The EMBO Journal vol.10 no.1 pp.11 – 15, 1991

The NMR structure of the activation domain isolated from porcine procarboxypeptidase B

Josep Vendrell, Martin Billeter, Gerhard Wider, Francesc X.Avilés[1] and Kurt Wüthrich

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule – Hönggerberg, CH-8093 Zürich, Switzerland and [1]Departament de Bioquímica i Biologia Molecular, Facultat de Ciències and Institut de Biologia Fonamental, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

Communicated by K.Wüthrich

The three-dimensional structure of the activation domain isolated from porcine pancreatic procarboxypeptidase B was determined using ^{1}H NMR spectroscopy. A group of 20 conformers is used to describe the solution structure of this 81 residue polypeptide chain, which has a well-defined backbone fold from residues 11 – 76 with an average root mean square distance for the backbone atoms of 1.0 ± 0.1 Å relative to the mean of the 20 conformers. The molecular architecture contains a four-stranded β-sheet with the polypeptide segments 11 – 17, 36 – 39, 50 – 56 and 75 – 76, two well defined α-helices from residues 20 – 30 and 60 – 70, and a 3_{10} helix from residues 43 – 46. The three helices are oriented almost exactly antiparallel to each other, are all on the same side of the β-sheet, and the helix axes from an angle of ~45° relative to the direction of the β-strands. Three segments linking β-strands and helical secondary structures, with residues 32 – 35, 39 – 43 and 56 – 61, are significantly less well ordered than the rest of the molecule. In the three-dimensional structure two of these loops (residues 32 – 35 and 56 – 61) are located close to each other near the protein surface, forming a continuous region of increased mobility, and the third disordered loop is separated from this region only by the peripheral β-strand 36 – 39 and precedes the short 3_{10} helix.

Key words: activation domain B/NMR/procarboxypeptidase/protein structure

Introduction

Procarboxypeptidases are the inactive precursors of carboxypeptidases, a class of proteolytic enzymes that degrade polypeptides from their carboxy terminal. By limited proteolysis the procarboxypeptidases are cleaved into the enzymatically active carboxypeptidase and an activation segment, which inhibits the catalytic activity of the protein in the zymogen state. Among the pancreatic digestive zymogens, procarboxypeptidases have outstandingly long activation segments, with 94 residues for procarboxypeptidase A (Quinto *et al.*, 1982; Vendrell *et al.*, 1986; Flogizzo *et al.*, 1988; Wade *et al.*, 1988) and 95 residues for procarboxypeptidase B (Clauser *et al.*, 1988). The degree of sequential homology among the activation segments is comparable to that observed for the active enzymes. Several previous studies have shown that the isolated activation polypeptides possess a stable conformation resistant to denaturation (Sánchez-Ruiz *et al.*, 1988; Vendrell *et al.*, 1990a) and act as inhibitors of the corresponding enzyme (San Segundo *et al.*, 1982). This led to the hypothesis that they may be described as separate globular domains within the whole proenzyme (Avilés *et al.*, 1982). These results correlate well with zymogen activation studies carried out both for procarboxypeptidase A (Vendrell *et al.*, 1990b) and procarboxypeptidase B (F.J.Burgos *et al.*, in preparation), which showed that, after limited tryptic proteolysis, the severed activation segments are remarkably stable to further degradation. Compared to the extensive biochemical studies, and also to the detailed information available on the three-dimensional structure of the active form of pancreatic carboxypeptidases (Schmid and Herriot, 1976; Christianson and Lipscomb, 1989), relatively little is known about the conformation of the activation segments either in the free form or in the intact procarboxypeptidases. The only report so far describes the secondary structure of the porcine activation domain B in solution (Vendrell *et al.*, 1990c). The present paper now describes the complete three-dimensional structure of the activation domain B in solution.

