Journal of Biomolecular NMR, 4 (1994) 123–128
ESCOM

J-Bio NMR 169

^{1}H, ^{13}C and ^{15}N NMR backbone assignments of the 269-residue serine protease PB92 from *Bacillus alcalophilus**

Rasmus H. Fogh[a], Dick Schipper[b], Rolf Boelens[a,**] and Robert Kaptein[a]

[a]*Department of Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands*
[b]*Gist-Brocades B.V., Research and Development, P.O. Box 1, 2600 MA Delft, The Netherlands*

Received 18 October 1993
Accepted 26 October 1993

Keywords: Triple-resonance spectroscopy; NMR spectrum assignment; Secondary structure

SUMMARY

The ^{1}H, ^{13}C and ^{15}N NMR resonances of the backbone of serine protease PB92 have been assigned. This 269-residue protein is one of the largest monomeric proteins assigned so far. The amount and quality of information available suggest that even larger proteins could be assigned with present methods. Measured chemical shifts show excellent agreement with the secondary structure.

The constant improvement of NMR techniques (for a review, see Bax and Grzesiek, 1993) has permitted the assignment of the NMR spectra of ever larger proteins (e.g. Grzesiek et al., 1992; Powers et al., 1993; Xu et al., 1993; Yamazaki et al., 1993). The use of $^{13}C/^{15}N$-labelled proteins in combination with heteronuclear 3D and 4D experiments alleviates the problem of spectral overlap, and allows specific assignment via heteronuclear J-coupling along the protein backbone, a pathway that is inherently less ambiguous than the use of NOEs. To illustrate the power of present-day methods, we present the backbone assignments of serine protease PB92. As a 269-residue protein with a molecular weight of 27 kDa, it is one of the largest protein monomers yet assigned (Wagner, 1993).

High-alkaline serine protease PB92 is a member of the subtilisin family of proteases (EC 3.4.21.14), which are of interest both as well-studied examples of enzyme catalysis (Kraut, 1977) and as molecules of industrial interest in their role as components of washing powders (Shaw,

*Supplementary material available from the authors: One table containing the backbone ^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, ^{13}CO and $^{1}H^{\alpha}$ assignments for serine protease PB92.

**To whom correspondence should be addressed.

Abbreviations: 2D/3D, two-/three-dimensional; HSQC, Heteronuclear Single Quantum Coherence; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser effect.

Experiment nomenclature (e.g. HNCO, HN(CO)(CA)HA) follows the conventions used elsewhere (e.g. Ikura et al., 1990).

1987). The protein shows a high degree of homology with subtilisin Carlsberg. The crystal structure of serine protease PB92 has been determined at 1.75 Å resolution (Van der Laan et al., 1992), but data derived from this structure were not used at any stage of the assignment presented in this paper.

Serine protease PB92 was isolated from the *Bacillus* strain PB92 (Van Eekelen et al., 1989; Van der Laan et al., 1992). The protein was produced as described by Teplyakov et al. (1992), except that the yeast extract was replaced either by ^{15}N-labelled yeast (cultured with $^{15}NH_4OH$) or by a doubly ^{13}C,^{15}N-labelled carbohydrate/amino acid medium. Purification involved cation exchange chromatography and ultrafiltration with a YM 10 membrane (Amicon). The protein was inhibited and aged as described by Van der Drift et al. (1985) and again submitted to ultrafiltration. NMR samples were 2 mM protein in 25 mM deuterated acetate buffer at pH 5.0 in H_2O : D_2O 95 : 5. A sample temperature of 42 °C was used throughout.

Backbone assignment of protease PB92 was carried out using [^{1}H,^{15}N,^{13}C] triple-resonance experiments (for a review, see Clore and Gronenborn, 1991; references for individual experiments are given in the legend to Fig. 1). The first step consisted in connecting the proton and nitrogen frequencies of each backbone NH group to frequencies in residues i and i–1 by means of HN-X 3D experiments. The frequencies of the HN,N pairs were sufficiently well resolved (if sometimes barely) to permit all the frequencies found to be connected to one unique NH pair. HNCA, HN(CO)CA, and HNCO experiments (Fig. 1) allowed C^α(i), C^α(i–1) and CO(i–1) frequencies to be determined for essentially all NH groups. The HN(CA)CO experiment (Fig. 1) had a less favourable signal-to-noise ratio, but still around 80% of the possible HN-N-CO(i) peaks could be found. Furthermore, analysis of a TOCSY-(^{15}N)HSQC experiment gave the H^α(i) resonance frequencies for over 90% of the NH groups. To expand this set of assignments further, an HCACO experiment was carried out, which proved to contain all but three non-glycine H^α-C^α-CO peaks (the experiment was set up to filter out CH_2 groups). These peaks provided a valuable check on the C^α-CO combinations found already, and — by fitting peaks to the known H^α and C^α(i) frequencies — allowed connection of all but a few of the NH groups to their CO(i) frequency. Peaks that could not be connected to any of the backbone NH groups yielded H^α-C^α-CO frequency triplets that accounted for all the prolines and the N-terminal residue.

