Journal of Biomolecular NMR (2005) 31: 115–127

A modified strategy for sequence specific assignment of protein NMR spectra based on amino acid type selective experiments**

Mario Schubert[a,c], Dirk Labudde[a], Dietmar Leitner[a], 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;* [c]*Present address: Laboratorium für Physikalische Chemie, ETH-Hönggerberg, HCl, CH-8093 Zürich, Switzerland*

Received 1 October 2004; Accepted 16 December 2004

Key words: amino acid type selective HSQC, backbone assignment, OPCA motif, PB1 domain, sequence specific assignment, triple-resonance experiments

Abstract

The determination of the three-dimensional structure of a protein or the study of protein–ligand interactions requires the assignment of all relevant nuclei as an initial step. This is nowadays almost exclusively performed using triple-resonance experiments. The conventional strategy utilizes one or more pairs of three dimensional spectra to obtain redundant information and thus reliable assignments. Here, a modified strategy for obtaining sequence specific assignments based on two dimensional amino acid type selective triple-resonance experiments is proposed. These experiments can be recorded with good resolution in a relatively short time. They provide very specific and redundant information, in particular on sequential connectivities, that drastically increases the ease and reliability of the assignment procedure, done either manually or in an automated fashion. The new strategy is demonstrated with the protein domain PB1 from yeast CDC24p.

Introduction

The assignment of almost all ^{1}H, ^{13}C and ^{15}N resonances of a protein is prerequisite for a subsequent structure determination or a study of protein–ligand interactions. Resonance assignment in proteins is nowadays almost exclusively performed using multidimensional triple-resonance experiments (Kay et al., 1990; Montelione and Wagner, 1990; Clore and Gronenborn, 1991; Sattler et al., 1999). Those experiments exhibit high sensitivity, yield well resolved spectra with easily predictable signal patterns and are therefore ideally suited for manual as well as automated resonance assignment. The starting point of the assignment procedure usually is the ^{1}H,^{15}N correlation, frequently called the seed spectrum. Then, a pair of 3D spectra is recorded: one technique correlates the frequencies of carbon or proton resonances of a particular amino acid with the amino proton and nitrogen frequency of the sequentially following amino acid. The other technique correlates the same frequencies with the amino proton and nitrogen frequency of the same as well as the sequentially following amino acid. The evaluation of those two types of spectra yields chains of spin systems. Usually a chain of 3 to 5 sequential amino acids is sufficient to match the residues to the amino acid sequence, given that information on the amino acid type is available. This information is frequently obtained from the chemical

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

**Dedicated to Rüdiger Winter († 06.04.2004)

shifts of the C^{α} and C^{β} nuclei. Theoretically, a complete backbone assignment should be obtained this way, which could subsequently be extended to the side chains using other well established techniques (lkura et al., 1991, Logan et al., 1992, Montelione et al., 1992, Grzesiek et al., 1993, Kay et al., 1993). In practice, however, due to overlap and imperfections in the spectra, one pair of three-dimensional correlations is usually not sufficient to obtain complete assignment. Several pairs of spectra are then required, in particular if the task of resonance assignment is performed in an automated fashion (Moseley and Montelione, 1999).

In a series of recent papers we have proposed a set of amino acid type selective triple resonance experiments that yield two-dimensional $^{1}H,^{15}N$ correlations (Schubert et al., 1999, 2000, 2001a,b, c). In contrast to the conventional ^{1}H, ^{15}N correlation, however, only signals from one or few types of amino acids appear in the spectrum. These triple resonance experiments are, in principle, based on the CBCA(CO)NNH (Grzesiek and Bax, 1992a) and the CBCANNH (Grzesiek and Bax, 1992b) techniques that are extended to select for a particular type of side chain. Two classes of spectra therefore exist. One class yields ^{1}H, ^{15}N correlations of amino acids sequentially following the selected amino acid type and these experiments are named X-(i + 1)-HSQC, with X being the selected amino acid type. The other class yields $^{1}H,^{15}N$ correlations of both amino acids, the one corresponding to the selected amino acid type and of that sequentially following it. These experiments are named X-(i, i + i)-HSQC, where X again indicates the selected amino acid type. As two-dimensional experiments they can be recorded with good resolution in a short time and the reduction of the number of peaks drastically reduces the probability of overlap in the spectra. The experiments were originally designed to provide information on the amino acid type and thus to aid the conventional strategy of resonance assignment in a time effective manner. However, the spectra also exhibit information that can directly be used for a sequential assignment. This has led us to develop a modified strategy for backbone assignment based on amino acid type selective experiments that is presented here. The modified strategy is extremely helpful in manual, computer-assisted resonance assignment, but it should also improve the reliability of automated assignment.

To further enhance the performance of the new procedure, we present another 3 pairs of amino acid selective experiments that yield amino acid type selective correlations for Gln, Asp and Glu and thus ease the reliable identification of those amino acid types compared to previously available experiments. The new strategy is applied to the protein domain PB1 of yeast CDC24p (PDB: 1PQS) containing an OPCA motif which was formerly known as OPR, PC or AID motif (Ponting et al., 2002). This motif is rich in acidic amino acids, and its consensus sequence is YxDEDGDxxxxSDED/E. Some PB1 domains have been reported to mediate protein-protein contacts via the formation of heterodimers (Terasawa et al., 2001; Noda et al., 2003; Wilson et al., 2003), in part using the OPCA motif, as those of the human proteins $p40^{phox}$ and $p67^{phox}$, and the PB1 domains of the yeast proteins BEM1p and CDC24p.

Materials and methods

Protein production and purification

The gene sequence (231 bp) corresponding to the C-terminal 77 aa of a *Saccharomyces cerevisiae* calcium regulatory protein, CDC24p (entry: M16809), was transferred into a modified pET17b vector (Novagen) at the 3′ end of the thrombin cleavage site. This yields the protein sequence SEIFTLLVEKVWNFDDLIMAINSKI SNTHNNNISPITKIKYQDEDGDFVVLGSDE-DWNVAKEMLAENNEKFLNIRLY.

The vector encodes for kanamycin resistance and for a GST fusion protein. GST appears at the N-terminus, followed in the 3′ direction by a 6 Gly-linker region and a thrombin cleavage site. ^{15}N- and $^{15}N/^{13}C$-labeled protein was made by growing the cells on M9 minimal medium with ^{15}N NH_4Cl and ^{13}C glucose (Campro Scientific), respectively, in shaking cultures at 30 °C during 4–5.5 h of expression which was induced by 1 mM lPTG. The cells from 1 l culture were resuspended in 25 ml PBS buffer and lysed by a French Press. After centrifugation and purification the yield from 1 l culture was at least 4 mg PB1 domain.

Data acquisition

NMR measurements were carried out at 27 °C on a Bruker DRX600 spectrometer in standard configuration. NMR experiments were set up with the program package PASTE/PAPST (Labudde et al., 2002). All spectra were acquired on a 1.4 mM uniformly ^{15}N or $^{13}C/^{15}N$ labeled sample containing 20 mM KH_2PO_4, 50 mM NaCl and 0.02% NaN_3 at pH 6.0 in 93% H_2O/7% D_2O. NMR-data were processed using the XWINNMR software (Bruker Biospin). A 3D CBCA(CO)NNH/CBCANNH (Grzesiek and Bax, 1992a,b) experiment pair was recorded using 512, 64 and 48 complex points and a spectral width of 10000, 10000 and 3012 Hz for proton, carbon and nitrogen, respectively. Amino acid type selective ^{15}N-HSQC-experiments were recorded using 512 and 48 complex data points and a sweep width of 10000 and 3012 Hz in the direct proton and the indirect nitrogen dimension, respectively. All spectra were recorded with a relaxation delay of 1.0 s, the number of scans and the resulting measurement time is given in Table 1.

