Journal of Biomolecular NMR, 5 (1995) 259–270
ESCOM

J-Bio NMR 244

Complete ^{1}H, ^{13}C and ^{15}N NMR assignments and secondary structure 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]*Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, NL-3584 CH Utrecht, The Netherlands*
[b]*Gist-Brocades B.V., Research and Development, P.O. Box 1, NL-2600 MA Delft, The Netherlands*

Received 16 September 1994
Accepted 29 September 1994

Keywords: Triple-resonance spectroscopy; Multidimensional NMR; ^{1}H, ^{15}N and ^{13}C NMR assignment; Serine protease

Summary

The ^{1}H, ^{13}C and ^{15}N NMR resonances of serine protease PB92 have been assigned using 3D triple-resonance NMR techniques. With a molecular weight of 27 kDa (269 residues) this protein is one of the largest monomeric proteins assigned so far. The side-chain assignments were based mainly on 3D H(C)CH and 3D (H)CCH COSY and TOCSY experiments. The set of assignments encompasses all backbone carbonyl and CH_n carbons, all amide (NH and NH_2) nitrogens and 99.2% of the amide and CH_n protons. The secondary structure and general topology appear to be identical to those found in the crystal structure of serine protease PB92 [Van der Laan et al. (1992) Protein Eng., **5**, 405–411], as judged by chemical shift deviations from random coil values, NH exchange data and analysis of NOEs between backbone NH groups.

Introduction

High alkaline serine protease PB92 (Maxacal®) is a subtilisin-like serine protease belonging to the subtilase family of proteases (Siezen et al., 1991), which are of interest both as well-studied examples of enzyme catalysis (Kraut, 1977) and as molecules of considerable industrial importance, being protein-degrading components of washing powders (Shaw, 1987). Subtilases are currently being used as model systems for protein engineering studies (see Teplyakov et al. (1992) and references cited therein) to improve stability and performance, and mutant enzymes are applied in modern detergent powders. For this purpose, knowing the solution structure and dynamics would be of immediate practical interest. The crystal structure of serine protease PB92 has been determined at 1.75 Å resolution (Van der Laan et al., 1992) and appears to have high similarity with various other members of the subtilase family, such as subtilisin BPN' (Bott et al., 1988), subtilisin Carlsberg (Bode et al., 1987), thermitase (Gros et al., 1989) and proteinase K (Betzel et al., 1988). However, both information regarding dynamics and a solution structure, as can be obtained from NMR studies, are lacking.

Serine proteases have received much attention as a subject for NMR studies (see Consonni et al. (1992) for references); however, due to the high autolytic capacity of these proteases in the concentration range normally used for NMR studies, some of these studies have been unintentionally performed on partially degraded enzymes. To prevent degradation during the (many) measurements necessary for the assignments of the resonances and the determination of the 3D structure of proteases in general, use has to be made of highly purified protein and a tightly bound inhibitor for successful NMR studies. Diisopropylfluorophosphate (used in this study) is such an inhib-

*To whom correspondence should be addressed.

Abbreviations: 2D/3D/4D, two-/three-/four-dimensional; HSQC, heteronuclear single-quantum coherence; HMQC, heteronuclear multiple-quantum coherence; COSY, correlation spectroscopy; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser enhancement (connectivity); NOESY, 2D NOE spectroscopy. Experiment nomenclature (H(C)CH, etc.) follows the conventions used elsewhere [e.g. Ikura et al. (1990) Biochemistry, **29**, 4659–4667].

Supplementary Material: Tables containing the ^{1}H, ^{15}N and ^{13}C assignments, the hydrogen exchange data and results of secondary structure and CSI analyses for serine protease PB92 are available from the authors.

itor that is stable for many weeks after the process of ageing (Van der Drift et al., 1985).

With a molecular weight of 27 kDa (269 residues), serine protease PB92 is one of the largest protein monomers studied in detail by NMR (Wagner, 1993). The constant improvement in NMR techniques (for a review, see Bax and Grzesiek (1993)) has permitted the assignment of the NMR spectra of such large 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, ultimately allowing the protons involved in an NOE to be distinguished on the basis of the frequencies of both the proton and the attached carbon/nitrogen. Assignment of protein backbone resonances can be done on the basis of triple-resonance $^{1}H,^{15}N,^{13}C$ experiments without recourse to the – inherently more ambiguous – NOE connectivities. For the assignment of amino acid side chains, the larger C-C J-couplings permit bond-bond connections to be observed, even for quite large proteins with correspondingly large line widths.

The assignment of the backbone resonances of serine protease PB92 has been described previously (Fogh et al., 1994). Similar results have been obtained for a closely related serine protease, subtilisin 309 from *Bacillus lentus*, by Remerowski et al. (1994). Here we present nearly complete ^{1}H, ^{13}C and ^{15}N assignments of serine protease PB92 using 3D triple-resonance NMR techniques. The secondary chemical shifts, NH exchange data and NOE data showed close correspondence with the secondary structure and the general structural topology of serine protease PB92 in the crystal. Data from the crystal structure, however, have not been used at any stage of the assignment procedure.

Materials and Methods

Serine protease PB92 was isolated from the *Bacillus alcalophilus* 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, obtained from H. Oschkinat (EMBL, Heidelberg). Purification involved only 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 contained 2 mM protein in 25 mM deuterated acetate buffer at pH 5.0 in H_2O:D_2O (95:5) or pure D_2O. Except where otherwise indicated a sample temperature of 315 K was used, and water suppression was done by presaturation.

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. 3D HNCA and HN(CA)CO (Clubb et al., 1992a; Grzesiek and Bax, 1992), 3D HNCO and HN(CO)CA (Grzesiek and Bax, 1992), 3D HN(CA)HA (Ikura et al., 1990a; Clubb et al., 1992b), and 3D TOCSY-($^{15}N,^{1}H$)-HSQC (Marion et al., 1989a) experiments were acquired with some modifications relative to the original pulse sequences (Vis et al., 1994). 3D HCACO experiments with water suppression and coherence selection using field gradients were modified from Kay et al. (1990). The pulse sequence for a 3D HA(CA)(CO)NH experiment was derived from the 3D HBHA(CBCA)(CO)NH experiment of Grzesiek and Bax (1993) by removing the pulses effecting magnetisation transfer between the C^{α} and C^{β} spins. 3D HCCH-COSY and HCCH-TOCSY experiments (DIPSI-3 spin-lock sequence, duration 15.5 ms), using pulsed field gradients for removal of artifacts and partial water suppression, were carried out according to the description by Bax et al. (1990a,b) with some modifications (Vis et al., 1994). In the 3D HA(CA)(CO)NH experiment, resonances normally suppressed by water presaturation were recovered using the SCUBA method (Brown et al., 1988). Two 3D CB-CA(CO)NH experiments were acquired according to Grzesiek and Bax (1993), with delays optimised for the observation of NH and NH_2 groups, respectively. 3D NOESY-($^{15}N,^{1}H$)-HSQC and NOESY-($^{13}C,^{1}H$)-HSQC experiments were acquired using the general scheme of Marion et al. (1989b) and Fesik and Zuiderweg (1988), and 3D ^{15}N and ^{13}C HMQC-NOESY-HMQC experiments were carried out according to Ikura et al. (1990b) and Frenkiel et al. (1990). NOESY mixing times were 145 ms for the 3D ^{15}N HMQC-NOESY-HMQC and 150 ms for other 3D NOESY spectra. 3D spectra were acquired with at least 16 scans per transient. Quadrature detection was obtained with either the TPPI (Marion and Wüthrich, 1983) or the States-TPPI (Marion et al., 1989c) method. In all cases, the numbers of real points acquired were 58–80 for nitrogen and aliphatic carbon, 80–96 for CO and typically 200–256 for indirectly detected protons. In carbon and nitrogen dimensions, the measured points were extended 50–100% using forward linear prediction with unmodified poles (Press et al., 1992), and Fourier transformed using cosine-square windows (typically shifted by 36°), zero-filling to 128 final points. In HCCH experiments the ω_2 carbon dimension was folded once, using a spectral width of 45 ppm. The carrier frequency was mostly positioned in the centre of the aliphatic carbon range (40.0 ppm), but the centre frequency of the final processed spectrum was moved to 22.3 ppm by applying a linear phase correction to the free induction decays before processing. Proton dimensions were typically processed to 256 final points in indirectly detected and 1024 points in directly detected dimensions, using Lorentzian–

