

# Comparison of protein structures determined by NMR in solution and by X-ray diffraction in single crystals

---

MARTIN BILLETER

*Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg CH-8093  
Zürich, Switzerland*

---

1. INTRODUCTION	325
2. COMPARISON TOOLS	326
3. DETERMINATION OF PROTEIN STRUCTURES	329
3.1 <i>Determination by NMR in solution</i>	329
3.2 <i>Determination by X-ray diffraction in single crystals</i>	330
4. PROTEINS SELECTED FOR COMPARISON	331
5. INDIVIDUAL COMPARISONS	334
5.1 <i><math>\alpha</math>-Amylase inhibitor (tendamistat)</i>	334
5.2 <i>Barnase</i>	339
5.3 <i>Calbindin</i>	343
5.4 <i>Chymotrypsin inhibitor 2</i>	346
5.5 <i>ColEI rop protein</i>	348
5.6 <i>Interleukin 1<math>\beta</math></i>	350
5.7 <i>Interleukin 8</i>	353
5.8 <i>Metallothionein</i>	357
5.9 <i>434 Repressor(I-69)</i>	360
5.10 <i>Ribonuclease</i>	362
5.11 <i>Thioredoxin</i>	365
5.12 <i>Trypsin inhibitor (BPTI)</i>	367
6. CONCLUSIONS	369
7. ACKNOWLEDGEMENTS	371
8. REFERENCES	371
APPENDIX: CRYSTALLOGRAPHIC PARAMETERS	377

## 1. INTRODUCTION

Following the first determinations of protein structures in the late 1950s and the early 1960s (see for example Kendrew *et al.* 1960; Perutz, 1964), the three-dimensional structures of several hundred proteins have been elucidated by X-ray

diffraction on single crystals. By the end of 1991, approximately 150 entries of proteins with substantially different sequences and a well resolved structure (Hobohm *et al.* 1992) were deposited in the Protein Data Bank (Bernstein *et al.* 1977; Abola *et al.* 1987). In addition, many structures of homologous proteins or of mutants have been described, bringing the total number of entries to about 600. While it was soon accepted that almost all of these structures do indeed give a correct picture of the fold of the active protein in spite of the non-physiological environment of single crystals, it is not clear to what extent structural details are reliably described by these structures. In particular the surface of a protein may be modified due to the dense packing of protein molecules in the crystal lattice. A detailed knowledge of the protein surface is, however, essential for the understanding of the function of the protein.

In the second half of the 1980s, NMR established itself as an alternative method for the determination of protein structures at atomic level, i.e. at a level of structural detail comparable to that of X-ray diffraction studies (in 1984 the structures of two globular proteins were obtained: BUSI (Williamson *et al.* 1985) and *lac* repressor headpiece (Zuiderweg *et al.* 1984; Kaptein *et al.* 1985)). In the past seven years many structures of proteins have been solved by an ever increasing number of NMR techniques (see Wüthrich, 1986, 1989; Clore & Gronenborn, 1991 *a*). By the end of 1991, about 40 protein structures determined by NMR were deposited or announced in the Protein Data Bank. The obvious advantage of the NMR technique is that the experiments are performed in solution, which approximates the conditions in which the protein exhibits its physiological activity better than a single crystal. The major disadvantage of NMR is its limitation to rather small proteins, or complexes with proteins, due to the increase in spectral complexity of large proteins.

In the present study, structures obtained by X-ray diffraction and by NMR were compared with the purpose of evaluating (*a*) the agreement between the two types of structures, (*b*) differences in the structures resulting from the experimental techniques employed as well as from the different tools used to describe the structures, and (*c*) differences due to the environment seen by the protein in the two types of experiments. Therefore, this study was restricted to cases where both an NMR and an X-ray structure are available for the same protein (or, in a few cases, where the two proteins differ minimally, e.g. in their oxidation state). In the following chapters, the comparison tools used in this study are described, some characteristic features of the two structure determination methods are highlighted, a set of proteins is selected for comparison, detailed comparisons are given for these, and the general observations from the comparisons are discussed.

## 2. COMPARISON TOOLS

Solution and crystal structures were compared using both interactive examination of the protein structures on a graphics workstation with the molecular graphics package MidasPlus (Ferrin *et al.* 1988), and quantitative automatic procedures. The latter consisted of measurements of the atomic deviations among the

structures, comparison of the dihedral angles  $\phi$ ,  $\psi$  and  $\chi^1$ , determination of networks of hydrogen bonds, as well as the examination of the crystallographic B-factors and intermolecular contacts in the crystal lattice. While all these comparison tools were applied to each protein, only those quantities are reported in Chapter 5 that best describe the observations for a given protein.

Two different questions were addressed regarding the quantification of atomic deviations. (a) How well can a set of atoms of one structure (e.g. the backbone) be superimposed onto the corresponding set of atoms of another structure? (b) After the superposition of the best defined parts of two structures according to (a), how much do other parts (e.g. the side-chains) differ in their position and orientation? In a first step of answering both questions, a subset of the protein atoms is selected for optimal superposition; these  $n$  atoms are denoted here with the index  $i$ . The two structures are then moved to the center of mass of the atoms  $i = 1, \dots, n$ , using the same weight factor for every atom. With  $\mathbf{r}_i$  and  $\mathbf{r}'_i$  describing the position of atom  $i$  in the two structures after the center of mass translation, a rotation  $\mathbf{R}$  is applied to the second structure such that the following expression becomes minimal:

$$\sum_{i=1}^n |\mathbf{r}_i - \mathbf{R}\mathbf{r}'_i|^2. \quad (1)$$

The following equation is then used to quantify structural differences for the group of atoms described by the index  $k$  with  $k = 1, \dots, m$ :

$$A = \left[ \frac{1}{m} \sum_{k=1}^m |\mathbf{r}_k - \mathbf{R}\mathbf{r}'_k|^2 \right]^{\frac{1}{2}}. \quad (2)$$

If the two subsets of atoms defined by the indices  $k = 1, \dots, m$  and  $i = 1, \dots, n$  coincide,  $A$  describes the usual RMSD value for these atoms (see Appendix in McLachlan, 1979), and yields an answer to question (a). For the case of different subsets of atoms  $k = 1, \dots, m$  and  $i = 1, \dots, n$ ,  $A$  is referred to as displacement  $D$  (Billeter *et al.* 1989), and answers question (b).

