

J-Bio NMR 050

## Comparison of the NMR solution structure with the X-ray crystal structure of the activation domain from procarboxypeptidase B

M. Billeter<sup>a</sup>, J. Vendrell<sup>b</sup>, G. Wider<sup>a</sup>, F.X. Avilés<sup>b</sup>, M. Coll<sup>c,d</sup>, A. Guasch<sup>b,c</sup>, R. Huber<sup>c</sup>  
and K. Wüthrich<sup>a,\*</sup>

<sup>a</sup>*Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg,  
CH-8093 Zürich, Switzerland*

<sup>b</sup>*Departament de Bioquímica i Biologia Molecular, Faculta de Ciències, Universitat Autònoma de Barcelona,  
E-08193 Bellaterra (Barcelona), Spain*

<sup>c</sup>*Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, Germany*

<sup>d</sup>*Unitat de Química Macromolecular (CSIC) ETS, Enginyers Industrials, Diagonal 647, E-08028 Barcelona, Spain*

Received 5 October 1991

Accepted 3 December 1991

**Keywords:** Protein structure; NMR solution structure; X-ray crystal structure; Procarboxypeptidase B; Activation domain B

---

### SUMMARY

The NMR solution structure of the activation domain isolated from porcine procarboxypeptidase B is compared with the X-ray crystal structure of the corresponding segment in the intact proenzyme. For the region of the polypeptide chain that has a well-defined three-dimensional structure in solution, i.e., the backbone atoms of residues 11-76 and 25 amino acid side chains in this segment that form a hydrophobic core in the activation domain, the root-mean-square distance between the two structures is 1.1 Å. There are no significant differences in average atom positions between the two structures, but only the NMR structure shows increased structural disorder in three outlying loops located along the same edge of the activation domain. These regions of increased structural disorder in the free domain coincide only partially with the interface to the enzyme domain in the proenzyme.

---

### INTRODUCTION

Pancreatic carboxypeptidases are digestive enzymes that hydrolyse the C-terminal peptide bond of polypeptide chains. While carboxypeptidase A shows moderate specificity for aromatic and aliphatic residues, carboxypeptidase B releases C-terminal arginines and lysines. The activa-

---

\* To whom correspondence should be addressed.

*Abbreviations:* RMSD, root-mean-square distance; ADB, activation domain of procarboxypeptidase B.

tion of the zymogen precursor, procarboxypeptidase B, occurs by tryptic removal of a 95-residue N-terminal activation segment (Clauser et al., 1988; Burgos et al., 1991). This segment is proteolytically degraded to a more stable species, i.e., the 81-residue activation domain B (ADB). A structure determination by NMR (Wüthrich, 1986,1990) has shown that in free ADB, in solution, the residues 11–76 form a stable globular fold (Vendrell et al., 1990,1991). The structure of the uncleaved procarboxypeptidase B was determined in single crystals with a resolution of 2.3 Å (Coll et al., 1991). In this structure the residues 7–95 of the activation segment are well characterized, while the N-terminal six residues are disordered and were not observed with the X-ray techniques. In the present paper, the solution structure for residues 11–76 is compared with the structure of the corresponding polypeptide segment in the intact proenzyme in single crystals. Compared to previous comparisons of corresponding or homologous globular protein structures determined by NMR in solution and by X-ray diffraction in single crystals (e.g., Billeter et al., 1989; Clore et al., 1987; Clore et al., 1990; Neri et al., 1992; Wagner et al., 1987; Williamson et al., 1985) this report is less focused on the assessment of differences in protein structure descriptions derived with the two methods. Instead, the main emphasis is on a comparative view of the structural properties of the surface of the activation domain B that contacts the rest of the proenzyme, and an analysis with regard to possible clues about the molecular interactions controlling the enzymatic activity.

## METHODS

Differences between pairs of three-dimensional structures of the same polypeptide chain were quantified by calculating mean deviations between the structures for various subsets of atoms using the following expressions:

$$A = [1/m \sum_{k=1}^m |r_k - R r'_k|^2]^{1/2} \quad (1)$$

$R$  is the rotation matrix for which the sum  $\sum_{l=1}^n |r_l - R r'_l|^2$  is minimal. The indices  $k = 1 \dots m$  and  $l = 1 \dots n$ , define two subsets of atoms that occur in both molecules.  $r_i$  and  $r'_i$  are the position vectors of corresponding atoms in the two molecules after these have been translated so that the average position of the atoms  $l = 1 \dots n$  coincides with the origin of the coordinate system. If the subset of atoms  $k = 1 \dots m$  is equal to the subset  $l = 1 \dots n$ , then  $A$  is the usual RMSD value for the selected subset of atoms (McLachlan, 1979). Equation (1) with different subsets  $k = 1 \dots m$  and  $l = 1 \dots n$  is typically used to measure the variations in the amino acid side-chain orientations after superposition of selected backbone fragments, and  $A$  is then referred to as the *displacement*,  $D$  (Billeter et al., 1989).

