

J-Bio NMR 185

^1H , ^{13}C and ^{15}N NMR backbone assignments and secondary structure of the 269-residue protease subtilisin 309 from *Bacillus lentus**

M.L. Remerowski^a, T. Domke^{b,**}, A. Groenewegen^b, H.A.M. Pepermans^{b,***},
C.W. Hilbers^a and F.J.M. van de Ven^{a,***}

^aNSR Center, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

^bUnilever Research Laboratory, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands

Received 4 November 1993

Accepted 24 November 1993

Keywords: Subtilisin; 3D NMR; Sequential assignment; Protein

SUMMARY

^1H , ^{13}C and ^{15}N NMR assignments of the backbone atoms of subtilisin 309, secreted by *Bacillus lentus*, have been made using heteronuclear 3D NMR techniques. With 269 amino acids, this protein is one of the largest proteins to be sequentially assigned by NMR methods to date. Because of the size of the protein, some useful 3D correlation experiments were too insensitive to be used in the procedure. The HNCO, HN(CO)-CA, HNCA and HCACO experiments are robust enough to provide most of the expected correlations for a protein of this size. It was necessary to use several experiments to unambiguously determine a majority of the α -protons. Combined use of HCACO, HN(COCA)HA, HN(CA)HA, ^{15}N TOCSY-HMQC and ^{15}N NOESY-HMQC experiments provided the H^α chemical shifts. Correlations for glycine protons were absent from most of the spectra. A combination of automated and interactive steps was used in the process, similar to that outlined by Ikura et al. [(1990) *J. Am. Chem. Soc.*, **112**, 9020–9022] in the seminal paper on heteronuclear backbone assignment. A major impediment to the linking process was the amount of overlap in the C^α and H^α frequencies. Ambiguities resulting from this redundancy were solved primarily by assignment of amino acid type, using C^α chemical shifts and 'TOCSY ladders'. Ninety-four percent of the backbone resonances are reported for this subtilisin. The secondary structure was analyzed using 3D ^{15}N NOESY-HMQC data and C^α secondary chemical shifts. Comparison with the X-ray structure [Betzel et al. (1992) *J. Mol. Biol.*, **223**, 427–445] shows no major differences.

INTRODUCTION

Subtilisin proteases constitute a family of serine endopeptidases of molecular mass 26–29 kDa,

*Supplementary material available from F.J.M. van de Ven: the source code (PASCAL) for the computer program described in this paper.

**Present address: Gesellschaft für Biotechnologische Forschung, Maschenroder Weg 1, D-38124 Braunschweig, Germany.

***To whom correspondence should be addressed.

secreted by various *Bacilli* (Markland and Smith, 1971). These enzymes have been widely studied, and their kinetics and mechanism of activity are well understood (Markland and Smith, 1971; Kraut, 1977). Collectively, over 50 X-ray structures of subtilisins have been determined (Wells and Estell, 1988). Because of their broad substrate specificity and stability in the neutral-to-alkaline pH range, they are industrially useful as protein-degrading additives in washing powders.

The amount of available information on subtilisins, combined with their industrial application, makes them an ideal model system for protein engineering. Extensive engineering studies have been carried out and are still being done to optimize subtilisin performance in cleaning solutions, and to advance the science of protein engineering.

Subtilisins are representatives of a subclass of serine proteases that have a mechanism of action identical to the trypsin-like family of serine proteases, but are evolutionarily and structurally distinct (Markland and Smith, 1971). The catalytic apparatus in subtilisin consists of the well-documented catalytic triad (Fink, 1987), involving the side chains of Ser²²¹, His⁶⁴ and Asp³², an oxyanion binding site (Asn¹⁵⁵ side chain and the main-chain amide of Ser²²¹) and other binding determinants further away from the scissile peptide bond of the substrate (Wells and Estell, 1988).

High-resolution X-ray structures of several subtilisins (Wright et al., 1969; Bode et al., 1987; Betzel et al., 1988; Bott et al., 1988; Teplyakov et al., 1990) have provided detailed structural information on this family of molecules. The subtilisin family differs from the trypsin family in that the subtilisins are single-domain molecules with no sulfur bridges. Their overall shape is a half-sphere with the active site located on the flat surface of the hemisphere. The core of the molecule is composed of a twisted parallel β -sheet with α -helices running antiparallel to the β -sheet and parallel to each other. Calcium ions enhance the thermal stability of subtilisins (Pantoliano et al., 1988), which in turn increases their resistance to autolysis (Briedigkeit and Frömmel, 1989). Most subtilisins have the capacity to bind two calcium ions, and their binding sites are similarly situated.

The subtilisin studied in this paper is secreted by *Bacillus lentus*, an alkalophilic bacterium which grows at a higher pH than is common for most *Bacilli*. This enzyme is highly active and stable at basic pH, and is marketed under the trade name SavinaseTM by Novo/Nordisk A.S., as a protein-degrading additive in washing products. SavinaseTM contains 269 amino acids (28 kDa), but the numbering used throughout this paper is that of BPN', a closely related subtilisin containing 275 amino acids, and the first subtilisin for which a 3D structure was reported (Wright et al., 1969). The numbering used for SavinaseTM is as follows: A1–I35 (36) S37–P55 (56) S57–G157 (158 159) A160–S162 (163 164) I165–R275. The numbers in parentheses have been excluded from the numbering scheme in order to obtain the optimal alignment between the homologous portions of BPN' and SavinaseTM.

NMR spectroscopy can provide additional information, not obtainable using X-ray crystallography, which is important for the further study of this family of molecules. Motional aspects of structure–function relations such as conformational changes, internal motions and kinetics are subjects suited for study by NMR spectroscopy. The first step toward a 3D NMR solution structure is the sequential assignment of the backbone atoms of the protein.

One of the impediments to determination of a solution structure of subtilisin is its autolytic capacity. Aryl boronic acids are potent competitive inhibitors of serine proteases (Antonov et al., 1971; Philipp and Bender, 1971). The boronic moiety inhibits enzyme activity by forming a covalent complex with residues in the catalytic triad. Different R-groups on the acids determine

the affinity of the boronic acid to the subtilisin, with electron-withdrawing side groups enhancing the affinity (Philipp and Maripuri, 1981). Halogen-substituted arylboronic acids show a very good affinity (K_i in the μM range at pH 7.0); 2,4-dichlorobenzeneboronic acid (dcp) was used as the inhibitor in this study.

Since subtilisin is a large protein for analysis by NMR, it was necessary to sequentially assign this molecule using triple-resonance 3D NMR techniques. With 269 amino acids, SavinaseTM is one of the largest monomeric proteins to be assigned by NMR to date.

MATERIALS AND METHODS

Doubly labelled (99% ¹⁵N, 99% ¹³C) SavinaseTM was a gift from Novo/Nordisk A.S., and was dissolved in a 10-mM deuterated acetate buffer (pH 5.0) with 20 mM CaCl₂ and 5 mM 2,4-dichlorobenzeneboronic acid (nondeuterated from Alfa Chemical). H₂O and D₂O samples were prepared similarly. Protein originally in H₂O/borate buffer at pH 6.5 was exchanged into the final sample conditions by passage over a pre-equilibrated G-25 Sephadex column, and concentration of the effluent using a Centricon system. H₂O samples were 95% H₂O/5% D₂O, 1 mM protein concentration. The D₂O sample was nearly 100% D₂O, 0.8 mM protein concentration. All experiments were carried out at 30 °C.

NMR spectroscopy

Most experiments were performed on a Bruker 600 MHz AMX spectrometer, equipped with a 5-mm inverse triple-resonance probe (¹H/¹⁵N/¹³C) and three-channel interface without further modifications. A single 3D ¹⁵N TOCSY-HMQC was recorded on a Bruker 400 MHz AM spectrometer.

3D CT-HNCO, CT-HNCA and CT-HN(CO)CA experiments (Kay et al., 1990; Bax and Ikura, 1991; Grzesiek and Bax, 1992) were recorded on the H₂O sample. The numbers of complex points, numbers of scans, acquisition times per FID, and total acquisition times for these and all

TABLE 1
ACQUISITION PARAMETERS FOR THE VARIOUS EXPERIMENTS

Experiment	Heteronucleus			No. of complex points			Acquisition time (ms)			Scans	Total ^a
	F1	F2	F3	F1	F2	F3	F1	F2	F3		
HNCO	¹⁵ N	¹³ C	¹ H	46	64	512	22.2	35.6	56.3	32	137
HNCA	¹⁵ N	¹³ C	¹ H	46	32	512	22.2	7.7	56.3	64	137
HN(CO)CA	¹⁵ N	¹³ C	¹ H	40	32	512	19.3	7.9	56.3	64	120
HN(COCA)HA	¹⁵ N	¹ H α	¹ H ^N	40	32	512	19.3	10.7	56.3	64	109
HN(CA)HA	¹⁵ N	¹ H α	¹ H ^N	45	32	512	21.8	10.7	56.3	64	122
CBCA(CO)NH	¹³ C	¹⁵ N	¹ H ^N	52	32	512	6.2	15.2	56.3	64	108
HCACO	¹³ C	¹³ CO	¹ H α	28	64	512	6.8	28.2	56.3	32	83
HCA(CO)N	¹³ C	¹³ CO	¹ H α	28	64	512	6.8	28.2	56.3	48	130
TOCSY	¹ H	¹⁵ N	¹ H ^N	110	41	512	15.6	19.5	72.7	16	104
NOESY	¹ H	¹⁵ N	¹ H ^N	64	62	512	7.0	30.0	56.3	16	91

^a Total acquisition time (h).

experiments described below are listed in Table 1. 3D HN(CA)HA (Clubb et al., 1992a; Kay et al., 1992) and HN(COCA)HA (Kay et al., 1992) experiments were performed as modified 3D versions (no C α frequency labelling) of the 4D experiments published by Olejniczak et al. (1992). The modifications were as follows: (1) labelling of the ^{15}N chemical shift during the τ_3 period; (2) replacement of the MLEV-16 pulse sequence with 180° pulses for proton decoupling; and (3) replacement of the GARP pulse sequence with a single 180° pulse during ^{15}N frequency labelling for carbonyl decoupling.

3D CT-HCACO and CT-HCA(CO)N experiments (Kay et al., 1990; Powers et al., 1991) were recorded on a D $_2$ O sample. A 3D clean ^{15}N TOCSY-HMQC (Griesinger et al., 1988; Marion et al., 1989a) was recorded using a 25-ms mixing time. 3D ^{15}N NOESY-HMQC experiments (Fesik and Zuiderweg, 1988; Marion et al., 1989c; Messerle et al., 1989) were recorded on the doubly labelled sample, using mixing times of 75 and 150 ms. The doubly labelled sample was used in this case because the ^{15}N -labelled sample had deteriorated. High-power 180° pulses centered on the C α region were included in the ^{15}N NOESY-HMQC pulse sequence for ^{13}C - ^1H decoupling. A 3D CBCA(CO)NH experiment (Grzesiek and Bax, 1993) was also recorded.