During the tryptic activation process of porcine pancreatic procarboxypeptidase B, an 81 residue polypeptide resistant to further degradation was isolated from the reaction mixture (F.J.Burger and F.X.Avilés, in preparation). This so-called 'activation domain B' contains most of the 95 residue N-terminal activation segment severed from the proenzyme after the initial activating tryptic hydrolysis. The aforementioned paper by Vendrell *et al.* (1990c) reported the sequence-specific ^{1}H NMR assignments and the identification of secondary structure elements for this activation domain B. Using the empirical pattern recognition approach (Wüthrich *et al.*, 1984), two α-helices and a three-stranded antiparallel β-sheet could be unambiguously assigned. These contain ~50% of the total of 81 residues in the molecule. The present determination of the complete three-dimensional structure confirmed the presence of these secondary structures and identified a fourth β-strand with residues 36 – 39, which contains a β-bulge centered on residue 38, and a short fragment of 3_{10} helix with residues 43 – 46.

The timing of the present publication coincides with a publication on the determination of the crystal structure of the intact procarboxypeptidase B (Coll *et al.*, 1990), which was started after the secondary structure and the global polypeptide fold of the isolated activation domain B in solution had been known, but was otherwise conducted completely independently of the NMR studies. It will be of special interest to compare in detail the conformations of the activation domain B in the free form and in the intact proenzyme.

Results

Survey of the NMR data collected for the structure determination

The input for the structure calculations was derived from two NOESY spectra recorded with a mixing time of 60 ms in H_2O and ^2H_2O solutions of the protein respectively. This mixing time was found to be an acceptable compromise for obtaining a good signal-to-noise ratio and minimizing the contributions from spin diffusion to the NOESY cross-peak intensities (Wüthrich, 1986). For example, different intensities for the intra-residual cross-peaks between the amide proton and the two β-methylene protons were generally observed. The analysis of the 60 ms NOESY spectra was supported by cross-examination with the previously recorded 120 ms NOESY data (Vendrell *et al.*, 1990c). After several cycles of spectral analysis and structure calculation (see Materials and methods), a total of 1136 NOESY cross-peaks had been identified and assigned. Among these the number of NOEs representing effective conformational constraints was 757. Figure 1 presents a distribution of these NOE distance constraints along the polypeptide chain. It can be seen that there are no medium-range or long-range NOEs involving the chain-terminal residues 1–8 and 79–81. A scarcity of medium-range and long-range contacts is also observed for some central segments, in particular 30–35, 41–42 and 55–60. The density of medium-range NOEs is clearly highest in the regions 20–30 and 60–70, where α-helical secondary structures had previously been identified (Vendrell *et al.*, 1990c). Stereospecific assignments could be established for the β-methylene protons of the six residues Asp19, Asn21, His27, Glu28, Asp66 and Phe67, for the δ-methylene protons of Pro48, for the γ-methyl groups of Val15 and Val43, and for the δ-methyl groups of Leu29. The main reason for the low number of stereospecific assignments lies in the fact that no supplementary information to the NOE distance constraints in the form of spin–spin coupling constants could be obtained due to extensive broadening of the NMR lines.

The polypeptide backbone conformation in the activation domain B

A set of 20 conformers used to characterize the solution structure of the activation domain B was computed by a combination of distance geometry calculations using the program DIANA (Güntert *et al.*, 1991a) and restrained simulated annealing with the program X-PLOR (Brünger *et al.*, 1987) (see Materials and methods). The following is a brief summary of the results of a statistical analysis of the residual violations of the experimental input data and the uniqueness of the structure determination (Wüthrich, 1986, 1989; for recent examples, see Widmer *et al.*, 1989; Billeter *et al.*, 1990): each of the 20 structures represents a good fit of the experimental constraints, since no residual distance constraint violation exceeds 0.3 Å. A superposition of the 20 conformers showed that the chain-terminal segments 1–10 and 77–81 are not well defined by the experimental data and show extensive disorder. This was to be expected from the small number of medium-range and long-range constraints observed for these residues (Figure 1). (The chain-terminal segments are not further considered in the following description of the three-dimensional structure of the activation domain B.) Different ways of optimally superposing the polypeptide backbone structures of the aforementioned group of 20 conformers (e.g. Figure 3) revealed that even in the segment 11–76 there are variations in the local root mean square distances (RMSDs) along the polypeptide chain. In particular, the three segments 32–35, 39–43 and 56–61 show significantly increased disorder. This is also reflected in the global RMSD (Table I), which is 1.0 Å when all backbone atoms of residues 11–76 are considered, and 0.8 Å for the segments 11–31, 36–38, 44–55 and 62–76. These latter regions in the solution structure of the activation domain B contain all the regular secondary structures, most of which were previously identified (Vendrell *et al.*, 1990c).