Gaussian windows optimised for line widths of 30–45 Hz. The point-to-point resolutions in the final 3D spectra were 0.35 ppm/point for nitrogen, 0.11 ppm/point for carbonyl, 0.22 ppm/point for C^{α}, 0.35 (0.5) ppm/point for folded (unfolded) aliphatic carbon dimensions and 0.04 (0.01) ppm/point for indirectly (directly) detected protons.

Constant-time 2D (^{13}C,^{1}H)-HSQC spectra (Vuister and Bax, 1992) were carried out with a constant-time carbon evolution period of 26.6 ms, so that signals from carbons with odd and even numbers of carbon coupling partners had opposite sign. Edited versions of this experiment, showing only carbons coupled to nitrogen, carbonyl carbon or aromatic carbon, were acquired according to Grzesiek and Bax (1993). A modified version of this experiment was designed to select carbon bound to either an odd or an even number of protons. Two half-experiments were acquired, differing only in the position of a proton 180° pulse, such that CH coupling was either decoupled or allowed to evolve for a period of J_{CH}^{-1}. Adding or subtracting the resulting experiments then yielded spectra containing respectively CH_2 or CH/CH_3 carbons. A ^{15}N HMQC experiment (Müller, 1979; Bax et al., 1983; Bendall et al., 1983) with water suppression using presaturation was acquired on a protonated sample immediately after dissolving in D_2O. 2D experiments were acquired with varying resolution in ω_1, and were generally processed as described above for 3D experiments.

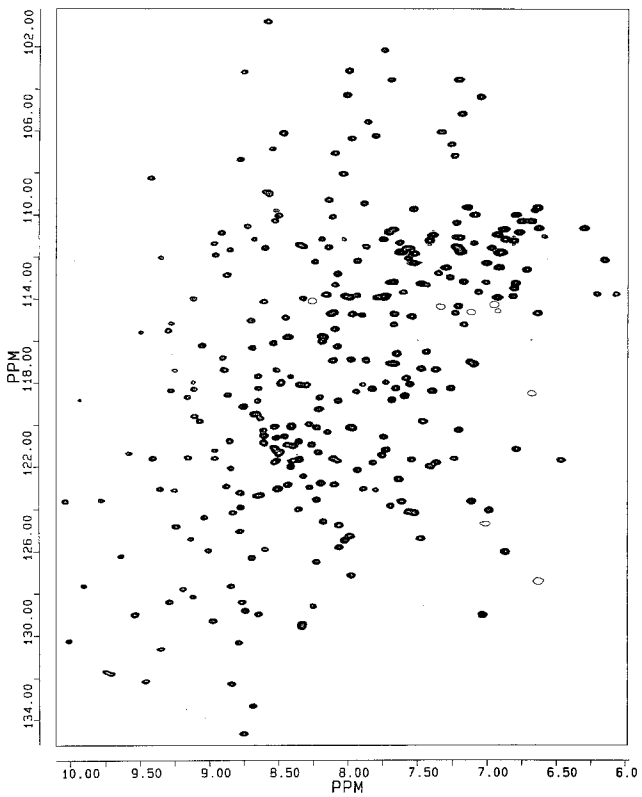

Fig. 1. 2D (^{15}N,^{1}H)-HSQC spectrum of serine protease PB92. Positive levels are drawn with a full set of contours; negative levels (showing folded-in peaks from Arg-N^{ε} and Lys-C^{ζ}) with one contour only.

Frequencies of NH groups were measured in a 2D HSQC spectrum (Bax et al., 1990c; Norwood et al., 1990) with the following parameters: ω_1/ω_2 2500/9615 Hz spectral width, 384 real/512 complex points acquired, ω_1 extended to 896 real points by linear prediction, 1024/1024 points final size. H^{α}, C^{α} and CO frequencies were measured in 3D HCACO or 3D HNCO experiments with resolutions as given above. Side-chain proton and carbon frequencies were first measured in 3D HCCH experiments (see above), but were to a large extent remeasured from a constant-time 2D HSQC (without decoupling of carbonyl carbon) with the parameters ω_1/ω_2 10560/5882 Hz spectral width, 274 complex/320 complex points acquired, ω_1 extended to 448 complex points by linear prediction, 1024/1024 points final size.

Proton chemical shifts were calibrated relative to H_2O at 315 K at 4.62 ppm. Due to variations in the actual sample temperature, the spectrum calibration that reproduced the protein chemical shifts varied by ±0.02 ppm between different spectra. Carbon chemical shifts were calibrated relative to dioxane in (pure) H_2O at 69.43 ppm, using the spectrometer frequency-generating system as a secondary reference. Nitrogen chemical shifts were calibrated by setting the zero point of the ^{15}N frequency scale equal to 0.10132914 times the proton zero frequency. Measured chemical shifts will be deposited in the BioMagResBank.

Processing and spectrum analysis were done on Unix workstations using software developed in this laboratory.

Results and Discussion

Backbone assignments

The assignment of the protein backbone resonances has been described elsewhere (Fogh et al., 1994) and will be treated only briefly. 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 the 3D HNCA, HNCO, HN(CO)CA and HN(CA)CO experiments. These experiments yielded all the N,HN pairs, and 100% of the connected CO(i – 1) and C^{α}(i – 1), 99.8% of the C^{α}(i) and ca. 80% of the CO(i) frequencies. The dispersion of the N,NH pairs can be judged from Fig. 1. A 3D TOCSY-(^{15}N,^{1}H)-HSQC experiment further gave the H^{α}(i) resonance frequencies for over 90% of the NH groups, while a 3D HCACO spectrum was found to contain all but three non-glycine H^{α}-C^{α}-CO peaks (the experiment was set up to filter out CH_2 groups). The H^{α}-C^{α}-CO frequency triples measured in this experiment provided a valuable check on the C^{α}-CO combinations already found, and – by fitting peaks to the known C^{α}(i) and H^{α}(i) or CO(i) frequencies – allowed all but a couple of the missing H^{α}(i) and CO(i) frequencies to be connected to their NH groups. Furthermore, the experiment gave