Software availablility

All pulse sequences presented here are available for Bruker Avance spectrometers under http://www.fmp-berlin.de/~schubert/pulseprograms.html, SMASH is available under http://www.fmp-berlin.de/~labudde

Table 1. Number of scans and experiment time for the two three-dimensional and all amino-acid selective experiments performed on the PB1 domain

Name of the experiment	Scans	Experiment time
3D-CBCA(CO)NNH	8	1 day 7 h
3D-CBCANNH	16	2 days 14 h
G-(*i* + 1)-HSQC	8	15 min
G-(*i*, *i* + 1)-HSQC	16	30 min
A-(*i* + 1)-HSQC	48	1 h 30 min
A-(*i*, *i* + 1)-HSQC	96	3 h
S-(*i*, *i* + 1)-HSQC	32	1 h
S-(*i* + 1)-HSQC	64	2 h
TA-(*i* + 1)-HSQC	48	1 h 30 min
TA-(*i*, *i* + 1)-HSQC	96	3 h
VIA-(*i* + 1)-HSQC	48	1 h 30 min
VIA-(*i*, *i* + 1)-HSQC	96	3 h
LA-(*i* + 1)-HSQC	96	3 h
LA-(*i*, *i* + 1)-HSQC	192	6 h
N-(*i* + 1)-HSQC	64	2 h
N-(*i*, *i* + 1)-HSQC	128	4 h
QN-(*i* + 1)-HSQC	128	4 h
Q-(*i*, *i* + 1)-HSQC	256	8 h
DE-(*i* + 1)-HSQC	32	1 h
D-(*i* + 1)-HSQC	64	2 h
DE-(*i* + 1)-HSQC	64	2 h
E-(*i*, *i* + 1)-HSQC	128	4 h
FHY-(*i* + 1)-HSQC	64	2 h
FHY-(*i*, *i* + 1)-HSQC	128	4 h
W-(*i* + 1)-HSQC	64	2 h
W-(*i*, *i* + 1)-HSQC	128	4 h
P-(*i* + 1)-HSQC	128	4 h
P-(*i*, *i* − 1)-HSQC	128	4 h

Results and discussion

New amino acid type selective experiments

So far, 15 pairs of amino acid type selective experiments have been presented (Schubert et al., 1999, 2000, 2001a,b,c), each pair consisting of an (*i* + 1)-HSQC and an (*i*, *i* + 1)-HSQC experiment. Using those experiments 8 amino acids (Gly, Ala, Asn, Ser, Trp, Pro, Arg and Lys) can immediately be identified from dedicated experiments while another 5 (Thr, Leu, Gln, Asp and Glu) require the combination of two or three experiments. The amino acids Val and Ile can be identified as a group but not distinguished, the same is true for Phe, Tyr and His. No experiments exist for Met and Cys. The experiments are based on three selection mechanisms which are used either alone or in combination: MUSIC (Schmieder et al., 1998), frequency selective pulses and the tuning of delays in the pulse sequence. MUSIC (MUltiplicity Selective In-phase Coherence transfer) accomplishes a transfer of magnetization from protons to a heteronucleus selective for a particular type of multiplicity (XH_2 or XH_3) resulting in in-phase magnetization. It can replace the INEPT transfer step that is conventionally used in most triple resonance experiments. Amino acids are thus selected according to multiplicity, chemical shift of side chain carbons as well as the topology of the side chain as depicted in the scheme in Figure 1 and described in Table 2.

For the use of those amino acid type selective experiments in resonance assignment, it would be

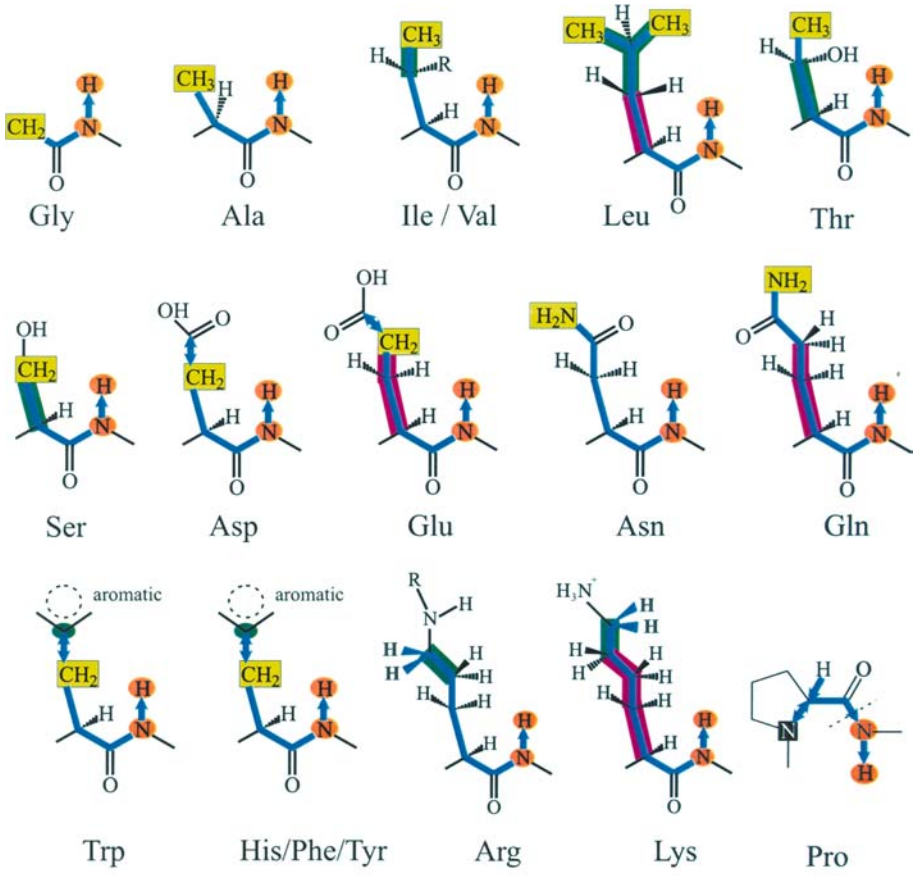


Figure 1. Schematic representation of the selection mechanism in the amino acid type selective experiments. 18 of the 20 natural amino acids can be selected. The blue lines represent the path of magnetization during the pulse sequence, the H^N,N pair that is detected is marked in orange. Two types of magnetization transfer are possible. The transfer depicted here leads from the side chain, where the actual selection is taking place, via the carbonyl exclusively to the nitrogen and the amino proton of the following amino acid, this is the (i + 1)-HSQC type of experiment. If the transfer does not lead via the carbonyl but from the C^α to the nitrogen nuclei directly, the nitrogen and the amino proton of both the same and the following amino acid will appear in the spectrum, this is the (i, i + 1)-HSQC-type of experiment. Three selection mechanisms are employed: the selection of multiplicity using the MUSIC pulse sequence building block is depicted by a yellow rectangle, the selection according to chemical shift is depicted in green and the selection according to topology is depicted in purple. In case of the Pro experiments the selection is based on the fact that the Pro nitrogen is not bound to a proton. Details about the selection mechanism are given in Table 2.

desirable to have as many dedicated experiments as possible to avoid ambiguities. In case of Gln the signals are identified by combining the N- and QN-HSQC. The identification of Asp and Glu is subsequently based on the combination of DNG-, N- and G-HSQC and EQG-, N-, QN-, and G-HSQC, respectively. Here we present three pairs of experiments (Q-, D- and E-HSQC) that are specific for Gln, Asp and Glu and designed to replace the less selective QN-, DNG-, and EQG-experiments. While the number of experiments is not reduced that way, 11 amino acids can be identified from dedicated experiments, resulting in a more reliable procedure.

