Drake, A. F., Siligardi, G., & Gibbons, W. A. (1988) *Biophys. Chem. 31*, 143-146.
Holzwarth, G., & Doty, P. J. (1965) *J. Am. Chem. Soc. 87*, 218-228.
Houghten, R. A., & DeGraw, S. T. (1987) *J. Chromatogr. 386*, 223-228.
Houghten, R. A., DeGraw, S. T., Bray, M. K., Hoffman, S. R., & Frizzell, N. D. (1986) *Biochemistry 4*, 522-528.
Levitt, M. (1978) *Biochemistry 17*, 4277-4285.
Marqusee, S., & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8898-8902.
Merutka, G., & Stellwagen, E. (1990) *Biochemistry 29*, 894-898.
Mihalyi, E. (1968) *J. Chem. Eng. Data 13*, 179-182.
Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., & Baldwin, R. L. (1990) *Nature (London) 344*, 268-270.
Richardson, J. A., & Richardson, D. C. (1988) *Science 240*, 1648-1652.
Rossi, V., Grandi, C., Dalzoppo, D., & Fontana, A. (1983) *Int. J. Pept. Protein Res. 22*, 239-250.
Rudolph, R., Fuchs, I., & Jaenicke, R. (1986) *Biochemistry 25*, 1662-1669.
Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M., & Baldwin, R. L. (1987) *Nature (London) 326*, 563-567.
Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Konishi, Y., & Scheraga, H. A. (1984) *Macromolecules 17*, 148-155.
Tiffany, M. L., & Krimm, S. (1968) *Biopolymers 6*, 1379-1382.
Tiffany, M. L., & Krimm, S. (1972) *Biopolymers 11*, 2309-2316.
Vasquez, M., Pincus, M. R., & Scheraga, H. A. (1987) *Biopolymers 26*, 351-371.

# Sequence-Specific $^1$H NMR Assignments and Determination of the Secondary Structure for the Activation Domain Isolated from Pancreatic Procarboxypeptidase B†

Josep Vendrell,‡ Gerhard Wider,‡ Francesc X. Avilés,§ and Kurt Wüthrich*,‡

*Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule—Hönggerberg, CH-8093 Zürich, Switzerland, and Departament de Bioquímica i Biologia Molecular, Unitat de Ciències, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*

*Received January 25, 1990; Revised Manuscript Received April 18, 1990*

ABSTRACT: Nearly complete sequence-specific $^1$H NMR assignments are presented for amino acid residues 3-81 in the 81-residue globular activation domain of porcine pancreatic procarboxypeptidase B isolated after limited tryptic proteolysis of the zymogen. These resonance assignments are consistent with the chemically determined amino acid sequence. Regular secondary structure elements were identified from nuclear Overhauser effects and the sequence locations of slowly exchanging backbone amide protons. The molecule contains two $\alpha$-helices, including residues 20-30 and approximately residues 58-72, and a three-stranded antiparallel $\beta$-sheet with the individual strands extending approximately from 12 to 17, 50 to 55, and 75 to 77. The identification of these secondary structures and a preliminary analysis of additional long-range NOE distance constraints show that isolated activation domain B forms a stable structure with the typical traits of a globular protein. The data presented here are the basis for the determination of the complete three-dimensional structure of activation domain B, which is currently in progress.

Procarboxypeptidases are the inactive precursors of carboxypeptidases, a class of proteolytic enzymes that degrade polypeptides from their carboxy terminal. Two of these zymogens, A and B, are named after the corresponding active enzymes, which are classified according to their specificity for cleaving certain C-terminal peptide bonds. Upon activation by limited proteolysis, the zymogens are converted into the active enzymes through the release of a polypeptide of about 90-100 amino acid residues (Quinto et al., 1982; Vendrell et al., 1986; Flogizzo et al., 1988; Wade et al., 1988; Gardell et al., 1988; Clauser et al., 1988). This activation segment[1] has been shown to play an essential role in inhibiting the catalytic activity of the protein in its zymogen state (San Segundo et al., 1982), probably by blocking the interaction between substrates and the preformed catalytic site (Uren & Neurath, 1974). The so far unanswered question then arises whether the function of the activation segment depends on the formation of a defined tertiary structure in the proenzyme. In

† Financial support for this project was obtained from the Schweizerischer Nationalfonds (Project 31.25174.88), CICYT (Ministerio de Educación y Ciencia, Spain, Project BIO88/0456), and EMBO (long-term postdoctoral fellowship to J.V.).

‡ Eidgenössische Technische Hochschule—Hönggerberg.
§ Universitat Autònoma de Barcelona.

[1] Abbreviations: NMR, nuclear magnetic resonance; 2D, two dimensional; Bis-Tris, [bis(2-hydroxyethyl)amino]tris(hydroxymethyl)-methane; 2QF-COSY, two-dimensional two-quantum filtered correlation spectroscopy; 2Q spectra, two-dimensional two-quantum spectroscopy; TOCSY, two-dimensional total correlation spectroscopy; NOE, nuclear Overhauser effect; NOESY, two-dimensional NOE spectroscopy; activation domain B, trypsin-resistant N-terminal 81-residue portion of the activation segment isolated from porcine pancreatic procarboxypeptidase B; activation segment A, activation segment isolated from porcine pancreatic procarboxypeptidase A; $d_{AB}(i,j)$, distance between proton types A and B located in amino acid residues $i$ and $j$, respectively, where N, $\alpha$, and $\beta$ denote the amide proton, $C^{\alpha}H$, and $C^{\beta}H$, respectively.

0006-2960/90/0429-7515$02.50/0 

porcine procarboxypeptidase A, which is the more extensively studied of the two proenzymes, qualitative evidence was obtained that the isolated activation segment includes a stable globular conformation with a high content of regular secondary structure (Avilés et al., 1982; Sánchez-Ruiz et al., 1988), but no structure determination in single crystals or in solution is available either for a procarboxypeptidase or for an isolated activation segment. As part of the ongoing studies on structure–function correlations in procarboxypeptidases and their activation segments in one of our laboratories (Avilés et al., 1982; San Segundo et al., 1982; Vendrell et al., 1989), we have started a $^1$H NMR determination of the three-dimensional structure of the polypeptide fragment consisting of the N-terminal 81 residues of activation segment B. In the following we refer to this 81-residue polypeptide fragment as *activation domain B*. It is resistant to further proteolysis after tryptic activation of the proenzyme. Procarboxypeptidase B was selected for this study because its activation segment and activation domain can be prepared with higher yields and better homogeneity than the corresponding fragments of procarboxypeptidase A and because the long-term stability of isolated activation domain B makes it suitable for NMR experiments. In this paper we report the sequence-specific $^1$H NMR assignments for this protein (Wüthrich, 1986) and the regular secondary structure elements determined from the observation of interresidual NOE's and the identification of the sequence locations of slowly exchanging amide protons (Wüthrich et al., 1984).