Water suppression in these experiments was typically done by presaturation of the water signal. In the D $_2$ O experiments and 3D ^{15}N NOESY-HMQC, the water suppression was followed by a SCUBA (Brown et al., 1988) pulse sequence to allow for re-equilibration of saturated α -protons.

Quadrature detection in the indirectly detected dimensions was accomplished using the States-TPPI acquisition method (Marion et al., 1989b). Spectra were processed and peak-picked on a SUN SPARC or Silicon Graphics Indigo workstation, using the NMRZ software package (New Methods Research, Inc., Syracuse, NY), or the MNMR/PRONTO software package (PRONTO Software Development and Distribution, Copenhagen). Linear prediction was used in the C α dimension of the CT-HCACO and CT-HCA(CO)N experiments. Zero-filling was used to achieve the final spectrum sizes of 512 (F3) \times 256 (F2) \times 256 (F1) points for the CT-HNCO, CT-HNCA and CT-HN(CO)CA spectra, and 256 \times 256 \times 256 points for the HNCAHA and HN(COCA)HA spectra, where the final F3 size contains only the extracted amide proton region, i.e., half of the original 15.1 ppm sweep width. CT-HCACO and CT-HCA(CO)N spectra were also 256 \times 256 \times 256 points final size; here the H α region was extracted between 2.5–6.3 ppm, i.e., one quarter of the original sweep width. TOCSY- and NOESY-HMQC spectra were 512 (F3) \times 128 (F2) \times 512 (F1) points in size, also with extracted amide regions in F3.

Strategy and software

Our assignment strategy generally followed that first described by Ikura et al. (1990), where several heteronuclear 3D NMR experiments are compared and cross peaks correlated to build 'clusters', each of which is composed of the correlated backbone atom chemical shifts for an amino acid and some chemical shifts of the preceding or following amino acid to give enough overlap of chemical shift information to link the clusters sequentially. It is necessary to make some assignment of amino acid type to uniquely position stretches of linked clusters within the primary structure. Differences in our strategy lie in the atoms that were chosen to compose the cluster, and in the set of experiments that was used to build the clusters. Our choices were based on the available set of experiments (some new experiments have been developed since the paper by Ikura et al. (1990) appeared, e.g. CBCA(CO)NH), and the limited utility of the less sensitive experiments, imposed by the size of our system. The complexity of the system required using all

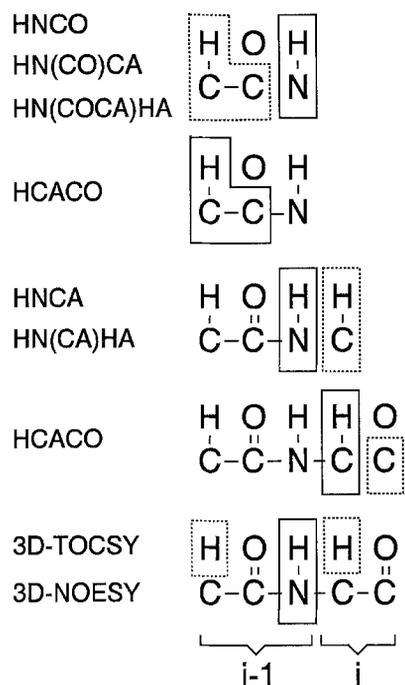
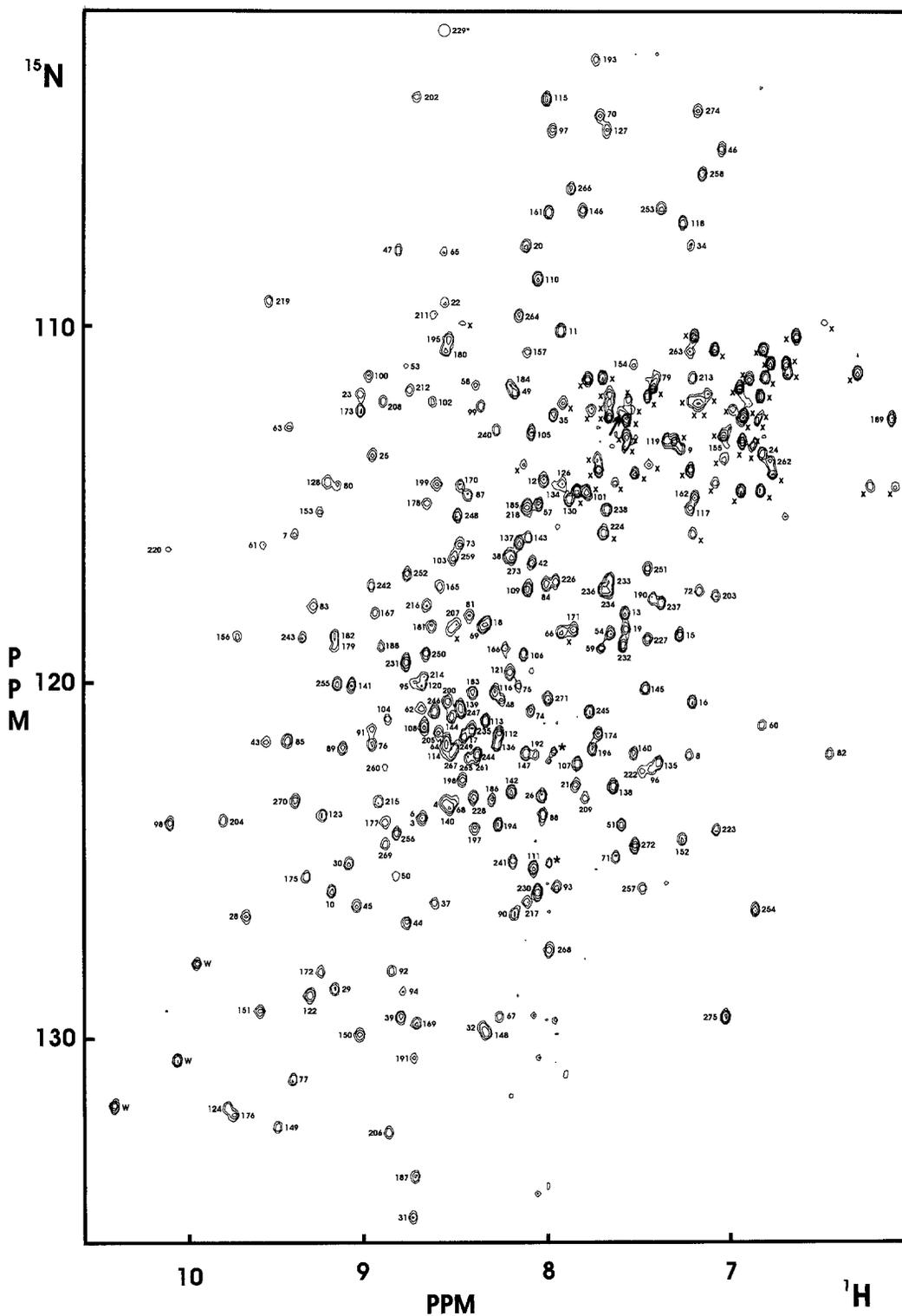


Fig. 1. Schematic representation of the way in which the chemical shifts of heteroatoms, correlated in 3D NMR experiments, were compared to build 'clusters' of correlated chemical shifts. Backbone atoms boxed by solid lines represent the heteroatoms whose frequencies are compared. Backbone atoms boxed by dashed lines are heteroatoms which are added to the cluster when common frequency axes are compared. For example, the first five backbone atoms in the cluster are correlated by comparing the ^1H and ^{15}N axes which are common to the HNCO, HN(CO)CA and HN(COCA)HA experiments. The third heteroatom correlation is different in each of these three experiments. After the cluster is begun, it can be extended or reinforced by comparing the correlated atoms from another experiment to any already in the cluster. For example, in the second step the three correlated heteroatoms in the HCACO experiment can be compared to the newly added H_{i-1}^α , C_{i-1}^α and CO_{i-1} backbone atoms added to the cluster in the first step. In the final step, the 3D ^{15}N TOCSY-HMQC and 3D ^{15}N NOESY-HMQC experiments are used to add α -protons missing in other experiments.

possible sources of information to complete the assignment. The group of backbone atoms composing a cluster was chosen, based on the sources of data available. The combining strategy is schematized in Fig. 1, and is further detailed in the Results and Discussion section.

Much of this assignment procedure is amenable to automation. For this work we wrote a program which performs the three basic steps of the assignment procedure: (1) building clusters; (2) linking clusters; and (3) placing stretches of linked clusters in the primary structure, using amino acid-type information provided by the user. The program uses peak lists as input and builds clusters containing peaks from each spectrum used in the analysis. The clusters can be linked by comparing any single or multiple combinations of chemical shift values among the peaks in the cluster. There are several advantages to this program. First, it is a complete book-keeping procedure for the sources of the chemical shifts of each atom in the cluster. It is easy to check the reliability of a chosen backbone atom chemical shift value within a cluster by checking its sources, i.e., all peaks in the cluster contributing a value for the given atom. The linking procedure can match clusters on several prioritized combinations. For example, the linking session can begin by matching on the chemical shifts of three different backbone atoms, but also on all combinations of two out of the three, if one of the atoms being compared has no value in a given cluster. It is also possible to prioritize the spectra compared in the linking procedure. If a chemical shift for a chosen atom is not present in the best spectrum for comparison, then the program can consider another spectrum, containing peaks with that atom's frequency for a matching possibility. After the linking phase, the linked stretches of clusters can be placed in the primary sequence. The user can enter any number of potential amino acid type assignments per cluster; the algorithm will then give the possible placements in the sequence for any stretch of partially spin-typed clusters. For interactive work with the spectra, e.g., manual peak picking,



checking assignments and comparing multiple possibilities, the PRONTO program was used extensively.

RESULTS AND DISCUSSION

Spectral quality

The 2D ^{15}N - ^1H HSQC spectrum (Fig. 2) shows the linewidths and degree of overlap for amide proton and nitrogen resonances in subtilisin. The dispersion in both the ^{15}N and ^1H dimensions is quite good, except in the 120–124 ppm (^{15}N) region and from 8–8.5 (^1H) ppm. There are 32 asparagine and glutamine residues, whose side-chain cross peaks appearing in the HSQC have been marked by x's. Two unusual pairs of side-chain cross peaks are found; the most upfield shifted pair, at a ^{15}N chemical shift of 110.88 ppm, has a 2-ppm separation in their proton resonances (6.46 and 8.45 ppm), the other pair is very much downfield shifted, to a ^{15}N frequency of 118.45 ppm. The side chains can be eliminated by using NH_2 -filtered versions of 3D spectra, so they did not present a problem in the assignment. An NH_2 -filtered HNCQ was used in conjunction with the unfiltered version to identify the asparagine/glutamine side-chain cross peaks.