Q-($i + 1$)- and Q-($i, i + 1$)-HSQC.

The pulse sequences for the HQ-(i + 1)-HSQC (Figure 2a) and the HQ-(i, i + 1)-HSQC (Figure 2b) are identical to those of the previously published QN-HSQCs (Schubert et al., 1999), except that the choice of delays is different. The selection is based on the fact that Asn and Gln are the only amino acids that have an NH_2 group next to a carbonyl carbon. After selection

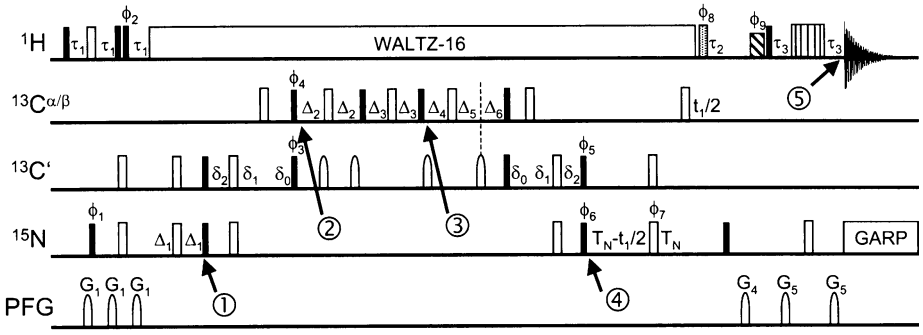
Table 2. Selection mechanisms of the 15 amino acid selective experiments depicted in Figure 1 and used for the new strategy

G	MUSIC-CH_2, next to carbonyl adjacent to NH	Schubert et al. (1999)
A	MUSIC-CH_3, CH_3 in C^{β} position	Schubert et al. (1999)
TA	MUSIC-CH_3, CH_3 in C^{β}/C^{γ} position, sel. pulse on T-C^{β}	Schubert et al. (1999)
VIA	MUSIC-CH_3, CH_3 in C^{β}/C^{γ} position, sel. pulse on VI-C^{β}	Schubert et al. (2001a)
LA	MUSIC-CH_3, CH_3 in $C^{\beta}/C^{\gamma}/C^{\delta}$ position, sel. pulses on L-carbons	Schubert et al. (2001a)
S	MUSIC-CH_2, CH_3 in C^{β} position, sel. pulse on S-C^{β}	Schubert et al. (2001a)
N	MUSIC-NH_2, next to carbonyl	Schubert et al. (1999)
D	MUSIC-CH_2, next to (CO_2^-)-carbonyl	This paper
Q	MUSIC-NH_2, next to carbonyl, suppression of N using tuned delays	This paper
E	MUSIC-CH_2, next to (CO_2^-)-carbonyl, suppression of D using tuned delays	This paper
FHY	MUSIC-CH_2, next to aromatic carbon, sel. pulse on C^{γ}	Schubert et al. (2001b)
W	MUSIC-CH_2, next to aromatic carbon, sel. pulse on C^{γ}	Schubert et al. (2001b)
R	Sel. pulses on C^{γ}/C^{δ}	Schubert et al. (2001c)
K	Sel. pulses on C^{δ}/C^{ϵ}, suppression of R using tuned delays	Schubert et al. (2001c)
P	No proton attached to nitrogen	Schubert et al. (2000)

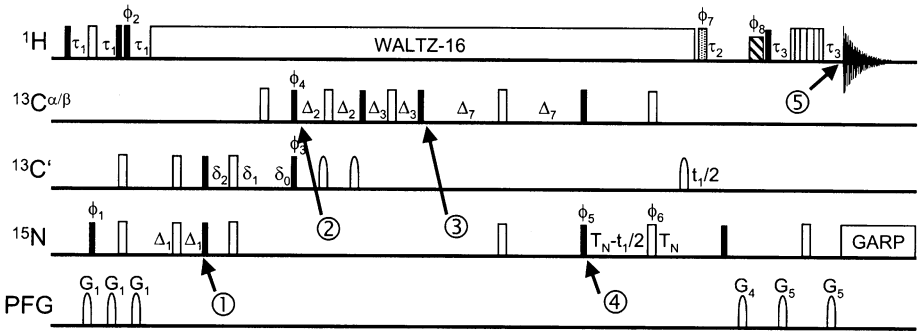
of the NH_2 moiety using MUSIC, the magnetization is transferred to the adjacent carbonyl carbon and then along the side chain using carbon–carbon, carbon–nitrogen and nitrogen–proton one-bond couplings to the amide proton of the sequential amino acid ((i + 1)-HSQC) or to the amide proton of the same and the sequential amino acid ((i, i + 1)-HSQC). Using the product operator formalism and omitting trigonometric factors the experiments for Q can be described as follows: Using the MUSIC sequence magnetization is transferred from the protons of the NH_2 group to the N^{ε} nitrogen. Up to point ① (Figures 2a and 2b) in the sequence, anti-phase magnetization $2N_y^{\varepsilon}C_z^{\delta}$ evolves. During the two subsequent lNEPT-type transfers it is converted to $2C_y^{\gamma}C_z^{\delta}$ (point ②). In the next part of the sequence up to point ③ this anti-phase magnetization is refocused and the magnetization transferred from C^{γ} to C^{β} and C^{α} via continuous defocusing/refocusing of carbon–carbon one-bond couplings. At point ③ anti-phase magnetization of the type $2C_y^{\alpha}C_z^{\beta}$ is present. From that point on the two sequences differ. In the (i + 1)-HSQC the coupling between C^{β} and C^{α} is refocused and that between C^{α} and C' creates anti-phase magnetization of the type $2C_y^{\alpha}C_z'$. The rest of the sequence are three subsequent INEPT-type transfers that create $2N_y^{(i+1)}C_z'$ at point ④ and $H_x^{N(i+1)}$ at point ⑤ where the detection begins. In the (i, i + 1)-HSQC the coupling between C^{β} and C^{α} is refocused and that between C^{α} and N creates anti-phase magnetization of the type $2C_y^{\alpha}N_z^{(i+1)}$ and $2C_y^{\alpha}N_z^{(i)}$. Then two lNEPT-type transfer steps create $2N_y^{(i+1)}C_z^{\alpha}$ and $2N_y^{(i)}C_z^{\alpha}$ at point and $H_x^{N(i+1)}$ and $H_x^{N(i)}$ at point ⑤ where the detection begins.

In case of the N-(i + 1)-HSQC and N-(i, i + 1)-HSQC (Schubert et al., 1999) a clean selection of Asn amino acids is achieved since only one transfer via $^1J_{cc}$ is performed in the pulse sequence and Gln cannot appear in the spectrum. The introduction of a second transfer step that enables the transfer of magnetization from the C^{γ} to the C^{α} in Gln resulted in the QN-pair of experiments. Magnetization transfer is still possible from the C^{β} to the C^{α} in Asn. There are, however, two possible magnetization pathways for Asn. While two transfers of magnetization between the aliphatic carbons are possible but only one transfer is necessary (from C^{β} to C^{α}), the magnetization may remain in-phase during one of the transfer delays, either $2\Delta_2$ or (Δ_4 + Δ_5 + Δ_6) in the (i + 1)-HSQC and either $2\Delta_2$ or $2\Delta_7$ in the (i, i + 1)-HSQC. As we have shown previously (Schubert et al., 2001a), proper choice of the delays will cause the transfer functions of the two pathways to acquire opposite sign. The two pathways will then cancel and no signals of Asn will appear in the spectrum while the signals of Gln will retain their intensity. This can easily be derived from an analysis of the three transfer functions that are shown here for the (i, i + 1)-HSQC (Figure 2b) (only relevant terms are given):