An NMR structure is usually described by a group of about 20 conformers. Whenever such a group of conformers was involved in an RMSD or displacement calculation, individual comparisons with all conformers were performed and the resulting values were averaged. Different comparisons were made for the backbone atoms N, C $\alpha$  and C', and for all heavy atoms. For the RMSD and displacement measurements it was assumed that the nitrogen and the oxygen in the side-chains of asparagines and glutamines cannot be distinguished in the electron density maps. In the crystal structures, the amide in these side-chains was therefore flipped by 180° around the bond preceding it whenever a better fit with the set of NMR conformers could be obtained. (For reasons of atom nomenclature, the carboxylic groups of aspartic acid and glutamic acid, as well as the rings of phenylalanine and tyrosine were sometimes also flipped by 180° with respect to the original data files.)

RMSD values were evaluated for the following two choices of residues. Initially, all residues were selected and the RMSD was calculated for all backbone atoms N, C $\alpha$  and C', and for all heavy atoms. The second choice consisted of those

residues with smallest displacements  $D$  among the set of conformers describing the NMR structure. A threshold for  $D$  was determined in the following way. For each residue, the displacement of the backbone atoms was evaluated after superposition of the complete backbone. These  $D$  values were then sorted according to their size, and the largest gap in this series of numbers determined the threshold, with the condition that between  $\frac{2}{3}$  and  $\frac{4}{5}$  of the residues were selected. The choice then consisted of all residues which had a displacement of the backbone atoms smaller than the threshold. For the heavy atom comparisons, side-chains with displacements smaller than 1.5 times the threshold were added.

Structural deviations were calculated for each protein using the above procedures within the group of NMR conformers, and between this group and the single coordinate set reported for the crystal structure. In addition, a mean NMR structure was obtained by superposition of all NMR conformers using the selection of  $\frac{2}{3}$  to  $\frac{4}{5}$  of the protein backbone (see above), and averaging the coordinates of corresponding atoms of the NMR conformers.

Square roots of B-factors were used to indicate the expected structural deviations within the crystal structures according to the relation  $B \approx \langle \Delta x^2 \rangle$ , with  $\langle \Delta x^2 \rangle$  describing the square of the uncertainty of an atom position along the diffraction vector (Glusker & Trueblood, 1985; Buerger, 1980). Similar to the displacements  $D$ , B-factors were averaged over the backbone atoms N, C $^\alpha$  and C', or over the heavy atoms, of a residue. Because displacements  $D$  describing atomic deviations in NMR structures and B-factors describing atomic deviations in X-ray structures do not exactly match in their information content, these measures of precision were normalized in the following way prior to quantitative comparisons between NMR and X-ray structures:

$$D_n = \frac{D - \langle D \rangle}{\Delta D} \quad (3a)$$

and

$$\sqrt{B}_n = \frac{\sqrt{B} - \langle \sqrt{B} \rangle}{\Delta \sqrt{B}}. \quad (3b)$$

The average,  $\langle D \rangle$ , and the standard deviation,  $\Delta D$ , of the displacements  $D$  of each residue were calculated. The average  $\langle D \rangle$  was subtracted from the value  $D$ , and this difference was divided by the standard deviation  $\Delta D$ . Thus the normalized displacement,  $D_n$ , for a given residue is negative when the displacement is smaller than the average displacement over all residues, and a value of  $D_n = 1.0$  indicates a displacement that is one standard deviation worse than the average. A similar treatment yielded normalized values  $\sqrt{B}_n$  for the square roots of the B-factors.

The most local view of structural differences is given by the values of individual dihedral angles, which represent the basic degrees of freedom when folding a polypeptide. The backbone angles  $\phi$  and  $\psi$  and the first side-chain angle  $\chi^1$  were plotted versus the protein sequence. For a given dihedral angle, the single value measured in the X-ray structure was plotted together with up to three distinct ranges of values observed in the NMR conformers.

On the level of the secondary structure, it is useful to examine the network of hydrogen bonds since these represent the most important interactions in helices and  $\beta$ -sheets. In the present context, a hydrogen bond was defined whenever the distance between a donor heavy atom and an acceptor heavy atom was smaller than 3.4 Å, and the angle defined by the position of the acceptor atom, the hydrogen and the donor heavy atom did not exceed 35°. For this purpose, a hydrogen atom is attached to the donor, either in the unique position defined by the donor heavy atom and two other heavy atoms bound to the latter, or by the optimal position for hydrogen bonding when the donor heavy atom is bound to only one other heavy atom.

The above data was complemented by a list of contacts shorter than 3.0 Å observed for the X-ray structure between different molecules in the crystal lattice. These crystal contacts were checked within a unit cell by application of crystallographic symmetry operations, and between unit cells by using periodic boundary conditions.