A number of weak peaks appear in the ^{15}N - ^1H HSQC spectrum shown, which also appear in the heteronuclear 3D spectra. These have been attributed to proteolysis products, since most do not appear in the HSQC spectrum taken on a freshly prepared sample. Most of these weak peaks appear in the downfield part of the ^{15}N dimension between 125 and 135 ppm, around 8.0 ppm in F2. The unlabelled weak peaks appearing in the side-chain region do not appear in the 3D spectra. Two peaks, also attributed to proteolysis products, have been labelled with asterisks because they are more intense than the others of that kind. They appear weakly in the spectrum of the freshly prepared sample, but their intensities increased as the sample aged. Neither has NOE peaks associated with them.

Gly²²⁹, the most upfield shifted peak in F1, is represented by a circle at its actual chemical shift values. It appears in the spectra as a negative (once-folded) peak at the high-field end of F1, because it lies outside the spectral width setting. The three peaks labelled 'W' in the lower left corner of the spectrum are cross peaks from the aromatic 1N protons of the three tryptophan residues.

Three peaks are not shown, since they occur downfield of 10.5 ppm in F2, and size considerations were important for the readability of this figure. These are the correlations for Lys²⁷, Thr³³ and Asp⁴¹, whose chemical shift values are given in Table 2.

In addition to the N-terminal amino acid, five residues remain unidentified. Ser⁷⁸ (in the high-affinity calcium-binding site), Ser¹²⁵ (in the substrate recognition site), Ser²²¹ (in the catalytic

←

Fig. 2. 2D ^1H - ^{15}N HSQC spectrum of doubly labelled SavinaseTM, recorded at 600 MHz ^1H frequency. Cross peaks marked with an 'x' correspond to glutamine and asparagine side-chain NH_2 groups. Peak 229* is folded negatively to the high-field F1 region in our spectrum. Peaks marked 'W' correspond to the three N1 protons of the tryptophan residues. Three downfield (F2) shifted (> 10.5 ppm) cross peaks are not shown, but their chemical shift values are given in Table 2. The peaks labelled with an asterisk are intense, but likely due to proteolysis products. The peak labelled with an arrow is a candidate for assignment to residue Ser¹³². The peaks are numbered by residue assignment. Although there are only 269 residues, the numbering goes up to 275 because we used the BPN' numbering scheme.

triad), and the adjacent pair Ser¹³²/Ala¹³³ (on the rim of the substrate recognition site) do not have HSQC peaks assigned to them. This absence of ¹⁵N-¹H cross peaks may be due to exchange or relaxation effects, caused by their interactions with ligands. Alternatively, they could be affected by saturation transfer effects, since all are in solvent-accessible areas. HSQC spectra recorded with little or no water presaturation exhibit some new peaks, but all are weak and could also belong to proteolysis products.

It is possible that Ser¹³² does appear in the spectra; there is an unassigned HSQC peak, labelled with an arrow in Fig. 2, which is a possible candidate. It is a broad peak, with a $d_{\alpha\text{N}}$ -type NOE associated with it. It also appears to be the only remaining real unassigned peak. It is considered as a candidate for assignment to Ser¹³², because it has correlated C_{i-1}^{α} , C_{i-1}^{β} and C_i^{α} chemical shift values, consistent with a proline-serine pair (Grzesiek and Bax, 1993). Since Ser¹³² is preceded by a proline residue, it cannot from our present data be sequentially attached from that side. If the following residue, Ala¹³³, has peaks which are too broad to be identified, then it is not possible to confidently insert Ser¹³² into the sequential order. If this peak does belong to Ser¹³², then the fact that it is broad, coupled with the fact that Thr¹³⁴ is also broad, is evidence in favour of the supposition that the Ala¹³³ resonance has disappeared due to exchange broadening. It has also been suggested that this is a flexible region of the molecule (Bech et al., 1993).

A very weak peak in the ¹⁵N-¹H HSQC spectrum, which is not visible in Fig. 2 because it was impractical to contour the spectrum to that level, has been tentatively assigned to Ser¹²⁵, based on three pieces of evidence. This cross peak has correlations in the HN(CO)CA and HNCO spectra which give values for C_{i-1}^{α} and CO_{i-1} matching those for the preceding leucine. Also a C_{i-1}^{β} correlation at 42 ppm appears in the CBCA(CO)NH spectrum, compatible with an assignment of a preceding leucine residue. All of these associated cross peaks are weak, and no cross peaks appear in other spectra to correlate this HSQC cross peak to the following Leu¹²⁶ residue. This peak appears much more intensely in an HSQC spectrum recorded without water presaturation; its ¹⁵N and H^N chemical shifts are given in Table 2.

The absence of the Ser²²¹ amide resonance is interesting, because it is still unclear exactly what type of covalent complex is formed between boronic acid inhibitors and subtilisin-type serine proteases. ¹⁵N NMR studies (Bachovchin et al., 1988; Tsilikounas et al., 1992) have determined two different types of adducts to exist in trypsin-like serine proteases. In one type of complex the boron is directly bonded to the N^{ε2} of the catalytic histidine, and may or may not form a second covalent link to the β-oxygen of the catalytic serine. The second type is a transition-state-like complex, in which the boron is covalently bound to the active-site serine. In this complex boronic hydroxyl moieties hydrogen-bond to the histidine N^{ε2}, as well as to the serine amide proton and the side chain of Asp¹⁵⁵.

Early ¹H NMR studies on subtilisin BPN' complexed with aryl boronic acids (Robillard and Shulman, 1974) show one peak in the low-field ¹H spectrum, indicative of the histidine adduct (Tsilikounas et al., 1992). The question of the type of complex formed by the subtilisins is still unresolved, because slight differences between these spectra and those recorded for histidine-adduct complexes formed by trypsin-like serine proteases, in addition to results from X-ray diffraction studies which support the serine-adduct-type complex (Robertus et al., 1972; Matthews et al., 1975), prevent a definitive statement to be made based on the low-field ¹H NMR evidence. Our 1D ¹H spectrum of SavinaseTM, complexed with 2,4-dichlorobenzeneboronic acid, shows no low-field cross peaks.

TABLE 2
BACKBONE ASSIGNMENT^a OF SUBTILISIN 309

Residue	¹⁵ N	HN	¹³ CO	¹³ C ^α	H ^α	Residue	¹⁵ N	HN	¹³ CO	¹³ C ^α	H ^α
A1	nd ^b	nd	nd	nd	nd	S49	111.74	8.15	174.45	55.32	4.79
Q2	nd	nd	174.92	55.12	4.11	F50	125.48	8.80	174.73	57.58	4.66
S3	123.74	8.65	172.30	56.94	4.61	V51	124.01	7.57	174.47 ^c	59.94	4.28
V4	123.34	8.53	174.94 ^c	60.37	4.37	P52	145.15 ^c		177.84	64.02	nd
P5	143.05 ^c		178.14	63.29	4.59	G53	110.94	8.74	174.12	45.25	3.79,4.20
W6	123.74	8.65	179.12	60.80	nd	E54	118.54	7.63	172.64 ^c	53.71	4.92
G7	115.74	9.35	175.20	46.83	nd	P55 ^d	135.35 ^c		177.30	64.24	4.41
I8	122.01	7.20	179.26	61.99	3.82	S57	114.94	8.01	175.24	56.94	4.69
S9	113.34	7.24	179.22	61.88	4.14	T58	111.48	8.37	173.84	62.85	3.95
R10	125.88	9.15	179.08	56.83	4.43	Q59	118.94	7.69	174.82	56.40	3.97
V11	110.01	7.90	173.61	61.66	4.55	D60	121.08	6.80	176.18	52.74	4.98
Q12	114.14	8.00	174.68	56.94	4.08	G61	116.14	9.53	173.05	45.41	nd
A13	118.01	7.55	175.41 ^c	55.97	3.90	N62	120.68	8.67	174.50	56.09	4.82
P14	133.04 ^c		177.58	66.02	3.92	G63	112.68	9.38	173.75	46.94	nd
A15	118.54	7.26	181.13	54.89	4.12	H64	121.61	8.52	175.85	62.09	3.90
A16	120.54	7.18	180.43	55.11	4.14	G65	107.74	8.53	176.23	48.55	nd
H17	121.48	8.43	181.46	57.47	4.67	T66	118.54	7.90	175.76	67.79	3.74
N18	118.28	8.31	176.65	55.40	4.63	H67	129.48	8.24	179.54	61.13	4.85
R19	118.48	7.54	176.04	56.50	4.55	V68	123.34	8.47	179.03	65.86	nd
G20	107.61	8.09	174.26	45.33	4.30,3.80	A69	118.41	8.34	179.87	55.46	3.75
L21	122.81	7.82	173.94	53.71	4.73	G70	103.88	7.69	178.70	46.83	nd
T22	109.21	8.53	175.67	60.91	4.33	T71	124.94	7.60	175.01	67.25	nd
G23	111.74	8.99	171.37	44.90	nd	I72	117.34	7.14	178.28	66.72	3.36
S24	113.48	6.80	175.71	59.30	3.88	A73	116.01	8.46	175.48	52.85	4.45
G25	113.48	8.93	173.84	45.54	3.79,4.13	A74	120.81	8.06	179.31	54.03	4.65
V26	123.21	8.00	175.01	61.99	4.57	L75	120.14	8.13	178.24	55.22	3.95
K27	131.88	11.38	176.27	57.21	5.10	N76	121.61	8.93	172.91	52.64	4.98
V28	126.54	9.61	173.70	60.80	4.90	N77	131.21	9.36	174.11 ^c	52.74	4.61
A29	128.68	9.13	175.95	49.95	5.36	S78	nd	nd	173.75	57.74	4.77
V30	125.08	9.05	175.10	62.74	4.26	I79	111.34	7.38	178.10	59.30	4.73
L31	135.08	8.71	174.68	56.40	5.81	G80	114.41	9.11	174.92	48.87	nd
D32	129.75	8.34	177.30	54.68	4.59	V81	118.01	8.40	173.56	58.98	5.71
T33	120.14	11.70	173.47	63.62	5.09	L82	122.01	6.44	175.62	54.68	4.39
G34	107.61	7.20	172.67	43.72	4.45,3.20	G83	117.74	9.24	171.97	45.01	nd
I35 ^d	112.41	7.94	174.96	62.09	4.00	V84	117.21	7.98	176.69	66.93	3.49
S37	126.28	8.59	172.67	59.30	4.67	A85	121.61	9.39	nd	49.30	4.71
T38	116.28	8.19	174.92	64.24	3.94	P86	nd		178.52	64.86	4.01
H39	129.48	8.78	nd	55.65	4.69	S87	114.68	8.41	173.09	57.47	4.79
P40	nd		176.69	64.76	4.43	A88	123.74	8.00	177.54	52.74	4.51
D41	124.41	11.69	175.67	51.99	5.61	E89	121.74	9.10	174.17	55.65	4.30
L42	116.54	8.06	176.83	53.93	4.81	L90	126.54	8.16	175.90	54.79	4.86
N43	121.61	9.51	173.80	52.31	4.98	Y91	121.43	8.94	174.68	56.90	4.80
I44	126.68	8.74	176.97	61.99	3.96	A92	128.15	8.83	173.56	51.45	4.22
R45	126.28	9.02	176.37	55.54	4.16	V93	125.74	7.92	175.01	60.37	4.63
G46	104.81	7.02	171.74	45.22	4.42,4.04	K94	128.68	8.77	175.48	57.04	nd
G47	107.74	8.78	171.23	45.44	5.17,3.27	V95	119.88	8.70	172.63	60.16	4.39
A48	120.41	8.22	174.82	51.56	4.08	L96	122.41	7.41	175.39	52.10	nd