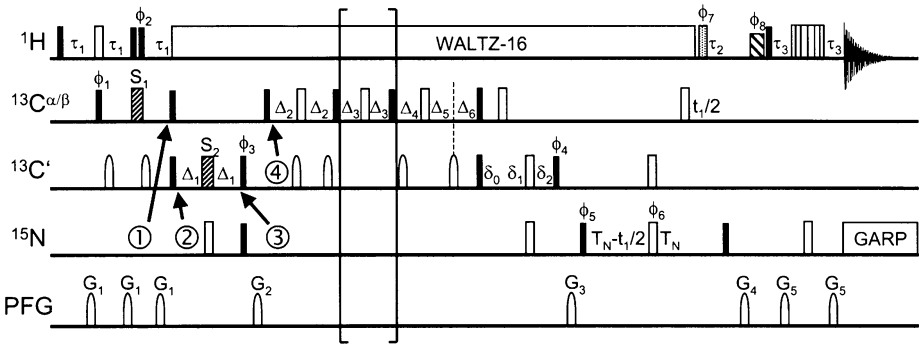
(a) **Q-(i+1)-HSQC**



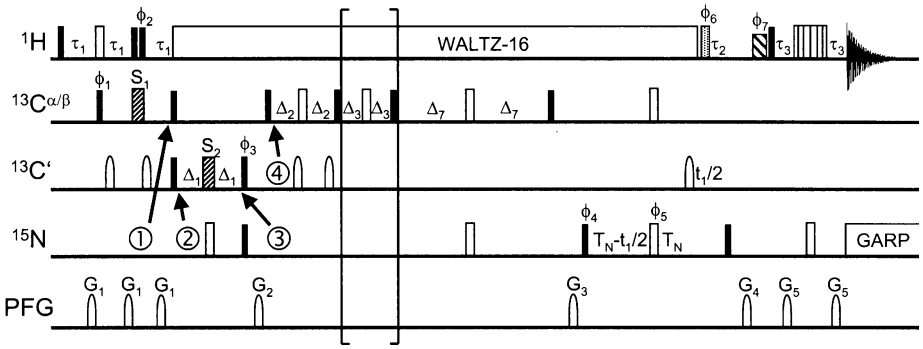
(b) **Q-(i,i+1)-HSQC**



(c) **D-(i+1)-HSQC and E-(i+1)-HSQC**



(d) **D-(i,i+1)-HSQC and E-(i,i+1)-HSQC**



1. First Asn-pathway:
   $\sin(\pi J_{C\beta C\alpha} 2\Delta_2) \sin(\pi J_{C\beta C\alpha} 2\Delta_3) \cos(\pi J_{C\beta C\alpha} 2\Delta_7)$.
2. Second Asn-pathway:
   $\cos(\pi J_{C\beta C\alpha} 2\Delta_2) \sin(\pi J_{C\beta C\alpha} 2\Delta_3) \sin(\pi J_{C\beta C\alpha} 2\Delta_7)$.
3. Gln-pathway:
   $\sin(\pi J_{C\gamma C\beta} 2\Delta_2) \sin(\pi J_{C\gamma C\beta} 2\Delta_3) \sin(\pi J_{C\beta C\alpha} 2\Delta_3) \sin(\pi J_{C\beta C\alpha} 2\Delta_7)$.

As can be seen from these expressions the requirement for the disappearance of the signals of Asn is $(\Delta_2 + \Delta_7) = 1/2J_{CC}$. A similar relation can be derived for the $(i + 1)$- HSQC (Figure 2a) by replacing Δ_7 with $(\Delta_4 + \Delta_5 + \Delta_6)$. The proper choice of the delays is given in the figure caption of Figure 2.

There are two caveats, however. Since the $^1J_{C\beta C\alpha}$ coupling constants of Asn vary by $\pm 10\%$ (Cornilescu et al., 2000) the suppression might not work perfectly and small residual signals might remain in the spectrum. In addition, it should be noted that the suppression of Asn in the Gln spectra lengthens the pulse sequence in case of the $(i + 1)$ experiment by 10 ms. Relaxation will therefore reduce the intensity in the Q- $(i + 1)$-HSQC but there is still the option to revert to the QN-$(i + 1)$-experiment. The QN- $(i + 1)$-HSQC of the of the PB1 domain is given in Figure 3a, the Q-$(i, i + 1)$-HSQC in Figure 3b.

D-$(i + i)$- and D-$(i, i + i)$-HSQC.

The identification of Asp has so far been accomplished via a selection of CH_2-groups (using MUSIC) adjacent to a carbonyl carbon. This, however, also selects for Asn and Gly, resulting in the DNG-pair of experiments (Schubert et al., 2001a). To create an experiment selecting for Asp alone, the fact is used that in Asn as well as

