Journal of Biomolecular NMR, **20:** 379–384, 2001.
KLUWER/ESCOM

Amino acid type-selective backbone ^{1}H-^{15}N-correlations for Arg and Lys

Mario Schubert[a,b], Hartmut Oschkinat[a,b] & Peter Schmieder[a,*]
[a]*Forschungsinstitut für Molekulare Pharmakologie, Robert-Rössle-Str. 10, D-13125 Berlin, Germany*
[b]*Freie Universität Berlin, Takustr. 3, D-14195 Berlin, Germany*

Received 20 March 2001; Accepted 5 June 2001

Key words: amino acid type-selective experiments, automated assignment, HSQC, proteins, triple-resonance

Abstract

Four novel amino acid type-selective triple resonance experiments to identify the backbone amino proton and nitrogen resonances of Arg and Lys and of their sequential neighbors in (^{13}C,^{15}N)-labeled proteins are presented: the R(i+1)-HSQC and R(i,i+1)-HSQC select signals originating from Arg side chains, the K(i+1)-HSQC and K(i,i+1)-HSQC select signals originating from Lys side chains. The selection is based on exploiting the characteristic chemical shifts of a pair of carbon atoms in Arg and Lys side chains using selective 90° pulses. The new experiments are recorded as two-dimensional ^{1}H-^{15}N-correlations and their performance is demonstrated with the application to a protein domain of 83 amino acids.

Triple-resonance experiments (Kay et al., 1990; Montelione and Wagner, 1990; Clore and Gronenborn, 1991; Sattler et al., 1999) are standard tools for resonance assignment in (^{13}C,^{15}N)-labeled proteins (McIntosh and Dahlquist, 1990) and are well suited for an automated assignment (Moseley and Montelione, 1999). The usual strategy is to correlate a (^{1}H^{N},^{15}N) frequency pair with C$^{\alpha}$, C$^{\beta}$ or carbonyl carbon frequencies of the same and a sequential amino acid and thus form chains of signal patterns that can be matched to the amino acid sequence of the protein once the type of amino acid corresponding to the pattern has been identified. The likelihood for the type of amino acid can conveniently be derived from the C$^{\alpha}$ and C$^{\beta}$ chemical shifts which form a two-dimensional landscape with characteristic areas for the different amino acid types (Grzesiek and Bax, 1993). In some cases the amino acid type is readily identified via this procedure. In other cases several amino acid types have to be considered since the probabilities are not sufficiently different. Moreover, the correct type might not even be among the most probable ones. This will be a particular problem in automated assignment and additional information is highly desirable.

We have recently presented amino acid type-selective ^{1}H-^{15}N-correlations (Schubert et al., 1999, 2000, 2001a, b) that help to identify the amino acid type of a certain residue and of its preceding neighbor and can thus provide this information. The majority of these new experiments is based on the MUSIC sequence (Schmieder et al., 1998; Schubert et al., 1999) that replaces the initial INEPT step in a triple-resonance pulse sequence. It can also be combined with selective pulses and carefully tuned delays. Thus the topology of the side chain and the chemical shifts of carbon atoms other than C$^{\alpha}$ and C$^{\beta}$ are used as additional criteria to identify the amino acid type.

Here we present amino acid type-selective ^{1}H-^{15}N correlations selective for Arg and for Lys. The selection criteria used in these sequences are the chemical shifts of the C$^{\gamma}$ and C$^{\delta}$ in case of Arg and the C$^{\delta}$ and C$^{\epsilon}$ in case of Lys. For both amino acids the sufficiently distinct chemical shifts of these carbons allow to treat the nuclei independently with selective pulses. The pathway of magnetization in the new experiments is depicted in Figure 1: after the selection with soft pulses the magnetization is transferred along the side chain to the C$^{\alpha}$. The magnetization is relayed either via the carbonyl carbon to the (i+1) nitrogen or di-