### 3. DETERMINATION OF PROTEIN STRUCTURES

#### 3.1 Determination by NMR in solution

This chapter describes several factors that limit the structural accuracy that can be obtained using the two methods of structure determination of proteins. NMR data on proteins result mainly from constraints on proton-proton distances derived from NOEs (nuclear Overhauser effects), and from dihedral angle information deduced from vicinal coupling constants (Wüthrich, 1986). The density of the NOEs observed per proton decreases on the protein surface, and is also lowered by internal mobility. Thus, the structure at the surface of a protein, in particular of surface residues, is often less well determined due to the lower density of the experimental data and to the lack of a rigid structure.

In contrast, the presence of very local data given primarily by the coupling constants, but also by intraresidual and sequential NOEs, may determine the local structure very precisely. For example, in residues with a single  $\beta$ -proton the three different conformers of the  $\chi^1$ -angle can often be distinguished. For many other residues, the two protons of the  $\beta$ -methylene groups can be individually assigned (Güntert *et al.* 1989). These stereospecific assignments may in turn allow to discriminate the different conformers of the  $\chi^1$ -angle. Other dihedral angles in the side-chains may in a similar way be well determined by the NMR data.

NMR data determine the position of a proton relative to the positions of other, nearby protons. Errors in the position of a proton will therefore spread out to other protons whose positions are determined by NOEs to the first one. This is different for X-ray structures where the position of each atom is given directly by experimental data, namely a peak in the electron density map. As the numerous correct protein structures determined by NMR show, the small errors of the local structure do usually not add up to large errors of the global structure in globular proteins due to the requirements of the covalent structure and of the dense packing. The fact that correct overall dimensions are usually found for protein

structures determined by NMR does not, however, prevent occasional distortions of loops protruding from the globular part of the protein.

The need for converting the experimental distance data into atomic coordinates raises the question of possible biases of the distance geometry algorithms used for the calculation of the NMR structure. These have been discussed several times (e.g. Wagner *et al.* 1987; Havel, 1990; Kuszewski *et al.* 1992), and the important conclusion for this study is that in the best defined parts of a protein, none of the newer algorithms appear to produce significantly biased structures.

The distance geometry calculations implicitly assume a rigid structure by the use of static distance constraints. This is in contrast to the dynamic behaviour of real proteins. The averaging of the experimental data due to the internal mobility of the protein will in many cases yield an average structure, which may correspond to the most highly populated state. Difficulties arise when the protein structure jumps between two, or a small number, of conformations. The average structure described by the set of distance constraints might in this case be highly strained and thus be scarcely populated (Torda *et al.* 1990).

### 3.2 *Determination by X-ray diffraction in single crystals*

The fundamental steps, and difficulties, of solving the structure of a protein by X-ray diffraction in single crystals are the crystallization of the protein, the recording of a data set to high resolution, the determination of the phases in order to obtain the electron densities, the construction of an initial model, and the refinement of this model. A detailed discussion of potential pitfalls of X-ray diffraction studies on proteins, and of ways around them, was published recently by Brändén & Jones (1990). The need of a crystallized form of the protein may require strongly non-physiological solution conditions from which the crystal is grown, e.g. with respect to the concentration of salts or added co-solvents. Furthermore, the introduction of heavy atoms in order to solve the phase problem is often accompanied by small structural changes of the protein. Phases are obtained from small differences between large numbers, and their errors may distort the electron density map. The subjectivity of the crystallographer plays an important role in the construction of the initial model into the electron density map. Today, constructing and testing hypothetical chain traces in the electron density map constitutes largely interactive work on a computer graphics system, and thus lacks the reproducible character of an automatic procedure. Knowledge of various features of the protein under investigation like content of regular secondary structure, known structure of related proteins and other chemical or physical information is used at this step. Difficulties in the construction of the initial model are more likely to occur in more mobile regions, e.g. on the protein surface and outside of regular secondary structure elements. Finally, an automatic refinement of the model attempts to modify both the structure and the phases to obtain the best agreement between the observed amplitudes and the amplitudes calculated from the structures. Additional constraints may describe the covalent structure. Optimization algorithms in high dimensions usually have a limited range of convergence. Thus, the refinement may fail in correcting errors of the initial

phases or of the initial model, and instead find values for the phases and a structure for which the target function of the optimization, the R-factor, reaches only a local minimum.

In single crystals, the protein molecules are densely packed and thus often form intermolecular contacts. These may immobilize surface side-chains that exhibit high mobility in the isolated protein, and may even affect the backbone conformation. For example, the comparison of the polypeptide backbone conformation of two different crystal forms of BPTI (Deisenhofer & Steigemann, 1975; Wlodawer *et al.* 1987*a*; entries 4PTI and 6PTI, respectively, of the Protein Data Bank) yields a global RMSD value of 0.4 Å with displacements of the backbone atoms of individual residues reaching 1.0 Å (for Arg17 after superposition of the backbone of residues 1–56). Often, the active site of a protein is located at the surface and therefore subject to distortions due to crystal contacts, thus making the interpretation of the active-site conformation of this protein difficult.