Graphically, circles around the  $C^\alpha$  atoms oriented perpendicular to the direction of the backbone are used to represent the precision with which the structure of the polypeptide backbone is locally determined (Billeter et al., 1989). For NMR solution structures, the radius of the circle about  $C^\alpha$  of a given residue is equal to the average of the displacements,  $D$ , of the atoms N,  $C^\alpha$  and  $C'$  of this residue in the 20 NMR conformers relative to the mean coordinates after global superposition of the backbone atoms of residues 11–76 of the 20 conformers on their mean coordinates. For the crystal structure only a single average B-factor for the backbone atoms of each residue was obtained during the refinement. These values are converted to mean expected deviations,

$\langle \Delta x^2 \rangle^{1/2}$ , of the atom positions around their equilibrium values using the following relation (Glusker and Trueblood, 1985):

$$B = \frac{8\pi^2}{3} \langle \Delta x^2 \rangle \quad (2)$$

The value  $\langle \Delta x^2 \rangle^{1/2}$  thus obtained for each residue is used as the radius of the circle around the C $^\alpha$  atom of this residue.

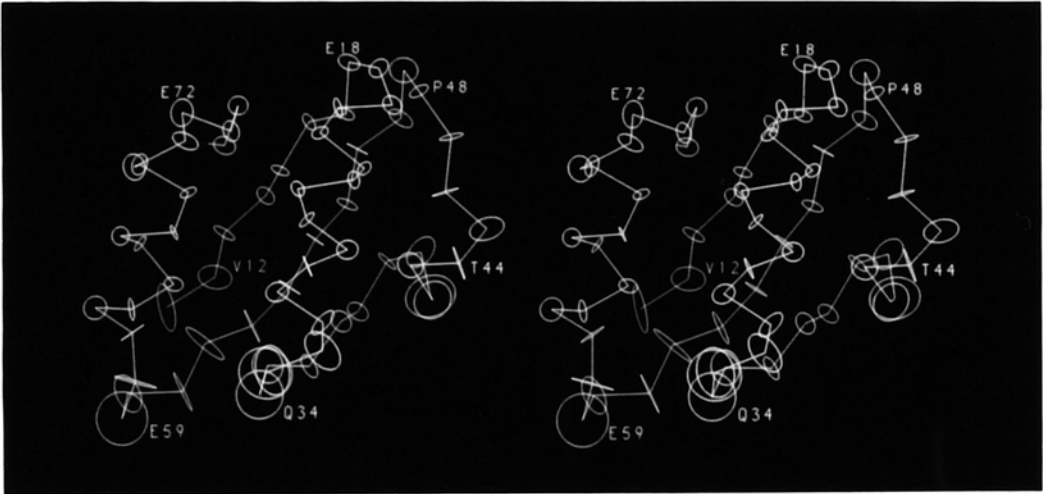
## RESULTS AND DISCUSSION

The NMR solution structure of the N-terminal 81-residue polypeptide segment obtained as a final product of the proteolytic activation of procarboxypeptidase B (Vendrell et al., 1991) has two unstructured, flexible chain ends with residues 1–10 and 77–81. The structured part of the molecule starts and ends with the first and the last strand of a four-stranded  $\beta$ -sheet consisting of residues 11–17, 36–39, 50–56 and 75–76. It further contains two  $\alpha$ -helices with residues 20–30 and 60–70, and a  $_3$ 10-helix with residues 43–46. The backbone fold of the NMR structure is shown in Fig. 1A, where the straight lines connect the average C $^\alpha$  positions calculated from the group of 20 conformers used to represent the solution conformation (Vendrell et al., 1991). The radii of the circles represent the displacements,  $D$ , of the 20 NMR conformers around their averaged coordinates (see Methods section). Figure 1A shows that there is significantly increased structural disorder in three loops connecting helical secondary structures with  $\beta$ -strands. These three loops with residues 32–35, 39–43 and 56–61 are located at the bottom of the molecule in the orientation of Fig. 1. The X-ray crystal structure of the ADB used for the present comparison is part of the structure determined for the intact procarboxypeptidase B (Coll et al., 1991). In addition to the ADB this structure includes a segment with residues 82–95 and a domain with 306 residues, which corresponds to the enzymatically active carboxypeptidase B obtained after the proteolytic activation. It has previously been established that there is an empirical, qualitative correlation between the structural disorder indicated by these local displacements in the NMR structure and that implicated by the atomic B-factors in the corresponding X-ray crystal structure (e.g., Billeter et al., 1989; Wüthrich, 1990). Figure 1B shows the C $^\alpha$  positions of residues 11–76 of the crystal structure with circles indicating the mean expected atomic deviations  $\langle \Delta x^2 \rangle^{1/2}$ . These were obtained by using the mean B-factors for the N, C $^\alpha$  and C $^\beta$  atoms of each residue in Eq. (2) to define the radii  $\langle \Delta x^2 \rangle^{1/2}$  of the circles around the C $^\alpha$  atoms. In contrast to the solution structure (Fig. 1A) only slightly increased structural disorder is observed for the loops at the bottom of the molecule, although otherwise the variation of local variability along the polypeptide chain is very similar in the two structures (Figs. 1A,B).