*To whom correspondence should be addressed. E-mail: schmieder@fmp-berlin.de

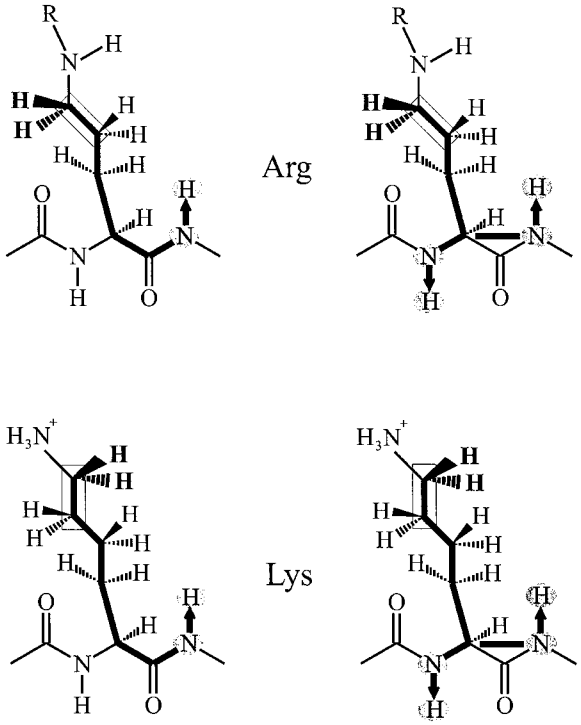

Figure 1. Schematic representation of the magnetization pathway in the new experiments. After proton excitation, magnetization is transferred to ^{13}C. With selective pulses the C^{δ}/C^{γ} pair in Arg and the $C^{\varepsilon}/C^{\delta}$ pair in Lys are selected (indicated by the rectangle). The magnetization is transferred along the side chain to the C^{α} carbon as indicated by the thick lines. The two magnetization pathways differ in the transfer from the C^{α} carbon to the nitrogen. In the left column the transfer of magnetization from the C^{α} carbon to the carbonyl and then to the nitrogen and amide proton in the (i+1)-HSQCs is depicted. The right column represents the (i,i+1)-HSQCs, where the magnetization is transferred from the C^{α} carbon to either the nitrogen of the same amino acid or that of the (i+1) neighbor.

rectly to both nitrogens coupled to the C^{α} (i or i+1), and finally detected on the respective amide proton.

The novel experiments are derived from the CBCA(CO)NH (Grzesiek and Bax, 1992a) and the CBCANH (Grzesiek and Bax, 1992b). Two and three relay steps are implemented in the original sequences to obtain experiments for Arg (R(i+1)- and R(i,i+1)-HSQC) and Lys (K(i+1)- and K(i,i+1)-HSQC), respectively. Despite an inevitable loss of magnetization due to relaxation during the additional delays, these reduce the intensity only slightly relative to the original sequences since the side chains of Arg and Lys are linear chains of carbon atoms with almost identical carbon-carbon one-bond couplings. The desired selectivity for Arg and Lys is achieved via the implementation of soft pulses in the beginning of the sequence as described below. The pulse sequences are shown in Figure 2. The new experiments are applied to the chicken EphB2 SAM domain (Smalla et al., 1999; PDB: 1SGG), a mostly helical protein of 83 amino acids for which nitrogen relaxation measurements yielded a rotational correlation time of 6.5 ns.

Characteristic chemical shifts of Arg C^{δ} and C^{γ} are 43.1 ± 1.0 ppm and 27.3 ± 1.4 ppm, respectively (Seavey et al., 1991; http://www.bmrb.wisc.edu). In the initial INEPT step both carbon pulses are replaced by pulses selective for the C^{δ}. The 180° pulse (S_1) is performed as a REBURB (Geen and Freeman, 1991) pulse with a length of 4096 μs centered at 45 ppm; it covers a range of 8 ppm. Note that this selective pulse extends throughout the complete delay $2\tau_1$. Only an effective J-coupling is therefore active between the proton and carbon spins. It was determined experimentally to be 90% of the true J-coupling and the delay of 4096 μs therefore has the proper length. During an I-BURP (Geen and Freeman, 1991) pulse, that has been designed for the inversion of z-magnetization, the effective coupling is smaller due to its unsymmetrical shape. It would thus have required a longer delay to evolve the coupling and we used the REBURP shape instead.