The molecular architecture of the activation domain B is dominated by the arrangement of a cluster of two antiparallel α-helices over a four-stranded antiparallel β-sheet (Figure 2). The first well-defined polypeptide segment (Figure 3) includes the first strand of the β-sheet (residues 11–17) and

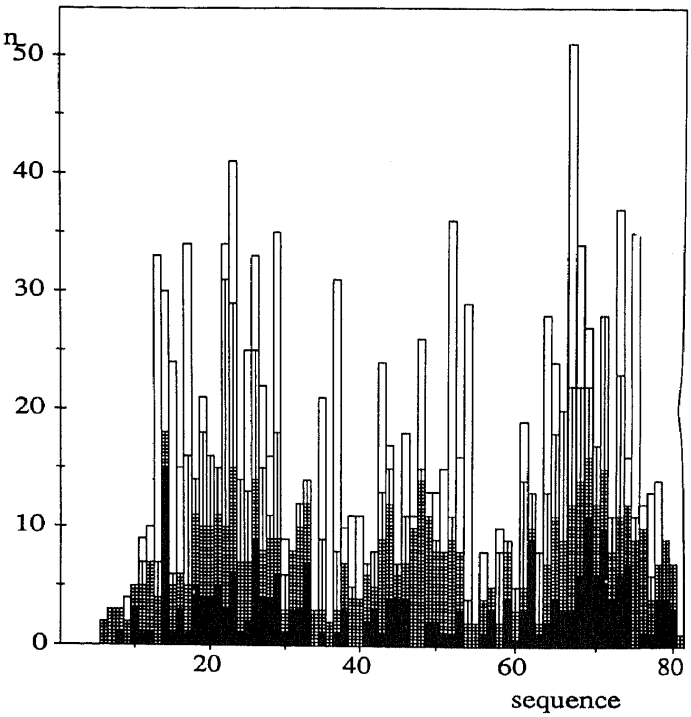

Fig. 1. Plot of the number of NOE distance constraints per residue, *n*, versus the amino acid sequence of the activation domain B. Four types of constraints are specified as follows: black, intra-residual; cross-hatched, constraints between protons in sequentially neighboring residues; vertically hatched, constraints between protons located in residues separated by 2–5 positions along the sequence; white, all other, longer-range constraints. Note that all interresidual NOEs are plotted twice, for each of the two interacting residues.

Table I. Average of the RMSD values obtained from comparison of the 20 refined structures of the activation domain B with the mean of these structures

Atoms used for the RMSD calculation	RMSD (Å)
N, C^α, C′ of residues 11–76	1.0 ± 0.1
With core side chains[a]	1.0 ± 0.1
With all side chains	1.7 ± 0.1
N, C^α, C′ of residues 11–31, 36–38, 44–55, 62–76	0.8 ± 0.1

[a]Core side chains are those of residues 13, 15, 17, 21, 23, 24, 26, 27, 29, 30, 37, 43, 46, 48, 50, 51, 52, 54, 63, 64, 66, 67, 68, 73 and 75 (see text).