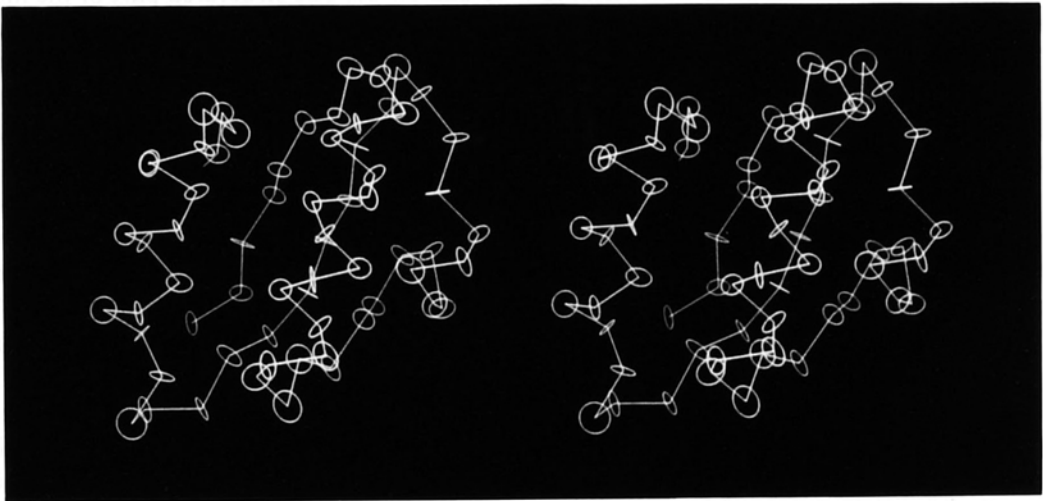
### *Comparison of the polypeptide backbone*

The figures 1A and B show that the polypeptide fold of the ADB free in solution and in the crystal structure of intact procarboxypeptidase B is very similar, the only readily apparent difference being the increased structural disorder for the three loops at the bottom of the molecule which is seen only in the NMR structure. This first visual impression is confirmed by Fig. 1C,