## Materials and Methods

Procarboxypeptidase B was isolated from porcine pancreatic acetone powders (Vilanova et al., 1985a), and activation domain B was prepared from the purified proenzyme by limited proteolysis with trypsin and anion-exchange chromatography, as will be described in detail elsewhere (F. J. Burgos and F. X. Avilés, unpublished results). Low molecular weight impurities present after the initial preparation were removed by ultrafiltration. When the protein samples used for the NMR experiments were kept in solution for several weeks, we noticed the appearance of weak extra peaks in the fingerprint region of the 2D NMR spectra. These samples could be repurified by ion-exchange chromatography on a FPLC system (Pharmacia), with a Mono-Q 5/5 column (Pharmacia) equilibrated with 20 mM Bis-Tris buffer and 0.20 M NaCl (pH 6.6), applying a salt gradient to 0.26 M NaCl in the same buffer in 20 min at a flow rate of 1 mL/min.

The solution conditions for the structure determination of activation domain B were selected on the basis of the following observations. Circular dichroism spectroscopy showed that the protein conformation was maintained over the pH range 2–13 and at temperatures from 20 to 60 °C. However, in the pH range 2.5–6.0 the solubility was found to be too small for 2D NMR measurements. Therefore, we decided to use 3.5 mM protein solutions in 20 mM phosphate buffer at pH 6.5. pH meter readings were used without any further correction for isotope effects. Because of the relatively high rates of amide proton exchange at pH 6.5 when compared to those at more acidic pH values (Wüthrich, 1986), the NMR data were recorded at 15 and 25 °C. No significant line broadening was observed at the lower temperature. Samples for NMR spectroscopy were prepared either in a mixture of 90% $H_2O$ and 10% $^2H_2O$ or in 99.99% $^2H_2O$. For the preparation of a $^2H_2O$ sample with nearly complete exchange of the labile protons, the protein was kept in the deuterated solvent for 35 h at 20 °C and then twice freeze-dried and redissolved in 99.99% $^2H_2O$.

All 2D $^1$H NMR spectra were recorded on a Bruker AM600 spectrometer. 2QF-COSY (Rance et al., 1983), 2Q spectra (Wagner & Zuiderweg, 1983), clean TOCSY (Griesinger et al., 1988), and NOESY (Jeener et al., 1979; Anil Kumar et al., 1980) experiments were performed in the phase-sensitive mode with time-proportional phase incrementation of the initial pulse (Marion & Wüthrich, 1983). Quadrature detection was used in both dimensions, and the carrier was placed in the center of the spectrum. The mixing time used for the NOESY spectra was 120 ms. In $H_2O$ samples the strong water signal was suppressed by selective saturation (Wider et al., 1983). The spectral width was 7.8 kHz in $H_2O$ and 6.5 kHz in $^2H_2O$. The COSY and NOESY spectra used for sequential assignments were measured with an interleaved mode to eliminate artifacts from medium-term instabilities (Neuhaus et al., 1985). The number of $t_1$ values varied from 560 for NOESY in $^2H_2O$ to 740 for NOESY in $H_2O$. The 2D NMR data sets were multiplied in both dimensions with phase-shifted sine bell functions (De Marco & Wüthrich, 1976). Prior to Fourier transformation, the time-domain data were zero-filled to 4096 points in $t_2$ and 2048 points in $t_1$. Base-line distortions were eliminated by a base-line correction with a second- or third-order polynomial. Spectra were recorded at both 15 and 25 °C to resolve overlap in crowded spectral regions.

For the identification of slowly exchanging amide protons, the fully protonated protein sample was lyophilized from 20 mM phosphate buffer in $H_2O$ at pH 6.5. The protein was then redissolved in 99.99% $^2H_2O$ and immediately inserted into the preshimmed spectrometer with a probe temperature of 15 °C, and a first NOESY spectrum was obtained in 14 h. The sample was then again lyophilized twice from $^2H_2O$, redissolved in 99.99% $^2H_2O$, and kept at 15 °C. Recording of a second NOESY spectrum was started after the protein had spent a total of 55 h in $^2H_2O$ solution at 15 °C, with a measuring time of 30 h.

## Results

Sequence-specific $^1$H NMR assignments were obtained by standard procedures (Wagner & Wüthrich, 1982; Wider et al., 1982). In a first step, 2QF-COSY and TOCSY spectra recorded in $^2H_2O$ solutions at 15 or 25 °C were used to classify the nonaromatic spin systems of the individual amino acid residues into eight classes, Gly, Ala, Thr, Val, Ile, Leu, AMX spin systems, and all other residues, which all have long side chains (Wüthrich, 1986). These spin systems were completed by connecting the labile protons, by use of NMR spectra in $H_2O$, and by combining the aromatic systems with their AMX systems, by use of NOESY spectra in $^2H_2O$. Sequential NOE's corresponding to $d_{\alpha N}$, $d_{NN}$, and $d_{\beta N}$ were then used to establish sequential connectivities between two or several of these spin systems (Billeter et al., 1982). The resulting oligopeptide segments were matched with the chemically determined sequence of the protein to obtain sequence-specific assignments. Finally, spin-system identification was completed for residues with complex side chains, which had previously been classified as AMX or as residues with long side chains.

*Spin System Identifications.* Activation domain B contains 2 Gly, 3 Ala, 3 Thr, 8 Val, 5 Ile, and 6 Leu, 29 AMX spin systems corresponding to the $C^\alpha H$–$C^\beta H_2$ fragments of 7 Asp, 5 Asn, 5 Ser, 5 His, 5 Phe, 1 Tyr, and 1 Trp, and 25 spin systems corresponding to "long side chains" of 12 Glu, 4 Gln, 2 Pro, 4 Lys, and 3 Arg. The identification of the spin systems started with the study of the 2QF-COSY spectra, where most of the $C^\alpha H$–$C^\beta H$ cross peaks were identified (Figure 1). Combined use with TOCSY spectra permitted the following initial classification of the spin systems.