the first α-helix (residues 20–30), together with the connecting turn. The pattern of H-bonds implicated by the observed polypeptide fold indicates that the residues Leu29–Arg33 might form a short 3_{10} helix at the end of the first α-helix. However, the low number of distance constraints for residues 31–35 (Figure 1) prevents a meaningful description of this structural region. Residues 36–39 are again better defined by the experimental data and represent a short β-strand. This strand was not included in the earlier description of the secondary structure of the activation domain B based on empirical pattern analysis (Vendrell *et al.*, 1990c), since only three interstrand backbone–backbone NOEs are observed with the partner residues in the neighboring strand, i.e. Phe37–C^{α}H to Phe54–C^{α}H, Trp38–NH to Phe54–C^{α}H and Lys39–NH to Asp53–NH. In the present structure, however, which was calculated from a more extensive data set, this β-strand is clearly defined. It was further possible to define the presence of a β-bulge at residue 38, since a H-bond is observed between the amide proton of Asp53 and the carbonyl oxygen of Lys39 (if the conformation were a regular antiparallel β-sheet, the carbonyl oxygen of Trp38 should act as the H-bond acceptor). A *cis*-peptide bond is found between Lys39 and Pro40, and the structure is again disordered between residues 40 and 43. The residues 43–46 form a short 3_{10} helix characterized by the medium-range NOE contacts and the H-bond formed by the amide group of Ile46. The residues 50–56 form the third strand of the β-sheet, and this is followed by another flexible loop from residues 57–59. The H-bonds involving the amide protons of residues 63 and 64 correspond to the pattern of a 3_{10} helix. This is continued by a regular α-helical H-bond pattern for the residue 65–70, a tight turn at the end of this helix, and the fourth, short strand of the β-sheet with residues 75–76.

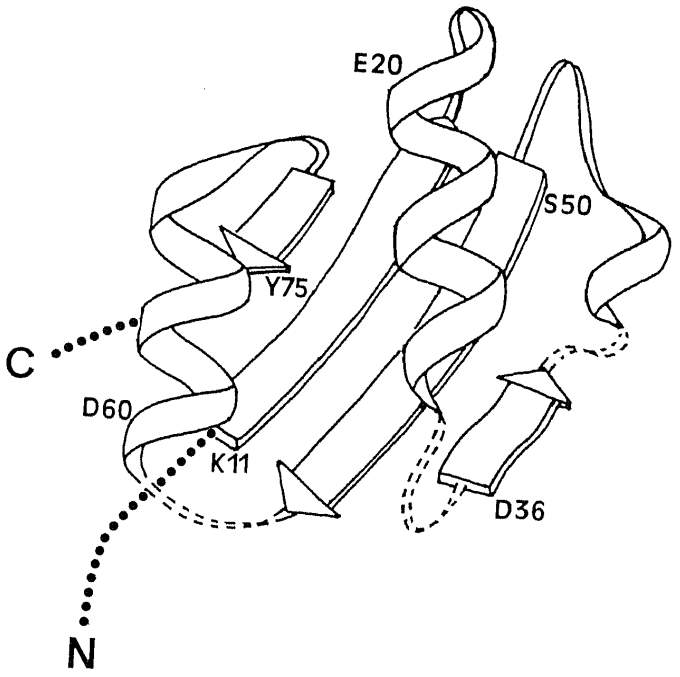

Fig. 2. Schematic drawing of the solution structure of the activation domain B. The β-strands are shown as arrows, the α-helices and the 3_{10} helix as helical bands, the well-defined turns and loops by narrow bands, the flexible loops with broken lines, and the two disordered chain ends with dotted lines. Some residue locations are identified by the one-letter symbols and sequence positions.

Amino acid side chain conformations in the activation domain B

The identification of the side chains to be used for the description of the molecular core was based on the average displacement of the side chain heavy atoms in the 20 conformers of Figure 3. Twenty-five side chains (listed in the footnote to Table I) had an average displacement of <2.0 Å. Most of these residues are hydrophobic, six are polar (Asn21, Ser24, His27, Ser50, Thr51 and Tyr75) and Asp66 is the only charged residue. These core side chains are well defined by the NMR data; the global RMSD values calculated for the backbone atoms of residues 11–76 plus all heavy atoms of these 25 side chains is the same as for the backbone atoms, i.e. 1.0 Å. If all side chains are included, the RMSD is significantly increased to 1.7 Å (Table I).