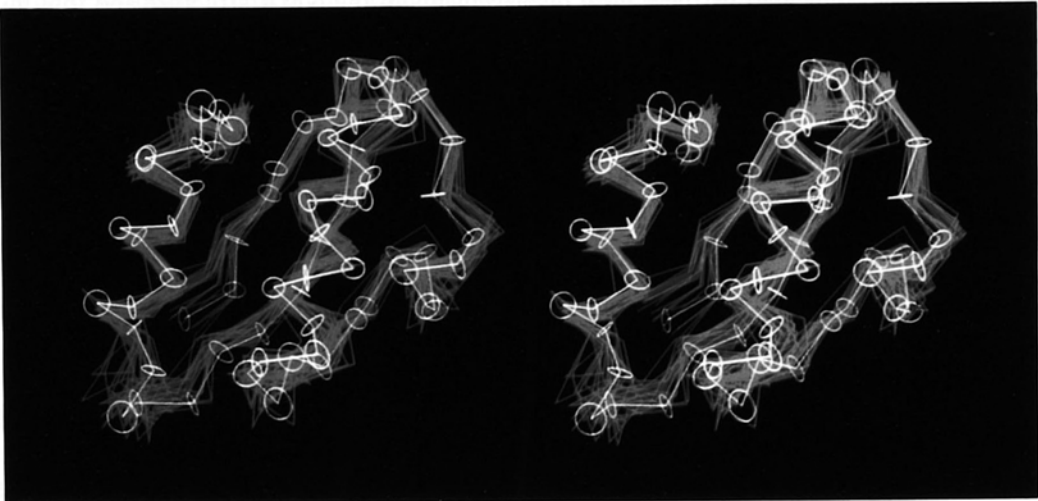
A



B



C



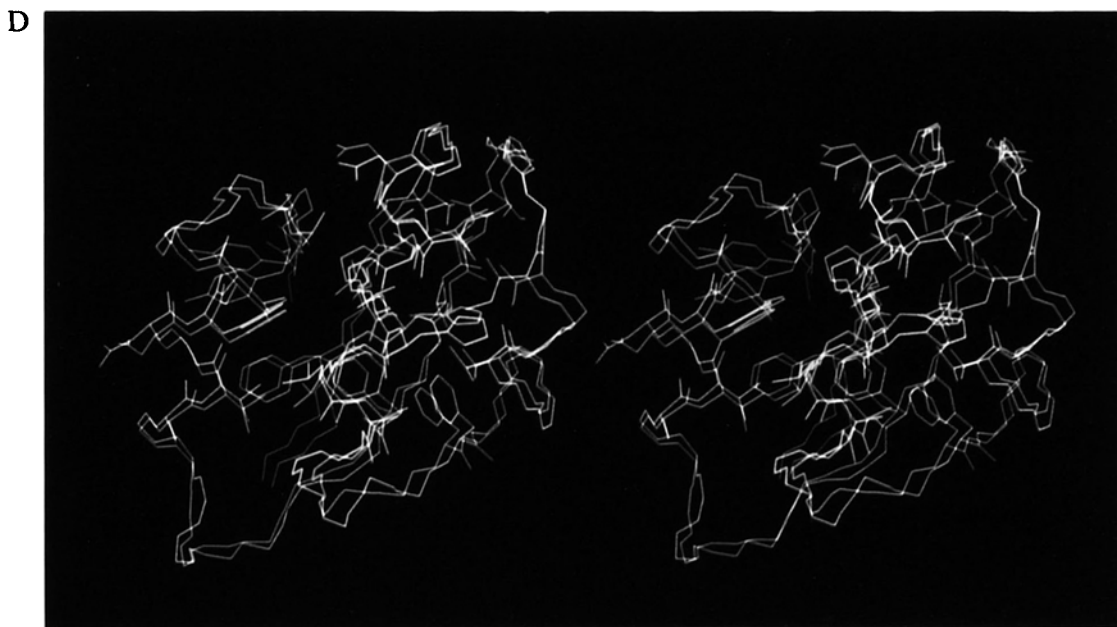


Fig. 1. Comparison of the NMR solution structure of the isolated activation domain B with the corresponding polypeptide segment in the X-ray crystal structure of intact procarboxypeptidase B. (A) Blue:  $C^\alpha$  atom positions of the averaged coordinates of the NMR structure connected by virtual  $C^\alpha$ - $C^\alpha$  bonds. The circles are displacements of the backbone atoms of each residue between the averaged coordinates and the 20 NMR conformers used to represent the solution structure. Some residues have been identified with the one-letter amino acid code and the sequence position. (B) Yellow: Same presentation of the  $C^\alpha$  atoms of the crystal structure as in (A), with circles indicating the mean expected deviation of the atomic positions. (See the text for a detailed description of the calculation of the circles in A and B.) (C) Global superposition using the backbone atoms N,  $C^\alpha$  and  $C'$  of residues 11–76 of the 20 NMR conformers (blue) on the crystal structure (yellow). The latter is displayed with yellow circles corresponding to those in (B). (D) Stereo view of the superposition of the averaged coordinates of the 20 NMR conformers (blue) and the crystal structure (yellow). The backbone atoms N,  $C^\alpha$  and  $C'$  of residues 11–76 and the 25 core side chains listed in the footnotes to Table 1 (Vendrell et al., 1991) are shown. The same orientation of the molecule is shown in (A)–(D).

which shows a superposition of the 20 NMR conformers and the crystal structure, and by the data in Figs. 2A and B, and Tables 1 and 2.

The RMSD values in Table 1 show that the difference between the averaged coordinates of the 20 NMR conformers and the crystal structure is the same as that between the individual NMR conformers and their averaged coordinates, namely 1.0 Å for the backbone atoms of residues 11–76. Table 2 shows that 25 hydrogen bonds were identified in both structures, 23 of which are located in regular secondary structures. In the crystal structure, 12 additional hydrogen bonds were assigned, 6 of which are part of regular secondary structures. In the NMR structure there is one additional  $\alpha$ -helix hydrogen bond that was not identified in the X-ray structure. Overall, although a significantly larger number of hydrogen bonds was identified in the crystal structure, there is good agreement between the patterns of hydrogen bonds in the two structures.