A self-refocusing L0S2-0 (Lunati et al., 1998) pulse with a length of 3072 μs applied at 45 ppm is used as the 90° pulse (S_2) and covers a range of 11 ppm. In the following delay $2\Delta_1$ the chemical shift evolution of the C^{δ} is refocused while the coupling between the C^{δ} and the C^{γ} evolves. This is accomplished using a nonselective rectangular 180° pulse in the center of the delay. The transfer of magnetization from the C^{δ} to the C^{γ} is then achieved by two subsequent selective 90° pulses (S_3 and S_4), one selective for the C^{δ}, the other for the C^{γ}. Both are performed as L0S2-0 pulses: S_3 has a time inverted shape and is centered at 45 ppm, S_4 is applied at 29 ppm. These two selective pulses, which are phase cycled independently, are of central importance for the new experiments. They transfer magnetization only between carbon spins that are connected via a one-bond coupling and have chemical shifts in the correct ranges. This is true solely for Arg (C^{δ} and C^{γ}) and Lys (C^{ε} and C^{δ}). Signals from the latter will not appear in the Arg spectra, since two relay steps are not sufficient to transfer magnetization from the C^{δ} to the C^{α}. The resulting spectra in Figure 3a and b demonstrate the clean selection achieved with the selective pulses, only signals from Arg are present.