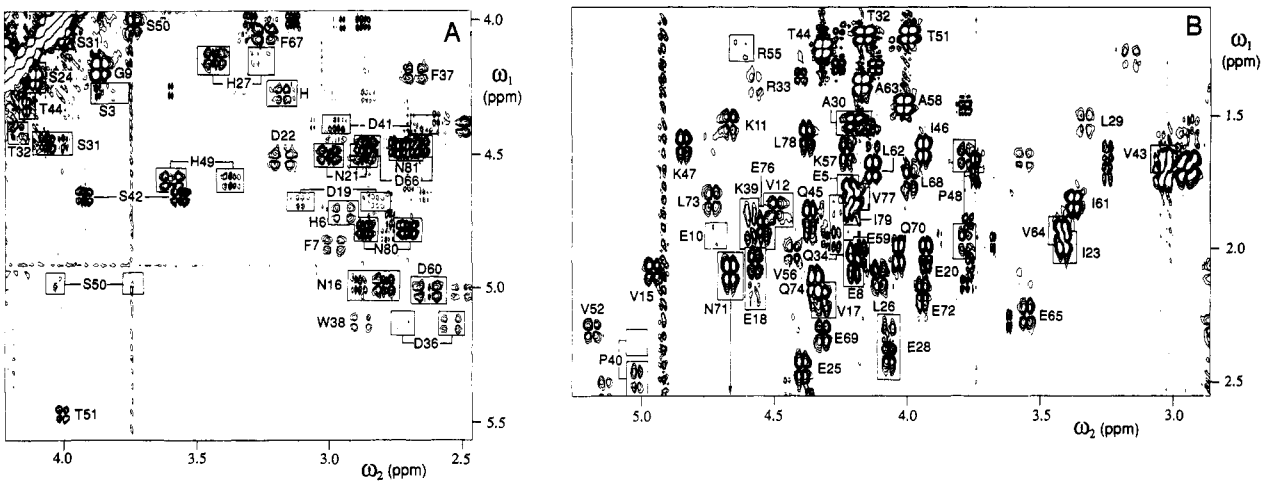


FIGURE 1: Proton 2QF-COSY spectrum of activation domain B ($^1$H frequency 600 MHz, solvent $^2H_2O$, p$^2$H 6.5, $T$ = 15 °C). (A) Contour plot of the region containing the $C^\alpha H$–$C^\beta H$ cross peaks for most of the AMX spin systems and the threonines and $C^\alpha H$–$C^\alpha H$ of Gly. H defines the position of a $C^\alpha H$–$C^\beta H$ cross peak corresponding to a His spin system, which corresponds either to His 1 or His 2. (B) Contour plot of the region containing the $C^\alpha H$–$C^\beta H$ cross peaks for the residue types not seen in panel A and the $C^\beta H$–$C^\gamma H_3$ cross peaks of Thr. The arrow indicates that the position of one of the β-protons of Asn 71 lies outside of the area shown. Positive and negative values are shown without distinction. Cross peaks are identified with the one-letter code for the amino acid type and the sequence position deduced after completion of the sequence-specific assignments. For improved ease of identification by the reader, some peaks are enclosed in boxes.

Only one of the two glycines was identified from the characteristic, large active coupling constant. (Evidence for the second Gly in position 4 was obtained from the sequential assignment procedure, but only one $C^\alpha H$ resonance could be identified; see Table I.) Ala $C^\alpha H$–$C^\beta H_3$ cross peaks could be distinguished from Thr $C^\beta H$–$C^\gamma H_3$ cross peaks on the basis of Thr $C^\alpha H$–$C^\beta H$ cross peaks observed close to the diagonal (Figure 1A) and the relayed $C^\alpha H$–$C^\gamma H_3$ TOCSY cross peaks for Thr. The other methyl-containing residues were also completely identified either with 2QF-COSY alone or from the combined analysis of 2QF-COSY and TOCSY spectra. The only exceptions to this were two Ile spin systems, for which only one of the methylene protons could be observed (see Table I) and which were only completed after the sequential connectivities had been established.

A first distinction between the $C^\alpha H$–$C^\beta H_2$ moieties of AMX spin systems and longer side chains results from the chemical shifts (Wüthrich, 1986). As is seen from Table I, 24 out of the 26 identified AMX spin systems have $C^\beta H$ chemical shifts at lower field than 2.3 ppm, and 21 out of 24 identified long side chains have $C^\beta H$ shifts at higher field than 2.3 ppm. The two AMX spin systems with smaller $C^\beta H$ chemical shifts than 2.3 ppm belong to Phe 13 and Phe 54, and they were readily identified through the observation of strong $C^\beta H$–$C^\delta H$ cross peaks in the NOESY spectra. In Figure 1B we present a spectral region of a 2QF-COSY spectrum which contains the $C^\alpha H$–$C^\beta H$ cross peaks for most of the residues with long side chains. Because of the large number of Glu and Gln residues, this region is very crowded, which made the use of TOCSY spectra essential for discrimination between the different spin systems. For most of the residues with long side chains, complete identification of the spin systems was achieved only after sequential assignments of the backbone resonances. Exceptions are the two Pro residues, which had chemical shift nonequivalent $C^\delta$ protons that could be identified through comparison of the NOESY and 2QF-COSY spectra. Among the AMX spin systems (Figure 1A) the Ser residues were tentatively identified from their low-field $C^\beta H$ chemical shifts (Bundi & Wüthrich, 1979).

The identification of the aromatic spin systems was achieved with 2QF-COSY and TOCSY in $^2H_2O$. The single Tyr was distinguished from the five Phe residues on the basis of the observation that it gave rise to a single TOCSY cross peak with the C2,6-protons, whereas two cross peaks were observed for the C2,6-protons of phenylalanine, i.e., direct connectivity with the C3,5-protons and relayed connectivity with the C4-proton. In the single Trp spin system the resonances of C6H and C7H were nearly degenerate, but they could be distinguished from the NOESY connectivity to the N1 proton. Five COSY cross peaks corresponding to the C2H–C4H connectivities in the His residues were found to match with outstandingly sharp lines in the one-dimensional $^1$H NMR spectrum. Four of these pairs of imidazole resonances could be connected with the AMX spin systems of their $C^\alpha H$–$C^\beta H_2$ moieties, and one set of His imidazole protons remained without further connectivities. From the comparison of spectra recorded in $H_2O$ and $^2H_2O$, it was possible to determine the positions of cross peaks corresponding to the side-chain amide protons of three Gln and five Asn. The individual connectivities with these peripheral labile protons were only established after the sequential assignments had been completed.

The connectivities between the spin systems of the nonlabile protons with the backbone amide protons were established with 2QF-COSY and TOCSY in $H_2O$. The COSY fingerprint region at 15 °C contained 71 NH–$C^\alpha H$ cross peaks (Figure 2) that could be attributed to one of the spin systems previously defined, either on the basis of unique chemical shifts of the $C^\alpha$ proton resonances or from relayed NH–$C^\beta H$ connectivities observed in TOCSY spectra. Further identifications of amide proton resonances were obtained with 2QF-COSY and TOCSY spectra recorded at 25 °C and from the NOESY spectra used for the sequential assignments (see below). Quite generally, the spectra recorded at the two different temperatures were useful for resolving spectral overlap in the amide proton chemical shift range. In addition, spectral overlap could in some instances also be resolved by inspection of the NOESY spectra recorded with the partially exchanged protein in $^2H_2O$ solution, which were used for the measurement of the amide proton exchange rates. Overall, virtually complete identification of the amide proton positions was achieved, the only exceptions being those of the N-terminal tripeptide segment.