A local comparison of the two structures on the level of corresponding individual residues is afforded by Fig. 2, where the ranges of the dihedral angles  $\phi$  (Fig. 2A) and  $\psi$  (Fig. 2B) observed in

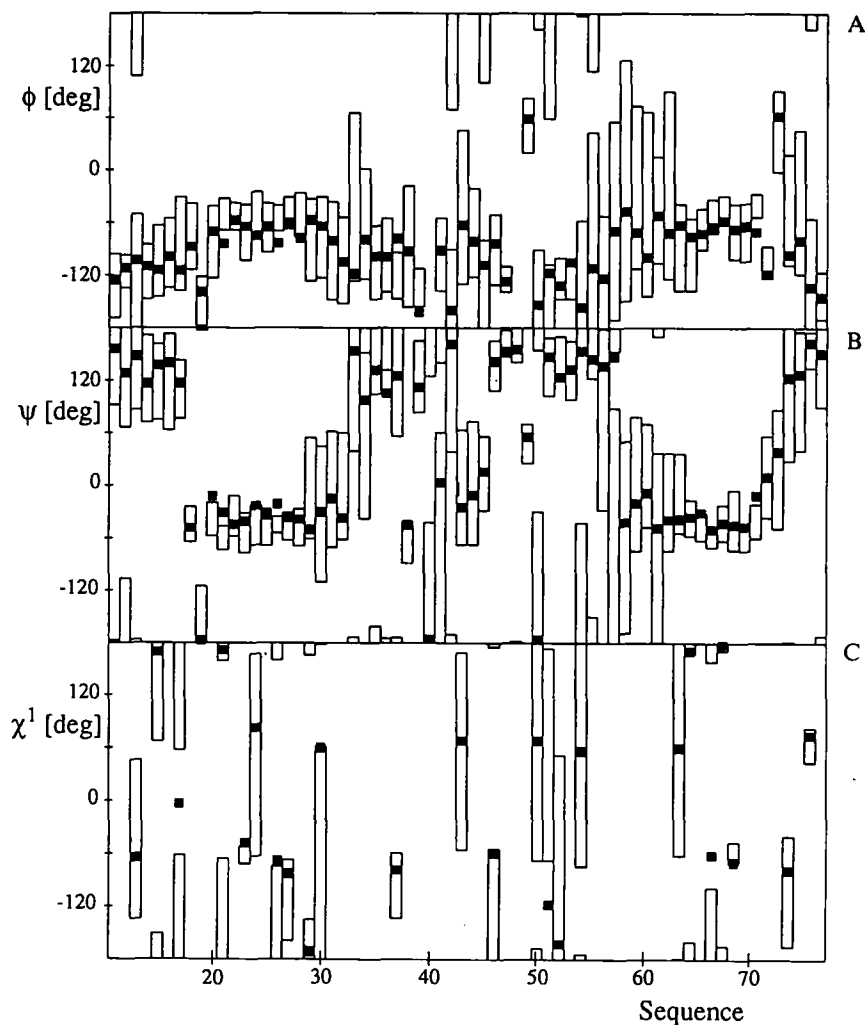


Fig. 2. In plots vs. the amino acid sequence, the ranges of values observed among the 20 NMR conformers for the dihedral angles  $\phi$ ,  $\psi$  and  $\chi^1$  are represented by vertical bars for each residue. The black squares indicate the values of the corresponding dihedral angles in the crystal structure. For  $\chi^1$  the plot includes only data for 24 core side chains (all except Pro<sup>48</sup>) described for the solution structure (Vendrell et al., 1991; see the footnote to Table I for a list of these side chains).

TABLE I

RMSD COMPARISONS BETWEEN THE NMR SOLUTION STRUCTURE OF THE ISOLATED ACTIVATION DOMAIN B AND THE ADB IN THE X-RAY CRYSTAL STRUCTURE OF THE INTACT PROCARBOXYPEPTIDASE B

Atoms used for the comparison	<NMR> <sup>a</sup> -X-ray (Å)	<NMR> <sup>a</sup> -NMR (Å)
N, C <sup>α</sup> , C', residues 11-76	1.0	1.0
Same + core side chains <sup>b</sup>	1.1	1.1

<sup>a</sup> <NMR> denotes the averaged coordinates of the 20 conformers that represent the solution structure of ADB.

<sup>b</sup> The core side chains are those of the following 25 residues: 13, 15, 17, 21, 23, 24, 26, 27, 29, 30, 37, 43, 46, 48, 50-52, 54, 63, 64, 66-68, 73, 75 (see Vendrell et al., 1991).