Figure 4 provides an overview of the distribution of different types of side chains in the solution structure of the activation domain B. In Figure 4(A) the hydrophobic residues are shown with heavy lines and the polar side chains with medium lines. A hydrophobic core in the interior of the protein is clearly apparent, and the polar residues are mostly

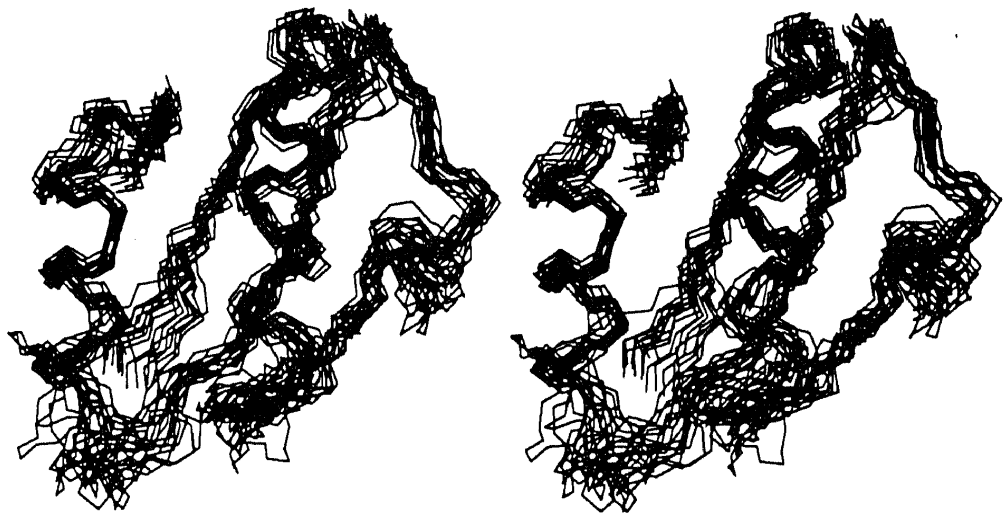

Fig. 3. Stereo view of the superposition of the polypeptide backbone from residues 11–76 in the 20 conformers of the activation domain B used to represent the solution structure. The structures 2–20 were superimposed with structure 1 for pairwise minimal RMSD of the backbone atoms N, C^{α} and C′ of the residues 11–31, 36–38, 44–55 and 62–76 (see text for the selection of these residues). The orientation of the molecule is the same as in Figure 2.