#### 4. PROTEINS SELECTED FOR COMPARISON

Detailed comparisons between the NMR and X-ray structures of a total of twelve proteins (Table 4.1) are given in Chapter 5. (Crystallographic parameters for the twelve proteins used for the calculation of the lattice contacts are listed in the Appendix.) A requirement for the selection of these proteins was, besides availability of the coordinates, that an NMR structure and an X-ray structure exist for the same protein. This excludes cases where a structure has been determined for proteins from different species, where one structure is for the free protein and the other for a complexed form, or where protein fragments do not have the same length. Examples of such proteins not meeting the above requirements are the zinc fingers, plastocyanin, cellobiohydrolase, hirudin, several proteinase inhibitors, the receptor FKBP, and the activation domain of procarboxypeptidase B. Also excluded are a few proteins that have a structure very similar to that of another protein already on the list. Thus, the proteinase inhibitor domain of Alzheimer's  $\beta$ -Amyloid precursor is not included due to its similarity to BPTI as quantified by a RMSD value for the backbone (residues 5–55) of 0.6 Å between the two proteins (Hynes *et al.* 1990). Leucine zippers are excluded, since their NMR structure consists of a simple helix; however, the Co1E1 rop protein is included due to the presence of a tertiary structure, a four-helix bundle. This selection should best serve the purpose of evaluating the degree of structural similarity achieved by the two methods, and the differences resulting from the two techniques for structure determination.

The first protein, tendamistat, serves the purpose of demonstrating the comparison tools used in this study and is thus presented in greater detail. For most other proteins in the study, genuine differences between the two structures are identified, and an attempt was made to explain these differences. For the 434 repressor(1–69) and for BPTI, comparisons based on very similar tools as used here have recently been published, so that the corresponding sections in the following Chapter represent a short summary.

Table 4.1. Protein structures used in the present comparison

Protein*	Source	Number of a.a.‡	NMR†		X-ray‡		
			Method	number	RMSD	resolution	R-factor
$\alpha$ -Amylase inhibitor	Streptomyces tendae	74	DISMAN	9	0.9¶ (69)	2.0	0.199 (74)
Barnase	B. amyloliquefaciens	110	DISGEO + X-PLO	20	1.11 (110)	2.2	
Ca-binding protein	Bovine	75	DISGEO + AMBER	33		2.3	0.178 (75)
Chymotrypsin inhib. 2	Barley	83	DISGEO + X-PLO	11	1.4 (62)	2.0	0.198 (65)
ColE1 rop protein	Plasmid ColE1	63*2	DISMAN + X-PLO	10	1.1 (47)	1.7	0.182 (56)
Interleukin 1 $\beta$	Human	153	DISGEO + X-PLO	32	0.4 (150)	2.0	0.19 (151)
Interleukin 8	Human	72*2	DISGEO + X-PLO	30	0.4 (67)	2.0	0.187 (69)
Metallothionein	Rat	30 + 31	DISMAN	10	2.0 + 1.7¶ (30 + 31)	2.0	0.176 (61)
434 repressor(1-69)	Phage 434	69	DIANA + AMBER	20	0.6 (63)	2.0	0.193 (63)
Ribonuclease A	Bovine	124	Own + GROMOS	8	1.8¶ (124)	1.26	0.15 (124)
Thioredoxin	E. coli	108	DISGEO + AMBER	12	0.6 (106)	1.68	0.165 (108)
Trypsin inhibitor	Bovine	58	DIANA + AMBER	20	0.4 (55)	1.0	0.200 (58)

\* See the text (Chapter 5) for references to these structures.

‡ This number corresponds to the complete protein or protein domain. The number of residues for which coordinates are reported for the NMR structure and the X-ray structure, and thus used in the comparison, is given in Table 4.2. ColE1 rop and interleukin 8 form dimers. Metallothionein consists of two domains that have no significant contacts in the solution or crystal structures.

† The entries under the heading NMR report the distance geometry and refinement programs used for the calculation of the solution structure, the number of conformers that describe this structure, and the average RMSD value as reported in the original publications with, in parentheses, the number of residues used for the RMSD calculation. DISMAN (Braun & Go, 1985) and DIANA (Güntert *et al.* 1991) are implementations of the variable target function algorithm. DISGEO (Havel & Wüthrich, 1984) is an implementation of the EMBED algorithm. X-PLO (Brünger, 1990), AMBER (Singh *et al.* 1986) and GROMOS (van Gunsteren, 1987) are Molecular Dynamics programs. 'Own' indicates not further specified code written by the authors of the structure.

‡ The entries under the heading X-ray list the resolution of the crystal structure, and the crystallographic R-factor with, in parentheses, the number of residues for which coordinates are reported.

¶ These RMSD values are averages of pairwise comparisons rather than of comparisons from each conformer to a mean structure. They are about 1.5 times larger than those calculated to a mean structure.



Table 4.2. RMSD values from the mean NMR structure to the NMR conformers and to the X-ray structure

Protein*	All residues†			Selected residues‡		
	RMSD [Å]†			RMSD [Å]†		
	Nr.	⟨NMR⟩- NMR	⟨NMR⟩- X-ray	Nr.	⟨NMR⟩- NMR	⟨NMR⟩- X-ray
α-Amylase inhibitor	74	1.35; 1.69	2.17; 2.50	56; 43	0.40; 0.56	0.69; 1.04
Barnase	107	1.00; 1.44	1.76; 2.21	72; 48	0.62; 0.77	1.27; 1.39
Ca-binding protein	75	0.98; 1.34	1.31; 1.71	58; 44	0.49; 0.66	0.71; 1.10
Chymotrypsin inhib. 2	64	0.87; 1.43	1.19; 1.85	49; 29	0.47; 0.59	0.86; 1.09
ColE1 rop protein	56	0.90; 1.45	1.16; 1.95	42; 18	0.59; 0.65	1.01; 1.07
Interleukin 1β	153	0.48; 0.85	0.99; 1.52	118; 67	0.26; 0.29	0.64; 0.74
Interleukin 8	69	0.51; 1.04	2.00; 2.65	50; 27	0.24; 0.31	0.95; 1.26
Metallothionein: β	30	1.36; 1.96	2.03; 2.58	21; 15	0.86; 1.05	1.34; 1.73
α	31	1.16; 1.80	1.47; 2.36	23; 14	0.83; 0.87	1.28; 0.83
434 repressor(1-69)	63	0.61; 1.10	1.09; 1.62	49; 19	0.40; 0.45	0.66; 0.77
Ribonuclease A	124	0.77; 1.14	1.11; 1.58	94; 72	0.44; 0.52	0.76; 1.25
Thioredoxin	108	0.76; 1.13	1.01; 1.53	76; 39	0.37; 0.43	0.75; 1.05
Trypsin inhibitor	58	0.73; 1.07	1.16; 1.44	44; 24	0.32; 0.34	0.64; 0.81

\* See footnote to Table 1.