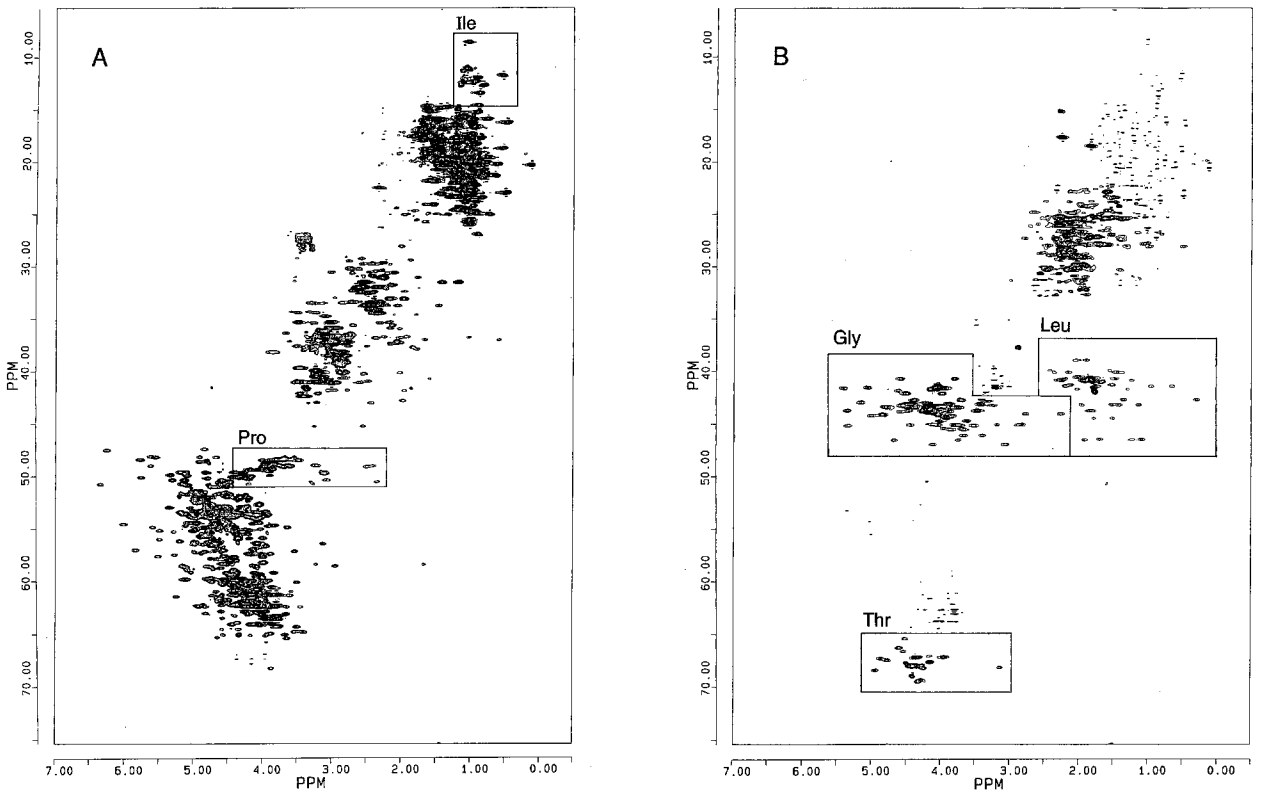

Fig. 2. 2D constant-time (^{13}C,^{1}H)-HSQC spectrum of serine protease PB92, acquired with a constant-time evolution period of 26.6 ms. (A) Positive levels, showing peaks with an odd number of carbon coupling partners. The regions of the spectrum containing Ile-C^{δ} and Pro-C^{δ} resonances are enclosed in boxes. (B) Negative levels, showing peaks with an even number of carbon coupling partners. The regions of the spectrum containing Gly-C^{α}, Thr-C^{β} and Leu-C^{β} resonances are enclosed in boxes.

H^{α}, C^{α} and CO frequencies for all the prolines and the N-terminal residue, which were seen here for the first time.

With almost complete knowledge of $C^{\alpha}(i)$, $C^{\alpha}(i-1)$, CO(i) and CO(i − 1) frequencies, adjacent NH groups in the sequence could be connected on the basis of two different frequency comparisons. The tolerance for deciding that two frequencies were different was as low as 66% of the point-to-point resolution, reflecting the high spectrum-to-spectrum frequency reproducibility and the accurate peak positions measured by computer interpolation. A simple program that performed the frequency comparison, followed by some pruning of low-probability candidates, unambiguously determined the predecessor residue for 90% of the non-proline residues, while giving multiple possibilities for the remainder. These sequential assignments, when combined with minimal residue-type information (the identity of glycine, proline and methionine residues), were sufficient to assign uniquely the backbone resonances of all 269 residues. The assignments were checked against data from 3D HN(CA)HA and HA(CA)-(CO)NH experiments, against the side-chain protons observed in the 3D TOCSY-(^{15}N,^{1}H)-HSQC experiment and against 2D (^{15}N,^{1}H)-HSQC spectra of selectively labelled proteins. A 3D CBCA(CO)NH experiment, acquired after publication of the original assignments, contained all expected peaks (except eight that were obscured by overlap), further confirming the correctness of the sequential assignments.

Side-chain assignments

The assignment of the side-chain resonances was based on 3D HCCH-TOCSY and -COSY experiments as described by Clore et al. (1990), with constant-time and edited 2D (^{13}C,^{1}H)-HSQC experiments used for checking and to provide starting points for further search. The (H)CCH and H(C)CH-TOCSY spectra were especially useful, since they frequently allowed essentially all resonances in a residue to be observed on a single C,H pair. The (H)CCH and HC(C)H-COSY, on the other hand, were more sensitive in some cases and helped, especially in cases of overlap, to assign frequencies to their correct position in the residue. Spectra in D_2O and at a lower temperature (305 K) were used to find resonances otherwise hidden by the residual water resonance, or bleached out by solvent presaturation.

The starting point for the assignment was the complete set of C^{α},H^{α} pairs in the molecule (only one was missing at this stage), plus a number of frequencies for β and γ protons that had been observed in the 3D TOCSY-(^{15}N,^{1}H)-HSQC experiment. The latter encompassed all

alanine β-CH_3 resonances (except for Ala[1]), as well as between zero and three resonances (1.5 on average) for other non-glycine, non-proline residues. These data, together with the knowledge of amino acid types, proved sufficient to distinguish all different side chains in the spectra, except for one case (see below). By systematically going through the C^{α},H^{α} pairs it proved possible to find most of the resonances for the majority of residues. These results could then be completed, and the resonances assigned to their proper nuclei, by checking the spectra at the, now known, positions of the other resonances. Where this procedure did not suffice to assign a residue completely, the 2D (^{13}C,^{1}H)-HSQC spectrum could be consulted to find new starting points for analysis, that could then be connected up to the assigned part of the molecule (Fig. 2). A 3D CBCA(CO)NH experiment that was acquired late in the procedure to help assign remaining serine and proline C^{β} resonances, yielded essentially 100% of the expected C^{α} and C^{β} peaks, confirming assignments already made. The results found for different types of amino acids are given below.