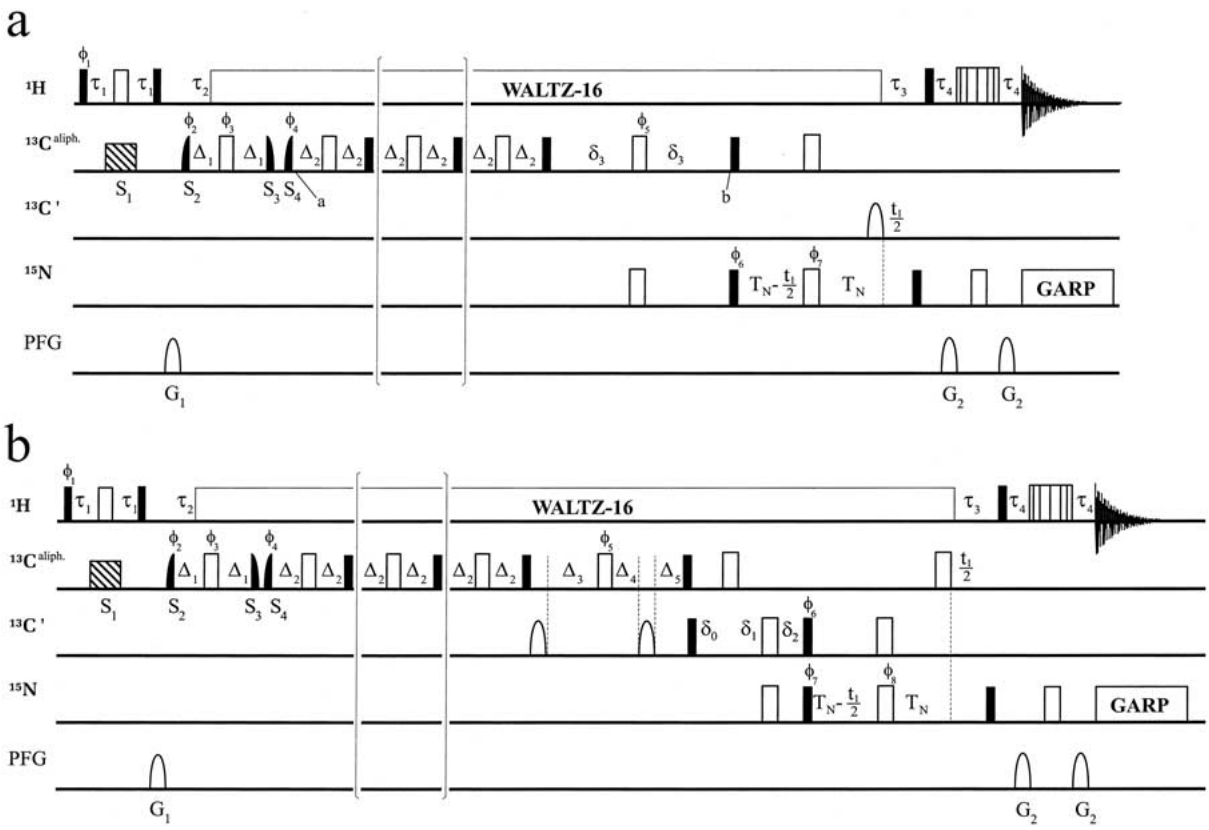

Figure 2. Pulse sequences of the amino acid type-selective ^{1}H-^{15}N-correlations selective for Arg and Lys. The pulse sequence for the (i,i+1)-HSQCs is shown in (a), that for the (i+1)-HSQCs in (b). Narrow filled and wide unfilled rectangles correspond to 90° and 180° rectangular pulses, respectively. Magnetic field gradients as well as shaped 180° ^{13}CO pulses are represented by sine shapes. Pulses applied at the $^{13}C^{\alpha/\beta}$ or ^{13}CO resonance frequencies were adjusted to provide a null at the corresponding ^{13}CO or $^{13}C^{\alpha}$ frequencies. The lengths of the pulses are given for a 600 MHz spectrometer. The rectangular $^{13}C^{\alpha/\beta}$ 90° and 180° pulses were set to 49 μs and 44 μs, respectively. The rectangular ^{13}CO 90° and 180° pulses were set to 54 μs and 108 μs, respectively. The shaped 180° ^{13}CO pulses were applied as G3 Gaussian cascades (Emsley and Bodenhausen, 1990) with a duration of 256 μs. The striped thick bar (S_1) stands for a band-selective 180° REBURP pulse (Geen and Freeman, 1991). Self-refocusing selective L0S2-0 excitation pulses (Lunati et al., 1998) are represented by filled half-sine shapes (S_2–S_4). The offsets and pulse lengths are given in the description of the particular experiments. The second L0S2-0 pulse (S_3) has a time inverted shape. Unless indicated otherwise, pulses are applied with phase x. Proton hard pulses were applied with 25 kHz field strength; WALTZ-16 (Shaka et al., 1983) of ^{1}H spins was achieved using a field strength of 3.1 kHz. The water-selective 90° square pulse had a duration of 1 ms. GARP-1 decoupling (Shaka et al., 1985) of ^{15}N was achieved using a field strength of 830 Hz. Water suppression was obtained using WATERGATE implemented with a 3-9-19 pulse (Sklenář et al., 1993). The gradients were applied as a sinusoidal function