Figure 2. Pulse sequences of the new amino acid type selective experiments. 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 length of the pulses is given for a 600 MHz spectrometer. The rectangular $13C^{\alpha/\beta}$ 90° and 180° pulses were set to 49 and 44 μs, respectively. The rectangular ^{13}CO 90° and 180° pulses were set to 54 and 108 μs, respectively. The carrier was placed at 45 and 175 ppm for $^{13}C^{\alpha/\beta}$ and ^{13}CO pulses, respectively. The shaped 180° ^{13}CO pulses were applied as G3 Gaussian cascades (Emsley and Bodenhausen, 1990) with a duration of 256 μs. 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 depicted by a striped rectangle had a duration of 1 ms, the water flip-back pulse at the end of the decoupling period depicted by a dotted rectangle had a duration of 80 μs. 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 and had the following duration and strength: G1 = 1 ms (28 G/cm), G2 = 1 ms (21 G/cm) G3 = 1 ms (28 G/cm), G4 = 1 ms (21 G/cm) G5 = 1 ms (28 G/cm). The carrier frequencies were centered at ^{1}H = 4.8 ppm, 15N = 119.6 ppm, $^{13}C^{\alpha/\beta}$ = 45 ppm and ^{13}CO = 175 ppm. The following delays were used in all sequences: $\tau_2 = 5.5$ ms, $\tau_3 = 2.25$ ms, $\delta_0 = 4.5$ ms, $\delta_1 = 6.9$ ms, $\delta_2 = 11.4$ ms, $T_N = 11$ ms. To achieve quadrature detection in the indirect dimension the States-TPPI States protocol (Marion et al., 1989) was used in all experiments. The circled numbers refer to the discussion of the pulse sequences in the text. (a) Q-$(i + 1)$-HSQC: The following delays were used in this pulse sequence (the values in brackets indicated the delays used in the QN-$(i + 1)$-HSQC): $\tau_1 = 5.5$ ms, $\Delta_1 = 8$ ms, $\Delta_2 = 3.7$ ms (4.5 ms), $\Delta_3 = 4.5$ ms, $\Delta_4 = 9.3$ ms (4.5 ms), $\Delta_5 = 4.8$ ms (0 ms), $\Delta_6 = 4.5$ ms. The phase cycling was: $\phi_1 = 16$ (x), 16 $(-x)$; $\phi_2 = 2$ (45°), 2 (135°), 2 (225°), 2 (315°); $\phi_3 = 310°$; $\phi_4 = 32$ (x), 32 $(-x)$; $\phi_5 = 50°$; $\phi_6 = x, -x$; $\phi_7 = 8$ (x), 8 (y), 8 $(-x)$, 8 $(-y)$; $\phi_8 = 4$ $(-y)$, 4 (y); $\phi_9 = -x$; $\phi_{rec} = 2$ $(x, 2$ $(-x), x)$, 4 $(-x, 2$ $(x), -x)$, 2 $(x, 2$ $(-x), x)$, 2 $(-x, 2$ $(x), -x)$, 4 $(x, 2$ $(-x), x)$, 2 $(-x, 2$ $(x), -x)$. States-TPPI phase cycling was applied to phase ϕ_6. (b) Q-$(i,i + 1)$-HSQC: The following delays were used in this pulse sequence: $\tau_1 = 5.5$ ms, $\Delta_1 = 8$ ms, $\Delta_2 = 3.7$ ms, $\Delta_3 = 4.5$ ms, . $\Delta_7 = 9.3$ ms. The phase cycling was: $\phi_1 = 16$ (x), 16 $(-x)$; $\phi_2 = 2$ (45°), 2 (135°), 2 (225°), 2 (315°); $\phi_3 = 310°$; $\phi_4 = 32$ (x), 32 $(-x)$; $\phi_5 = x, -x$; $\phi_6 = 8$ (x), 8 (y), 8 $(-x)$, 8 $(-y)$; $\phi_7 = 4$ $(-y)$, 4 (y); $\phi_8 = -x$; $\phi_{rec} = 2$ $(x, 2$ $(-x), x)$, 4 $(-x, 2$ $(x), -x)$, 2 $(x, 2$ $(-x), x)$, 2 $(-x, 2$ $(x), -x)$, 4 $(x, 2$ $(-x), x)$, 2 $(-x, 2$ $(x), -x)$. States-TPPI phase cycling was applied to phase ϕ_5. (c) D-$(i + 1)$-HSQC and E-$(i + 1)$-HSQC (including the part in parenthesis): The following delays were used in this pulse sequence (the values in brackets indicated the delays used in the D and ED-$(i + 1)$-HSQC): $\tau_1 = 3.5$ ms, $\Delta_1 = 16$ ms, $\Delta_2 = 4.5$ ms, $\Delta_3 = 4.5$ ms, $\Delta_4 = 9$ ms (4.5 ms), $\Delta_5 = 4.5$ ms (0 ms), $\Delta_6 = 4.5$ ms. In case of the D-$(i + 1)$-HSQC the striped rectangles S1 and S2 represent band-selective 180° REBURP (Geen and Freeman, 1991) pulses of 2048 and 409.6 μs length executed at 42.5 and 175 ppm, respectively. In case of the E-$(i,i + 1)$-HSQC S1 is simply a rectangular pulse executed at the normal carrier position while S2 is performed as in the D-$(i,i + 1)$-HSQC. The phase cycling was: $\phi_1 = 16$ (x), 16 $(-x)$; $\phi = 2$ (45°), 2 (135°), 2 (225°), 2 (315°); $\phi_3 = x, -x$; $\phi_4 = 50°$; $\phi_5 = x$, $\phi_6 = 8$ (x), 8 (y), 8 $(-x)$, 8 $(-y)$; $\phi_7 = 4$ $(-y)$, 4 (y); $\phi_8 = -x$; $\phi_{rec} = 2$ $(x, 2$ $(-x), x)$, 4 $(-x, 2$ $(x), -x)$, 2 $(x, 2$ $(-x), x)$. States-TPPI phase cycling was applied to phase ϕ_5. (d) D-$(i,i + 1)$-HSQC and E-$(i,i + 1)$-HSQC (including the part in parenthesis): The following delays were used in this pulse sequence: $\tau_1 = 3.5$ ms, .1 = 16 ms, .2 = 4.5 ms, .3 = 4.5 ms, .7 = 9 ms. In case of the D-$(i,i + 1)$-HSQC the striped rectangles S1and S2 represent band-selective 180° REBURP (Geen and Freeman, 1991) pulses of 2048 and 409.6 μs length, respectively, S1 is executed at 42.5 ppm. In case of the E-$(i,i + 1)$-HSQC S1 is simply a rectangular pulse executed at the normal carrier position while S2 is performed as in the D-$(i,i + 1)$-HSQC. The phase cycling was: $\phi_1 = 16$ (x), 16 $(-x)$; $\phi_2 = 2$ (45°), 2 (135°), 2 (225°), 2 (315°); $\phi_3 = x, -x$, $\phi_4 = 32$ (x), 32 $(-x)$; $\phi_5 = 8$ (x), 8 (y), 8 $(-x)$, 8 $(-y)$; $\phi_6 = 4$ $(-y)$, 4 (y); $\phi_7 = -x$; $\phi_{rec} = 2$ $(x, 2$ $(-x), x)$, 4 $(-x, 2$ $(x), -x)$, 2 $(x, 2$ $(-x), x)$, 2 $(-x, 2$ $(x), -x)$, 4 $(x, 2$ $(-x), x)$, 2 $(-x, 2$ $(x), -x)$. States-TPPI phase cycling was applied to phase ϕ_4.