TABLE 2  
 HYDROGEN BONDS BETWEEN BACKBONE ATOMS IN THE NMR SOLUTION STRUCTURE OF THE ACTIVATION DOMAIN B AND THE CORRESPONDING POLYPEPTIDE SEGMENT IN THE X-RAY CRYSTAL STRUCTURE OF PROCARBOXYPEPTIDASE B<sup>a</sup>

Amide group	Carbonyl group	X-ray <sup>b</sup>	NMR <sup>c</sup>	Secondary structure <sup>d</sup>
13 Phe	54 Phe	+	10	β
14 Arg	76 Glu	+	14	β
15 Val	52 Val	+	16	β
16 Asn	74 Gln	+	17	β
17 Val	50 Ser	+	11	β
23 Ile	19 Asp	+		
24 Ser	20 Glu	+	20	α
25 Glu	21 Asn	+	19	α
26 Leu	22 Asp	+	19	α
27 His	23 Ile	+	20	α
28 Glu	24 Ser	+	11	α
29 Leu	25 Glu	+		α
30 Ala	26 Leu	+		α
30 Ala	27 His		10	α
32 Thr	29 Leu	+	11	
33 Arg	29 Leu	+		
36 Asp	55 Arg	+		β
38 Trp	53 Asp	+	17	β
39 Lys	53 Asp	+		
45 Gln	42 Ser	+		
50 Ser	17 Val	+		β
52 Val	15 Val	+	17	β
53 Asp	39 Lys	+	19	β
54 Phe	13 Phe	+	11	β
55 Arg	36 Asp	+	10	β
56 Val	11 Lys	+		β
60 Asp	57 Lys	+		
64 Val	60 Asp	+		α
65 Glu	61 Ile	+	16	α
66 Asp	62 Leu	+	16	α
67 Phe	63 Ala	+	16	α
68 Leu	64 Val	+	16	α
69 Glu	65 Glu	+	15	α
70 Gln	66 Asp	+	18	α
71 Asn	68 Leu	+		
72 Glu	69 Glu	+	17	
73 Leu	68 Leu	+	16	
76 Glu	14 Arg	+	20	β

<sup>a</sup> Hydrogen bonds are identified by the following criteria: the distance between the proton and the acceptor atom must be smaller than 2.4 Å, and the angle between the bond from the donor heavy atom to the hydrogen atom and a straight line connecting the donor heavy atom with the acceptor atom must be smaller than 35°.

<sup>b</sup> Hydrogen atoms were added to the crystal structure using standard bond lengths and bond angles. The identification of a hydrogen bond in the crystal structure is indicated by a '+' sign.

<sup>c</sup> The number of NMR conformers, for which the hydrogen bond was observed, is given. Only those hydrogen bonds are listed that occur in 10 or more out of the 20 conformers used to represent the solution structure.

<sup>d</sup> Both the donor group and the acceptor group of the hydrogen bond belong to residues for which the regular secondary structure indicated was attributed in Vendrell et al. (1991).

the 20 NMR conformers are given for each residue together with the values adopted in the crystal structure. Because the coupling constants  $J_{\text{HN}\alpha}$  could for technical reasons not be measured in the determination of the solution structure (Vendrell et al., 1991), many dihedral angles show a rather large spread of values among the NMR conformers. For all but six  $\phi$  angles and eight  $\psi$  angles, the values adopted in the crystal structure lie within the range of values observed in the NMR conformers. The maximal deviations of the values for the crystal structure from the range of values for the NMR conformers is  $17^\circ$  for the  $\phi$  angles and  $26^\circ$  for the  $\psi$  angles. In three cases the differences in the  $\psi$  angles are compensated by differences with opposite sign in the  $\phi$  angles of the following residues, resulting in different orientations of the intervening peptide planes.

Overall, within the precision of the two structure determinations (Figs. 1A,B), there are no significant deviations of average atom positions between the ADB polypeptide backbone fold in solution and in the crystal structure of procarboxypeptidase B (Fig. 1C). As was mentioned above the increase of flexible structural disorder in the three loops of residues 32–35, 39–43 and 56–61 in the NMR structure has no parallel in the X-ray structure, although the B-factors are somewhat increased for the residues 32–35, 43, 56 and 59. Otherwise, the variation of local disorder along the polypeptide chain is very similar; for example, in both structures the two helices with residues 20–30 and 60–70 are particularly well defined, and both structures show somewhat increased disorder in the loop of residues 71–74.