At certain moments the appearance of some weak additional resonances in the 2QF-COSY fingerprint, which are marked with circles in Figure 2, caused some ambiguities during this

Table I: $^1$H Chemical Shifts for Activation Domain B at pH 6.5 and 15 °C

| residue | chemical shift, δ (ppm)[a] | | | |
|---|---|---|---|---|
| | NH | αH | βH | others |
| His 1[b] / His 2 | | 4.27 | 3.23, 3.13 | δ$^2$H 7.25; ε$^1$H 7.50 / δ$^2$H 7.11 ε$^1$H 7.25 |
| Ser 3 | | 4.28 | 3.85, 3.72 | |
| Gly 4 | 8.50 | 3.89, – | | |
| Glu 5 | 8.04 | 4.20 | 1.79, 1.79 | γCH$_2$ 2.07, 1.95 |
| His 6 | 8.58 | 4.71 | 3.10, 2.93 | δ$^2$H 7.17; ε$^1$H 7.72 |
| Phe 7 | 8.71 | 4.82 | 3.30, 2.95 | ring[c] 7.243 |
| Glu 8 | 9.11 | 4.19 | 2.05, – | γCH$_2$ 2.34, – |
| Gly 9 | 9.25 | 4.18, 3.85 | | |
| Glu 10 | 7.64 | 4.71 | 1.95, – | γCH$_2$ 2.57, 2.41 |
| Lys 11 | 8.83 | 4.67 | 1.53, 1.33 | γCH$_2$ 1.13, –; δCH$_2$ 1.70, 1.53; εCH$_2$ 2.80, – |
| Val 12 | 8.75 | 4.47 | 1.85 | γCH$_3$ 0.60, – |
| Phe 13 | 8.83 | 4.74 | 1.94, 1.61 | δH 6.74, –; εH 6.86, –; ζH 7.22 |
| Arg 14 | 8.89 | 5.69 | 1.69, 1.65 | γCH$_2$ 1.59, 1.53; δCH$_2$ 3.19, 2.98; εNH 9.78; ηNH$_2$ 6.70 |
| Val 15 | 9.88 | 4.95 | 2.08 | γCH$_3$ 1.16, 1.07 |
| Asn 16 | 8.72 | 4.99 | 2.88, 2.78 | δNH$_2$ 7.54, 7.02 |
| Val 17 | 8.86 | 4.32 | 2.18 | γCH$_3$ 0.88, 0.79 |
| Glu 18 | 10.38 | 4.57 | 2.17, 2.05 | γCH$_2$ 2.43, 2.27 |
| Asp 19 | 7.51 | 4.67 | 3.10, 2.82 | |
| Glu 20 | 8.88 | 3.92 | 2.08, 2.01 | γCH$_2$ 2.38, 2.34 |
| Asn 21 | 8.38 | 4.49 | 2.99, 2.86 | δNH$_2$ 8.18, 7.13 |
| Asp 22 | 8.21 | 4.50 | 3.15, 2.63 | |
| Ile 23 | 7.42 | 3.40 | 1.97 | γCH$_2$ 1.98, 1.74; γCH$_3$ 0.84; δCH$_3$, 0.59 |
| Ser 24 | 8.18 | 4.20 | 4.09, 4.09 | |
| Glu 25 | 8.27 | 4.38 | 2.44, – | γCH$_2$ 1.98, – |
| Leu 26 | 7.95 | 4.09 | 2.10, 2.02 | γH 1.27; δCH$_3$ 0.95, 0.82 |
| His 27 | 8.36 | 4.13 | 3.42, 3.26 | δ$^2$H 7.10; ε$^1$H 8.03 |
| Glu 28 | 8.27 | 4.06 | 2.37, 2.32 | γCH$_2$ 2.38, 2.25 |
| Leu 29 | 8.43 | 3.32 | 1.51, 0.71 | γH 0.92; δCH$_3$ 0.35, 0.18 |
| Ala 30 | 8.03 | 4.19 | 1.53 | |
| Ser 31 | 7.65 | 4.44 | 4.05, 3.98 | |
| Thr 32 | 7.61 | 4.40 | 4.16 | γCH$_3$ 1.20 |
| Arg 33 | 8.00 | 4.57 | 1.41, 1.30 | γCH$_2$ 1.42, 1.30; δCH$_2$ 3.15, 2.81; εNH 7.22; ηNH$_2$ 6.53 |
| Gln 34 | 8.75 | 4.31 | 1.98, 1.86 | |
| Ile 35 | 8.26 | 4.39 | 1.30 | γCH$_2$ 0.76, 0.65; γCH$_3$ 0.04; δCH$_3$ −0.25 |
| Asp 36 | 8.00 | 5.13 | 2.79, 2.53 | |
| Phe 37 | 8.93 | 4.19 | 3.56, 2.66 | δH 7.27, –; εH 6.98, –; ζH 6.65 |
| Trp 38 | 9.19 | 5.13 | 3.21, 2.87 | δ$^1$H 6.93; ε$^3$H 7.47; ε$^1$NH 9.72; ζ$^3$H 7.10; ζ$^2$H 7.25; η$^2$H 7.27 |
| Lys 39 | 8.09 | 4.58 | 1.92, 1.87 | |
| Pro 40 | | 5.01 | 2.49, 2.30 | γCH$_2$ 1.76, 1.69; δCH$_2$ 3.79, 3.61 |
| Asp 41 | 8.05 | 4.42 | 2.98, 2.64 | |
| Ser 42 | 7.03 | 4.65 | 3.92, 3.55 | |
| Val 43 | 8.82 | 3.05 | 1.60 | γCH$_3$ 0.63, 0.35 |
| Thr 44 | 7.42 | 4.12 | 4.31 | γCH$_3$ 1.25 |
| Gln 45 | 7.30 | 4.36 | 1.87, – | γCH$_2$ 2.39, 2.31; εNH$_2$ 7.57, 7.11 |
| Ile 46 | 7.00 | 3.93 | 1.63 | γCH$_2$ 1.67, –; γCH$_3$ 0.80; δCH$_3$ 0.68 |
| Lys 47 | 8.10 | 4.83 | 1.87, 1.61 | |
| Pro 48 | | 3.77 | 1.97, 1.65 | γCH$_2$ 2.11, 1.66; δCH$_2$ 3.71, 3.54 |
| His 49 | 8.15 | 4.58 | 3.57, 3.35 | δ$^2$H 7.32; ε$^1$H 8.69 |
| Ser 50 | 8.50 | 4.98 | 4.01, 3.71 | |
| Thr 51 | 8.30 | 5.46 | 3.98 | γCH$_3$ 1.20 |
| Val 52 | 9.15 | 5.17 | 2.30 | γCH$_3$ 1.30, 1.16 |
| Asp 53 | 9.43 | 6.33 | 2.74, 2.41 | |
| Phe 54 | 8.97 | 5.05 | 2.33, 2.08 | δH 6.62, –; εH 6.67, –; ζH 6.53 |
| Arg 55 | 8.65 | 4.61 | 1.26, – | |
| Val 56 | 9.21 | 4.40 | 2.00 | γCH$_3$ 0.99, 0.86 |
| Lys 57 | 9.37 | 4.21 | 1.63, 1.54 | εCH$_2$ 2.58, – |
| Ala 58 | 9.63 | 4.00 | 1.46 | |
| Glu 59 | 9.59 | 4.16 | 2.06, 2.01 | γCH$_2$ 2.28, 2.20 |
| Asp 60 | 8.14 | 5.01 | 2.60, 2.60 | |
| Ile 61 | 7.20 | 3.36 | 1.82 | γCH$_2$ 2.02, –; γCH$_3$ 0.98; δCH$_3$ 1.03 |
| Leu 62 | 8.27 | 4.12 | 1.71, 1.59 | γH 1.72; δCH$_3$ 0.98, 0.92 |
| Ala 63 | 8.27 | 4.16 | 1.39 | |
| Val 64 | 8.30 | 3.40 | 1.93 | γCH$_3$ 0.77, 0.32 |
| Glu 65 | 8.82 | 3.54 | 2.29, 2.24 | γCH$_2$ 2.83, – |
| Asp 66 | 8.32 | 4.46 | 2.85, 2.65 | |
| Phe 67 | 7.90 | 4.04 | 3.22, 3.13 | δH 7.05, –; εH 7.12, –; ζH 7.38 |
| Leu 68 | 8.79 | 3.97 | 1.75, 1.67 | γH 2.01; δCH$_3$ 0.86, 0.03 |
| Glu 69 | 8.44 | 4.31 | 2.32, 2.20 | γCH$_2$ 2.72, 2.51 |
| Gln 70 | 8.92 | 4.03 | 2.02, 1.99 | γCH$_2$ 2.45, –; εNH$_2$ 7.72, 6.80 |
| Asn 71 | 7.29 | 4.66 | 2.75, 2.08 | γNH$_2$ 6.55, 6.25 |
| Glu 72 | 7.68 | 3.93 | 2.23, 2.18 | |
| Leu 73 | 8.35 | 4.72 | 1.81, 1.37 | γH 1.62; δCH$_3$ 1.00, 0.87 |
| Gln 74 | 9.32 | 4.34 | 2.13, 2.13 | γCH$_2$ 2.49, 2.41; εNH$_2$ 7.58, 6.89 |
| Tyr 75 | 8.29 | 5.88 | 2.93, 2.86 | δH 7.04, –; εH 6.81, – |