from 0 to π. The carrier frequencies were centered at ^{1}H = 4.8 ppm, ^{15}N = 119.6 ppm, $^{13}C^{\alpha/\beta}$ = 45 ppm and ^{13}CO = 175 ppm. The following delays were used: τ_1 = 2.0 ms, τ_2 = 2.1 ms, τ_3 = 5.5 ms, τ_4 = 2.25 ms, δ_0 = 4.5 ms, δ_1 = 6.9 ms, δ_2 = 11.4 ms, δ_3 = 9 ms, T_N = 11 ms, Δ_1 = 3.5 ms, Δ_2 = 5.5 ms, Δ_3 = 8.8 ms for the K-(i+1)-HSQC and 4.5 ms in case of the R-(i+1)-HSQC, Δ_4 = 4.3 ms for the K-(i+1)-HSQC and 0 ms in case of the R-(i+1)-HSQC, Δ_5 = 4.5 ms. Alternative values for Δ_3, Δ_4 and Δ_5 in case of the K-(i+1)-HSQC are given in the text. To achieve quadrature detection in the indirect dimension the States-TPPI-States protocol (Marion et al., 1989) was used in all experiments. All spectra were processed using XWINNMR (Bruker AG). The pulse programs and the L0S2-0 shape in Bruker format are available under http://www.fmp-berlin.de/ schubert. (a) R-(i,i+1)-HSQC (omitting the part in parentheses) and K-(i,i+1)-HSQC (including the part in parentheses). The phase cycling was: $\phi_1 = y$; ϕ_2 = x, −x; ϕ_3 = 8 (x), 8 (y), 8 (−x), 8 (−y); ϕ_4 = 2 (x), 2 (−x); ϕ_5 = 4 (x), 4 (−x); ϕ_6 = 4 (x), 4 (−x); ϕ_7 = 8 (x), 8 (−x); ϕ_{rec} = x, 2 (−x), x, 2 (−x, 2 (x), −x), x, 2 (−x), x. States-TPPI phase cycling was applied to phase ϕ_6. The gradients had the following duration and strength: G_1 = 2 ms (28 G/cm), G_2 = 1 ms (21 G/cm). (b) R-(i+1)-HSQC (omitting the part in parentheses) and K-(i+1)-HSQC (including the part in parentheses). The phase cycling was as follows: $\phi_1 = y$; ϕ_2 = x, −x; ϕ_3 = 8 (x), 8 (y), 8 (−x), 8 (−y); ϕ_4 = 2 (x), 2 (−x); ϕ_5 = 4 (x), 4 (−x); $\phi_6 = 50°$; ϕ_7 = 4 (x), 4 (−x); ϕ_8 = 8 (x), 8 (−x); ϕ_{rec} = x, 2 (−x), x, 2 (−x, 2 (x), −x), x, 2 (−x), x. States-TPPI phase cycling was applied to phase ϕ_7. The gradients had the following duration and strength: G_1 = 2 ms (28 G/cm), G_2 = 1 ms (21 G/cm).