Glycine, alanine, asparagine/aspartate and aromatic amino acids

Essentially, these residues could be completely assigned straight from connectivities to the C^{α},H^{α} pairs. Aromatic resonances were identified in a 2D CT-HSQC experiment with parameters optimised for this purpose, and were connected with each other and with the rest of the residues using 3D NOESY-(^{13}C,^{1}H)-HSQC and ^{13}C HMQC-NOESY-HMQC experiments, optimised for aromatic resonances. Asparagine (and glutamine) side-chain amide resonances were connected using a 3D CBCA(CO)NH experiment, optimised for NH_2 groups. Of the 33 side-chain NH_2 groups, 31 could be found and assigned in this spectrum for the expected nitrogen chemical shift range of 109–116 ppm. Cases of near-overlap in the carbon frequencies could be resolved by reference to NOEs in a 3D NOESY-(^{15}N,^{1}H)-HSQC experiment. At this stage, peaks remaining in the 'NH_2' region of the spectrum did not fit with the two asparagines that remained unassigned; since no corresponding NOESY-HSQC peaks were found, the peaks in question presumably arose from low concentrations of breakdown products of the protein. The NH_2 groups of the remaining asparagines 75 and 153 were found at an unusually high ^{15}N chemical shift of 118 ppm in the 3D CBCA(CO)NH experiment, where similar resonances had also been observed (but not assigned) by Remerowski et al. (1994). All the assigned NH_2 groups gave rise to observable peaks in the 3D NOESY-(^{15}N,^{1}H)-HSQC spectrum, and all but three were assigned on the basis of two independent frequency comparisons.

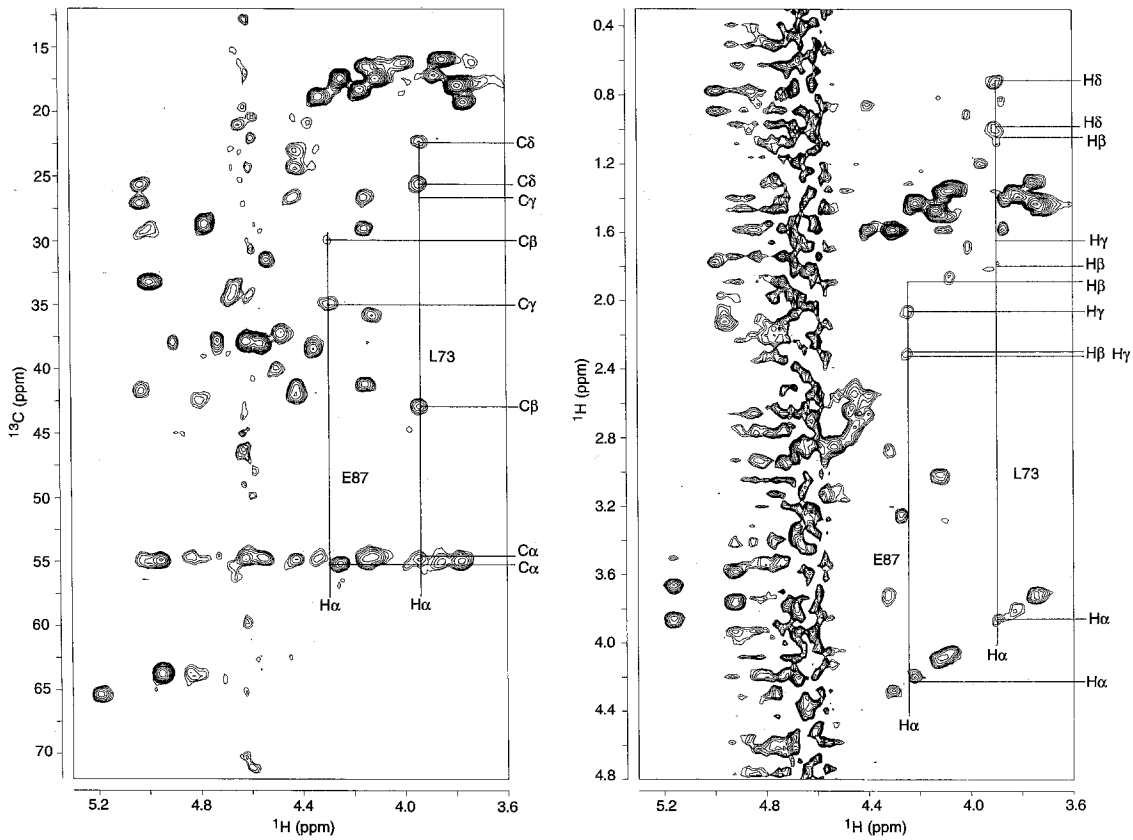

Fig. 3. 3D (H)CCH-TOCSY in D_2O (left) and 3D H(C)CH-TOCSY in H_2O (right), ω_2 = 55.4 ppm. The spin systems of Glu^{87} and Leu^{73}, both belonging in this plane, are marked.

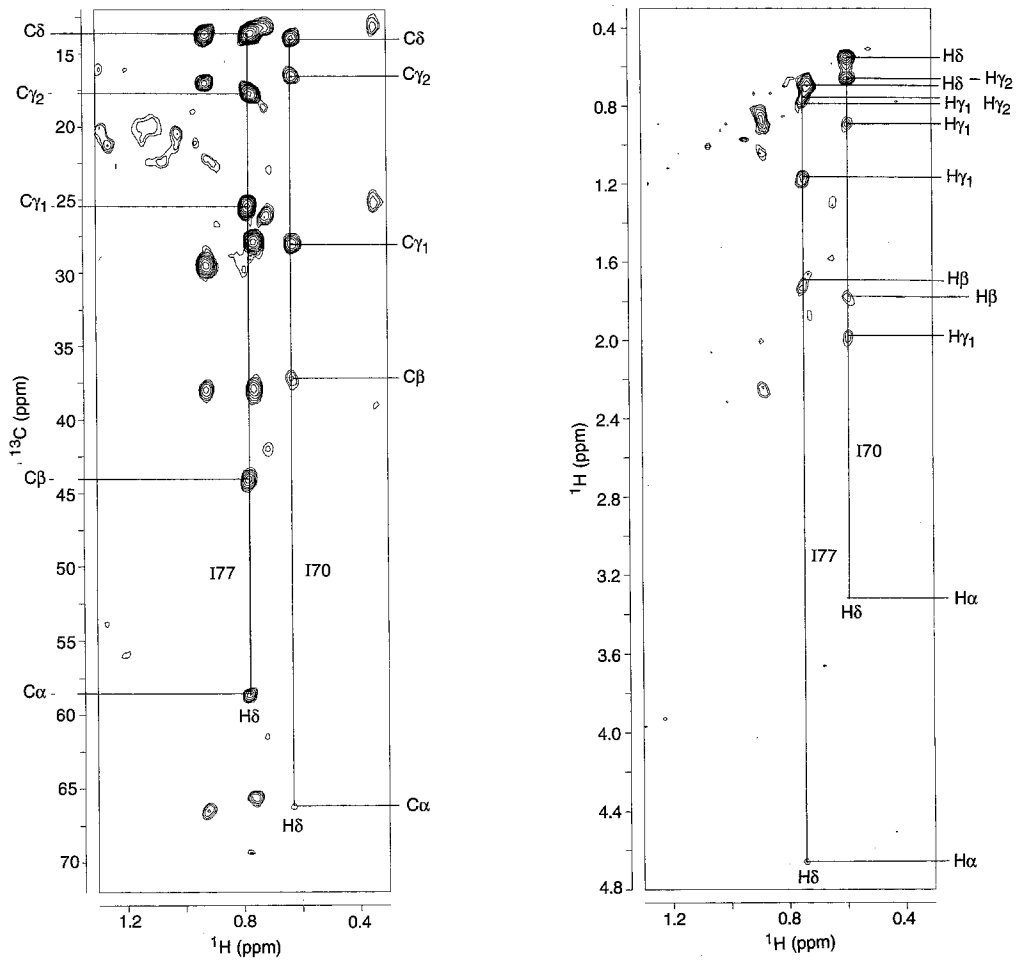

Fig. 4. 3D (H)CCH-TOCSY in D_2O (left) and 3D H(C)CH-TOCSY in H_2O (right), ω_2 = 14.4 ppm. The spin systems of Ile^{70} and Ile^{77}, both belonging in this plane, are marked.