Table I (Continued)

| residue | chemical shift, δ (ppm)[a] NH | αH | βH | others |
|---|---|---|---|---|
| Glu 76 | 9.00 | 4.53 | 1.92, 1.92 | $\gamma CH_2$ 2.10, - |
| Val 77 | 8.72 | 4.18 | 1.82 | $\gamma CH_3$ 0.66, 0.59 |
| Leu 78 | 8.74 | 4.36 | 1.56, - | γH 1.41; $\delta CH_3$ 0.79, 0.67 |
| Ile 79 | 8.32 | 4.21 | 1.84 | $\gamma CH_2$ 1.45, 1.19; $\gamma CH_3$ 0.90; $\delta CH_3$ 0.86 |
| Asn 80 | 8.60 | 4.78 | 2.85, 2.70 | $\delta NH_2$ 7.76, 6.95 |
| Asn 81 | 8.13 | 4.48 | 2.71, 2.68 | $\delta NH_2$ 7.56, 6.79 |

[a] For methylene groups two chemical shifts are given only when two resolved signals were observed or when the presence of the two degenerate signals had unambiguously been established by the observation of a remote peak in the two-quantum experiment in $^2H_2O$. [b] The chemical shifts for the protons belonging to either His 1 or His 2 could not be sequentially assigned. [c] Ring denotes that only one resonance could be observed for the ring protons of Phe 7.

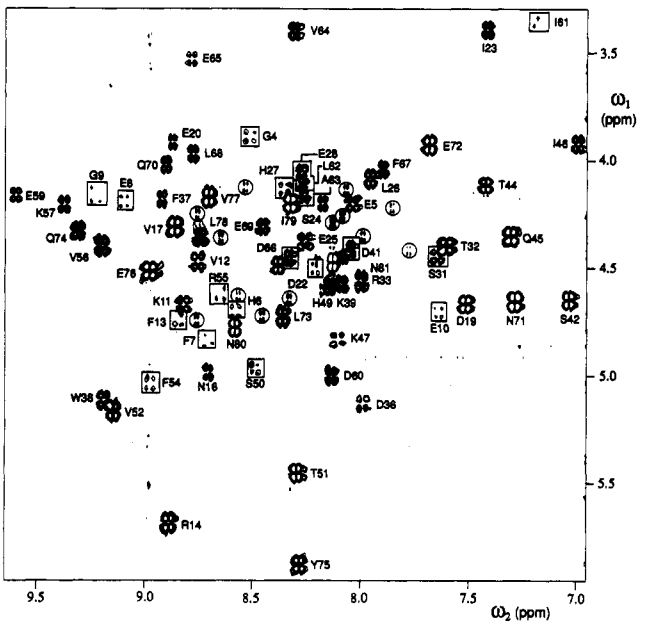


FIGURE 2: Contour plot of the central fingerprint region of a $^1H$ 2QF-COSY spectrum of activation domain B ($^1H$ frequency 600 MHz, mixed solvent of 90% $H_2O$ and 10% $^2H_2O$, pH 6.5, $T$ = 15 °C). NH–C$^\alpha$H cross peaks are labeled with the one-letter symbol for the amino acid type and the sequence position. For improved ease of identification by the reader, some peaks are enclosed in boxes. Circles identify peaks that arise from minor protein contaminations in the preparation of activation domain B.

part of the spectral analysis. These extra peaks probably arise from slow deterioration of the protein. FPLC anion-exchange chromatography of samples kept in solution during several weeks revealed the presence of a minor second protein component in preparations that were chromatographically homogeneous at the outset of the NMR experiments. These extra peaks were too weak to cause difficulties for the sequential assignments, since they were not visible in the NOESY spectra.

*Sequential Resonance Assignments.* Sequential assignments were obtained from NOESY spectra recorded in $H_2O$ solution with a mixing time of 120 ms. The usual procedures (Wüthrich, 1986; Wagner & Wüthrich, 1982; Wider et al., 1982) were followed. Once the intraresidual cross peaks had been identified, a search was made for sequential connectivity $d_{\alpha N}$, $d_{\beta N}$, or $d_{NN}$. Figure 3 illustrates the assignment pathways for the polypeptide segment Asp 60–Leu 73. To optimize the reliability of the sequential assignments (Billeter et al., 1982), we searched the spectra for all three types of connectivities. At the same time, the medium-range connectivities used to identify helical secondary structures (Wüthrich et al., 1984) were also used to further ascertain the assignments. Figure 4 shows a survey of all the sequential and medium-range NOE connectivities that could be identified.