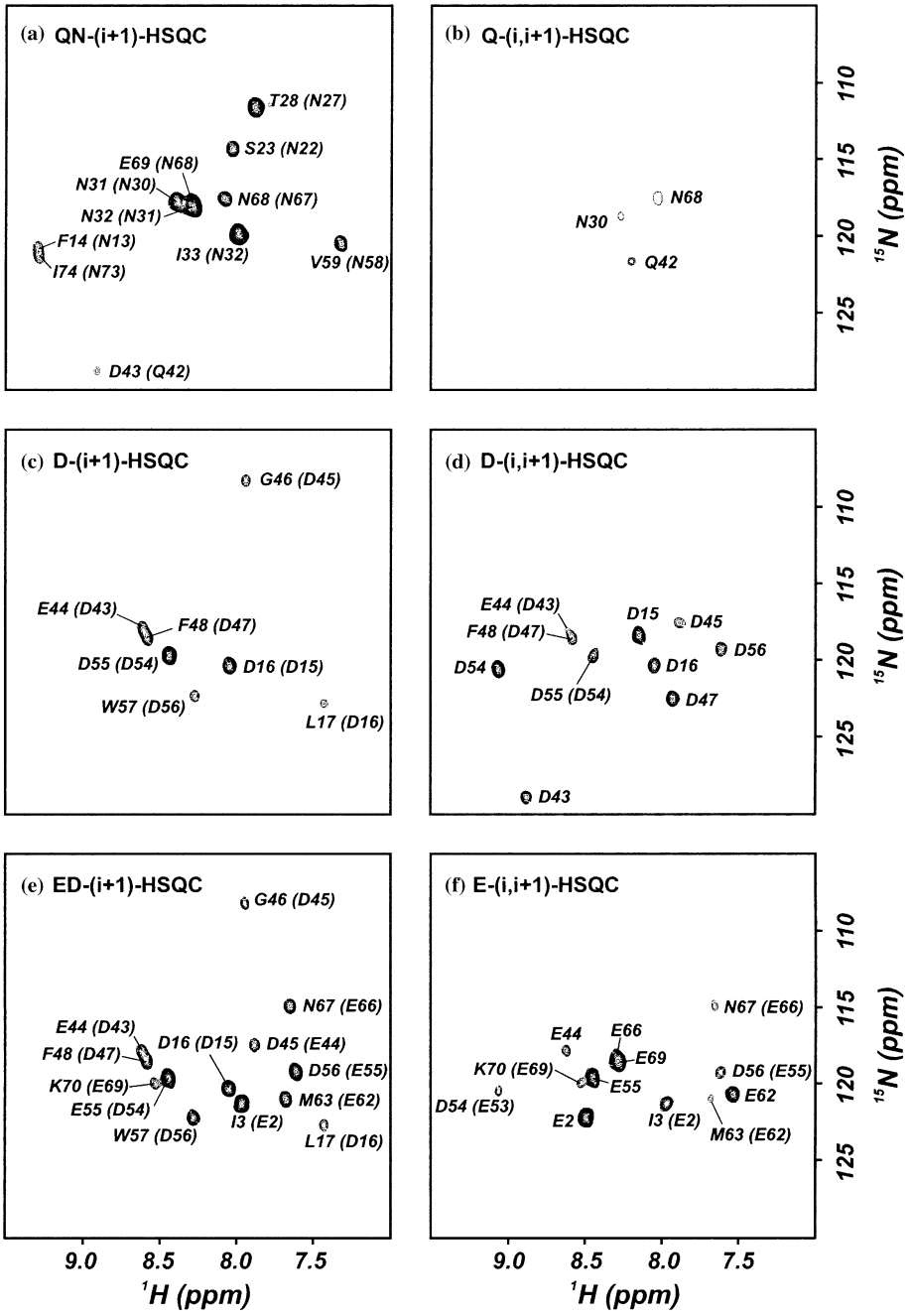


Figure 3. Spectra of the PB1 domain resulting from the new amino acid type selective experiments. The peaks are labeled according to the manually verified resonance assignment. In the case of an $i + 1$ magnetization transfer pathway the selected amino acid is indicated in parenthesis. (a) QN-($i + 1$)-HSQC recorded in 4 h using 128 scans. (b) Q-($i,i + 1$)-HSQC recorded in 8 h using 256 scans. Note the two negative signals resulting from incomplete suppression of Asn residues due to variations in the $^1J_{C\beta C\alpha}$ coupling constant. (c) D-($i + 1$)-HSQC, recorded in 1 h using 32 scans. (d) D-($i,i + 1$)-HSQC recorded in 2 h using 64 scans. (e) ED-($i + 1$)-HSQC recorded in 2 h using 64 scans. (f) E-($i,i + 1$)-HSQC recorded in 4 h using 128 scans.

in Gly a nitrogen is bound directly to the carbonyl adjacent to the CH_2 group. Suppression of Asn and Gly is then accomplished using the coupling between the nitrogen and that carbonyl carbon to convert magnetization of those two amino acids into undetectable magnetization

(Pellecchia et al., 1997). The pulse sequences for the D-(i + 1)-HSQC (Figure 2c) and D-(i, i + 1)-HSQC (Figure 2d) are similar to the sequences of the DNG experiments. During the initial MUSIC transfer step magnetization is transferred from CH_2-protons to the attached carbons and in addition anti-phase magnetization $2C_y^{\beta}C_z^{\gamma}$ and $2C_y^{\alpha}C_z'$ (in case of Gly) is created at point ① (Figures 2c and 2d). Instead of the HMQC-type sequence that was used to select for the CH_2-groups adjacent to a carbonyl carbon in the DNG experiments an HSQC-type sequence is now employed. At point ② anti-phase magnetization of the type $2C_y^{\gamma}C_z^{\beta}$ is present for Asp and Asn, while $2C_y'C_z^{\alpha}$ is present in case of Gly. In case of Asp this magnetization is also present at point ③, in case of Gly and Asn it has evolved to $4C_x'C_z^{\alpha}N_z^{(i-1)}$and $4C_x^{\gamma}C_z^{\beta}N_z^{\delta}$, respectively. The resulting magnetization with the nitrogen anti-phase with respect to the carbonyl carbon is then converted into carbon–nitrogen multiple quantum coherence by a 90° pulse on nitrogen and subsequently dephased using a pulsed field gradient. In case of Asp magnetization of the type $2C_z^{\gamma}C_z^{\beta}$ has been created that is insensitive to the gradient and at time point ④ magnetization of the type $2C_z^{\gamma}C_y^{\beta}$ will be present. Note that the evolution of carbon–nitrogen coupling will lengthen the pulse sequence by 30 ms. Because of the relaxation properties of the carbonyl carbons this might become unfavorable at large magnetic fields. The rest of the sequence corresponds to the Q sequences, transferring the magnetization along the side chain to the backbone for detection, except that there is only one transfer via carbon–carbon coupling. Spectra of the D-HSQCs of the PB1 domain are given in Figures 3c and 3d.

E-($i+1$)- and E-($i, i+1$)-HSQC.

Glu has previously been identified using the EQG experiments that use the same type of selection as employed DNG experiments but include an additional transfer via $^1J_{CC}$ (Schubert et al., 2001a). To create an E-pair of experiments out of the EQG-pair Gln, Asn and Gly are again suppressed based on the coupling between the carbonyl carbon and the adjacent nitrogen, with the same lengthening of the sequence and identical consequences for the execution of the experiment at high magnetic field as discussed above. Compared to the D-HSQCs the sequence has been extended by one additional transfer via carbon–carbon coupling. Asp is suppressed in a similar manner as Asn is suppressed in the Q-HSQCs by tuning the delays to values such that the two possible transfer pathways of magnetization in Asp cancel. The same caveats as in the Q experiments apply here, namely the possibility of breakthrough peaks due to variation of the $^1J_{C\alpha C\beta}$ and of relaxation losses in the E-(i + 1)-HSQC due to long delays. The experiment may therefore also be performed as a ED(i + 1)-HSQC to optimize sensitivity. An ED(i + 1)-HSQC spectrum of the PB1 domain is given in Figure 3e, the E-(i,i + 1)-HSQC in Figure 3f.

A modified strategy for backbone assignment in proteins

The key feature of the modified strategy is an extended use of the information contained in the amino acid selective experiments. Previously they were used for the determination of the amino acid type and did thus enhance the probability of matching chains of spin systems detected in three-dimensional triple-resonance experiments to the amino acid sequence. Since the experiments are, however, based on the CBCA(CO)NNH and CBCANNH techniques they also contain specific information suitable for sequence specific assignment. This is exploited by not only comparing the two spectra resulting from a pair of selective experiments but also spectra from different types of selective experiments. If an amino acid Z precedes amino acid X in the sequence, then both the Z-(i + 1)-HSQC and the X-(i,i + 1)-HSQC spectra will contain a signal corresponding to the ^1H,^{15}N-crosspeak of residue X. If the pair Z–X appears only once in the sequence and if for both amino acids an exclusive amino acid selective experiment exists, the signal of X is immediately assigned. The strategy is also robust towards signal overlap since only the superposition of ^1H,^{15}N-crosspeaks of the same type of amino acids may cause ambiguities. Overlap of ^1H,^{15}N-crosspeaks from different amino acids can be resolved using the selective experiments. Once it is known which amino acids correspond to particular chemical shift values, a transfer into the 3Ds can easily be accomplished using the knowledge about the expected carbon chemical shifts for a particular amino acid type.