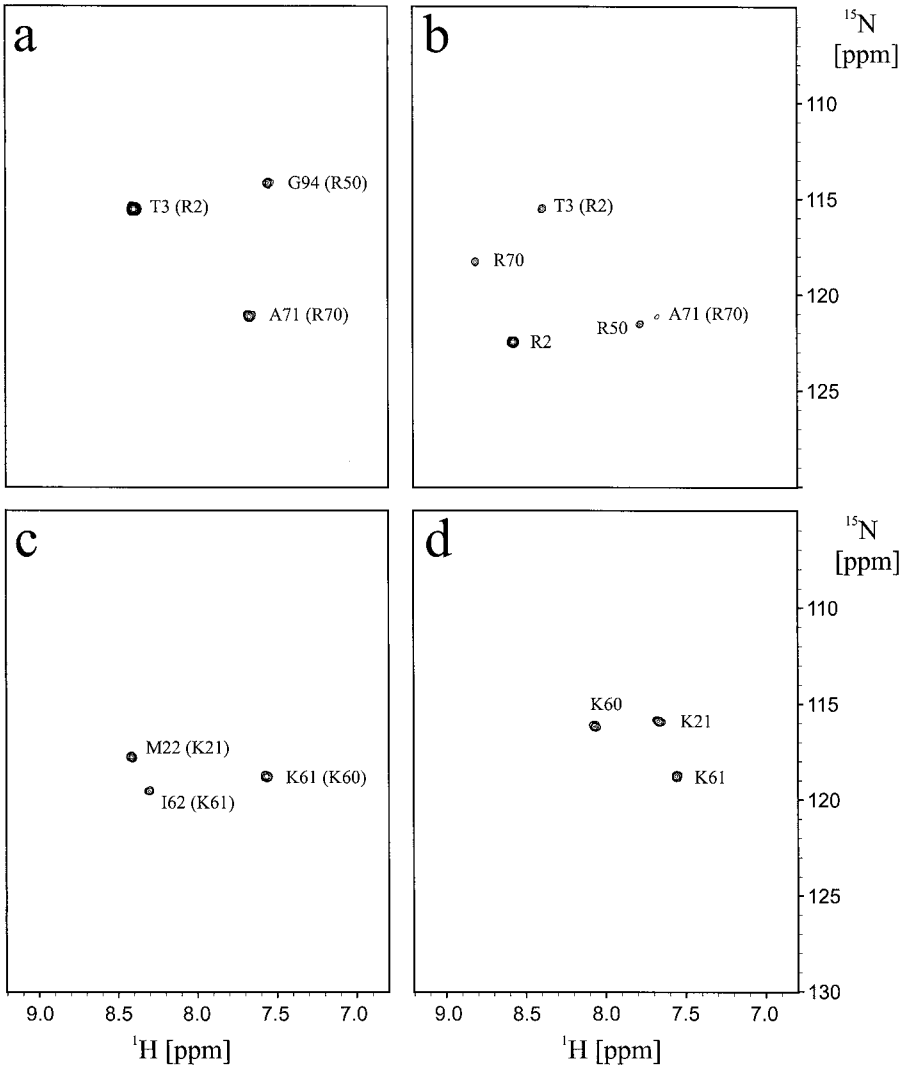

Figure 3. Arg and Lys selective ^{1}H-^{15}N-correlations of the chicken EphB2 receptor SAM domain (Smalla et al., 1999; PDB: 1SGG), which contains 3 Arg and 4 Lys residues. The spectra were recorded on a DRX600 in standard configuration using an inverse triple resonance probe equipped with three-axis self shielded gradient coils. The experiments were recorded at 300 K with a 1.5 mM ^{15}N/^{13}C labeled sample of the SAM domain at pH 5.7. A 5 mm ultra-precision sample tube was used. All spectra were recorded with 64 (t_1) × 512 (t_2) complex points in each dimension and spectral widths of 3012 Hz (^{15}N) × 10000 Hz (^{1}H). The data were processed using a squared sine-bell shifted by 90° as a window function in both dimensions. The ^{15}N t_1 interferograms were quadrupled in length by linear prediction using XWINNMR. The final spectra had a size of 512 (t_1) × 1024 (t_2) real points. (a) The R-(i+1)-HSQC was acquired in 3 h using 64 scans. All expected signals are visible. (b) The R-(i,i+1)-HSQC was acquired in 4.5 h using 96 scans and contains all Arg residues and smaller signals from the sequential neighbors. (c) The K-(i+1)-HSQC was acquired in 6 h using 128 scans. It contains three out of four K-(i+1) signals. E27, the sequential neighbor of K26, is missing due to the relaxation properties of the latter. In the shorter version of the experiment that does not suppress the Arg residues the signal is present. (d) The K-(i,i+1)-HSQC was acquired in 6 h using 128 scans and contains three out of four Lys residues. The signal resulting from K26 is again missing.

Characteristic chemical shifts of Lys C^{ε} and C^{δ} are 41.7 ± 0.9 ppm and 28.8 ± 1.4 ppm, respectively (Seavey et al., 1991; http://www.bmrb.wisc.edu). To record the spectra selective for Lys the same pulse sequences as in the Arg experiments are used but they are extended by an additional relay step. The selective pulses are of the same type and of the same length, but are applied at slightly different frequency positions: S_1 is centered at 44 ppm, S_2 and S_3 at 45 ppm and S_4 at 29 ppm.

Signals from Arg can potentially appear in the Lys spectra. It is, however, desirable to obtain spectra that contain only signals from Lys. This can be achieved by tuning the delays in the relay steps to appropriate values which can be derived from an inspection of the transfer functions of various magnetization pathways.

The transfer function for Lys between points a and b in the (i,i+1)-pulse sequence depicted in Figure 2a is:

$$\sin^6(\pi J_{CC} 2\Delta_2) \sin(\pi J_{CC} 2\delta_3) \sin(\pi J_{CN} 2\delta_3) \quad (1)$$

J_{CC} stands for all one-bond carbon-carbon couplings which are assumed to be uniform. During each of the three delays $2\Delta_2$ one carbon-carbon coupling is refocussed and one is evolving and each of the three 90° pulses following the delay accomplishes a transfer of magnetisation to the next carbon in the side chain.