With almost complete knowledge of C^α(i), C^α(i–1), CO(i) and CO(i–1) frequencies, adjacent NH groups in the sequence could be connected on the basis of two different frequency comparisons. This determined the predecessor residue unambiguously for 90% of the non-proline residues, allowing them to be connected into contiguous stretches between 3 and 35 residues in length. To combine these assigned stretches into a complete backbone assignment, we included as extra information (i) the identity of the prolines or N-terminus, as deduced from the absence of an amide proton; (ii) the identity of the glycines, as deduced from the shifts of C^α and peptide nitrogen, and the absence of a peak in the HCACO spectrum; and (iii) the identity of the three methionine residues, which were known from the spectrum of a specifically labelled protein (data not shown). Using no further information, we were able to assign uniquely the backbone resonances of all 269 residues.

The assignments were checked against HN(CA)HA and HA(CA)(CO)NH experiments, against the side-chain protons observed in the TOCSY-(^{15}N)HSQC experiment, and against 2D (^{15}N)HSQC spectra of semiselectively labelled proteins (data not shown). Furthermore, an HCACO experiment was set up to give peaks for both CH and CH_2 groups, in order to check the

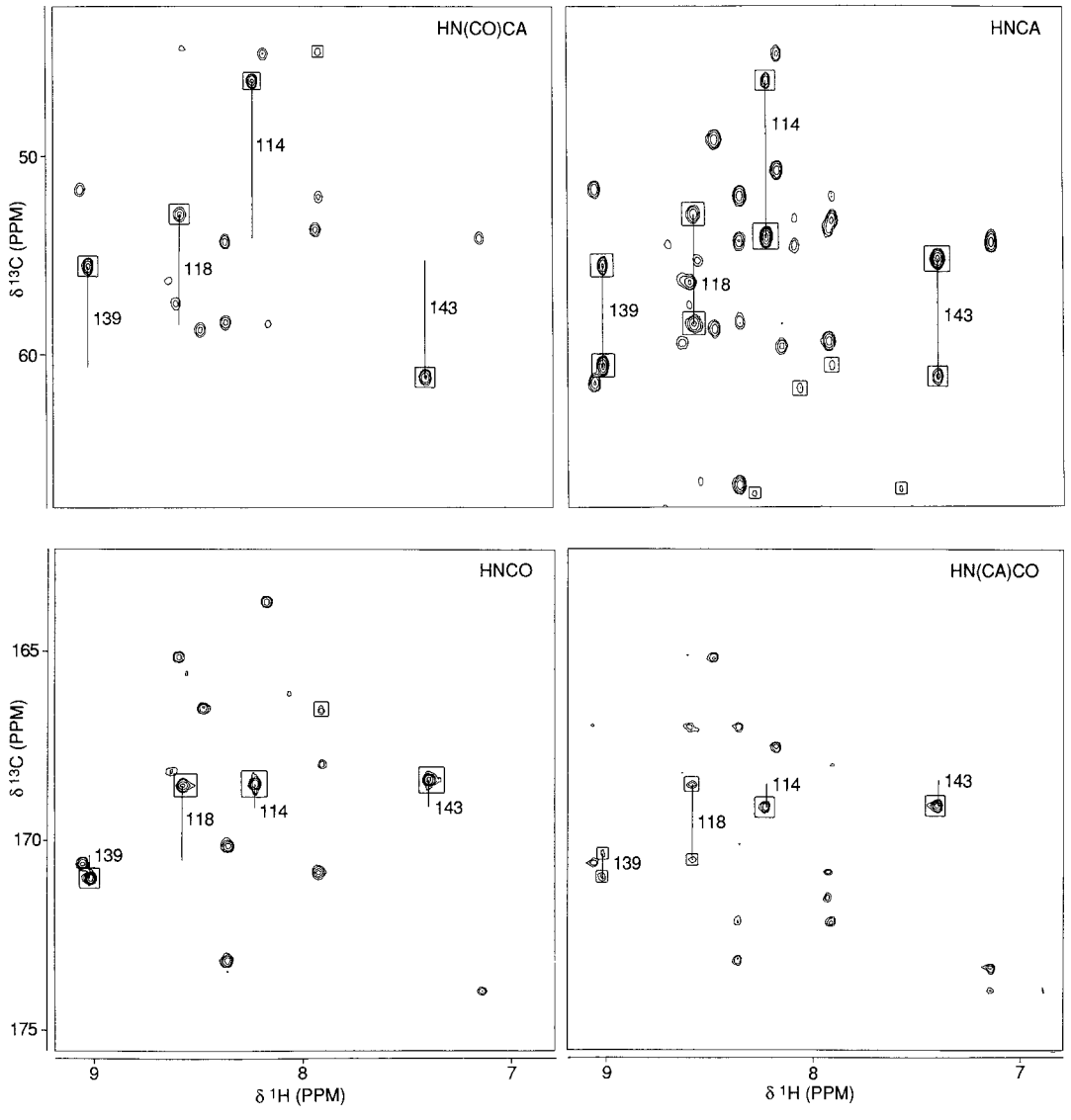

Fig. 1. Nitrogen plane at 122.2 ppm for HN(CO)CA, HNCA, HNCO, and HN(CA)CO experiments of fully enriched [^{13}C,^{15}N] PB92 serine protease. All peaks and all residues with a nitrogen resonance belonging in this plane are marked. Residues (number given) are marked with a line at the peptide NH frequency, connecting the intraresidue and preceding residue C^{α} (or CO) frequencies. Peaks belonging in this plane are enclosed in a box. All NMR experiments were performed on a Bruker 600 MHz AMX spectrometer, equipped with a three-channel NMR interface, and a triple-resonance ^{1}H/^{15}N/^{13}C probe with an additional gradient coil. HNCA