The following tripeptides, which were identified by the sequential assignment technique, could be unambiguously matched with unique positions in the chemically determined amino acid sequence. They thus served as starting points for the sequence-specific assignments (Wüthrich, 1986) (□ stands for AMX; ■, for a long side chain): Val 12–□–■, Val 15–□–Val, ■ 18–□–■, Ser 31–Thr–■, □ 36–□–□, Ser 50–Thr–Val, ■ 57–Ala–■, Leu 62–Ala–Val, and Leu 73–■–□. Starting from these tripeptide segments, the sequential assignments were continued in both directions. Overall, of the 80 pairs of sequentially neighboring residues in activation domain B, 65 were identified by more than one sequential NOE connectivity, and one sequential NOE was observed for 12 of the remaining 15 pairs of residues. The missing sequential contacts are between the first three residues and between residues Gln 34 and Ile 35 (Figure 4). The sequential NOE's with the two Pro residued helped to identify a trans and a cis X–Pro peptide bond (Wüthrich et al., 1984; Wüthrich, 1986). Sequential connection $d_{\alpha\alpha}$ for Lys 39-Pro 40 indicated a cis X–Pro bond, and $d_{\alpha\delta}$ and $d_{\beta\delta}$ contacts showed that there is a trans X–Pro bond between Lys 47 and Pro 48.

*Connectivities with Peripheral Side-Chain Protons.* NOESY spectra in $^2H_2O$ were used to attach the aromatic spin systems to the corresponding AMX moieties, mainly by use of the C$^\beta$H–C$^\delta$H cross peaks (Billeter et al., 1982). For the single Trp residue, the connectivity of 4H with C$^\beta H_2$ was used. Weak NOE's from the C$^\beta$ protons to the more peripheral ring protons of phenylalanine were also useful for ascertaining the spin-system identifications. For Phe 7 only a single resonance could be identified, and one imidazole ring system could not be connected with an AMX spin system. Sequence-specific assignments of the labile side-chain amide protons of five Asn and three Gln residues were identified from the NOE's to C$^\beta$H and C$^\gamma$H, respectively, which were observed in the NOESY spectra in $H_2O$ solution. Finally, NOESY spectra were also used to check, or in some cases to extend, the assignments of complex side-chain spin systems that had so far not been completely identified. The new assignments thus obtained include that the spin system identifications were completed for Arg 14, Ile 23, Arg 33, Ile 46, and Ile 61 and that some peripheral side-chain proton resonances could be identified for Lys 11 and Lys 57.

*Identification of Regular Secondary Structures.* The presence of both helical and antiparallel β-sheet secondary structures was implied by the observation of numerous $d_{NN}$ contacts typical of α-helix (Billeter et al., 1982; Wüthrich et al., 1984; Wüthrich, 1986) and some strong sequential $d_{\alpha N}$ and long-range $d_{\alpha\alpha}(i,j)$ contacts involving protons in low-field positions relative to the random coil chemical shifts (Bundi & Wüthrich, 1979), which are typical of β-sheets (Pardi et al., 1983). Helical structures were further characterized by

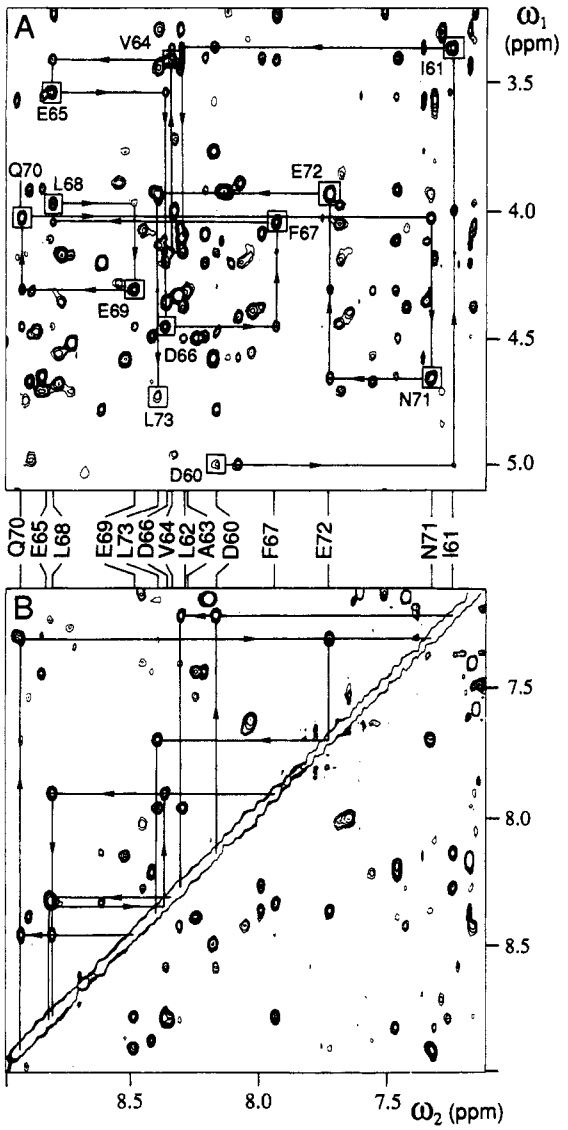


FIGURE 3: Illustration of sequential connectivities in activation domain B. Two regions of a NOESY spectrum are shown ($^1$H frequency 600 MHz, mixed solvent of 90% $H_2O$ and 10% $^2H_2O$, pH 6.5, $T$ = 15 °C, mixing time $\tau_m$ = 120 ms), and connectivities $d_{\alpha N}$ and $d_{NN}$ for the polypeptide segment from Asp 60 to Leu 73 are identified. (A) $d_{\alpha N}$ connectivities are shown with solid lines connecting squares that identify the NH–C$^\alpha$H COSY cross-peak positions with the sequential NOESY cross peaks [the squares are identified with the one-letter symbol for the amino acid residue and the sequence position; in all cases shown here they contain the intraresidual NOESY cross peak corresponding to $d_{N_\alpha}(i,i)$]. (B) Sequential assignment pathway of $d_{NN}$ connectivities. The amide proton chemical shifts are indicated in the space between parts A and B of the figure. The arrows indicate the direction of the assignment pathways from residues 60–73.