For Arg three magnetization pathways are possible. There are three relay steps after the selective pulse S_4, while it takes only two steps to transfer the magnetization from the C^{γ} to the C^{α}. Consequently, detectable signal does only result if during one of the delays $2\Delta_2$ coupling to the previous carbon in the side chain is refocussed while coupling to the next carbon in the chain does not evolve. No magnetization will then be transferred by the subsequent 90° pulse. Equations 2, 3 and 4 represent the situation if no transfer takes place after the first, second or third of the three delays $2\Delta_2$, respectively.

$$\begin{aligned} &\sin(\pi J_{CC} 2\Delta_2) \cos(\pi J_{CC} 2\Delta_2) \times \\ &\quad \cos(\pi J_{CC} 2\Delta_2) \sin(\pi J_{CC} 2\Delta_2) \times \\ &\quad \sin^2(\pi J_{CC} 2\Delta_2) \times \sin(\pi J_{CC} 2\delta_3) \sin(\pi J_{CN} 2\delta_3) \end{aligned} \quad (2)$$

$$\begin{aligned} &\sin^2(\pi J_{CC} 2\Delta_2) \times \sin(\pi J_{CC} 2\Delta_2) \cos(\pi J_{CC} 2\Delta_2) \times \\ &\quad \cos(\pi J_{CC} 2\Delta_2) \sin(\pi J_{CC} 2\Delta_2) \times \\ &\quad \sin(\pi J_{CC} 2\delta_3) \sin(\pi J_{CN} 2\delta_3) \end{aligned} \quad (3)$$

$$\begin{aligned} &\sin^2(\pi J_{CC} 2\Delta_2) \times \sin^2(\pi J_{CC} 2\Delta_2) \times \\ &\quad \sin(\pi J_{CC} 2\Delta_2) \times \cos(\pi J_{CC} 2\delta_3) \sin(\pi J_{CN} 2\delta_3) \end{aligned} \quad (4)$$

Note that the transfer functions (2) and (3) are identical. The delays Δ_2 and δ_3 can be tuned such that the signals from transfer functions (2), (3) and (4) cancel each other while the signal from transfer function (1) is optimal. After experimental optimization the best results were obtained with $\Delta_2 = 5.5$ ms and $\delta_3 = 9$ ms. The same type of suppression can be used in the (i+1)-sequence. The delays Δ_3, Δ_4 and Δ_5 are adjusted to 8.8, 4.3 and 4.5 ms, respectively, which is simultaneously optimal for the evolution of coupling between the C^{α} and the carbonyl and the cancellation of signals from Arg. Spectra are shown in Figure 3c and d. All expected resonances are present while Arg signals are suppressed. If the delays Δ_3, Δ_4 and Δ_5 are adjusted to 4.5, 0 and 4.5 ms, respectively, the sensitivity is higher but signals from Arg are not suppressed any more.

With the new experiments the signal of K21 in the SAM domain can be easily identified as a Lys residue and the signal of M22 as a sequential neighbor of Lys (Figure 3c and d). With C^{α} and C^{β} chemical shifts of 57.0 and 28.2 ppm for K21 both the program 'type_prob' (Grzesiek and Bax, 1993) and current methods used within 'AutoAssign' (Zimmermann et al., 1997; Moseley and Montelione, personal communication) yield less than 1% probability for Lys. We observed a similar problem in the EVH1 domain of VASP (Ball et al., 2000) in the case of K21 with C^{α} and C^{β} chemical shifts of 56.5 and 28.8 ppm. Even though the programs would retain Lys as a possible spin system assignment, the novel selective experiments would greatly enhance the robustness of the assignment procedure.

