]-kRas and the unlabeled ligand GDP. (D) 2D plane from a 3D NOESY spectrum, which is ^{13}C/^{15}N-edited in dimensions ω_1/ω_2 and ^{13}C/^{15}N-filtered in the direct dimension ω_3. An expansion of a plane at the GDP H8 resonance in ω_3 (7.9 ppm) is shown. (E) NOEs (yellow broken lines) between the hydrogen atoms of GDP and of methyl groups of kRas (white spheres) are shown in a structural model based on the solution structure (C; PDB 1aa9).[22]

the stretches 46−48 and 74−79 in kRas and two residues in protein A due to unfavorable protein dynamics in these segments. Further, for both proteins, nearly 90% of the Ala, Ile, Leu, Thr, and Val side chain methyl groups were assigned (Figure 4B).

These assignments provide the basis for the efficient characterization of different protein−ligand interactions. In the first example, the APSY assignments enabled the structural characterization of the kRas−GDP complex with intermolecular NOEs, determining the orientation of the ligand GDP in the binding pocket (Figure 4D and 4E). In the second example, a set of compounds that bind protein A had previously been identified in a large library screen.[23,24] Using the APSY resonance assignments, their precise binding site within the extensive interaction surface of the protein was determined by chemical shift mapping (Figures S1 and S2).

In summary, the CA-CB-CM resonance assignment procedure presented here relies on APSY experiments with an extended size limit. The successful application of the process to the 22.4 kDa protein kRas at 0.4 mM protein concentration implies that routine backbone and side chain methyl assignments can be expected for proteins up to at least 25 kDa at 1 mM concentration, thus covering the statistical majority of proteins currently studied by solution NMR spectroscopy (Figure S3). Five days of experiment time and a single protein sample are sufficient for the entire assignment process, thus overcoming the time-consuming bottleneck of resonance assignments even in high-throughput settings. The APSY assignments obtained from the CA-CB-CM procedure provide the basis for the subsequent characterization of protein−ligand and protein−protein interactions, notably enabling structure-based drug design by NMR. Further applications, such as for automated protein structure determination, are readily imaginable.

Acknowledgment. We thank Gerhard Wider, Chrystèle Henry, Armin Widmer, and Alain Dietrich for discussions and technical support and Andrea Saenger and René Hemmig for the protein A sample.

Supporting Information Available: Table on parameters leading to the choice of the overall strategy (Table S1); PDB statistics (Figure S3); HSQC spectra of protein A with its natural ligand and chemical shift mapping of fragment hits (Figures S1 and S2); experimental details on sample preparation, NMR spectroscopical parameters including projection angles and data analysis (Tables S2−S7). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Wüthrich, K. *NMR of proteins and nucleic acids*; Wiley: New York, 1986.
(2) Lescop, E.; Brutscher, B. *J. Biomol. NMR* **2009**, *44*, 43–57.
(3) Liu, G. H.; Shen, Y.; Atreya, H. S.; Parish, D.; Shao, Y.; Sukumaran, D. K.; Xiao, R.; Yee, A.; Lemak, A.; Bhattacharya, A.; Acton, T. A.; Arrowsmith, C. H.; Montelione, G. T.; Szyperski, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10487–10492.
(4) López-Méndez, B.; Güntert, P. *J. Am. Chem. Soc.* **2006**, *128*, 13112–13122.
(5) Moseley, H. N. B.; Monleon, D.; Montelione, G. T. *Methods Enzymol.* **2001**, *339*, 91–108.
(6) Narayanan, R. L.; Dürr, U. H.; Bibow, S.; Biernat, J.; Mandelkow, E.; Zweckstetter, M. *J. Am. Chem. Soc.* **2010**, *132*, 11906–11907.
(7) Staykova, D. K.; Fredriksson, J.; Bermel, W.; Billeter, M. *J. Biomol. NMR* **2008**, *42*, 87–97.
(8) Yee, A.; Gutmanas, A.; Arrowsmith, C. H. *Curr. Opin. Struct. Biol.* **2006**, *16*, 611–617.
(9) Jaravine, V. A.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2006**, *128*, 13421–13426.
(10) Jaravine, V. A.; Zhuravleva, A. V.; Permi, P.; Ibraghimov, I.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2008**, *130*, 3927–3936.
(11) Fiorito, F.; Hiller, S.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **2006**, *35*, 27–37.
(12) Hiller, S.; Fiorito, F.; Wüthrich, K.; Wider, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10876–10881.
(13) Volk, J.; Herrmann, T.; Wüthrich, K. *J. Biomol. NMR* **2008**, *41*, 127–138.
(14) Gossert, A. D.; Hiller, S.; Fiorito, F.; Wüthrich, K. *J. Biomol. NMR* **2007**, *38*, 195–195.
(15) Hiller, S.; Wasmer, C.; Wider, G.; Wüthrich, K. *J. Am. Chem. Soc.* **2007**, *129*, 10823–10828.
(16) Hiller, S.; Joss, R.; Wider, G. *J. Am. Chem. Soc.* **2008**, *130*, 12073–12079.
(17) Hiller, S.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **2008**, *42*, 179–195.
(18) Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 11655–11666.
(19) Leiting, B.; Marsilio, F.; O'Connell, J. F. *Anal. Biochem.* **1998**, *265*, 351–355.
(20) Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E.; Pawson, T.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 627–636.

(21) Otten, R.; Chu, B.; Krewulak, K. D.; Vogel, H. J.; Mulder, F. A. *J. Am. Chem. Soc.* **2010**, *132*, 2952–2960.
(22) Ito, Y.; Yamasaki, K.; Iwahara, J.; Terada, T.; Kamiya, A.; Shirouzu, M.; Muto, Y.; Kawai, G.; Yokoyama, S.; Laue, E. D.; Walchli, M.; Shibata, T.; Nishimura, S.; Miyazawa, T. *Biochemistry* **1997**, *36*, 9109–9119.
(23) Gossert, A. D.; Henry, C.; Blommers, M. J.; Jahnke, W.; Fernández, C. *J. Biomol. NMR* **2009**, *43*, 211–217.
(24) Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W. *J. Am. Chem. Soc.* **1997**, *119*, 12257–12261.

JA108383X