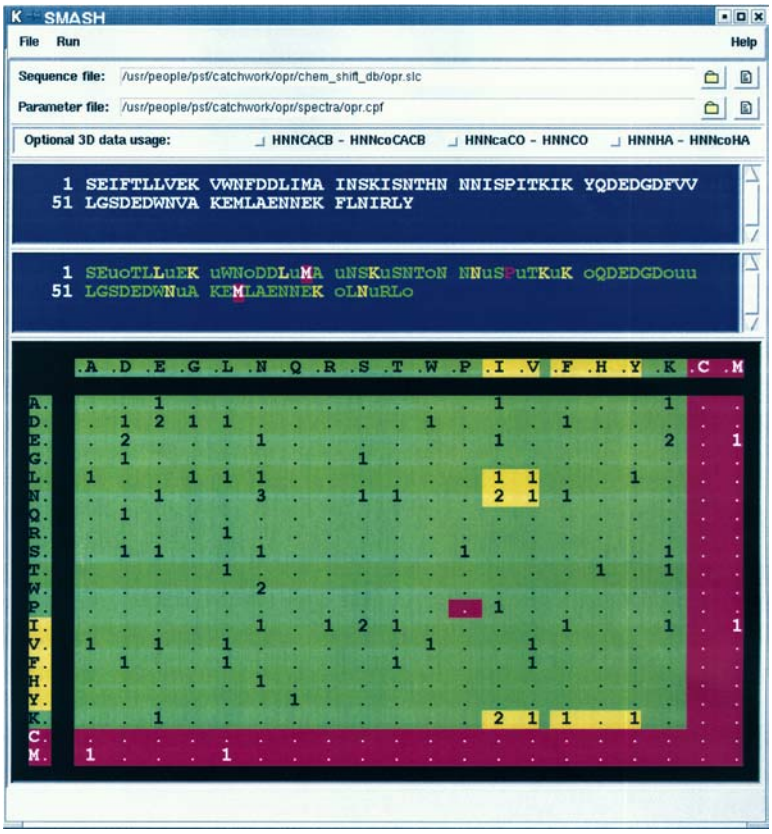
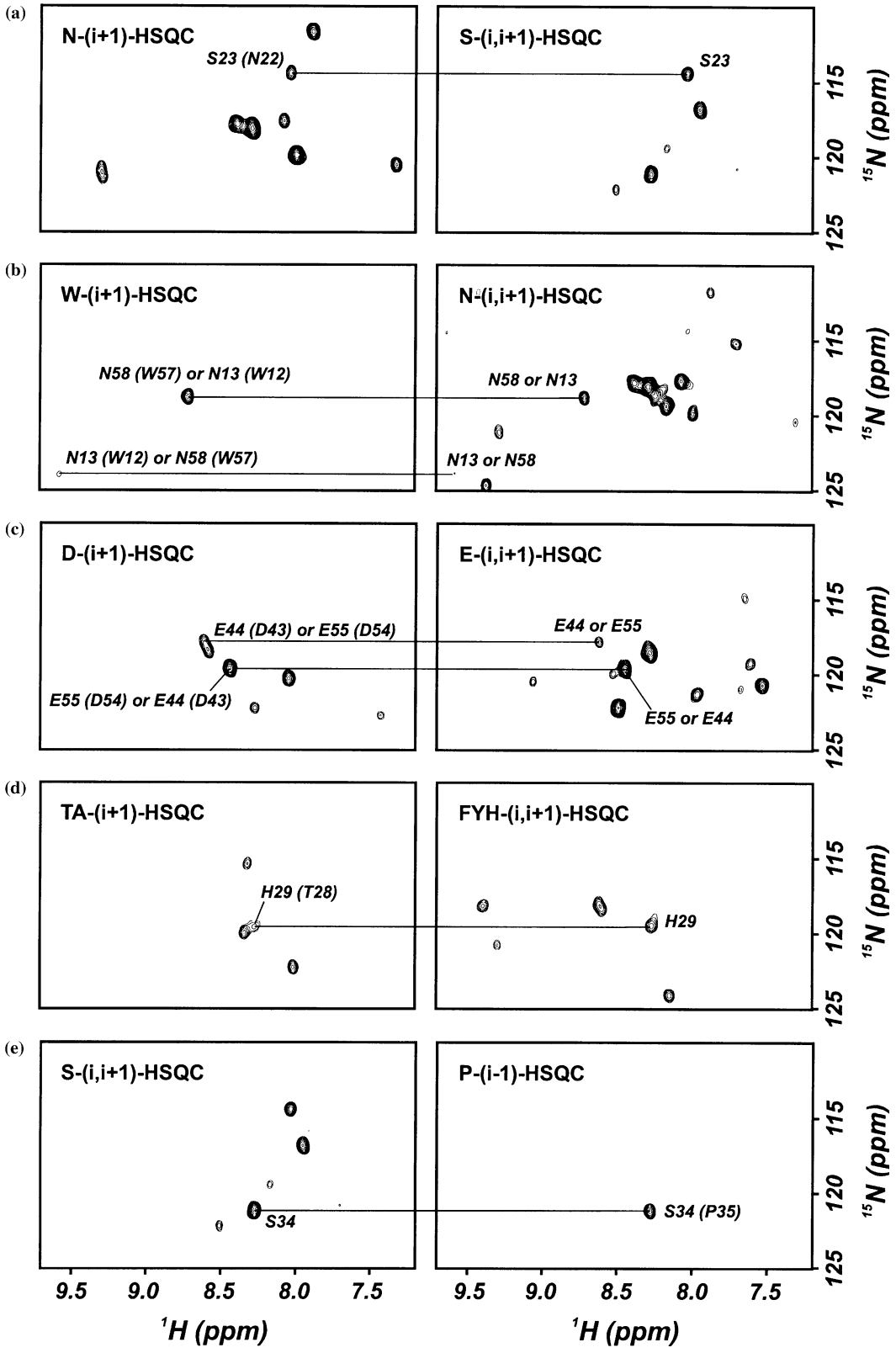


Figure 4. SMASH (Selective Matrix for Automated SearcHes) matrix of the protein domain PB1. The matrix results from an analysis of the amino acid sequence. Amino acids in the columns represent the (i + 1) amino acids while those in rows represent the i amino acids. The numbers in the matrix elements represent the number of occurrences of a particular pair of amino acids. Depending on the property of the employed amino acid type selective experiments the amino acid pairs get a different color code. Red indicates that no experiment exists for the particular amino acid type. Yellow indicates that for these amino acids no exclusive amino acid type selective experiment exists. Therefore the resulting amino acid pairs can not be unambiguously resolved. For example the results of the searches of the sequence pair Leu-lle contains also the results of the pair of Leu-Val. Amino acid pairs which are colored in green represent unambiguously detectable pairs. In addition two sequence windows are present. On top the mere amino acid sequence is given and below the sequence is colored according to the scheme described above. The symbols u and o define the aliphatic amino acids lle and Val and the aromatic amino acids Phe, His and Tyr, respectively. While Pro does not have an amino proton, resonances of its neighbors can be observed using an amino acid type selective experiments.

Figure 5. Examples for the assignment of amino acids based on the modified strategy for sequence specific assignment. In the case of an i + 1 magnetization transfer pathway the selected amino acid is indicated in parenthesis. (a) N-(i + 1)-HSQC and S-(i, i + 1)-HSQC. The marked cross peak in both spectra corresponds to the H^N and N frequencies of a Ser following an Asn. Since the only Asn-Ser pair in the amino acid sequence is N22-S23 (see Figure 4) an unambiguous assignment to S23 is possible. (b) W-(i + 1)-HSQC and N-(i, i + 1)-HSQC. The highlighted cross peaks indicate the H^N and N frequencies of the two Asn residues following Trp residues. Since there are two Trp-Asn pairs in the PB1 domain, the two Asn residues (N13 and N58) which follow a Trp can not be distinguished. (c) D-(i + i)-HSQC and E-(i, i + 1)-HSQC. The marked cross peaks correspond to H^N and N frequencies of Glu residues following Asp residues. Analogously to (b) two pairs exist in the PB1 domain. Hence the two Glu (E44 and E55) residues which follow an Asp could not be distinguished. (d) TA-(i + 1)-HSQC and FYH-(i, i + 1)-HSQC. The tagged cross peak in both spectra indicates the H^N and N frequencies of His following Thr. Despite the fact that between His, Tyr, and Phe no spectroscopic discrimination is possible using amino acid type selective experiments, the Thr-His pair could be unambiguously identified. (e) S(i, i+1)-HSQC and P-(i – 1)-HSQC. The labeled cross peak in both spectra depicts the H^N and N frequencies of a Ser preceding a Pro. Since there is only one Ser-Pro pair present in the amino acid sequence S34 could be unambiguously assigned.