#### *Comparison of the amino acid side chains*

The architecture of independently folding protein domains is usually stabilized by the formation of a hydrophobic core, with numerous side chains of internal residues having particularly well-defined conformations. Vendrell et al. (1991) have identified 25 ‘best-defined’ core residues in the solution structure of ADB. Table 1 and Fig. 1D show that after a global polypeptide backbone superposition, the conformations of these side chains are also in good agreement between solution and crystal structure. In Fig. 2C the  $\chi^1$  dihedral angles are compared for these same core

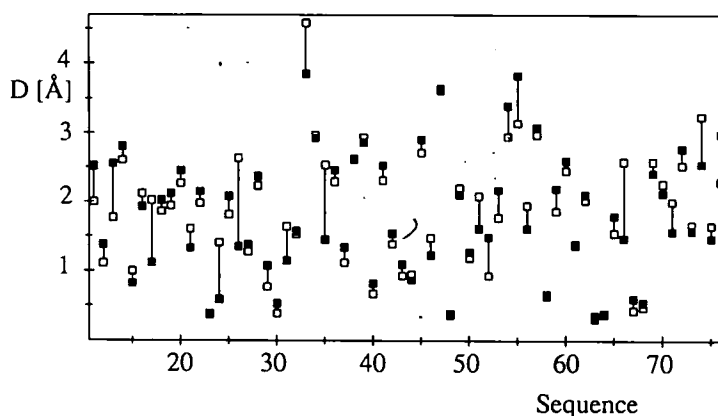


Fig. 3. Plots versus the amino acid sequence of the average displacements,  $D$ , within the group of 20 NMR conformers (black squares) and between the NMR conformers and the crystal structure (open squares). After local superposition of the backbone atoms N, C $^\alpha$  and C' of successive tripeptide segments, Eq. (1) was used to calculate the displacements for the side chain heavy atoms of the central residue.



residues. For all but three residues the crystal structure values fall within or very near the range of values covered by the 20 NMR conformers. The three exceptions are Val<sup>17</sup>, Thr<sup>51</sup> and Asp<sup>66</sup>, where the  $\chi^1$  range for residues 17 and 51 is very wide and thus hardly defined in the NMR structure. In Fig. 3, the comparison has been extended to all amino acid side chains. In this presentation, average displacements,  $D$ , of the side chain heavy atoms after local superposition of the backbone atoms N, C $^\alpha$  and C' of tripeptide segments are plotted for comparisons of the 20 NMR conformers (black squares), and comparisons of the 20 NMR conformers with the crystal structure (open squares). Therefore, significant structural differences are indicated for those residues where the open square is well above the black square. Overall, this figure shows that there is good agreement also for numerous side chains located outside of the core of the protein domain, although differences exceeding 0.5 Å are implicated for the residues 17, 24, 26, 33, 35, 66 and 74, where Arg<sup>33</sup> is, however, very poorly defined in both the NMR and the X-ray structure.

## CONCLUSIONS

In contrast to the comparisons between corresponding protein structures in solution and in crystals mentioned in the Introduction, the main interest of the present work is focused on the structure and function of the different species encountered in the carboxypeptidase B system rather than on the different methods used for the structure determinations. In this respect it is of keen interest that the isolated activation domain forms a stable three-dimensional structure of which

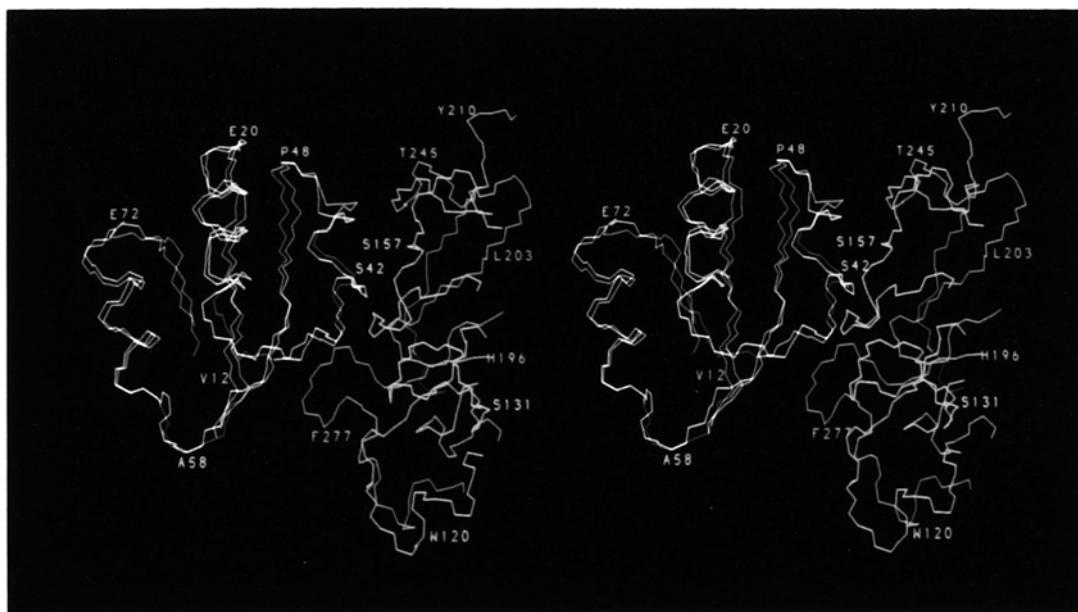


Fig. 4. Stereo plot showing the contacts of the activation domain B with the enzyme domain in the structure of pro-carboxypeptidase B. The polypeptide backbone of the mean of the 20 conformers used to represent the ADB solution conformation is drawn with a blue line, except that the less well-defined regions of residues 32–35, 39–43 and 56–61 are drawn in magenta. The crystal structure of the ADB in the proenzyme is drawn in yellow and the contact region of the enzyme domain is represented with green lines. Some residues are labeled with the one-letter amino acid code and the sequence position, where blue labels refer to residues of the ADB and green labels to residues of the enzyme domain.