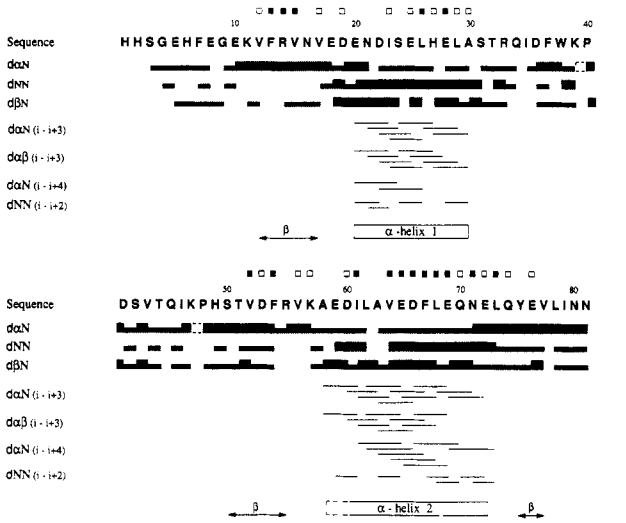


FIGURE 4: Amino acid sequence and survey of sequential and medium-range NOE connectivities and amide proton exchange data used to establish the sequence-specific $^1$H NMR assignments and to identify elements of regular secondary structure. Above the sequence, open squares identify residues with slowed exchange of the amide protons to the extent that the NH lines were visible in a NOESY spectrum of freshly dissolved activation domain B in $^2H_2O$ recorded during 14 h at p$^2$H 6.5 and 15 °C. Filled squares identify those amide protons for which the NMR lines were still observable after 85 h in $^2H_2O$ at p$^2$H 6.5 and 15 °C. Below the sequence, thin and thick bars indicate sequential $d_{\alpha N}$, $d_{\beta N}$, and $d_{NN}$ connectivities represented by weak and strong NOESY cross peaks, respectively. The $d_{\alpha\alpha}$ connectivity K39–P40 and the $d_{\alpha\delta}$ connectivity K47–P48 are shown with broken lines. The NOE connectivities $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$, $d_{\alpha\beta}(i,i+3)$, and $d_{NN}(i,i+2)$ are represented by lines that start and end at the positions of the two interacting residues. At the bottom the locations of the $\alpha$-helical and $\beta$-sheet segments in the polypeptide chain are indicated, which were identified from the NMR data of this figure and Figure 5. Only approximate bounds are indicated for the N-terminal end of $\alpha$-helix 2 (see the text).

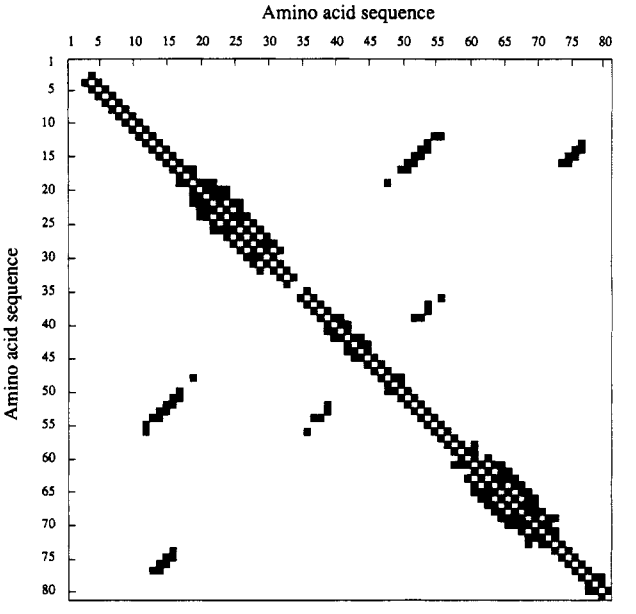


FIGURE 5: Diagonal plot representing the NOE's observed between different backbone protons in activation domain B, which were used for the delineation of the secondary structure. Both axes are calibrated with the amino acid sequence. A filled square in position $(i,j)$ indicates that one or several backbone proton–backbone proton NOE's were observed between the two residues in sequence locations $i$ and $j$.

medium-range NOE's, which arise from spatial proximity between pairs of residues in positions $i$ and $i$+3 and also $i$ and $i$+2 or $i$ and $i$+4. Figure 4 shows that all types of connectivities defining an $\alpha$-helix are found from Gly 20 to Leu 30, which corresponds to three helix turns, and from Ala 58 to Glu 72, which is equivalent to four helix turns. The presence of slowly exchanging amide protons in these two segments (Figure 4) is in agreement with the hydrogen-bonding patterns expected in $\alpha$-helices.

In a diagonal plot of backbone proton–backbone proton contacts (Figure 5) the helical regions stand out because of the crowding of medium-range backbone NOE's close to the diagonal. It is also apparent from Figure 5 that a number of the long-range backbone contacts $d_{\alpha\alpha}(i,j)$, $d_{\alpha N}(i,j)$, and $d_{NN}$-$(i,j)$, which are typical for antiparallel $\beta$-sheets (Wüthrich et al., 1984; Wüthrich, 1986), connect the polypeptide segments Val 12–Val 17, Ser 50–Arg 55, and Tyr 75–Val 77. These contacts are consistent with the three-stranded sheet shown in Figure 6, with the central strand formed by residues 12–17.

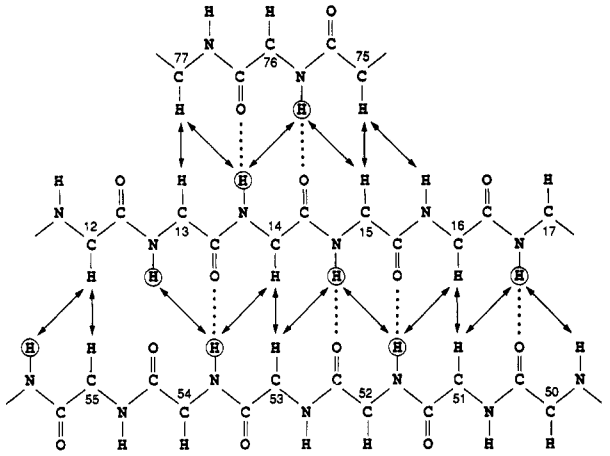

FIGURE 6: Schematic representation of the antiparallel $\beta$-sheet structure identified in activation domain B. In the three-stranded sheet the backbone atoms, the $C^{\alpha}$ protons, the amide protons, and the carbonyl oxygen atoms are drawn, and the sequence locations are indicated near the $C^{\alpha}$'s. Arrows indicate experimental interstrand NOE's between backbone protons, and circles identify amide protons for which slowed exchange with the solvent was observed. Hydrogen bonds are indicated by dotted lines only in those positions where both slow exchange of the amide proton and at least two interstrand long-range NOE's with the amide proton were observed.

The hydrogen bonds in this structure are independently implicated by the experimental data on slowly exchanging amide protons (Figure 4). Some additional contacts were also observed between backbone protons of residues Val 52, Asp 53, and Phe 54 in one of the peripheral strands (Figure 6) and backbone protons of residues Phe 37, Trp 38, and Lys 39 (Figure 5). Since no regular secondary structure appears to be compatible with these connectivities, their interpretation was deferred to the determination of the complete tertiary structure, which is currently in progress.

## Discussion

Overall, the sequence locations of regular secondary structure elements of activation domain B as shown in Figure 4 are well-defined by the nearly complete sequence-specific resonance assignments (Table I and Figure 4) and the additional medium-range and long-range NOE constraints shown in Figures 4 and 5 (Wüthrich et al., 1984; Wüthrich, 1986). Independently, these secondary structures are also evidenced by the slowed exchange of distinct amide protons with the solvent (Figure 4). In particular, all NOE's implicated by the $\beta$-sheet drawn in Figure 6 were observed, with the exception of cross-strand NOE $d_{\alpha N}(55,13)$, which could not be positively identified because of overlap with other cross peaks. The amide protons of Asn 16 and Ser 50 do not exchange slowly, since they are located at the edge of the $\beta$-sheet. In addition to the data shown in Figures 5 and 6, numerous long-range side-chain proton–backbone proton NOE's between residues in the three $\beta$-strands confirm the proposed structure. A number of backbone–backbone contacts between protons in tripeptides Val 52–Phe 54 and Phe 37–Lys 39 suggest that there may be a fourth strand to the $\beta$-sheet, but since the observed NOE's are not typical for a regular $\beta$-structure, we defer further investigation on this detail to the distance geometry computation of the complete tertiary structure. The length of helix 1 from residues 20–30 appears to be quite well-defined. For helix 2 it is uncertain whether residues 58–60 form a turn preceding the helix or are actually an integral part of a helix extending from Ala 58 to Glu 72. Clarification of this detail will again have to await the outcome of the tertiary structure calculations.