TABLE 2 (continued)

Residue	¹⁵ N	HN	¹³ CO	¹³ C α	H α	Residue	¹⁵ N	HN	¹³ CO	¹³ C α	H α
G97	104.41	7.94	175.39	43.93	4.36,3.81	R145	120.14	7.44	176.37	55.97	4.45
A98	123.88	10.03	180.06	55.32	4.08	G146	106.54	7.78	174.50	45.76	4.45,3.95
S99	112.14	8.34	176.51	59.19	4.36	V147	122.01	8.09	177.21	62.52	3.71
G100	111.21	8.95	172.86	44.90	4.33,3.23	L148	129.88	8.31	173.84	54.46	4.81
S101	114.54	7.75	173.61	57.26	4.79	V149	132.55	9.45	173.70	63.28	4.26
G102	112.01	8.59	171.88	45.22	nd	V150	129.88	9.01	173.24	61.66	4.65
S103	116.41	8.49	175.01	56.29	4.85	A151	129.21	9.54	174.59	50.27	5.67
V104	121.08	8.84	177.58	67.15	3.62	A152	124.41	7.24	178.84	52.10	4.75
S105	112.81	8.06	177.30	61.66	4.12	S153	115.21	9.21	175.10	61.81	3.89
S106	119.21	8.10	176.04	63.17	4.26	G154	110.94	7.51	173.34	44.25	nd
I107	122.28	7.81	177.82	66.29	3.32	N155	113.34	6.99	178.66	52.71	nd
A108	121.21	8.65	179.87	55.75	3.77	S156	118.68	9.66	176.88	59.73	4.51
Q109	117.34	8.07	179.64	59.30	3.71	G157 ^d	110.68	8.09	172.07	45.65	nd
G110	108.54	8.03	175.62	47.80	3.25,3.50	A160	122.01	7.51	175.81	52.10	4.24
L111	125.21	8.06	178.98	58.22	3.88	G161	106.68	7.97	176.23	45.87	nd
E112	121.48	8.24	179.31	59.30	4.05	S162 ^d	114.68	7.17	173.80	58.22	4.61
W113	121.08	8.31	180.06	62.52	3.98	I165	117.21	8.56	175.90	60.80	4.26
A114	121.74	8.52	178.80	55.32	3.86	S166	118.94	8.21	171.79	60.05	4.24
G115	103.48	7.98	175.85	46.94	3.85,3.65	Y167	118.01	8.92	nd	57.69	nd
N116	120.28	8.27	176.41	54.79	4.49	P168	129.15 ^e		174.92	62.87	4.54
N117	115.08	7.20	173.89	53.39	4.36	A169	129.61	8.70	177.16	56.83	3.69
G118	106.94	7.23	175.06	47.26	3.83,3.55	R170	114.28	8.44	178.10	56.90	3.48
M119	113.08	7.32	175.90	53.71	4.08	Y171	118.41	7.84	179.45	58.44	4.85
H120	120.01	8.65	177.77	59.30	4.83	A172	128.15	9.21	180.43	55.97	4.29
V121	119.61	8.18	173.33	60.37	5.00	N173	112.28	8.99	172.67	54.14	4.79
A122	128.81	9.27	173.42	49.95	5.47	A174	121.34	7.70	175.95	50.49	4.79
N123	123.74	9.20	173.88	54.25	4.85	M175	125.48	9.29	173.80	55.86	4.67
L124	132.01	9.72	175.38	55.42	nd	A176	132.15	9.67	177.77	51.45	5.10
S125	130.01 ^e	10.46 ^e	173.05	58.89	4.79	V177	123.88	8.86	176.09	61.77	4.97
L126	114.28	7.90	174.31	53.82	3.61	G178	114.81	8.62	172.11	43.29	nd
G127	104.28	7.64	172.11	46.94	nd	A179	118.94	9.13	177.21	49.30	6.08
S128	114.28	9.17	173.58 ^c	55.11	5.16	T180	110.54	8.52	175.81	59.30	5.09
P129	136.69 ^c		177.07	64.13	4.81	D181	118.41	8.61	179.45	51.45	4.95
S130	114.81	7.85	171.04 ^c	55.65	4.94	Q182	118.68	9.14	177.39	59.08	3.37
P131	135.60 ^c		nd	nd	nd	N183	120.14	8.38	174.36	52.74	4.92
S132	nd	nd	nd	nd	nd	N184	111.48	8.18	172.72	54.90	4.13
A133	nd	nd	180.06	55.64	nd	N185	114.94	8.09	175.85	51.78	5.18
T134	114.28	7.94	176.79	66.31	3.87	R186	123.21	8.28	176.74	57.15	4.00
L135	122.28	7.36	177.44	57.47	4.04	A187	134.01	8.70	179.78	51.99	3.96
E136	121.74	8.25	178.14	60.59	2.79	S188	118.94	8.89	175.29	62.09	3.75
Q137	116.01	8.13	178.61	58.87	3.96	F189	112.41	6.10	176.46	55.54	4.32
A138	122.94	7.61	180.39	55.00	4.34	S190	117.61	7.39	175.99	59.08	4.30
V139	120.68	8.46	179.31	67.43	3.47	Q191	130.55	8.70	176.41	58.22	4.57
N140	123.48	8.50	178.24	56.29	4.41	Y192	122.01	8.03	352.84	55.22	nd
S141	120.01	9.04	177.68	61.23	4.26	G193	102.41	7.72	179.68	42.94	3.68,4.45
A142	123.08	8.18	180.06	55.86	3.98	A194	124.01	8.25	178.47	54.03	4.04
T143	115.88	8.07	178.94	67.04	4.39	G195	110.28	8.50	174.45	44.68	3.56
S144	120.94	8.49	175.71	61.88	4.41	L196	121.88	7.73	176.69	56.51	3.86

TABLE 2 (continued)

Residue	¹⁵ N	HN	¹³ CO	¹³ C ^α	H ^α	Residue	¹⁵ N	HN	¹³ CO	¹³ C ^α	H ^α
D197	124.01	8.37	177.07	58.12	5.06	K237	117.74	7.35	175.24	58.55	3.77
I198	122.68	8.44	171.32	61.66	4.94	N238	115.08	7.64	nd	49.95	4.73
V199	114.28	8.58	173.84	59.51	5.34	P239	nd		179.12	65.39	4.39
A200	120.54	8.52	172.64 ^c	49.95	4.49	S240	112.81	8.25	175.95	59.19	4.53
P201	133.95 ^c		nd	63.92	nd	W241	125.08	8.16	176.83	56.72	4.94
G202	103.34	8.68	172.63	45.90	4.20	S242	117.21	8.93	174.50	56.31	4.77
V203	117.48	7.05	174.31	60.59	4.92	N243	118.68	9.30	176.93	56.98	4.24
N204	123.88	9.73	174.82	54.03	4.22	V244	122.01	8.35	177.82	66.29	3.57
V205	121.34	8.56	175.29	62.85	4.22	Q245	120.81	7.75	180.01	59.41	3.85
Q206	132.68	8.84	174.78	55.54	5.00	I246	120.81	8.59	177.02	67.04	3.81
S207	118.28	8.47	174.68	54.79	4.34	R247	120.81	8.44	177.96	60.16	3.83
T208	112.01	8.87	172.35	61.88	3.73	N248	115.21	8.46	177.44	55.97	4.47
Y209	123.21	7.78	172.99 ^c	56.40	4.94	H249	121.74	8.47	179.59	60.70	4.61
P210	134.81 ^c		175.40	63.82	nd	L250	119.08	8.64	178.94	57.80	3.83
G211	109.61	8.59	176.13	45.76	nd	K251	116.68	7.42	178.75	60.48	3.96
S212	111.61	8.73	173.66	58.44	2.96	N252	116.81	8.74	177.16	55.32	4.55
T213	111.34	7.18	172.58	58.22	4.85	T253	106.54	7.35	173.33	61.34	4.32
Y214	119.74	8.64	174.40	57.15	5.34	A254	126.41	6.84	177.86	52.53	4.36
A215	123.34	8.89	176.37	51.67	4.65	T255	120.01	9.11	174.26	62.09	4.26
S216	117.74	8.64	173.56	56.83	5.45	S256	124.28	8.80	175.62	59.94	4.30
L217	126.14	8.09	175.10	54.03	4.69	L257	125.74	7.47	176.09	53.60	4.50
N218	114.94	8.07	175.43	50.81	5.49	G258	105.61	7.12	173.46	45.01	3.75,4.27
G219	109.08	9.50	175.48	47.05	nd	S259	116.41	8.47	176.27	58.41	4.46
T220	116.14	10.04	nd	67.24	nd	T260	122.41	8.86	178.33	64.13	nd
S221	nd	nd	173.38	63.29	4.22	N261	122.14	8.37	174.40	56.80	4.48
M222	122.41	7.45	176.93	57.58	5.41	L262	113.61	6.75	176.60	56.72	3.73
A223	124.14	7.05	180.12	54.68	4.67	Y263	110.54	7.20	176.37	56.72	4.55
T224	115.74	7.67	nd	nd	nd	G264	109.61	8.13	175.62	46.62	3.87,3.45
P225	nd		179.92	64.86	4.67	S265	122.01	8.40	172.58	61.66	4.24
H226	117.08	7.92	178.28	62.42	4.08	G266	106.01	7.85	171.60	43.39	3.04,4.91
V227	118.81	7.42	177.35	66.18	3.63	L267	122.01	8.50	176.88	53.80	4.36
A228	123.21	8.38	179.68	56.08	3.90	V268	127.48	7.97	171.74	63.06	3.79
G229	101.34	8.59	175.24	47.26	3.71,3.53	N269	124.54	8.86	175.48	51.78	4.93
A230	125.88	8.03	179.17	55.86	3.85	A270	123.21	9.35	178.10	54.46	3.88
A231	119.34	8.74	178.84	55.22	3.79	E271	120.41	7.97	178.80	60.16	3.79
A232	118.81	7.55	179.03	55.10	4.14	A272	124.54	7.49	180.62	54.68	3.86
L233	117.08	7.63	178.52	58.26	3.88	A273	116.41	8.16	178.24	54.03	4.22
V234	117.34	7.64	177.77	66.72	3.22	T274	103.74	7.15	173.19	60.70	4.43
K235	121.34	8.38	176.51	56.94	3.85	R275	129.48	7.01	nd	59.62	3.81
Q236	117.34	7.67	178.24	58.58	3.83						

^a ¹H chemical shifts are expressed relative to TSP, ¹⁵N and ¹³C chemical shifts are expressed relative to hypothetical internal TSP, by multiplying the ¹H TSP frequency by 0.25144954 for ¹³C and by 0.10132914 for ¹⁵N. All values are in ppm.