Glutamate, glutamine and methionine

With the 15.5 ms mixing time used for the ^{13}C-^{13}C spin-lock in the 3D H(C)CH-TOCSY experiments, the C^{γ}-C^{α}-H^{α} peak was generally strong, whereas the C^{β}-C^{α}-H^{α} peak was quite weak, but still observable. The example shown in Fig. 3 is typical in this respect. It proved possible to observe all C^{β} and C^{γ} and more than half of the side-chain proton resonances in the glutamate, glutamine and methionine residues from the C^{α},H^{α} pairs, rendering the complete assignment of these residues straightforward. The methionine ε-CH_3 resonances were found in 2D constant-time HSQC spectra of fully and/or specifically labelled protein, and were later assigned based on NOEs to the H^{β} and H^{γ} protons observed in a 3D NOESY(^{13}C, ^{1}H)-HSQC experiment.

Valine and leucine

Starting on C^{α},H^{α} it is possible to observe all resonances for valine, and one or both of C^{δ} and H^{δ} for leucine in almost all cases. A typical leucine spin network is shown in Fig. 3. Almost the whole spin system can be found again on the methyl groups with their narrow intense peaks, so that it is easy to determine which methyl proton corresponds to which methyl carbon. Peaks involving the weak and frequently overlapped leucine γ-CH resonances were often hidden or absent on other CH_n groups, but peaks from γ-CH to other resonances could generally be observed.

Isoleucine

Here, only the β and γ-CH_3 resonances could be observed directly on C^{α}, but, as can be seen in Fig. 4, (almost) all resonances could be found starting on the δ-CH_3 groups. These can easily be identified in a 2D HSQC spectrum from their low carbon chemical shift (Fig. 2).

Threonine

Threonines, like isoleucines, could generally not be completely assigned from the α-CH group, partly because the large frequency difference between C^{β} and C^{γ} diminishes the efficiency of the TOCSY transfer between these nuclei, and partly because some resonances were lost through the effect of water presaturation. The threonine β-CH groups are, however, uniquely recognisable in a

constant-time HSQC spectrum (Fig. 2), and it proved simple to assign the threonines from this starting point. One of the threonines to be assigned in this way was Thr[218], whose C^{α} had previously been misassigned in the backbone assignments because it was not recognised that the resonance observed in the 3D HNCA experiment was folded from a position outside the spectral width.

Serine

The large number of serines (33) present in serine protease PB92 and the resulting high degree of overlap in the region of the spectrum that contains serine C^{α}, C^{β}, H^{α} and H^{β} resonances, made a complete assignment difficult. Ultimately, recourse was taken to a 3D CBCA(CO)NH experiment and a 2D HSQC experiment on a protein specifically labelled with ^{13}C at Ser-C^{β}. With the aid of these two experiments, the serines could be completely assigned.

Arginine, lysine and proline

These residues tend to be the more difficult to assign, because they consist of CH_2 groups that have broader lines, more carbon-carbon coupling partners and less chemical shift dispersion than e.g. α-CH groups. In about half of the cases it was possible to observe many proton resonances and all the carbon resonances directly on the α-CH, and these residues could be assigned fairly easily. In other cases very few side-chain resonances could be observed, making it necessary to use the end of the side chain as a starting point for analysis. Six of the eight arginines had observable ε-NH groups, from which H^{δ}, C^{δ} and sometimes H^{γ} resonances could be observed in various spectra and used for further analysis. After the assignment of these residues and the AMX spin systems was complete, the relatively few peaks that remained unassigned in the spectral region around 3 ppm (H), 40 ppm (C) could also be used, and four out of five observed (pairs of) peaks could eventually be connected with the four remaining, incompletely assigned arginines and lysines, chiefly on the basis of shared carbon chemical shifts. A few γ protons remained unassigned, because of overlap and the low intensity of peaks detected on CH_2 groups. For the prolines that remained unassigned, low sensitivity proved a problem. An example is seen in Fig. 5. Note that the figure depicts two prolines that were completely assigned relatively early in the procedure. Another com-

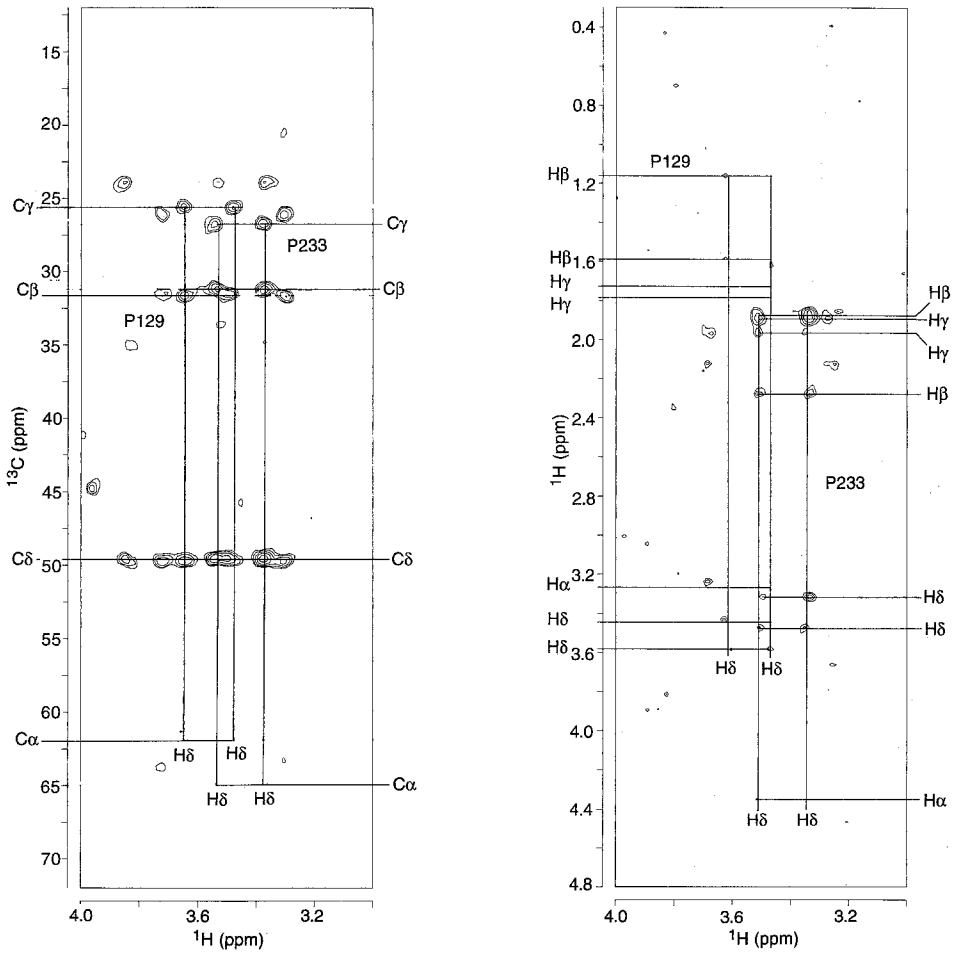

Fig. 5. 3D (H)CCH-TOCSY in D_2O (left) and 3D H(C)CH-TOCSY in H_2O (right), ω_2 = 50.0 ppm. The spin systems of Pro[129] and Pro[233], both belonging in this plane, are marked.