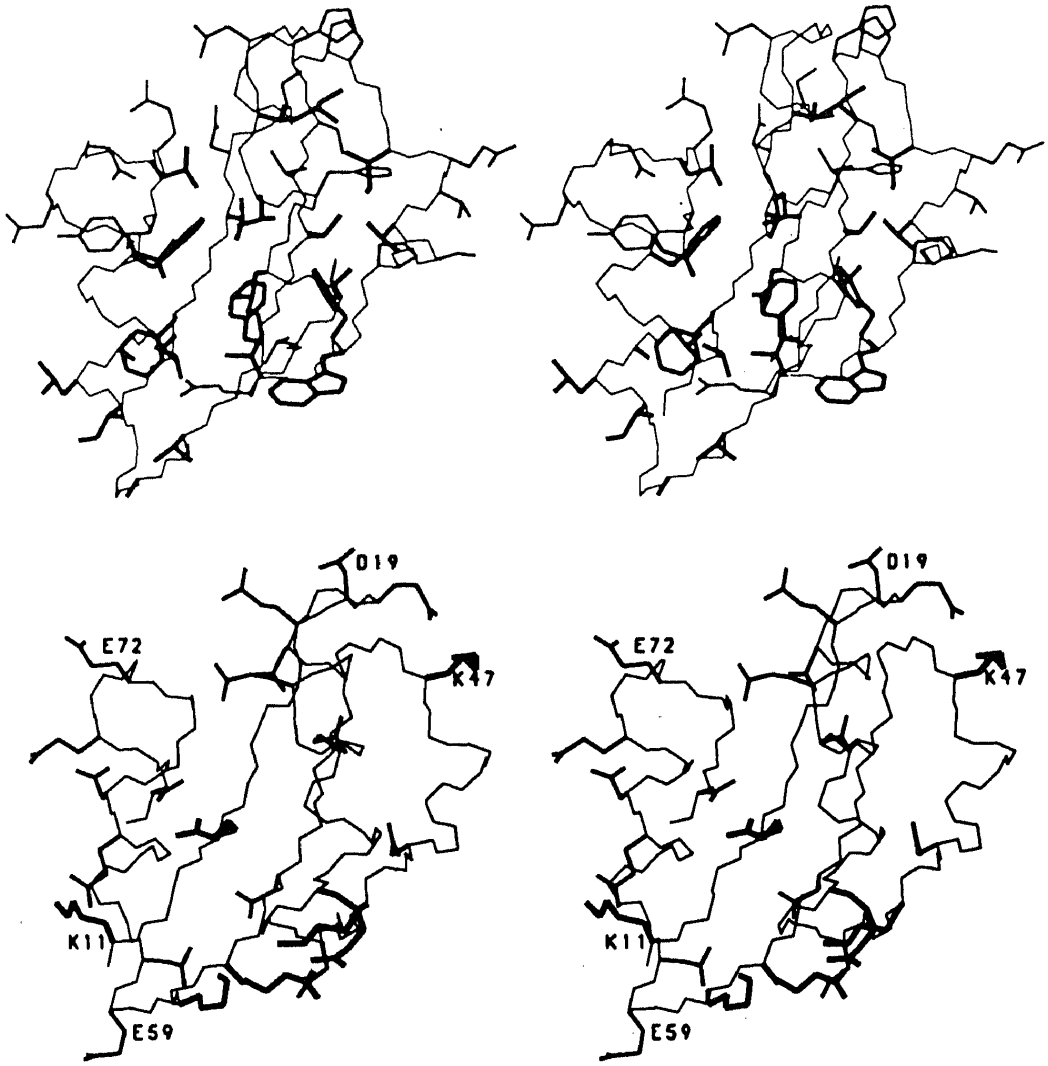

Fig. 4. Stereo view of the activation domain B illustrating the distribution of different types of residues (same orientation as in Figure 2). (**A**) Side chains of hydrophobic residues are displayed with thick lines, side chains of polar uncharged residues with medium lines. (**B**) Side chains of basic residues are displayed with thick lines, acidic side chains with medium lines. In both parts of the figure the backbone is drawn with thin lines, and some residues are identified by the one-letter symbols and the sequence positions.

near the surface. There is a hydrophobic patch located between the two α-helices that is probably partially exposed to the solvent (front of Figure 4(A)). Figure 4(B) shows that the charged groups are on or near the surface. It also gives a visual impression of the fact that the activation domain is an acidic protein, with a considerably larger number of acidic side chains (medium lines) compared to basic ones (heavy lines). Both types of charged residues are not evenly distributed. Acidic side chains are concentrated in the helices, whereas basic side chains are mostly found in or near the β-sheet. Overall, there is clearly an excess of negative charges in the upper left region, whereas a more even charge distribution prevails elsewhere in the molecule.

Discussion

The structure of the activation domain B determined here confirms earlier conclusions from biochemical evidence (Avilés *et al.*, 1982) and a preliminary analysis of NMR data (Vendrell *et al.*, 1990c) that this polypeptide forms a globular conformation. From inspection of the Figures 2–4 the following molecular architecture is apparent: a four-stranded antiparallel β-sheet forms the bottom of the molecule, and three antiparallel helical segments form a cover. In between these two walls a group of predominantly hydrophobic side chains form the β-strands and the helices interdigitate to form the molecular core (Figure 4A). Three loops connecting β-strands and helical secondary structure are flexibly disordered. Two of these loops are at the lower edge and the third one is on the right edge of the β-sheet in the orientation of the activation domain B shown in Figures 2–4. One of these loops includes as *cis* peptide bond between Lys39 and Pro40; a second Lys–Pro segment in positions 47 and 48 forms a *trans* peptide bond and is part of a well-defined loop at the top of the molecule (Figures 2 and 4B). Negatively charged amino acid side chains are crowded in the two α-helices, whereas the β-sheet and the flexible loops in the lower right have a more even distribution of charged, polar and hydrophobic side chains [Figure 4(A) and (B)].