^b Not determined.

^c Chemical shift determined in D₂O solution.

^d Some numbers (36, 56, 158, 159, 163, 164) have been excluded to follow the numbering scheme of subtilisin BPN¹ (see text).

^e Tentative assignment.

The fact that we find no cross peak for the amide proton of Ser²²¹ argues against an adduct which is hydrogen-bonded to Ser²²¹, although the hydrogen-bond formation may be a fast-exchange process, occurring within the framework of the slow-exchange process of the boron-serine adduct formation (Nakatani et al., 1975). The amide resonance of the adjacent residue, Thr²²⁰, is also broad.

C^α linewidths are 35 Hz or more, and ¹H linewidths are also in that range. The breadth of these lines, in addition to the degree of overlap in these frequency dimensions, was an impediment to the assignment procedure. Figure 3 is a projection along the carbonyl axis of the 3D HCACO spectrum onto the C^α-H^α plane and illustrates the degree of overlap in this plane.

Backbone assignment

The size of this protein challenges the limits of the 3D heteronuclear backbone assignment strategy. Although many experiments have now been published to correlate backbone atoms, only the more sensitive 3D experiments give interpretable spectra for proteins of this size. Several correlation experiments were needed to provide H^α chemical shifts unambiguously. The least sensitive experiments, i.e., HNCA and HN(CA)HA, provide chemical shifts crucial for linking the clusters. Peaks missing from these experiments, in addition to the severity of overlap in the H^α and C^α frequencies, created the greatest assignment obstacles. Although overlap was greatest for the C^α and H^α resonances, chemical shift overlap in all frequency axes exacerbated the assignment difficulties because of the sheer number of resonances in this protein. The overlap problem was compounded by limited resolution in the 3D spectra, and the fact that peak intensities in the spectra varied by factors of 10 (HNCA, HN(CO)CA) or 20 (HNCO, HCACO), causing problems for optimal processing and peak picking.

Figure 1 shows how the different experiments were combined to make a cluster of correlated chemical shifts. At each step the cluster contains the chemical shifts of the backbone atoms shown. The overlapping chemical heteroatom frequencies are boxed by solid lines. Chemical shifts of heteroatoms that are added by the experiment(s) in a given step are boxed by dashed lines.

The HNCO and HN(CO)CA experiments are the typical starting points for building clusters, because of their sensitivity. Most of the expected peaks appeared in these spectra. The resulting clusters contain chemical shift values for N_{*i*}, H_{*i*}^N, CO_{*i*-1} and C_{*i*-1}^α.

Next, the H_{*i*-1}^α values were added to complete the 'i - 1' subcluster. A few problems arose in obtaining H^α chemical shifts, so several experiments were used to identify these α-protons. The HCACO experiment is sensitive, but suffers from extensive overlap in the C^α and H^α regions for a protein of this size. Additionally, in our case the D₂O sample still contained sufficient water to cause the protons in the 4.7–4.8 ppm region to be obliterated (Fig. 3). Finally, correlations for most of the glycine protons did not appear in this spectrum. Even an HCACO experiment with coherence transfer delays, optimized for magnetization transfer in a glycine spin system, failed to produce these cross peaks.

Although the HN(COCA)HA experiment is one of the less sensitive experiments, it still provided approximately 200 peaks. It supplied H_{*i*-1}^α chemical shifts in the water resonance region that were missing from the HCACO spectrum, but it was also deficient in cross peaks from glycine α-protons. Finally, the 3D NOESY-HMQC and HN(CA)HA experiments were used to supply missing interresidual α-protons. Fifty-four peaks were picked manually from these two spectra. Because both HCACO and HN(COCA)HA were used, ambiguities for the C_{*i*-1}^α-H_{*i*-1}^α combinations

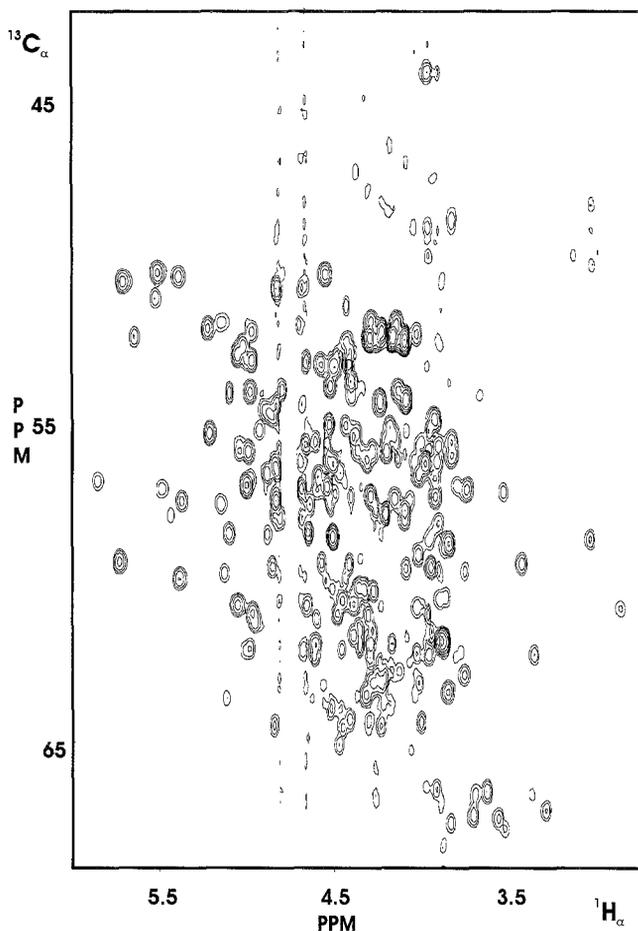


Fig. 3. The projection of the 3D HCACO experiment along the carbonyl axis onto the C^α - H^α plane, illustrating the degree of overlap present in the C^α and H^α frequencies. Peaks from most of the 35 glycines are absent or only weakly present, as seen from the dearth of peaks in the upfield region of the C^α axis. The amount of water still present in the sample obscures peaks in the 4.7–4.8 ppm H^α region of the spectrum.

were minimized. Matching initially on CO_{i-1} - H_{i-1}^α , using HCACO, still gave multiple matches in approximately 30 cases. When both HCACO and HN(COCA)HA were used for matching to the N_i , H_i^N and CO_{i-1} shifts in the cluster, about 10 clusters remained with multiple choices of HCACO peaks for H_{i-1}^α , primarily due to instances where there was no confirming peak from the HN(COCA)HA spectrum. In most cases remaining ambiguities were resolved using the HCA(CO)N experiment.

In order to sequentially link two clusters, the chemical shifts of the 'i - 1' portion of a cluster are matched to the C_i^α , H_i^α and CO_i of another cluster. Obtaining the 'i' chemical shifts to complete the clusters before sequential matching is more problematic, for several reasons. First, many C_i^α cross-peak correlations did not appear in the HNCA experiment. The H^α chemical shifts are obtainable from the HN(CA)HA and 3D TOCSY-HMQC experiments. The HN(CA)HA is another of the less sensitive experiments, and many correlations are also absent in this spectrum,

particularly those for glycine α -protons. About 200 computer-picked peaks from the HN(CA)HA spectrum were used for H_i^α chemical shifts, and another 55 weak peaks were manually chosen and used as likely candidates. Thirty-seven peaks were also manually picked and used from the 3D TOCSY-HSQC.

When linking the clusters, it is only practical to match on at least two chemical shifts; in practice, matching on three chemical shifts is best, especially when two out of the three are C^α and H^α . The HN(CA)CO experiment (Clubb et al., 1992b), which would give the CO_i chemical shift from comparison with the relatively well dispersed $H_i^N-N_i$ axes, is too insensitive to be applicable to this system. Therefore, in our case the only way to obtain three chemical shift values was to have both the C^α and H^α values, and match the HCACO to obtain the CO_i value, but it is worth noting that adding the CO_i chemical shift value in this way does not really alleviate the redundancy problem in the C^α and H^α axes, since the CO_i value is added by comparison on the C^α and H^α axes. As a result, approximately 45 HCACO cross peaks were multiply assigned to the 'i' end of different clusters.

In contrast, the 'i - 1' atoms in the clusters were more accurately ascertained. The chemical shifts for those three atoms were obtained by matching on frequencies with better chemical shift dispersion, and each chemical shift could theoretically be obtained from two different spectra, further eliminating ambiguities. It was crucial for the linking that at least one 'end' of the clusters was well determined. Such ambiguity at both ends of the clusters would have created too many branch points for any automated linking.

In order to resolve branch points in the sequential linking procedure, it was necessary to spin-type the clusters. Assignment of amino acid spin type has been aided by the development of the CBCA(CO)NH experiment (Grzesiek and Bax, 1993), which can be used to assign the preceding residue of the $H^N-^{15}N$ chemical shifts of each cross peak. The CBCA(CO)NH experiment was published during the later stages of our analysis, so we relied on C^α chemical shifts from the HN(CO)CA experiment, which pinpoint spin type less accurately but are useful nonetheless, and 'TOCSY ladders' to assign spin-type possibilities. The timing of the introduction of the CBCA(CO)NH experiment was fortuitous, and we used it to confirm our earlier assignments, as well as to resolve some sticky 'leftovers' where there were only two or three linked clusters. Approximately 70% of the possible C^β cross peaks appeared in the spectrum. The HCA(CO)N experiment was used to identify candidates for residues preceding prolines.

The presence of 35 glycines was both a help and a hindrance in the assignment. On the one hand, glycine residues can be easily spin typed. On the other hand, because of their CH_2 topology which increases broadening of C^α resonances and decreases coherence transfer efficiency, cross peaks correlating glycine H^α chemical shifts were often absent from spectra, thus causing breaks in the linking of clusters. Most of the H^α chemical shifts that remain undetermined belong to glycine residues.