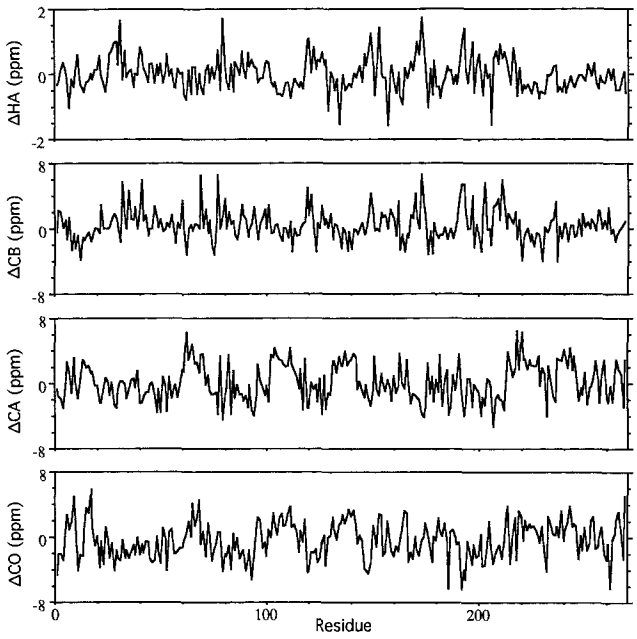

Fig. 6. Deviation Δ from random coil chemical shifts of H^{α}, C^{β}, C^{α} and CO resonances, as a function of residue number. Random coil chemical shifts are those of Wishart and Sykes (1994).

plication was the presence of three intense pairs of peaks in the proline δ-CH_2 region of the 2D HSQC spectrum, that proved in the end to be absent in the 3D NOESY-(^{13}C,^{1}H)-HSQC spectrum and thus must arise from breakdown products of low molecular weight. Ultimately, all the proline δ-CH_2 groups were found in the 2D HSQC spectrum, but three of them could only be assigned by observing the peaks they gave rise to in the 3D NOESY-(^{13}C,^{1}H)-HSQC and ^{13}C HMQC-NOESY-HMQC spectra. The side chains of prolines 51 and 195, whose C^{α}, C^{β} and H^{α} resonances overlapped almost completely, had to be assigned in the same way.

In total, the assignments covered all 1050 backbone carbonyl and CH_n carbon resonances in serine protease PB92, all 288 amide nitrogen resonances (NH and NH_2) and 1353 out of 1364 proton resonances from CH_n and amide groups. Six out of eight arginine ε-NH groups, five out of eight arginine C^{ζ} atoms, and all three tryptophan side-chain NH groups were assigned as well. The missing 11 protons were all from CH_2 groups: four proline H^{γ}, three lysine H^{γ}, two arginine H^{γ}, one asparagine H^{β} and one leucine H^{β}. (Almost) degenerate CH_2 protons were both assigned when the presence of the second, overlapping resonance could be inferred with confidence. This inference was made where the shape of isolated 2D HSQC peaks showed signs of a second resonance, but also (e.g. for lysine CH^{ε} protons) when the one observed CH_2 proton gave sharp lines in uncrowded regions of the (3D) spectra, so that a peak at a different frequency could not have remained unobserved. A table containing the chemical shifts is available as Supplementary Material and will be deposited with the BioMagResBank.

Chemical shifts and secondary structure

The assignment strategy used here does not employ sequential NOE connectivities, so that these are not available for an evaluation of the secondary structure. It has been shown, however (Wishart and Sykes, 1994), that deviations from 'random coil' chemical shifts of H^{α}, C^{α}, C^{β} and CO can be used as a precise indicator of secondary structure. Accordingly, we decided to compare the NMR observations with the secondary structure observed for serine protease PB92 in the crystal (Van der Laan et al., 1992).

The secondary structure was determined from the crystal structure according to the rules of Kabsch and Sander (1983), as incorporated in the program PROCHECK (Morris et al., 1992). The results were modified by requiring that all residues in a secondary structure have dihedral angles in the appropriate part of the Ramachandran plot and be enclosed between hydrogen bonds that form part of the secondary structure. The practical effect of this was to disallow residues involved in two β-bulges at positions 43–44 and 190–191, as well as a number of single residues at the termini of α-helices, and to remove residues 214–218 from the α-helix that extends up to residue 231. The secondary structure is shown in Fig. 7. Also shown are five residues (126, 161, 180, 249 and 261) which are, in effect, single-residue β-sheet strands, assuming an extended conformation and forming two β-sheet-type hydrogen bonds. These residues are not found as β-sheet by PROCHECK and are not counted when evaluating the success rate of the secondary structure determination. Discounting these five residues, we find that serine protease PB92 contains nine α-helices, encompassing 73 residues, and nine β-sheet strands, encompassing 44 residues.

We evaluated the observed secondary structure of serine protease PB92 using the chemical shift index (CSI) method of Wishart and Sykes, as incorporated in their program CSI (Wishart and Sykes, 1994). This method reduces the chemical shift deviations of each of H^{α}, C^{α}, C^{β} and CO to a three-valued classification, positive / intermediate / negative, and defines an overall index by comparing the results. The raw chemical shift deviations are shown in Fig. 6, while the results of the structure evaluation can be seen in Fig. 7. For α-helices the CSI correctly predicts 64 residues, gives false negatives for nine and false positives for four, while identifying eight of the nine helices in the molecule. The method misses the very short helix at residues 254–257, but it is arguable whether this part of the molecule is most accurately described as an α-helix; Van der Laan et al. (1992; personal communication) chose to describe it as containing a type 3_{10} turn at residues 254–255, followed by a type II turn at residues 258–259. The remaining errors in the prediction concern the precise beginning and end of the helices, a decision that remains somewhat arbitrary, even when the structure and its coordinates are known. The CSI method correctly

identifies 30 β-sheet residues, with 14 false negative and 16 false positive identifications. Again, many of the false positives and negatives reflect uncertainties in the exact beginning and end of the strands. The program correctly identifies the presence of seven sheet strands, with two false positive and two false negative identifications.

We also tried the intuitively more obvious method of summing the chemical shift deviations for the different nuclei and evaluating the secondary structure from the resulting curve (Fogh et al., 1994). This method gave results very similar to the CSI method as regards α-helices, but proved inferior in identifying β-sheet strands (results not shown). The application of a three-point 1:2:1 moving average filter gave similar results, as shown previously (Fogh et al., 1994).

The chemical shift deviations discussed above show that the secondary structure elements of serine protease PB92 appear to be very similar in the crystal and in solution, especially with regard to the α-helices. A table containing the results of the secondary structure analysis using PROCHECK and the CSI method is available as Supplementary Material.