A discussion on functional implications of the solution structure of the activation domain B is somewhat hypothetical, given the fact that in the natural state this polypeptide is not found in the isolated form but as part of the procarboxypeptidase B. A future detailed comparison with the corresponding structure contained in the crystal structure of the intact procarboxypeptidase B (Coll *et al.*, 1990) will be of keen interest with regard to a more profound insight into the protein–protein interactions regulating the

enzymatic activity. On the basis of the structure of the activation domain alone, the negatively charged surface formed by the two α-helices (Figure 4B) may be excluded as a possible site for interdomain interactions with the active enzyme, since both the activation domain and the enzyme are acidic proteins. In contrast, the outer surface of the β-sheet and the flexible loops have a normal distribution of side chains for a molecular region involved in interdomain contacts. A well-ordered structure for the flexible lower right edge of the molecule (Figure 2) might therefore be induced by contacts with the other parts of the intact procarboxypeptidase.

Materials and methods

Procarboxypeptidase B was isolated from porcine pancreatic glands as described before (Vilanova *et al.*, 1985). The purification of the activation domain B, the solution conditions and the preparation of the protein samples for the NMR measurements in H_2O and ^2H_2O were described previously (Vendrell *et al.*, 1990c). The data used as input for the structure calculations were obtained from two NOESY experiments recorded with a mixing time of 60 ms at 15°C, in a mixed solvent of 90% H_2O/10% ^2H_2O, and in 100% ^2H_2O, respectively. Details of data acquisition and processing have been described elsewhere (Vendrell *et al.*, 1990c).

The input for the structure calculations was prepared by well-established procedures. For the calibration of the NOEs to obtain upper limits for proton–proton distances, a similar approach to that described for the determination of the structure of the *Antp (C39 – S)* homeodomain (Güntert *et al.*, 1991b) was used, i.e. the experimental NOE intensities were fitted to empirical relations between NOEs and $^1H-^1H$ distances for the different types of protons. Supplementary conformational constraints were included in the form of upper and lower distance limits (Williamson *et al.*, 1985; Wüthrich, 1986) describing the 23 H-bonds previously identified in the secondary structure (Vendrell *et al.*, 1990c) From an initial set of distance constraints we determined stereospecific assignments of those β-methylene groups for which sufficient local NOE distance constraints had been observed, using the program HABAS (Güntert *et al.*, 1989). HABAS also established limited allowed ranges for certain dihedral angles ϕ, ψ and χ^1 (no measurements of spin–spin coupling constants were obtained because of the large line widths). Individual proton assignments were also determined for the β-methylene protons of Pro40 and Pro48, based on the different intraresidual NOE cross-peak intensities with $C^{\alpha}H$, and for 9 out of a total of 10 side chain amino groups of Asn and Gln, based on intraresidual NOEs with the nearest methylene group. A first round of structure calculation with the program DIANA (Güntert *et al.*, 1991a) produced a group of preliminary structures, which were used to identify additional NOE distance constraints by resolving ambiguities in the assignments of NOESY cross-peaks, and to determine additional stereospecific assignments of methylene and isopropyl groups with the program GLOMSA (Güntert *et al.*, 1991a). The group of 20 structures characterized by Table I and Figure 3 was calculated with the resulting final set of NOE distance constraints and without inclusion of H-bond constraints.

For the final structure calculations, two of the preliminary structures obtained with DIANA were energy minimized. Next, each of the two structures was submitted to rapid heating to 2000°C and kept at this temperature for 40 ps, using the molecular dynamics program X-PLOR (Brünger *et al.*, 1987). During this time period, a total of 10 conformers were collected in each of the two calculations, taken as snapshots at 4 ps intervals. Each of the 20 conformers thus generated was carefully cooled down and energy minimized. For these calculations the energy potential of X-PLOR had been augmented by a potential enforcing the NOE distance constraints. Since we had not used this protocol for structure calculations in earlier work, we generated a group of 10 energy-minimized DIANA structures from different starting conformations (see, for example, Güntert *et al.*, 1991b). The molecular geometry of these conformers coincides closely with those obtained with the procedures described in this work, but the conformational energies of around −759 kcal/mol are higher than those from the molecular dynamics treatment, which are in the range −1248 to −1539 kcal/mol for the 20 conformers of Figure 3.