We formed 274 clusters, for a possible 256 assignments. The additional clusters were formed from cross peaks stemming from the proteolysis products. The initial automated linking session yielded several sets of unambiguous sequences, ranging from five to ten clusters. Most of the breaks occurred at glycine and proline residues. Another typical cause for breaks was incorrect choice for some chemical shift in part of a cluster, due to cross-peak overlap. Approximately 40% of the protein was quickly assigned from the larger linked pieces, using C^α chemical shifts and 'TOCSY ladders' to assign spin type. After the initial linking session the assignment process was

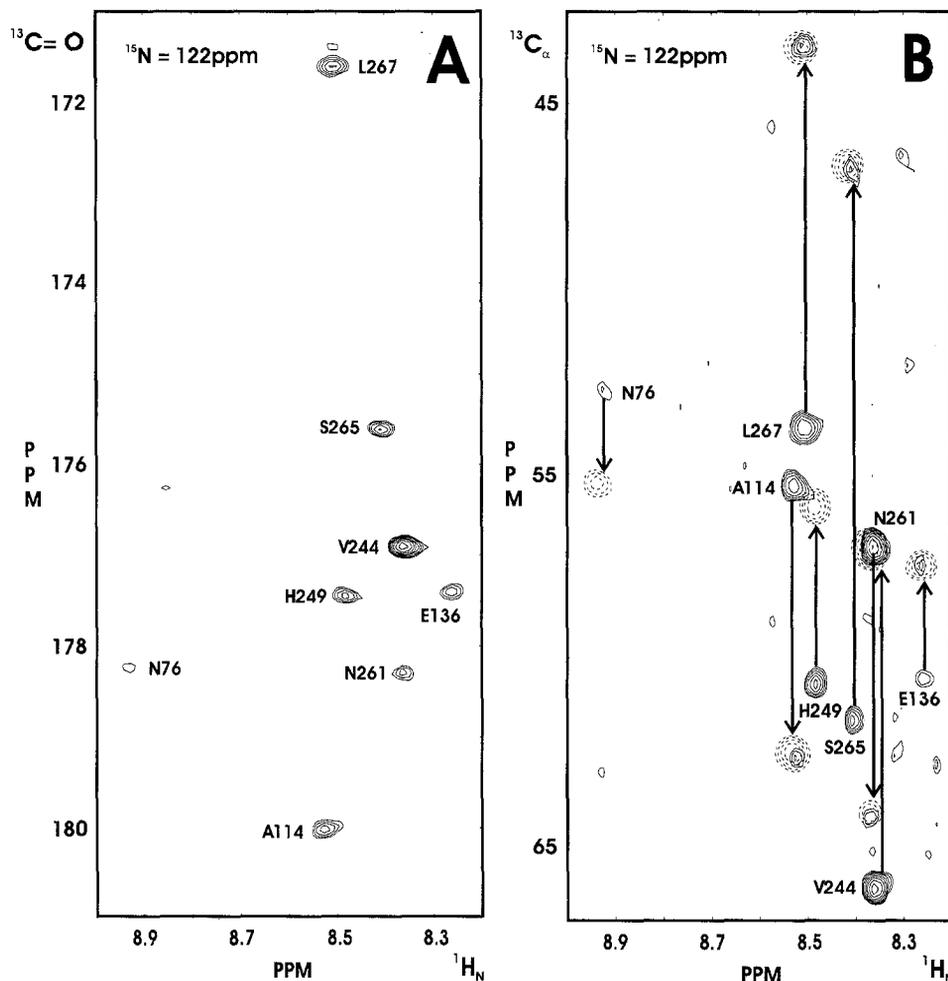


Fig. 4. Cross sections through the HNCA, HN(CO)CA and HNCO spectra taken at a ^{15}N chemical shift of 122 ppm. (A) HNCO; (B) HNCA (solid lines) and HN(CO)CA (dashed lines) are overlaid to show how the spectra can be aligned using the interresidual correlations which appear in both spectra. This is helpful, considering the degree of overlap in the $^1\text{H}^{\text{N}}$ dimension illustrated here. Both inter- and intraresidual correlations appear in the HNCA spectrum; lines are drawn connecting the two. Overlap of inter- and intraresidual correlations as shown by the two peaks labelled N261 (see text) can cause problems. Here the HN(CO)CA (dashed) peak belongs to Val²⁴⁴ and the overlapping HNCA peak belongs to Asn²⁶¹.

more interactive; we repeatedly used the computer linking program, followed by examination of the spectra to choose between branch points, and check break points for incorrect chemical shifts within the clusters.

Figure 4 shows slices from the HNCO, HNCA and HN(CO)CA spectra, taken at a single ^{15}N value. The HNCA (solid lines) and HN(CO)CA (dashed lines) are superimposed onto one picture to illustrate the redundancy of information in these two experiments, which was useful for comparing and assigning these spectra. Since both the HNCA and HN(CO)CA spectra have interresidual correlations, exact comparison of cross peaks is possible and helps in aligning the

spectra, which is necessary, given the degree of overlap illustrated by this example. Lines are drawn from the HNCA to the HN(CO)CA cross peaks, showing the correct correlations.

The peaks associated with Asn²⁶¹ in the figure illustrate an example where recourse to examination of the spectra was necessary to correct a cluster that was built using computers and chemical shift lists. Initially, the cluster containing this HN(CO)CA peak had no C^α correlation, since the intraresidual HNCA peak which gives that value was associated (by computer comparison of chemical shift values) with the interresidual HN(CO)CA cross peak later assigned to Val²⁴⁴, and was assigned on that basis as being an interresidual-type HNCA cross peak. Although it is not certain just by examination of the spectra whether the HNCA peak is inter- or intraresidual, the proximity to other peaks and the peak intensity (more characteristic of an intraresidual peak) are immediate clues to the possibility of an erroneous assignment.

Using these methods, 94% of the backbone atoms were assigned and the results are presented in Table 2.

Secondary structure results

Figure 5 summarizes the sequential NOE data, all taken from a 3D ¹⁵N NOESY-HMQC experiment with a 75-ms mixing time. Although spin-diffusion is already extensive at this mixing time for a protein of this size, this degree of spin-diffusion is acceptable for the determination of secondary structure (Grzesiek et al., 1992). Amide proton exchange data taken from a sample kept in D₂O for two weeks, and information on the deviation of C^α chemical shifts from random coil values (Spera and Bax, 1991) are included as additional indicators of secondary structure.

Because of the limited chemical shift dispersion of α-protons, overlap of these chemical shifts makes it difficult to determine d_{αN}-type NOEs unambiguously. Chemical shift redundancy of the amide protons also limits the determination of d_{NN} connectivities, i.e., if the correlated proton resonances lie within 0.15 ppm of one another, the cross peak is obscured by the diagonal. In Fig. 5, question marks show where some ambiguities occurred in reading the NOE spectrum; only ambiguous peaks which are found between residues in the neighbourhood of a 'run' of similar connectivities have been noted. Seventeen pairs of residues noted by question marks have no other connectivities. The d_{αN} (i,i + 2) and d_{βN} connectivities have not yet been determined.

Approximately 50 more residue pairs in the figure are without a d_{NN},d_{αN} connectivity, or a question mark. Out of these 50, about 15 apparently show neither of those connectivities. Most of the remainder do not have such connectivities because they are paired with one of the 13 prolines, or because one of the residues lacks an H^α assignment.

It is clear from the number of breaks in the sequential connectivities that the protein under investigation would have been difficult, if not impossible, to assign solely by the Wüthrich method (Wüthrich, 1986), but assignment of this protein using only backbone connectivities was also complicated, and required confirmation from NOE connectivities in some cases.

The deviations from random coil values of the C^α chemical shifts have been shown by Spera and Bax (1991) to be cautiously correlated to secondary structure and can serve as a valuable source of supplementary information about secondary structure. For each residue in Fig. 5 these deviations, δC^α, have been indexed as +1, -1 or 0 to represent a likelihood of the residue being in a helical (+1 value) or β-sheet (-1 value) stretch, as discussed below. Twenty-five of the 33 question marks shown for indeterminate NOE connectivities are for d_{NN}-type connectivities. The data from the secondary C^α chemical shifts indicate unbroken helical structure, i.e., consecutive

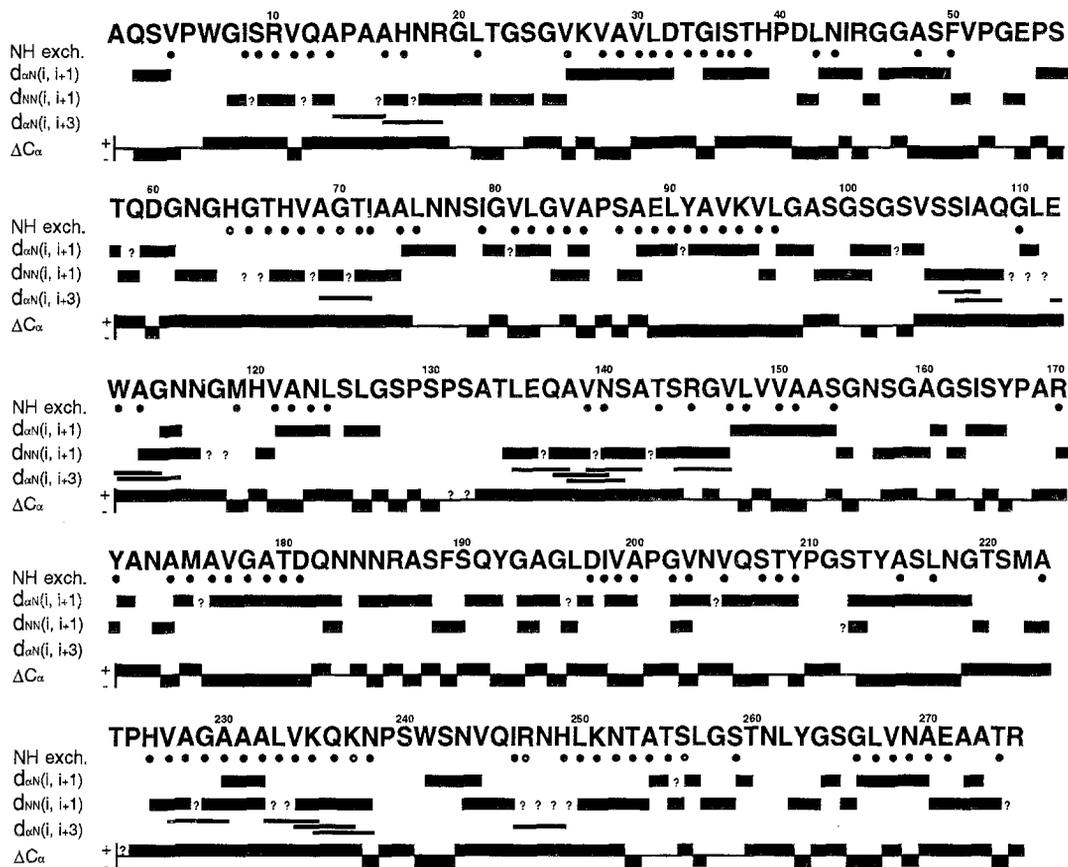


Fig. 5. Summary of $d_{\alpha N}(i,i+1)$ and $d_{\alpha N}(i,i+3)$ NOEs, amide exchange data, and C^α secondary shifts observed for SavinaseTM. Circles indicate amide protons which were still present after two weeks. Open circles are tentative assignments, because of overlap in the 2D 1H - ^{15}N HSQC spectrum. Question marks in the NOE lines indicate uncertainty due to overlap in the 3D ^{15}N NOESY-HMQC from which all NOEs were taken. Question marks only appear if there is uncertainty in a position where a cross peak might reasonably be expected because of the surrounding connectivities and C^α secondary shift. The C^α secondary chemical shifts from random coil are designated with -1, 0 or +1, according to criteria explained in the text. Numbering goes up to 275, because we are using the BPN' numbering system, as explained in the introduction.

residues with indices of +1 where these breaks in the continuity of NOE connectivities occur; therefore this source of supplementary information supports the continuity of a secondary structural element where the NOE data are missing.