and HN(CA)CO (Grzesiek and Bax, 1992; Clubb et al., 1992a), HNCO and HN(CO)CA (Grzesiek and Bax, 1992), HN(CA)HA (Ikura et al., 1990; Clubb et al., 1992b), HA(CA)(CO)NH (Kay et al., 1990), and TOCSY-(^{15}N)HSQC (Marion et al., 1989) experiments were acquired with some modifications relative to the original pulse sequences. HCACO experiments with water suppression and coherence selection using field gradients were modified from Kay et al. (1990). Indirectly detected dimensions were acquired with 58–112 real points, extended 50–100% using linear prediction with unmodified poles (Press et al., 1992), and Fourier transformed, zero-filling to 128 final points. Processing and spectrum analysis were done on Unix workstations, using software developed in our laboratory.

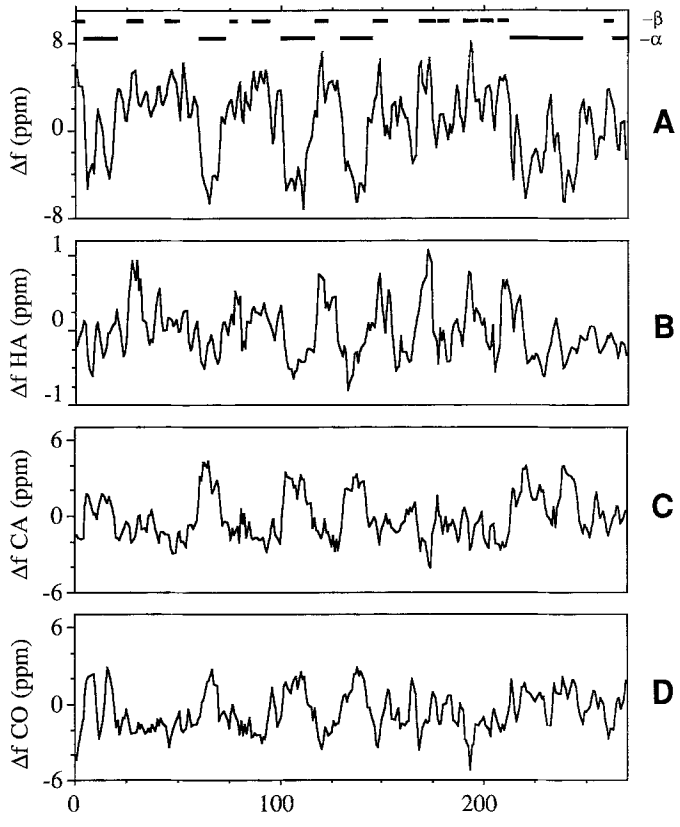

Fig. 2. Deviations Δf from random-coil chemical shifts for the linear combination of frequencies (H^{α}–C^{α}–CO) (A), and for the individual frequencies H^{α} (B), C^{α} (C), and CO (D). Note that the frequency deviations expected for carbons have an opposite sign to those expected for H^{α}, a fact that is reflected in the signs used in the linear combination in (A). Residues classed as α-helix (α) and β-sheet (β) by Van der Laan et al. (1992, and personal communication) are shown at the top of (A). Frequency values are smoothed with a three-point 1 : 2 : 1 moving average filter. For glycines the average of the two H^{α} frequencies was used.

frequencies of the glycines. The new experiments allowed several previously unobserved H^{α} protons to be identified, but did not lead to any changes in the assignments. These thus reflect an average of more than two independent pieces of information for each i–(i + 1) contact, plus considerable residue-type information. They encompass all backbone NH groups and C^{α} carbons, all but one CO carbon, and all but two H^{α} protons. The assigned frequencies will be deposited in the BioMagRes database (University of Wisconsin, Madison).