In conclusion, we have presented a set of four ^{1}H-^{15}N-correlations selective for Arg and Lys which can be used to identify Arg and Lys residues and their sequential neighbors. The R-(i+1)-HSQC and R-(i,i+1)-HSQC contain only signals originating from Arg, while the K-(i+1)-HSQC and K-(i,i+1)-HSQC contain only signals originating from Lys. Together with the previously published amino acid type-selective ^{1}H-^{15}N-correlations (Schubert et al., 1999, 2000, 2001a, b) we are now able to identify Ala, Gly, Thr, Val/Ile, Asn, Gln, Asp, Glu, Ser, Leu, Pro, Trp, Phe/Tyr/His, Arg and Lys residues and their sequential neighbors. The novel experiments can be used as additional information for manual or automated assignment procedures.

Acknowledgements

Support from the Forschungsinstitut für Molekulare Pharmakologie is gratefully acknowledged. Mario Schubert was supported by the DFG Graduiertenkolleg GRK 80 'Modellstudien'. The authors thank Maika Smalla for the preparation and assignment of the SAM domain, Rüdiger Winter and Wolfgang Bermel for helpful discussions and Grant Langdon for carefully reading the manuscript.

References

Ball, L.J., Kühne, R., Hoffmann, B., Häfner, A., Schmieder, P., Volkmer-Engert, R., Hof, M., Wahl, M., Schneider-Mergener, J., Walter, U., Oschkinat, H. and Jarchau, T. (2000) *EMBO J.*, **19**, 4903–4914.

Clore, G.M. and Gronenborn, A.M. (1991) *Prog. NMR Spectrosc.*, **23**, 43–92.

Emsley, L. and Bodenhausen, G. (1990) *Chem. Phys. Lett.*, **165**, 469–476.

Geen, H. and Freeman, R. (1991) *J. Magn. Reson.*, **93**, 93–141.

Grzesiek, S. and Bax, A. (1992a) *J. Am. Chem. Soc.*, **114**, 6291–6293.

Grzesiek, S. and Bax, A. (1992b) *J. Magn. Reson.*, **B99**, 201–207.

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.

Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496–514.

Lunati, E., Cofrancesco, P., Villa, M., Marzola, P. and Osculati, F. (1998) *J. Magn. Reson.*, **134**, 223–235.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.

McIntosh, L.P. and Dahlquist, F.W. (1990) *Quart. Rev. Biophys.*, **23**, 1–38.

Montelione, G.T. and Wagner, G. (1990) *J. Magn. Reson.*, **87**, 183–188.

Moseley, H.N. and Montelione, G.T. (1999) *Curr. Opin. Struct. Biol.*, **9**, 635–642 and references cited therein.

Sattler, M., Schleucher, J. and Griesinger, C. (1999) *Prog. NMR Spectrosc.*, **34**, 93–158.

Schmieder, P., Leidert, M., Kelly, M.J.S. and Oschkinat, H. (1998) *J. Magn. Reson.*, **131**, 199–201.

Schubert, M., Smalla, M., Schmieder, P. and Oschkinat, H. (1999) *J. Magn. Reson.*, **141**, 34–43.

Schubert, M., Ball, L.J., Oschkinat, H. and Schmieder, P. (2000) *J. Biomol. NMR*, **17**, 331–335.

Schubert, M., Oschkinat, H. and Schmieder, P. (2001a) *J. Magn. Reson.*, **148**, 61–72.

Schubert, M., Oschkinat, H. and Schmieder, P. (2001b) *J. Magn. Reson.*, in press.

Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J. (1991) *J. Biomol. NMR*, **1**, 217–236.

Shaka, A.J., Keeler, J., Frenkiel, T. and Freeman, R. (1983) *J. Magn. Reson.*, **52**, 335–338.

Shaka, A.J., Barker, P.B. and Freeman, R. (1985) *J. Magn. Reson.*, **64**, 547–552.

Sklenář, V., Piotto, M., Leppik, R. and Saudek, V. (1993) *J. Magn. Reson.*, **A102**, 241–245.

Smalla, M., Schmieder, P., Kelly, M.J.S., ter Laak, A., Krause, G., Ball, L.J., Wahl, M., Bork, P. and Oschkinat, H. (1999) *Protein Sci.*, **8**, 1954–1961.

Zimmermann, D.E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C., Powers, R. and Montelione, G.T. (1997) *J. Mol. Biol.*, **269**, 592–610.