Figure 6 compares the secondary structures of SavinaseTM, determined by X-ray crystallography (Betzel et al., 1992), 'NMR' (i.e., defined by NOE and amide exchange data), and C^α chemical shift differences. The criteria used to decide secondary structure from δC^α data in this example are analogous to those developed by Wishart et al. (1992) for using H^α chemical shift deviations to determine secondary structure. We adapted a form that has been applied to a similar phenomenon, i.e., H^α secondary chemical shifts, and arbitrarily adjusted the parameters so that the maximum number of structural elements would be recognized. Similar to Wishart et al. (1992), we have assigned a chemical shift index of -1, 0 or +1 to each residue, but in this application for

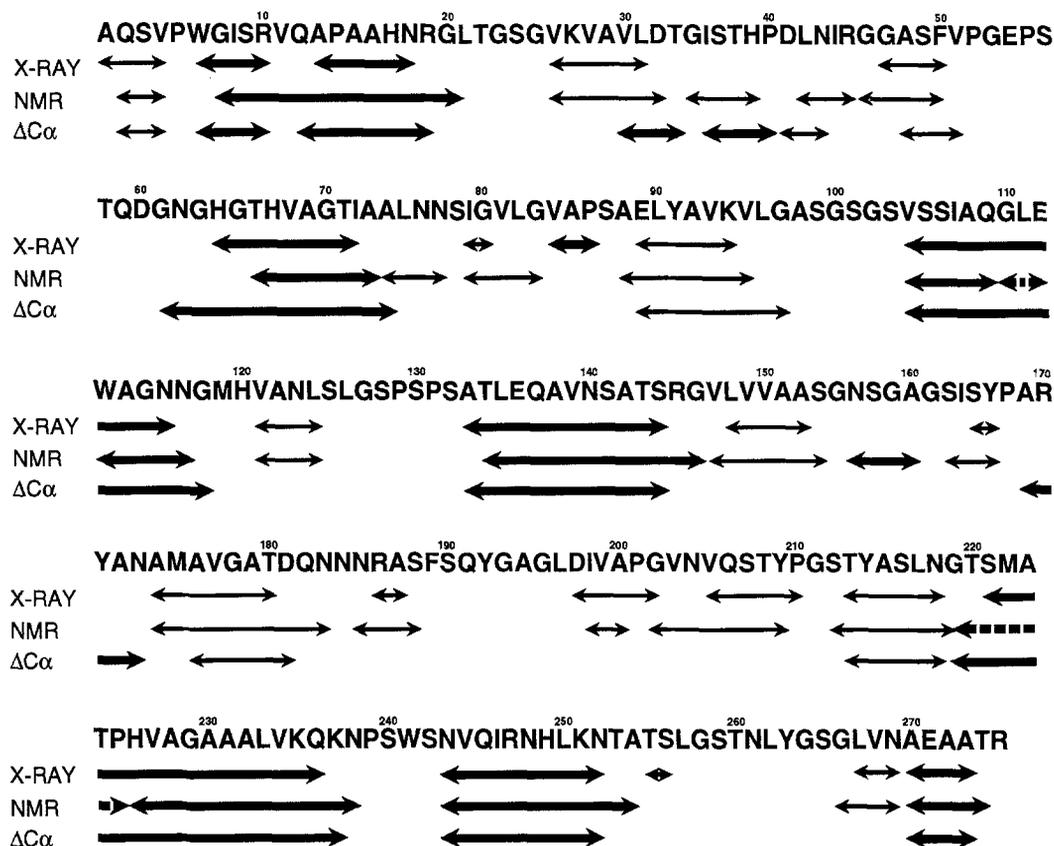


Fig. 6. Summary of helical/turn (thick lines) and extended (thin lines) types of structure in SavinaseTM from X-ray and NMR sources. The band labelled 'NMR' refers to structure determination using NOE and amide exchange data. The band labelled ΔC^α refers to structure deduced using C^α secondary chemical shifts from random coil. The three sources coincide in their determinations for the most part. NMR and X-ray diverge, due to differences in definition of these structures. The C^α chemical shifts method fails to indicate several regions of extended structure found by the other methods, and indicates two extra helical regions. Broken lines in the NMR summary indicate spots where missing data precludes putting in a solid line, but the surrounding evidence suggests that structure.

$\delta C^\alpha \geq 0.3$ ppm, $\delta C^\alpha \leq -0.3$ ppm or $-0.3 \leq \delta C^\alpha \leq 0.3$ ppm, respectively. The boundary value of ± 0.3 ppm was chosen as the smallest value which could be viewed as being outside the error of calculation. We also defined the helices and β -sheets as shown in Fig. 6 analogously to Wishart et al. (1992), i.e., a group of four or more +1 values is defined as a helix, and a group of three or more -1 values is defined as a β -sheet. In our case, any breaks in sign of a run of indices are shown as breaks in the structure type.

For the most part, the three methods show the same secondary structures. Although it is doubtful that gross perturbations between these structures will be observed, it is worth noting that the X-ray structure of SavinaseTM was calculated for the molecule not complexed with inhibitor.

NOE data indicate more extended structure than the X-ray data, primarily because of the more stringent definition of extended structure used in the X-ray structure. Only a β -sheet which shows interstrand hydrogen bonding is defined as such in the X-ray structure. There are four small

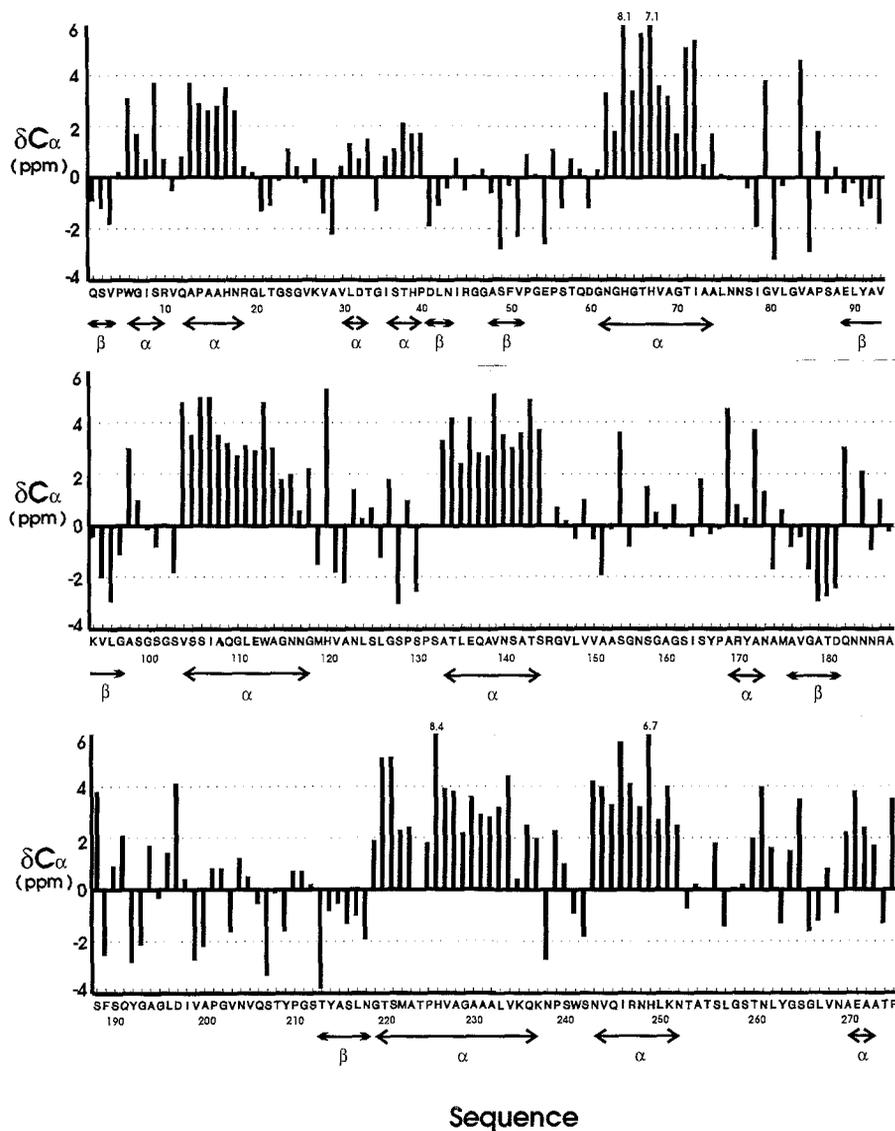


Fig. 7. Summary of C_{α} chemical shift deviations from random coil values for each residue in SavinaseTM. Secondary structural elements determined from these data, using criteria described in the text, are shown by the arrows under the sequence, and labelled α for α -helices and β for β -sheet secondary structures. δC_{α} values exceeding 6 ppm are indicated above the bars for those residues.

stretches of $d_{\alpha N}$ connectivities (also containing nonexchanged amide protons) in the NMR data which are not defined as extended structure by X-ray. Otherwise, the two are quite similar.