Acknowledgements

We thank the ETH Zürich for the use of a Cray X-MP/28. We thank Mrs E.Huber for her careful processing of the manuscript. Financial support by the Schweizerischer Nationalfonds (project 31.25174.88), CICYT (Ministerio de Educación y Ciencia, Spain, Project BIO88/0456) and EMBO (long-term fellowship to J.V.) are gratefully acknowledged.

References

Avilés,F.X., San Segundo,B., Vilanova,M., Cuchillo,C.M. and Turner,C. (1982) *FEBS Lett.*, **149**, 257–260.
Billeter,M., Qian,Y.Q., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1990) *J. Mol. Biol.*, **214**, 183–197.
Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) *Science*, **235**, 458–460.
Christianson,D.W. and Lipscomb,W.N. (1989) *Accts. Chem. Res.*, **22**, 62–69.
Clauser,E., Gardell,S.J., Craik,C.S., MacDonald,R.J. and Rutter,W.J. (1988) *J. Biol. Chem.*, **263**, 17837–17845.
Coll,M., Guasch,A., Avilés,F.X. and Huber,R. (1991) *EMBO J.*, **10**, 1–9.
Flogizzo,E., Bonicel,J., Kerfelec,B., Granon,S. and Chapus,C. (1988) *Biochim. Biophys. Acta*, **954**, 183–188.
Güntert,P., Braun,W., Billeter,M. and Wüthrich,K. (1989) *J. Am. Chem. Soc.*, **111**, 3997–4004.
Güntert,P., Qian,Y.Q., Otting,G., Müller,M., Gehring,W. and Wüthrich,K. (1991a) *J. Mol. Biol.*, in press.
Güntert,P., Braun,W. and Wüthrich,K. (1991b) *J. Mol. Biol.*, in press.
Quinto,C., Quiroga,M., Swain,W.F., Nikovits,W.C.Jr., Standring,D.N., Pictet,R.L., Valenzuela,P. and Rutter,W.J. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 31–35.
Sánchez-Ruiz,J.M., López-Lacomba,J.L., Mateo,P.L., Vilanova,M., Serra,M.A. and Avilés,F.X. (1988) *Eur. J. Biochem.*, **176**, 225–230.
San Segundo,B., Martínez,M.C., Vilanova,M., Cuchillo,C.M. and Avilés,F.X. (1982) *Biochim. Biophys. Acta*, **707**, 74–80.
Schmid,M.F. and Herriot,J.R. (1976) *J. Mol. Biol.*, **103**, 175–190.
Vendrell,J., Avilés,F.X., Genescà,E., San Segundo,B., Soriano,F. and Méndez,E. (1986) *Biochem. Biophys. Res. Commun.*, **141**, 517–523.
Vendrell,J., Vilanova,M., Avilés,F.X., Turner,C.H., Cary,P.D. and Crane-Robinson,C. (1990a) *Biochem. J.*, **267**, 213–220.
Vendrell,J., Cuchillo,C.M. and Avilés,F.X. (1990b) *J. Biol. Chem.*, **265**, 6949–6953.
Vendrell,J., Wider,G., Avilés,F.X. and Wüthrich,K. (1990c) *Biochemistry*, **29**, 7515–7522.
Vilanova,M., Vendrell,J., Lopez,M.T., Cuchillo,C.M. and Avilés,F.X. (1985) *Biochem. J.*, **229**, 605–609.
Wade,R.D., Hass,G.M., Kumar,S., Walsh,K. and Neurath,H. (1988) *Biochimie*, **70**, 1137–1142.
Widmer,H., Billeter,M. and Wüthrich,K. (1989) *Proteins*, **6**, 357–371.
Williamson,M.P., Havel,T.F. and Wüthrich,K. (1985) *J. Mol. Biol.*, **182**, 295–315.
Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.
Wüthrich,K. (1989) *Science*, **243**, 45–50.
Wüthrich,K., Billeter,M. and Braun,W. (1984) *J. Mol. Biol.*, **180**, 715–740.

Received September 11, 1990, revised November 2, 1990