The results presented here, as well as other data on the stability of procarboxypeptidase activation segments against denaturation by heat (Sánchez-Ruiz et al., 1988) and urea (Vilanova et al., 1985b; Vendrell et al., 1989), indicate that the activation segments of procarboxypeptidases have compactly folded, globular structures. The observation of numerous long-range NOE's between side-chain protons of residues located in the different regular secondary structure elements provides additional support for this conclusion. Compared to those of other small proteins, for example, basic pancreatic trypsin inhibitor (Wagner & Wüthrich, 1982b), the amide proton exchange rates in activation domain B are nonetheless considerably faster, suggesting that activation domain B may possess a flexible globular conformation similar to that of small, $\alpha$-helix-rich proteins devoid of S–S bridges (Williams, 1986). Further quantitative experiments on the kinetics of the amide proton exchange may be warranted to describe more precisely the internal flexibility of activation domain B.

The activation segments from different procarboxypeptidases show considerable sequence homology. This homology is higher between the same procarboxypeptidases in different species than between different proenzymes in the same species, indicating that genetic differentiation took place before speciation (Gardell et al., 1988). The comparison of the presently studied porcine protein with the rat protein, which is the only other available complete sequence of a procarboxypeptidase B activation segment (Clauser et al., 1988), shows that in the two species 60% of the residues are identical and 85% of the residues are either identical or differ by conservative substitutions. In view of this high degree of sequence homology it can be anticipated that the two proteins also share some conformational traits. Among the residues involved in the regular secondary structures identified in this paper, those implicated in the $\beta$-sheet formation have close to 100% homology or conservative substitutions in the two proteins. A similar degree of homology is observed for helix 2 from residues 58–72. In contrast, strict homology in helix 1 is only 36%, and homology or conservative substitutions are observed for 73% of the residues. This could indicate that sequence conservation of helix 1 is not a stringent requirement for the conservation of the overall conformation and function of the protein.

Sequence comparisons of porcine and rat procarboxypeptidase B also show that the peptide segment linking the activation domain with the N-terminal residue of the active enzyme contains approximately 14 amino acid residues. Studies with porcine procarboxypeptidase B in one of our laboratories (Burgos, 1989) and conformational predictions for the rat protein show that this linking segment has a low propensity to fold into a regular secondary structure and that it contains mainly polar residues. The proteolytic activation of porcine procarboxypeptidase B is very fast and involves the hydrolysis of several peptide bonds in this segment. From all this it may be hypothesized that the polypeptide link between the activation domain and carboxypeptidase B, which is the target for proteolytic activation, does not have stringent structural requirements. It may, for example, form a surface loop that is readily accessible for the activating enzyme. One could then imagine that the absence of this segment does not critically affect the global structure of either the active enzyme or the activation domain. This might be further investigated by NMR studies of a complete activation segment including the 14-residue linker peptide.

The sequence-specific resonance assignments (Table I), which are the principal result of this paper, provide the basis for the determination of the complete three-dimensional structure of activation domain B in solution, which is currently in progress.

ACKNOWLEDGMENTS

We thank Mr. R. Marani for careful processing of the manuscript.

**Registry No.** Procarboxypeptidase B, 37329-68-3.

REFERENCES

Anil-Kumar, Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.

Avilés, F. X., San Segundo, B., Vilanova, M., Cuchillo, C. M., & Turner, C. (1982) *FEBS Lett. 149*, 257–260.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285–298.

Burgos, F. J. (1989) Ph.D. Thesis, Universitat Autònoma de Barcelona.

Clauser, E., Gardell, S. J., Craik, C. S., MacDonald, R. J., & Rutter, W. J. (1988) *J. Biol. Chem. 263*, 17837–17845.

De Marco, A., & Wüthrich, K. (1976) *J. Magn. Reson. 24*, 201–204.

Flogizzo, E., Bonicel, J., Kerfelec, B., Granon, S., & Chapus, C. (1988) *Biochim. Biophys. Acta 954*, 183–188.

Gardell, S. J., Craik, C. S., Clauser, E., Goldsmith, E. J., Stewart, C., Graf, M., & Rutter, W. J. (1988) *J. Biol. Chem. 263*, 17828–17836.

Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

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

Neuhaus, D., Wagner, G., Vašák, M., Kägi, J. H. R., & Wüthrich, K. (1985) *Eur. J. Biochem. 151*, 257–273.

Pardi, A., Wagner, G., & Wüthrich, K. (1983) *Eur. J. Biochem. 137*, 445–454.

Quinto, C., Quiroga, M., Swain, W. F., Nikovits, W. C., Jr., Standring, D. N., Pictet, R. L., Valenzuela, P., & Rutter, W. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 31–35.

Rance, M., Sørensen, O., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1984) *Biochem. Biophys. Res. Commun. 117*, 479–485.

Sánchez-Ruiz, J. M., López-Lacomba, J. L., Cortijo, M., & Mateo, P. L. (1988) *Eur. J. Biochem. 176*, 225–230.

San Segundo, B., Martinez, M. C., Vilanova, M., Cuchillo, C. M., & Avilés, F. X. (1982) *Biochim. Biophys. Acta 707*, 74–80.

Uren, J. R., & Neurath, H. (1974) *Biochemistry 13*, 3512–3520.

Vendrell, J., Avilés, F. X., Genescà, E., San Segundo, B., Soriano, F., & Méndez, E. (1986) *Biochem. Biophys. Res. Commun. 141*, 517–523.

Vendrell, J., Vilanova, M., Avilés, F. X., Turner, C. H., Cary, P. D., & Crane-Robinson, C. (1990) *Biochem. J. 267*, 213–220.

Vilanova, M., Vendrell, J., López, M. T., Cuchillo, C. M., & Avilés, F. X. (1985a) *Biochem. J. 229*, 605–609.

Vilanova, M., Burgos, F. J., Cuchillo, C. M., & Avilés, F. X. 1985b) *FEBS Lett. 191*, 273–277.

Wade, R. D., Hass, G. M., Kumar, S., Walsh, K., & Neurath, H. (1988) *Biochimie 70*, 1137–1142.

Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 347–366.

Wagner, G., & Zuiderweg, E. R. P. (1983) *Biochem. Biophys. Res. Commun. 113*, 854–860.

Wider, G., Lee, K. H., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 367–388.

Wider, G., Hosur, R. V., & Wüthrich, K. (1983) *J. Magn. Reson. 52*, 130–135.

Williams, R. P. J. (1986) in *Calcium and the Cell*, Ciba Foundation Symposium *122*, pp 145–161, Wiley, Chichester, U.K.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.