Comparison of the secondary structural elements derived from our classification of the δC_{α} data with those derived from the other two methods shows interesting differences. In order to discuss those differences we have graphically shown the actual δC_{α} values obtained for each residue in Fig. 7. The structural elements determined by our adaptation of the Wishart method

are shown beneath the sequence. The δC^α -defined α -helices coincide with those defined by the other methods, with two exceptions. In the region from Val³⁰ to Pro⁴⁰ the δC^α method defines two small helical regions where NOEs define a β -sheet, while the X-ray structure shows no extended structure (Fig. 6). Examination of Fig. 7 reveals that the average positive shift difference for those nine residues is less than 1.5 ppm, whereas the average positive shift difference for a helical structure is 3 ppm (Spera and Bax, 1991). It can be seen from Fig. 7 that most of the other defined helices have greater average positive shifts. The other exception occurs from Ala¹⁶⁹ to Asn¹⁷³. In this case the actual C^α chemical shift differences show an uneven distribution of positive deviations, also at variance with the more even distribution seen in the other helices. Interestingly, this region and the section from Ser²⁵⁹ to Gly²⁶⁶ (most residues show a positive deviation) both consist of consecutive β -turns that 'give rise to almost helically condensed stretches of secondary structure' (Betzel et al., 1992).

Several small stretches of β -sheet, defined by NOEs and X-ray, are not indicated in the δC^α method. This may be due to several factors. One is that the average C^α shift difference is greater for α -helices (3.09 ± 1.00 ppm) than it is for β -sheets (-1.48 ± 1.23 ppm) (Spera and Bax, 1991). Moreover, the relative error associated with the β -sheet deviations is much more significant. The average deviations determined by Spera and Bax (1991) for helices and sheets coincide with those determined for SavinaseTM (3.0 ppm for the helices and -1.5 ppm for the sheets). Perturbations due to factors other than secondary structure could more easily obscure the smaller shifts observed for β -sheet structures.

CONCLUSIONS

It was shown by Grzesiek and Bax (1992) that it is possible to obtain enough types of 3D and 4D multinuclear NMR spectra of sufficient quality to assign a protein dimer in the 30-kDa range, using the backbone resonance assignment method. We have shown that it is also possible to sequentially assign a monomeric protein in that molecular weight range. With 269 residues, SavinaseTM is one of the largest monomeric proteins to be assigned by this method to date. Such an increase in the number of residues commensurately increases the complexity of the spectra, to the point that overlap is a problem even in three dimensions.

Two main problems hindered the assignment. One was obtaining H^α chemical shifts unambiguously and completely, and the other was to link clusters of chemical shifts on the basis of a 2D match of C^α and H^α chemical shift values. The first problem was partially solved by using HN(CA)HA and HN(COCA)HA experiments to unambiguously determine as many α -protons as possible. Some missing H^α resonances were found using TOCSY and NOESY experiments, but the use of only TOCSY and NOESY spectra appeared impractical because of the plethora of peaks in those spectra, and the inability to distinguish which of them correspond to α -proton resonances. Additionally, in the NOESY spectrum it is impossible to distinguish whether a correlation is intraresidual, sequential, or long-range.

The second problem exists because the size of the protein limits the 3D multinuclear experimental repertoire, excluding two experiments (CBCANH and HN(CA)CO) which could add a third chemical shift for use in linking clusters and alleviate the problem of multiple matches. It was therefore necessary to resort to spin-type identification early in the assignment, spin typing as many clusters as possible, similar to the Wüthrich method. This spin typing was done reasonably

accurately using only C^α chemical shift values and 'TOCSY ladders' as criteria, although the C^β chemical shift values obtained late in the analysis were crucial in making some final determinations.

It was also necessary to resort to the Wüthrich assignment method to complete the sequential assignment. Because of missing peaks and overlap ambiguities, inherent in multinuclear multidimensional spectra for a protein of this size and complexity, as many different correlations as possible had to be used to obtain the amount of information needed to sequentially assign this molecule.

Solution structure determination by NMR hinges upon the identification of NOE constraints. Overlap of NOESY peaks in three dimensions suggests moving into four dimensions to resolve ambiguities. Some progress may be made by using a 4D $^{15}N/^{13}C$ -separated NOESY, but as noted by Bax and Grzesiek (1993), recording of 4D spectra for these larger proteins is also problematic because of the loss of sensitivity. However, more $d_{\alpha N}$ connectivities may be found from this experiment. For example, about 15 $d_{\alpha N}(i, i + 3)$ connectivities were identified, but another 15 possible connections could not be verified because of the uncertainty caused by chemical shift overlap. Additional d_{NN} connectivities may be found from a 3D $^1H-^{15}N$ HMQC-NOESY-HMQC experiment (Frenkiel et al., 1990; Ikura et al., 1990). Our initial efforts with these experiments yielded a rather low signal-to-noise ratio. We expect more recent versions of experiments of this type, e.g. Vuister et al. (1993), to be more successful. Other missing connectivities are due to as yet unassigned H^α resonances, which will probably be forthcoming after further analysis. Work is in progress on the assignment of the side-chain resonances and a determination of the 3D solution structure of SavinaseTM.

Some residues remain unassigned, but may be absent due to ligand interactions or saturation transfer. This is unfortunate, since precisely these interacting residues are of interest. Still, the fact that they are not present may be informative. Perhaps under different conditions they will be resolved. We hope to study solvent and ligand interactions of the molecule after determination of a complete 3D solution structure. We are currently working on an analysis of the molecule's flexibility, using amide relaxation and exchange data.

ACKNOWLEDGEMENTS

M.L.R. was supported by a grant from Unilever Research. We thank W. Musters and P. Ravesteyn for the production and purification of the ^{15}N -labelled sample of subtilisin, and S. Hastrup from Novo/Nordisk for the $^{13}C/^{15}N$ -labelled sample of subtilisin. We gratefully acknowledge Dr. S. Grzesiek for sharing his pulse program for the CBCA(CO)NH experiment with us. We also wish to thank J.J.M. Joordens for his technical assistance, and C.H.M. Papavoine and R.H.A. Folmer for their help in the preparation of the manuscript.

REFERENCES

- Antonov, V.K., Ivaniva, T.V., Berezin, I.V. and Martinek, K. (1971) *FEBS Lett.*, **7**, 23–25.
Bachovchin, W.W., Wong, W.Y.L., Farr-Jones, S., Shenvi, A.B. and Kettner, C.A. (1988) *Biochemistry*, **27**, 7689–7697.
Bax, A. and Ikura, M. (1991) *J. Biomol. NMR*, **1**, 99–104.
Bax, A. and Grzesiek, S. (1993) *Acc. Chem. Res.*, **26**, 133–138.
Bech, L.M., Sørensen, S.B. and Breddam, K. (1993) *Biochemistry*, **32**, 2845–2852.

- Betzal, C., Pal, G.P. and Saenger, W. (1988) *Eur. J. Biochem.*, **178**, 155–171.
- Betzal, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S. and Wilson, K.S. (1992) *J. Mol. Biol.*, **223**, 427–445.
- 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, S. (1988) *J. Biol. Chem.*, **263**, 7895–7906.
- Briedigkeit, L. and Frömmel, C. (1989) *FEBS Lett.*, **253**, 83–87.
- Brown, S.C., Weber, P.L. and Mueller, L. (1988) *J. Magn. Reson.*, **77**, 166–169.
- Clubb, R.T., Thanabal, V. and Wagner, G. (1992a) *J. Biomol. NMR*, **2**, 203–210.
- Clubb, R.T., Thanabal, V. and Wagner, G. (1992b) *J. Magn. Reson.*, **97**, 213–217.
- Fesik, S.W. and Zuiderweg, E.R.P. (1988) *J. Magn. Reson.*, **78**, 588–593.
- Fink, A.L. (1987) In *Enzyme Mechanisms* (Eds, Page, M.I. and Williams, A.) Royal Society of Chemistry, London, pp. 159–177.
- Frenkiel, T., Bauer, C., Carr, M.D., Birdsall, B. and Feeney, J. (1990) *J. Magn. Reson.*, **90**, 420–425.
- Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.
- 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., Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) *J. Am. Chem. Soc.*, **112**, 9020–9022.
- Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496–514.
- Kay, L.E., Wittekind, M., McCoy, M.A., Friedrichs, M.S. and Mueller, L. (1992) *J. Magn. Reson.*, **98**, 443–450.
- 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. (1989a) *Biochemistry*, **28**, 6150–6156.
- Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989b) *J. Magn. Reson.*, **85**, 393–399.
- Marion, D., Kay, L.E., Sparks, S.W., Torchia, D. and Bax, A. (1989c) *J. Am. Chem. Soc.*, **111**, 1515–1517.
- Markland, F.S. and Smith, E.L. (1971) In *The Enzymes*, Vol. 3 (Ed, Boyer, R.D.) Academic Press, New York, NY, pp. 561–608.
- Matthews, D.A., Alden, R.A., Birktoft, J.J., Freer, S.T. and Kraut, J. (1975) *J. Biol. Chem.*, **250**, 7120–7126.
- Messerle, B.A., Wider, G., Otting, G., Weber, C. and Wüthrich, K. (1989) *J. Magn. Reson.*, **85**, 608–613.
- Nakatani, H., Uehara, Y. and Hiromi, K. (1975) *J. Biochem.*, **78**, 611–616.
- Olejniczak, E.T., Xu, R.X., Petros, A.M. and Fesik, S.W. (1992) *J. Magn. Reson.*, **100**, 444–450.
- Pantoliano, M.W., Whitlow, M., Wood, J.F., Rollence, M.L., Finzel, B.C., Gilliland, G.L., Poulos, T.L. and Bryan, P.N. (1988) *Biochemistry*, **27**, 8311–8317.
- Philipp, M. and Bender, M.L. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 478–480.
- Philipp, M. and Maripuri, S. (1981) *FEBS Lett.*, **133**, 36–38.
- Powers, R., Gronenborn, A.M., Clore, G.M. and Bax, A. (1991) *J. Magn. Reson.*, **94**, 209–213.
- Robertus, J.D., Kraut, J., Alden, R.A. and Birktoft, J.J. (1972) *Biochemistry*, **11**, 4293–4303.
- Robillard, G. and Shulman, R.G. (1974) *J. Mol. Biol.*, **86**, 541–558.
- Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.
- Tepljakov, A.V., Kuranova, I.P., Harutyunyan, E., Vainshtein, B.K., Frömmel, C., Höhne, W.E. and Wilson, K.S. (1990) *J. Mol. Biol.*, **214**, 261–279.
- Tsilikounas, E., Kettner, C.A. and Bachovchin, W.W. (1992) *Biochemistry*, **31**, 12839–12846.
- Vuister, G.W., Clore, G.M., Gronenborn, A.M., Powers, R., Garrett, D.S., Tschudin, R. and Bax, A. (1993) *J. Magn. Reson. Ser. B*, **101**, 210–213.
- Wells, J.A. and Estell, D.A. (1988) *Trends Biochem. Sci.*, **13**, 291–297.
- Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.
- Wright, C.S., Alden, R.A. and Kraut, J. (1969) *Nature*, **221**, 235–242.
- Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.