Hydrogen exchange

Further information on the secondary structure can be gained from a comparison between the hydrogen–deuterium exchange behaviour of the molecule, and the hydrogen bonds observed in the crystal structure. The results are summarised in Fig. 7. The backbone NH groups of 175 residues exchange slowly enough to be observable in a freshly prepared D_2O sample, and 158 of these are hydrogen bonded in the crystal structure. Another 76 NH protons exchange too fast to be observed under these conditions, and 54 of these are not hydrogen bonded in the crystal structure. A further 18 NH groups could not be identified with certainty in the 2D HSQC experiments used to monitor H/D exchange. Of the 22 NH groups that are hydrogen bonded in the crystal structure but exchange fast in solution, more than half involve side-chain hydrogen-bond acceptors, and none are involved in the α-helices or in the parallel β-sheet structure of the molecule. Only three of the 22 NH groups have any relevance for the defined structural elements. The hydrogens involved in the NH-95–CO-99 and NH-98–CO-95 hydrogen bonds – which close off the type II turn at residues 96–97 in the crystal structure – are in fast exchange. This is also the case for the NH in the NH-161–CO-126 hydrogen bond, which is one of the two hydrogen bonds defining the one-residue antiparallel β-sheet between residues 126 and 161. None of these residues is close to the two calcium molecules found in the crystal structure, or to the catalytic site where the inhibitor is bound. The turn at residues 96–97 sticks out into the solvent in the crystal structure, and the mini-β-sheet between residues 126 and 161 is also on the surface, so that fast exchange of the hydrogens involved could be possible with only minor local fluctuations.

NOE analysis

To characterise further the β-sheets and general topology of the molecule, we assigned the NH-NH NOEs of the molecule. In order to avoid the lower resolution, lower signal-to-noise ratio and technical complications associated with 4D spectra, we used 3D NOESY-(^{15}N,^{1}H)-HSQC and ^{15}N HMQC-NOESY-HMQC experiments. A 4D cross peak Ha-Na-Nb-Hb is associated with four 3D cross peaks Ha-Nb-Hb, Hb-Na-Ha, Na-Nb-Hb and Nb-Na-Ha in the two 3D experiments, each pair of which share two frequency coordinates. If any two of these cross peaks can be uniquely paired (i.e., if only those two cross

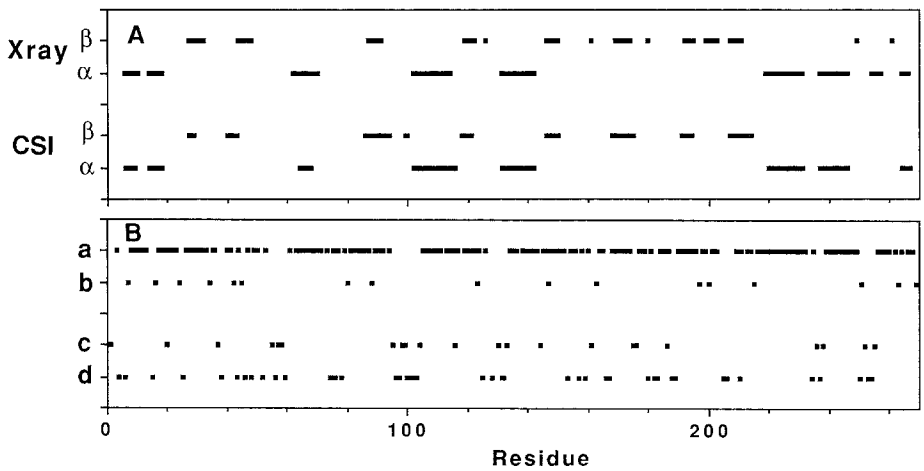

Fig. 7. Secondary structure, chemical shifts and hydrogen exchange. (A) β-sheet and α-helix in the crystal structure of serine protease PB92 (Van der Laan et al., 1992), and determined by the chemical shift index method of Wishart and Sykes (1994). (B) Backbone hydrogen exchange versus residue number: (a) slow exchange, hydrogen bonded in crystal; (b) slow exchange, not hydrogen bonded in crystal; (c) fast exchange, hydrogen bonded in crystal; (d) fast exchange, not hydrogen bonded in crystal.

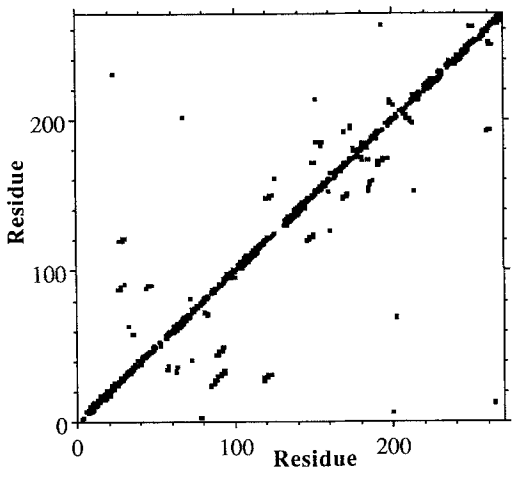

Fig. 8. Backbone NH-NH NOEs and molecular topology. Above the diagonal: backbone NH-NH NOEs between residues. Experimental details are given in the text. Below the diagonal: backbone NH-NH distances shorter than 5 Å in the crystal structure of serine protease PB92 (Van der Laan et al., 1992). Note the similarity between the patterns of crystal structure distances and of observed NOEs.

peaks fit with e.g. the Ha,Hb frequency pair), it is possible to reconstruct all four frequencies of the 4D cross peak. This then allows both the NH groups involved in the NOE to be assigned on the basis of two frequencies. Further assignments can be made by removing already assigned peaks from consideration as 'pairing partners', and by considering the intensities of the peaks being compared. The procedure was carried out by interactive graphics using in-house written software, and could be completed within a week. It should be emphasized that assignments were made (and rejected) solely on the basis of NMR data (although the structure was used as a guide to peaks that needed attention at the cross-checking stage). In total, we identified 267 distance constraints involving only backbone NH groups. Of these, 182 were sequential, 42 were medium range ($1<i-j<5$) and 43 long range. All the NOEs corresponded with distances in the crystal of less than 6.3 Å, a distance that is just within the possible range, considering the spin diffusion expected for a molecule of this size at the relatively long mixing time of 150 ms.

The observed NH-NH NOEs are shown in Fig. 8, contrasted with the NH-NH distances shorter than 5 Å as measured in the crystal structure. The presence, if not the precise extent, of the parallel β-sheet including (strand by strand) residues 44–48, 87–92, 27–32, 119–122 and 146–150 is clearly visible from the measured NOEs. The parallel sheet connecting residues 169–174 with 192–195, as well as the antiparallel sheet connecting residues 199–203 with 207–211 can likewise be identified, and the remaining β-sheet strands found in the X-ray structure (Fig. 7) are all connected by at least one observed NOE. The agreement between the two halves of Fig. 8 is not perfect, partly because not all observable NOEs could be assigned. Still, it is immediately clear that there can be only minor differences between the β-sheet structure and the general topology of serine protease PB92 in solution and in the crystal.

Homologous proteins

It is interesting to compare the backbone chemical shifts of serine protease PB92 with the backbone shifts published for the homologous protein subtilisin 309 from *Bacillus lentus* by Remerowski et al. (1994). This protein differs from serine protease PB92 only in a single mutation ($Asn^{85} \rightarrow Ser$). It should be noted that the numbering system used by Remerowski et al., which is that of subtilisin BPN', is different from the one used in the present paper. The spectra of Remerowski et al. were also acquired at pH 5, but there were differences in temperature (303 vs. 315 K), buffer composition (10 mM acetate, 20 mM $CaCl_2$, vs. 25 mM acetate), and inhibitor used (2,4-dichlorobenzene boronic acid vs. diisopropylfluorophosphate). The chemical shifts show a constant difference that must be caused by differences in calibration (see Table 1), but when this effect is disregarded a limited number of differences in chemical shifts stand out. Differences in residues 84–86 (strong) and 22–24 (weak) can be attributed to the effect of the mutation at residue 85, since residues 85 and 25 contact each other in the crystal. Other differences in residues 124–128, 152–154 and 213–218 probably arise from the effect of the different inhibitors, since all these residues are found near the active site of the protein crystal structure, where the inhibitors are thought to bind to the catalytic Ser^{215}. The C^{α} of the catalytic histidine, His^{62}, shows a shift difference of 0.5 ppm between the two structures that must arise from the same cause.