(a)
N-(i+1)-HSQC
S23 (N22)
S-(i,i+1)-HSQC
S23
(b)
W-(i+1)-HSQC
N58 (W57) or N13 (W12)
N13 (W12) or N58 (W57)
N-(i,i+1)-HSQC
N58 or N13
N13 or N58
(c)
D-(i+1)-HSQC
E44 (D43) or E55 (D54)
E55 (D54) or E44 (D43)
E-(i,i+1)-HSQC
E44 or E55
E55 or E44
(d)
TA-(i+1)-HSQC
H29 (T28)
FYH-(i,i+1)-HSQC
H29
(e)
S-(i,i+1)-HSQC
S34
P-(i-1)-HSQC
S34 (P35)
115
120
125
15N (ppm)
9.5
9.0
8.5
8.0
7.5
1H (ppm)

The modified strategy thus starts with an analysis of the protein sequence in form of a matrix as depicted in Figure 4 for the PB1 domain. Amino acids in the columns represent the (i + 1) amino acids while those in rows represent the i amino acids. The numbers in the matrix elements represent the number of occurrences of a particular pair of amino acids. It is quite common in protein domains that a particular pair of amino acids does appear only once or twice in the sequence and the signals are then either unambiguously assigned using the amino acid type selective experiments or the number of possibilities is reduced to the number of occurrences.

The most common situation where an unambiguous assignment is immediately possible is illustrated for the amino acid pair Asn-Ser. The PB1 domain of CDC24p contains 4 Ser and 10 Asn. The sequence pair Asn-Ser, however, is present only once in the sequence (N22-S23). A comparison of the spectra from the N-(i + 1)-HSQC and the S-(i, i + 1)-HSQC is shown in Figure 5a. The only cross peak appearing in both spectra necessarily corresponds to S23, which is therefore immediately assigned.

An example where the number of possible assignments is reduced to two is shown in Figure 5b. Two Trp residues are preceding an Asn in the sequence and a comparison between the W-(i + 1) and the N-(i, i + 1) will therefore yield two peaks in each spectrum that cannot be assigned unambiguously. A twofold ambiguity in the assignment of the Asn residues remains. The two Trp residues follow different amino acids and can thus easily be assigned, this assignment, however, can not be transferred to the Asn residues. The source of the problem is that it is not possible to transfer the information about the assignment from the W-(i + 1)-HSQC to the W-(i, i + 1)-HSQC. A similar case involves the new experiments presented above which is shown in Figure 5c. Two Glu residues follow two Asp residues in the sequence and can not be distinguished. The Asp residues, however, can be assigned due to their correlation to the preceding E42 and S53.

Another example is shown in Figure 5d. His can, in principle, not be assigned unambiguously since it appears in the spectra together with Phe and Tyr. In case of H29, however, an assignment is possible. H29 is the only aromatic amino acid that follows Thr or Ala. A comparison between the TA(i + 1) and the FHY(i, i + 1) therefore allows the identification of H29 in a straightforward manner.

Assignments are also possible when Pro is involved. The experiments for Pro do not follow the conventional pattern, but yield peaks for the preceding (P-(i–1)-HSQC) or the following (P-(i + 1)-HSQC) residue (Schubert et al. 2000). An example is the pair S34-P35 which is shown in Figure 5e. Here, a comparison between the P-(i–1) and the S-(i, i + 1) experiments provides an assignment of S34.

In case of the PB1 domain the majority of the 77 residues were assigned using only amino acid type selective experiments. A complete assignment using exclusively the selective experiments will not be possible since there are not selective experiments for all types of amino acids and because ambiguity remains in some of the selective experiments. However, a full assignment of the ^{1}H,^{15}N-HSQC was achieved using only the CBCA(CO)NNH/CBCANNH pair of three-dimensional experiments in addition to the selective experiments. The assigned stretches derived from the selective experiments were easily confirmed by analyzing strip plots at the assigned ^{1}H,^{15}N-chemical shift pairs. Note that the assignment of the remaining gaps using the 3D spectra is straightforward since gaps often consist of only a few residues and the number of unassigned spin systems is drastically reduced.

As can be seen from the experimental times given in Table 1 the more elaborate amino-acid selective experiments exhibit a reduced sensitivity compared to the shorter standard experiments. While experiments for Gly, Ala and Ser have similar signal-to-noise as the standard 3Ds they are derived from, the extension of the sequence in the other experiments will restrict the applicability to protein domains up to 150 amino acids. Deuterium decoupling can easily be added to some of the experiments to make them applicable to larger systems, the majority of the sequences, however, uses the MUSIC selection and thus require non-deuterated proteins. The presented assignment procedure will be robust and reliable since the information from the selective experiments is very specific. While the new strategy has already proven to drastically speed up manual assignment, its reliability should be particularly

helpful to improve the results of an automated assignment procedure.

Conclusion

In conclusion we have developed a modified strategy for the assignment of protein $^{1}H,^{15}N$-HSQC spectra. The strategy is based on a set of amino acid type selective triple resonance experiments, which may be used in conjunction with a minimum set of conventional triple resonance experiments. Increased sensitivity using cryogenic probes will allow for a reduction in the measurement time of the 2D experiments by reducing the number of scans without sacrificing resolution. Analysis of the amino acid sequence and combined evaluation of the selective experiments provide information for many unambiguous assignments. While this strategy has been successfully used in manual assignments, the selective and redundant information in amino acid type selective experiments increases the reliability and thus the robustness of automated assignment procedures. Results about a new algorithm and the utilization of the new strategy in automated assignment will be presented elsewhere.

Acknowledgements

Support from the Forschungsinstitut für Molekulare Pharmakologie is gratefully acknowledged. The work was supported by a grant of the BMBF (01 GG 9812, Leitprojekt 'Proteinstrukturfabrik').

References

Cornilescu, G., Bax, A. and Case, D.A. (2000) *J. Am. Chem. Soc.*, **122**, 2168–2171.

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.*, **B 99**, 201–207.

Grzesiek, S., Anglister, J. and Bax, A. (1993) *J. Magn. Reson.*, **B 101**, 114–119.

lkura, M., Kay, L.E. and Bax, A. (1991) *J. Biomol. NMR*, **1**, 299–304.

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

Kay, L.E., Xu, G.-Y., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J.D. (1993) *J. Magn. Reson.* B, **101**, 333–337.

Labudde, D., Leitner, D., Schmieder, P. and Oschkinat, H. (2002) *BRUKER-Rep.*, **150**, 8–11.

Logan, T.M., Olejniczak, E.T., Xu, R.X. and Fesik, S.W. (1992) *FEBS Lett.*, **314**, 413–418.

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

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

Montelione, G.T., Lyons, B.A., Emerson, S.D., and Tashiro, M. (1992) *J. Am. Chem. Soc.*, **114**, 10974–10975.

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

Noda, Y., Kohjima, M., Izaki, T., Ota, K., Yoshinaga, S., Inagaki, F., Ito, T. and Sumimoto, H. (2003) *J. Biol. Chem.*, **278**, 43516–43524.

Pellecchia, M., Iwai, H., Szyperski, T. and Wüthrich, K. (1997) *J. Magn. Reson.*, **124**, 274–278.

Ponting, C.P., lto, T., Moscat, J., Diaz-Meco, M.T., Inagaki, F. and Sumimoto, H. (2002) *Trends Biochem. Sci.*, **27**, 10.

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., 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.*, **153**, 186–192.

Schubert, Oschkinat, H. and Schmieder, P. (2001c) *J. Biomol. NMR*, **20**, 379–384.

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

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

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

Sklenář, V., Piotto, M., Leppik, R. and Saudek, V. (1993) *J. Magn. Reson.*, **A 102**, 241–245.

Terasawa, H., Noda, Y., lto, T., Hatanaka, H., lchikawa, S., Ogura, K., Sumimoto, H. and Inagaki, F. (2001) *EMBO J.*, **20**, 3947–3956.

Wilson, M.l., Gill, D.J., Perisic, O., Quinn, M.T. and Williams, R.L. (2003) *Mol. Cell*, **12**, 39–50.