† 'All residues' are those residues which are found in both the NMR and the X-ray structure. 'Selected residues' are the ones selected by the automatic procedure described in Chapter 2.

‡ '⟨NMR⟩' denotes the mean NMR structure, 'NMR' the set of individual conformers describing the NMR structure, and 'X-ray' the X-ray structure. The first number is for the selected backbone fragments, the second for the selected heavy atoms.

Table 4.2 lists, for each protein, the total number of residues for which coordinates are reported in both the NMR and the X-ray structure. Furthermore, the number of residues forming the well defined regions in the NMR structure is given (see Chapter 2 for their selection). These residues are used for all superpositions shown in the figures of Chapter 5. For both residue selections, RMSD values are listed for the comparison of the NMR conformers to their mean and between this mean structure and the crystal structure. The two values in each column are comparisons of the backbone atoms N, C $\alpha$  and C', and of heavy atoms. Heavy atoms are either all non-hydrogen atoms in the protein or only those from the selected backbone fragments and the selected side-chains (see Chapter 2).

All crystal structures have a resolution of 2.3 Å or better, with R-factors of 0.2 or below (Table 4.1). The RMSD values for all residues between the mean NMR structure and the individual NMR conformers vary from 0.5 Å to 1.4 Å; those between the mean NMR structure and the X-ray structure are in the range from 1.0 Å and 2.2 Å. In contrast, the backbone of the well defined regions is determined in the NMR structures for all but two proteins with a RMSD value not exceeding 0.5 Å, whereas the differences to the X-ray structure are often larger than 0.5 Å. If the best defined side-chains are included in the calculation of the global RMSD values, a threshold of about 0.75 Å delimits the RMSD values of the mean NMR structure to the individual NMR conformers from those of the mean NMR structure to the X-ray structure.

## 5. INDIVIDUAL COMPARISONS

### 5.1 $\alpha$ -Amylase inhibitor (*tendamistat*)

The  $\alpha$ -amylase inhibitor Hoe-467A from *Streptomyces tendae* 4158 (also called *tendamistat*) is a small, very stable protein resistant to denaturation and proteolytic attack. It consists of two triple-stranded  $\beta$ -sheets (residues 12–17, 20–25 and 52–58; 41–49, 30–37 and 67–73) with all neighbouring strands arranged in an antiparallel fashion (Fig. 5.1.1). The structure is further stabilized by two disulfide bridges (Cys11–Cys27 and Cys45–Cys73). The tripeptide Trp18–Arg19–Tyr20 is assumed to play a major role in the binding to  $\alpha$ -amylase. The determination of the solution structure (Kline *et al.* 1986, 1988) yielded a well defined conformation for the complete backbone with the exception of about five residues at the N-terminus and of the C-terminal residue. This structure was based on 842 distance constraints from NOEs and about 70 constraints on dihedral angles. In the crystal structure (Pflugrath *et al.* 1986), unique positions are given for all atoms except for the aromatic ring of Tyr15, for which no electron density could be observed. The NMR and X-ray structures of *tendamistat* have been compared earlier (Billeter *et al.* 1989). Fig. 5.1.1 shows the close coincidence of the backbone fold of the two structures. In the following, the precision of the structures determined by both NMR and crystallography is described in detail, and the differences between the two structures are evaluated in terms of this structural precision.

For the present comparison, the NMR structure of *tendamistat* described by

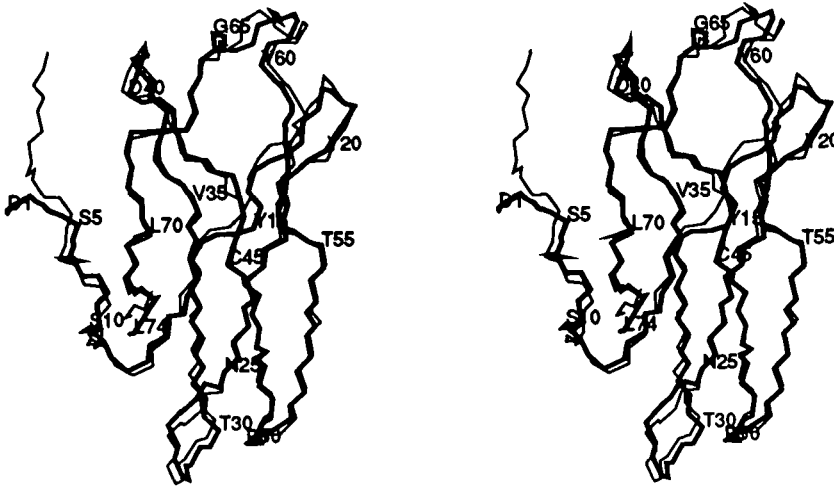


Fig. 5.1.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of the  $\alpha$ -amylase inhibitor tendamistat. Distortions of the covalent geometry for the terminal residues of the mean NMR structure are caused by the averaging of the coordinates. The superposition was calculated for selected residues which are identified in Fig. 5.1.2.

the nine conformers obtained after restrained energy-refinement using a large weight for the NOE distance constraints is considered (Billeter *et al.* 1990). A comparison of the complete backbone of the nine NMR conformers with their mean structure yields an average RMSD value of 1.35 Å; for all heavy atoms a value of 1.69 Å is obtained (Table 4.2). The selection procedure described in the Methods section identified 56 residues with a structurally well defined backbone in the solution structure, to which 43 side-chains with a well defined conformation were added. With this selection, the average RMSD values drops to 0.40 Å and 0.56 Å, respectively (see Fig. 5.1.2 for this selection).