A difference in chemical shift of 0.4 ppm has been observed for the NH proton of Thr^{33}, together with a difference of 0.65 ppm (corrected for calibration differences) for the NH nitrogen. The chemical shift of this NH

TABLE 1
CHEMICAL SHIFT DIFFERENCES BETWEEN SERINE PROTEASE PB92 FROM *Bacillus alcalophilus* AND SUBTILISIN 309 FROM *Bacillus lentus*

	Nucleus				
	N	HN	CO	C^{α}	H^{α}
Number compared	232	232	238	246	238
Average Δfreq.[a]	0.60	0.043	0.36	0.20	0.018
SD Δfreq.	0.19	0.036	0.11	0.09	0.019

[a] Difference Δfreq. (in ppm) between chemical shifts measured for serine protease PB92 and for subtilisin 309 from *Bacillus lentus*, as published by Remerowski et al. (1994). Resonances from residues 22–24, 84–86, 124–128, 152–154, 213–218, as well as Thr^{33}-N, Thr^{33}-HN, Tyr^{186}-CO and Gly^{187}-CO are excluded from the calculation for reasons given in the text. The chemical shifts measured by Remerowski et al. were higher in all cases.

proton is not only high (11.28 ppm) but also quite variable, and has been observed to vary up to 0.15 ppm between different samples under what should have been similar conditions. The hydrogen bond between Thr^{33}-NH and the Asp^{32} side-chain carboxyl group might well cause both the unusual shift and its variability. Chemical shifts for a number of resonances were excluded from the comparison, since they were not reported by Remerowski et al. (the N and HN chemical shifts of four residues; the CO chemical shifts of 10 residues; the H^{α} chemical shifts of 32 residues; and all chemical shifts for residues 1, 129 and 130), while proline nitrogen shifts were not measured in the present study.

The agreement between the chemical shifts measured by us and by Remerowski et al. is quite remarkable (Table 1). If the results are corrected for systematic differences and the residues with major differences described above are disregarded, the measured shifts agree within about 0.2 ppm for nitrogen, 0.1 ppm for carbon, 0.04 ppm for NH and 0.02 ppm for H^{α} protons. This not only confirms that the structure of the two molecules is essentially identical, but demonstrates a reproducibility in frequency measurements that could not, a priori, be taken for granted.

Conclusions

We have presented the assignment of the complete ^{1}H, ^{15}N and ^{13}C NMR spectra of serine protease PB92. Although the molecule is one of the largest proteins to be assigned so far, the assignment did not require the exhaustive (and exhausting) use of every available bit of information. Notably, NOE information was used only cursorily in the assignment process. Molecules of this size (27 kDa) thus seem to be well within the limits of the newest experimental NMR techniques, at least for reasonably well behaved proteins under optimal measuring conditions.

In the absence of the NOE information usually acquired during assignment, it was possible to determine the secondary structure of the molecule quite accurately using the deviation from random coil of the measured chemical shifts. The inclusion of NOEs between backbone NH groups – a well-defined and easily analysed subset of the total set of NOEs – confirmed the location for the β-sheet parts of the structure, and allowed the topology of the sheet to be determined. From the evidence available, it appears that any differences between the crystal and solution structures are likely to be local.

Acknowledgements

We thank the group of A. Bax at the NIH, Bethesda, MD, for making available the pulse sequence of the 3D CBCA(CO)NH experiment.

References

Bax, A., Griffey, R.H. and Hawkins, B.L. (1983) *J. Magn. Reson.*, **55**, 301–315.

Bax, A., Clore, G.M., Driscoll, P.C., Gronenborn, A.M., Ikura, M. and Kay, L.E. (1990a) *J. Magn. Reson.*, **87**, 620–627.

Bax, A., Clore, G.M. and Gronenborn, A.M. (1990b) *J. Magn. Reson.*, **88**, 425–431.

Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. (1990c) *J. Magn. Reson.*, **86**, 304–318.

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

Bendall, M.R., Peggy, D.T. and Doddrell, D.M. (1983) *J. Magn. Reson.*, **52**, 81–117.

Betzel, Ch., Pal, G.P. and Saenger, W. (1988) *Acta Crystallogr.*, **B44**, 163–172.

Bode, W., Papamakos, E. and Musil, D. (1987) *Eur. J. Biochem.*, **166**, 673–692.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B. and Powers, D. (1988) *J. Biol. Chem.*, **263**, 7895–7906.

Brown, S.C., Weber, P.L. and Mueller, L. (1988) *J. Magn. Reson.*, **77**, 166–169.

Clore, G.M., Bax, A., Driscoll, P.D., Wingfield, P.T. and Gronenborn, A.M. (1990) *Biochemistry*, **29**, 8172–8184.

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.

Consonni, R., Molinari, H., Greco, F., Zannoni, G., Zetta, L., Carrea, G. and Riva, S. (1992) *Biochim. Biophys. Acta*, **1119**, 39–44.

Fesik, S.W. and Zuiderweg, E.R.P. (1988) *J. Magn. Reson.*, **78**, 588–593.

Fogh, R.H., Schipper, D., Boelens, R. and Kaptein, R. (1994) *J. Biomol. NMR*, **4**, 123–128.

Frenkiel, T., Bauer, C., Carr, M.D., Birdsall, B. and Feeney, J. (1990) *J. Magn. Reson.*, **90**, 420–425.

Gros, P., Betzel, Ch., Dauter, Z., Wilson, K.H. and Hol, W.G.J. (1989) *J. Mol. Biol.*, **210**, 347–367.

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

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

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

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

Ikura, M., Bax, A., Clore, G.M. and Gronenborn, A.M. (1990b) *J. Am. Chem. Soc.*, **112**, 9020–9022.

Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.

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. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.

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

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989b) *J. Am. Chem. Soc.*, **111**, 1515–1517.

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

Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992) *Proteins*, **12**, 345–364.

Müller, L. (1979) *J. Am. Chem. Soc.*, **101**, 4481–4484.

Norwood, T.J., Boyd, J., Heritage, J.E., Soff, N. and Campbell, I.D. (1990) *J. Magn. Reson.*, **87**, 488–501.

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)*, 2nd ed., Cambridge University Press, Cambridge, pp. 559–563.

Remerowski, M.L., Domke, T., Groenewegen, A., Pepermans, H.A.M., Hilbers, C.W. and Van de Ven, F.J.M. (1994) *J. Biomol. NMR*, **4**, 257–278.

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

Siezen, R.J., De Vos, W.M., Leunissen, J.A.M. and Dijkstra, B.W. (1991) *Protein Eng.*, **4**, 719–737.

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.J. (1985) *Biochemistry*, **24**, 6894–6903.

Van Eekelen, C.A., Van der Laan, J.C., 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.

Vis, H., Boelens, R., Mariani, M., Stroop, R., Vorgias, C.E., Wilson, K. and Kaptein, R. (1994) *Biochemistry*, **33**, 14858–14870.

Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson.*, **98**, 428–435.

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

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.

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.