the backbone fold and the core side chains are virtually identical to the ADB in the intact procarboxypeptidase B.

Figure 4 shows the contacts of the ADB and the enzyme domain in the intact procarboxypeptidase B. One of the less well-defined loops in the solution structure of ADB, residues 39–43, is in contact with the enzyme domain of the zymogen and Asp<sup>41</sup> forms a salt bridge with Arg<sup>145</sup> in the carboxypeptidase B domain. This interaction, which tightens the conformation of the side chain of Asp<sup>41</sup>, was considered to be an explanation for the complete lack of activity of the zymogen versus small molecular weight substrates (Coll et al., 1991), since Arg<sup>145</sup> binds the C-terminal carboxyl group of substrates in the active carboxypeptidase. It is further seen that none of the 25 best-defined core side chains of the ADB solution structure listed in a footnote to Table 1 is involved in the contacts with the enzyme domain, which confirms again that the ADB forms a self-contained protein domain. Overall, in concert with the observations by Coll et al. (1991), Fig. 4 indicates that the contacts between the two domains are not very tight. Although the ADB shields the active site from access of substrates, it does not bind tightly to it or even penetrate the other domain. Actually, several water molecules in the domain–domain interface mediate intermolecular side chain–side chain interactions (Coll et al., 1991). This observation then also provides a plausible explanation for the finding that although its three-dimensional structure is preserved after proteolytic cleavage of the procarboxypeptidase B, the isolated ADB is inefficient as an inhibitor of carboxypeptidase B (Burgos et al., 1991). Most contacts between the activation domain and the enzyme are indeed provided by the connecting segment 82–95, which is absent in the free ADB (Coll et al., 1991). This segment, which in contrast to the ADB is susceptible to proteolysis, seems to be required for proper binding of the activation segment 1–95 and the inhibition of the procarboxypeptidase.

## ACKNOWLEDGEMENTS

Financial support by the Schweizerischer Nationalfonds (Project 31.25174.88) and CICYT (Ministerio de Educación y Ciencia, Spain, Project BIO88/0456) is gratefully acknowledged.

## REFERENCES

- Billeter, M., Kline, A.D., Braun, W., Huber, R. and Wüthrich, K. (1989) *J. Mol. Biol.*, **206**, 677–687.
- Burgos, F.J., Salva, M., Villegar, V., Soriano, F., Mendez, E. and Avilés, F.X. (1991) *Biochemistry*, **30**, 4082–4089.
- Clauser, E., Gardell, S.J., Craik, C.S., MacDonald, R.J. and Rutter, W.J. (1988) *J. Biol. Chem.*, **263**, 17837–17845.
- Clore, G.M., Gronenborn, A.M., James, M.N.G., Kjær, M., McPhalen, C.A. and Poulsen, F.M. (1987) *Protein Eng.*, **1**, 313–318.
- Clore, G.M., Driscoll, P.C., Wingfield, P.T. and Gronenborn, A.M. (1990) *J. Mol. Biol.*, **214**, 811–817.
- Coll, M., Guasch, A., Avilés, F.X. and Huber, R. (1991) *EMBO J.*, **10**, 1–9.
- Glusker, J.P. and Trueblood, K.N. (1985) *Crystal Structure Analysis*, Oxford University Press, New York.
- McLachlan, A.D. (1979) *J. Mol. Biol.*, **128**, 49–79.
- Neri, D., Billeter, M. and Wüthrich, K. (1992) *J. Mol. Biol.*, in press.
- Vendrell, J., Wider, G., Avilés, F.X. and Wüthrich, K. (1990) *Biochemistry*, **29**, 7515–7522.
- Vendrell, J., Billeter, M., Wider, G., Avilés, F.X. and Wüthrich, K. (1991) *EMBO J.*, **10**, 11–15.
- Wagner, G., Braun, W., Havel, T.F., Schaumann, Th.M., Gö, N. and Wüthrich, K. (1987) *J. Mol. Biol.*, **196**, 611–639.
- 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. (1990) *J. Biol. Chem.*, **265**, 22059–22062.