A plot of displacements,  $D$ , versus the sequence is given in Fig. 5.1.2. The conformation of the four N-terminal residues is completely undefined in the NMR structure, and the next two residues still show a large structural spread. The following four residues clearly adopt a single conformation with Ser10 having both backbone and side-chain displacements,  $D$ , just above the threshold used for the selection of the best defined protein fragments. With the start of the first  $\beta$ -strand, the peptide chain becomes very well defined. This remains true for the backbone of all six  $\beta$ -strands except for the exposed Thr41 and for Cys73, the next to last residue. Other well defined backbone fragments are the loops 26–29 and 50–51 which connect the two  $\beta$ -sheets, and the extension of the  $\beta$ -strand 52–58, i.e. residues 59–61. Besides residue 10, all backbone fragments with an increased structural spread are located on one edge of the molecule (Fig. 5.1.1). The unstructured side-chains of residues with a well defined backbone part contain polar or charged groups located on the protein surface, namely Tyr20, Glu29, Gln52, Asp58, Arg68 and Arg72; the only exception is Ile61 with a  $D$ -value for the side-chain just above the selection threshold.

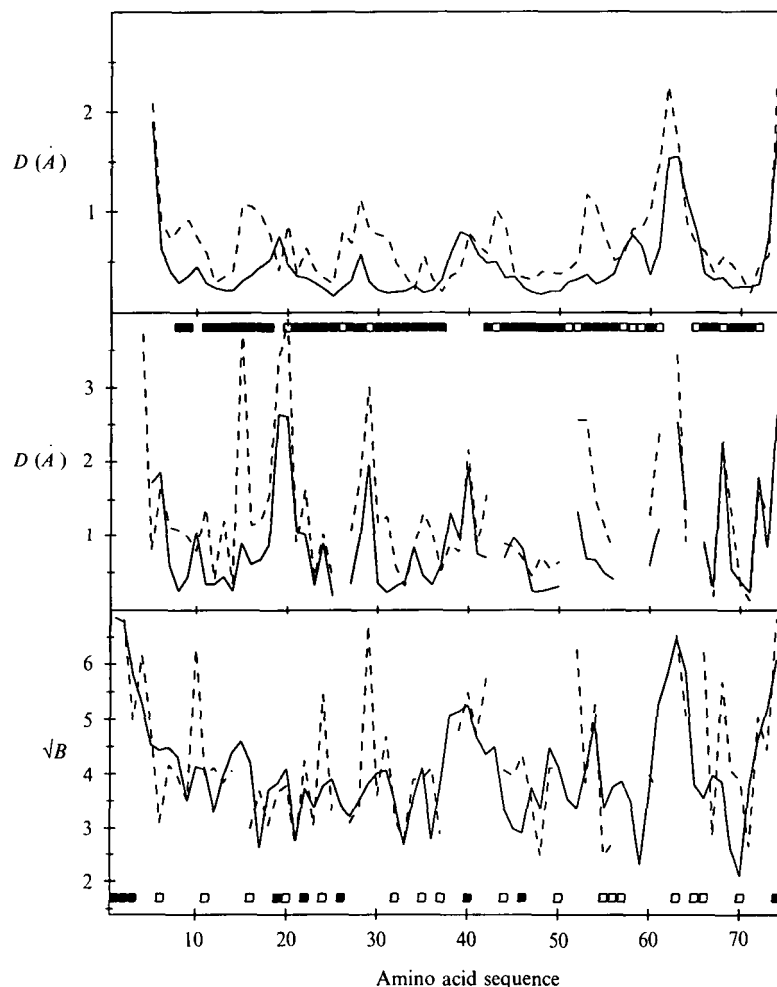


Fig. 5.1.2. Displacements and square roots of the B-factors are plotted versus the sequence for the  $\alpha$ -amylase inhibitor tendamistat. Top panel: Displacements from the individual NMR conformers to their mean structure (solid line), and from this mean structure to the X-ray structure (broken line). All displacements are averaged over the atoms N, C $^{\alpha}$  and C $^{\prime}$ . Centre panel: Same for the heavy atoms of the side-chains with averaging over all heavy atoms of the side-chains. The squares at the top of this panel identify the residues forming the best defined backbone fragments (for the selection procedure see Chapter 2). Filled squares indicate that the side-chain of the residue is also well defined. Lower panel: Square roots of the B-factors for the backbone (solid line) and the side-chains (broken line) of the X-ray structure. Squares at the bottom identify the residues with contacts in the crystal lattice to other molecules, where filled squares indicate intermolecular protein-protein contacts.