Even without analysing the NOE interactions, the present stage of assignment allows a qualitative determination of the secondary structure through analysis of the chemical shifts. As described by Wishart et al. (1991), H^{α} protons have *higher* resonance frequencies in β-sheet and *lower* frequencies in α-helix than in random coil, regardless of residue type. C^{α} and CO show a similar systematic variation, except that the frequency deviations here have opposite sign. Plots of the frequency deviations versus residue number for each of the three nuclei showed parallel behaviour (Figs. 2B–D), with deviations for H^{α} being smaller not only in absolute terms, but

relative to the random variation. Thus, an unweighted linear combination of the H^{α}, C^{α} and CO frequency deviations seemed the most convenient indicator of secondary structure (Fig. 2A). The correspondence between the chemical-shift deviations and the secondary structure as observed in the crystal (Van der Laan et al., 1992) is remarkable, even if the distinction between structure and coil is a little less clear-cut for β-sheet than for α-helix. The empirical nature of the model does not allow any conclusions about differences in secondary structure between PB92 in the crystal and in solution, except that they are (not surprisingly) small. It is tempting to speculate, however, that the rather small chemical-shift differences seen in residues 258–269, where the X-ray structure contains a β-sheet strand and a helix, might be a sign of increased flexibility in solution for these residues.

Although protease PB92 is one of the largest proteins assigned so far, the assignment procedure was remarkably simple on account of the complete set of backbone resonances that could be determined for each NH pair. This was in part due to the success of the HN(CA)CO experiment in our hands. It is worth noting that this experiment, although less sensitive than e.g. the HNCA experiment, can be applied successfully to proteins larger than 250 residues. It has been shown by Grzesiek et al. (1992) that the inclusion of NOEs, $H^{\alpha}(i-1)$ frequencies and more extensive residue-type information allows the assignment of proteins where the linewidth does not permit the HN(CA)CO experiment to be used. It appears that the limits of present experimental methods have not yet been reached, and that the assignment of proteins significantly larger than the one presented here might well be possible.

REFERENCES

Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 131–138.

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

Clubb, R.T., Thanabal, V. and Wagner, G. (1992a) *J. Magn. Reson.*, **97**, 213–217.

Clubb, R.T., Thanabal, V. and Wagner, G. (1992b) *J. Biomol. NMR*, **2**, 203–210.

Grzesiek, S., Döbeli, H., Gentz, R., Garotta, G., Labhardt, A.M. and Bax, A. (1992) *Biochemistry*, **31**, 8180–8190.

Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **96**, 432–440.

Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry*, **29**, 4659–4667.

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

Kraut, J. (1977) *Annu. Rev. Biochem.*, **46**, 331–358.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) *Biochemistry*, **28**, 6150–6156.

Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. and Clore, G.M. (1993) *Biochemistry*, **32**, 6744–6762.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1992) *Numerical Recipes (FORTRAN Version)*, Cambridge University Press, Cambridge.

Shaw, W.V. (1987) *Biochem. J.*, **246**, 1–17.

Teplyakov, A.V., Van der Laan, J.M., Lammers, A.A., Kelders, H., Kalk, K.H., Misset, O., Mulleners, L.J.S.M. and Dijkstra, B.W. (1992) *Protein Eng.*, **5**, 413–420.

Van der Drift, A.C.M., Beck, H.C., Dekker, W.H., Hulst, A.G. and Wils, E.R. (1985) *Biochemistry*, **24**, 6894–6903.

Van Eekelen, C.A., Van der Laan, J.M., Mulleners, L.J.S.M., Misset, O., Cuperus, R.A. and Lensink, J.H.A. (1989) European Patent Application 0 378229–A1.

Van der Laan, J.M., Teplyakov, A.V., Kelders, H., Kalk, K.H., Misset, O., Mulleners, L.J.S.M. and Dijkstra, B.W. (1992) *Protein Eng.*, **5**, 405–411.

Wagner, G. (1993) *J. Biomol. NMR*, **3**, 375–385.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.

Xu, R.X., Nettesheim, D., Olejniczak, E.T., Meadows, R., Gemmecker, G. and Fesik, S.W. (1993) *Biopolymers,* **33**, 535–550.

Yamazaki, T., Yoshida, M. and Nagayama, K. (1993) *Biochemistry*, **32**, 5656–5669.