A closer look at the crystal structure and at the reported B-factors show a lower than average precision for a number of backbone fragments (Fig. 5.1.2). The N-terminal five residues are again among the structurally most disordered parts. In the first  $\beta$ -strand, the backbone of Tyr<sub>15</sub> has a slightly increased B-factor. Other less well defined fragments are the loop with residues 38–41 connecting the first

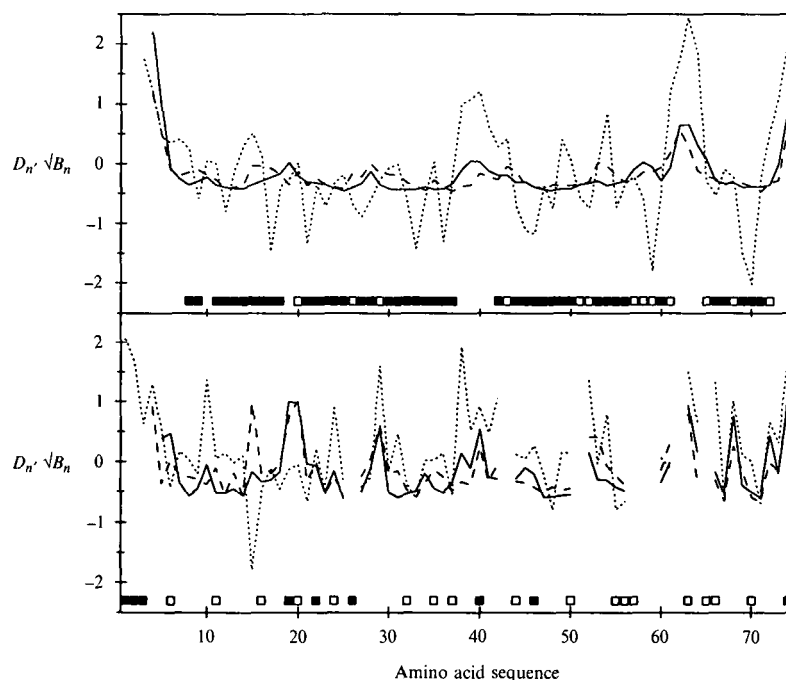


Fig. 5.1.3. Normalized displacements  $D_n$  (equation 3a) and normalized square roots of the B-factors  $\sqrt{B_n}$  (equation 3b) are plotted versus the sequence for the  $\alpha$ -amylase inhibitor tendamistat. Solid line: Normalized displacements  $D_n$  between the NMR conformers and their mean structure. Dashed line: Normalized displacements  $D_n$  between the mean NMR structure and the X-ray structure. Dotted line: Normalized square roots of the B-factors  $\sqrt{B_n}$  of the X-ray structure. Top panel: backbone, lower panel: side-chains. The squares in the top panel identify the residues forming the best defined backbone fragments (for the selection procedure see Chapter 2). Filled squares indicate that also the side-chain of the residue is well defined. The squares in the lower panel identify the residues with contacts in the crystal lattice to other molecules, where filled squares indicate intermolecular protein-protein contacts.

two strands of the second  $\beta$ -sheet, Thr54 on one of the outer strands of the first sheet, an exposed loop at one end of the molecule with residues 61–64, and the three C-terminal residues. Again, all but one of the disordered side-chains attached to backbone parts with small B-factors are polar or charged and located on the protein surface: Ser10, Asp24, Glu29, Glu42, Gln52, His66 and Arg68. The exception is the side-chain of Val31 with a marginally higher than average B-factor.

The backbone displacements between the mean NMR structure and the crystal structure after global superposition are smaller than 0.5 Å for 27 residues; the side-chain displacements are smaller than 1.0 Å for 28 side-chains (Fig. 5.1.2). However, the global RMSD values of Table 4.2 are about 1.6 times larger for the comparisons of the crystal structure with the mean solution structure than for the comparison of the individual NMR conformers to the mean solution structure. In Fig. 5.1.3, normalized values for the displacement after global superposition (equation 3a) are plotted together with the normalized square roots of the B-

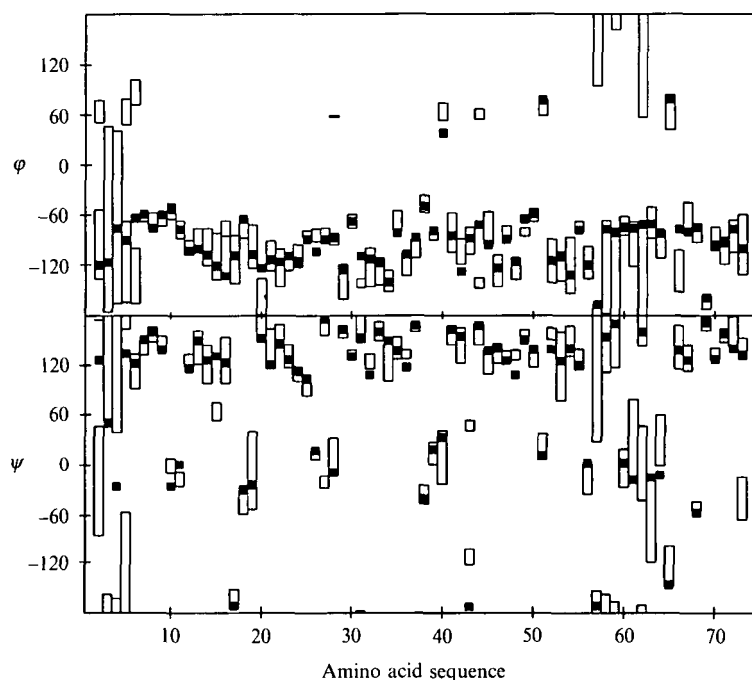


Fig. 5.1.4. Plot of the dihedral angles  $\phi$  and  $\psi$  versus the sequence for the  $\alpha$ -amylase inhibitor tendamistat. The open bars delineate the range of values adopted by the dihedral angles in the individual NMR conformers. The filled squares indicate the value of the corresponding dihedral angle in the X-ray structure.

factors (equation 3*b*). The displacements between the solution and the crystal structure follow closely those describing the solution structure for both the backbone and the side-chains, whereas the curve for the normalized square roots of the B-factors shows the largest variations. Differences between the two structures exceeding the normalized precision of each structure by more than 0.4 s.d. occur for the side-chains of Tyr15 and Ile53. The difference for Tyr15 can be explained by the absence of electron density in the crystal structure for this side-chain. For Ile53, which is well defined by both methods, no obvious explanation is available. This side-chain is located on the protein surface, next to Tyr15. In the crystal structure, its  $\delta$ -methyl group is where the  $\gamma$ -methyl group is in the solution structure, and in the NMR structure, the  $\delta$ -methyl group is about 2 Å away from the ring of Tyr15. Two side-chains, Arg19 and Tyr20, are clearly better defined in the X-ray structure. For each of these side-chains, contacts to other protein molecules and to crystal water molecules are detected (Fig. 5.1.3).

Local backbone differences described by the dihedral angles (Fig. 5.1.4) are observed for residues 5–6, 15–16 and 43–44, and thus do not coincide with the major differences detected after global superposition. In most cases, these local differences are due to the rotation of individual peptide planes, which occur mostly on the protein surface.

The network of hydrogen bonds is very similar, in particular for those involving backbone–backbone interactions. Of the 34 hydrogen bonds observed, only four

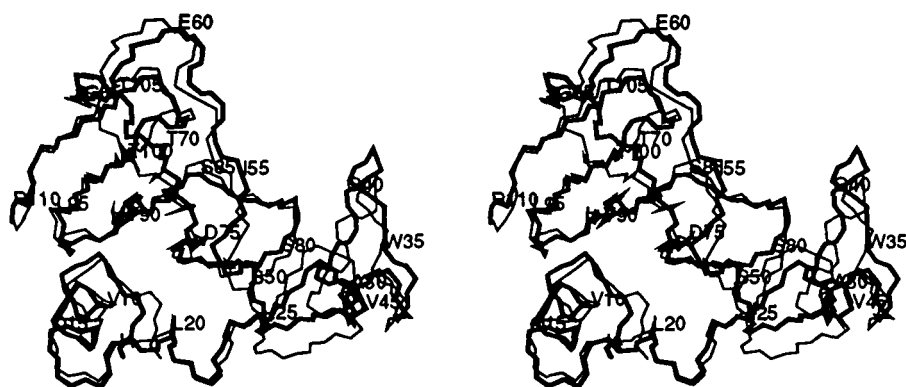


Fig. 5.2.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of barnase. The superposition was calculated for selected residues identified in Fig. 5.2.2.

have different bonding partners in the two structures: The amide proton to carbonyl oxygen hydrogen bonds Gly26–Asp24 and Asp40–Tyr37 of the solution structure was replaced by Gly26–Cys11 and Asp40–Glu38 in the crystal structure. In the NMR conformers, the amide proton of Gly62 is sometimes bound to Tyr60 rather than to Gly59 as in the crystal. The carbonyl oxygen of Val36 pairs with the amide of Tyr69 in the solution structure and to the OH of Tyr69 in the crystal structure. With the hydrogen bonds involving side-chain atoms, 22 are common to both structures, 11 occur in the solution structure only and 13 in the crystal structure only.

In summary, there is a strong similarity of the two structures which holds for practically the entire polypeptide backbone and for many side-chains. This similarity often extends to the degree of local precision with which the two methods have determined the structure. Interestingly, the largest difference in side-chain precision is observed for the expected active site (Arg19 and Tyr20). The only difference between the two structures that cannot be explained neither by the lower precision of one or the other method nor by crystal contacts (Fig. 5.1.3), is the side-chain conformation of Ile53.

## 5.2 Barnase

Barnase is a purine-specific ribonuclease from *Bacillus amyloliquefaciens* with 110 residues. It consists of two helices with residues 6–18 and 26–34, and a five-stranded antiparallel  $\beta$ -sheet with residues 50–55, 70–75, 85–91, 94–101 and 105–108 (Fig. 5.2.1). Its structure differs from that of pancreatic ribonuclease (Mauguen *et al.* 1982). The active site involves His102, which is a conserved residue among several bacterial ribonucleases. In the folded protein two additional conserved residues, Glu73 and Arg87, are located next to His102. The crystal structure at a resolution of 2.2 Å was presented by Mauguen *et al.* (1982), and the

NMR structure for this protein was recently reported based on 853 distance constraints from NOEs, 72 constraints on  $\phi$ -angles and 53 constraints on  $\chi^1$ -angles (Bycroft *et al.* 1991).

According to the procedure explained in Chapter 2, the backbone of 72 residues (see Fig. 5.2.1) was selected as well defined and superimposed to obtain a mean structure. These fragments include all regular secondary structure elements. Restriction to these residues reduces the average backbone RMSD between the 20 NMR conformers and their mean structure from 1.0 Å to 0.62 Å (Table 4.2). When including 48 well defined side-chains (see Fig. 5.2.1), a RMSD of 0.77 Å is obtained.

In the Protein Data Bank entry for the crystal structure, a different coordinate set was reported for each of the three protein molecules in the asymmetric unit (Mauguen *et al.* 1982). These are all very similar with a RMSD for the complete backbone of only 0.38 Å. One of the three coordinate sets contains no entries for the first three residues (the other two sets give coordinates for the third residue), and it has two conformations reported for the side-chain of Lys49. For this structure, which is used in the following comparison, occupancy factors smaller than 1.0 are given for several peripheral side-chain atoms of Lys19, Lys27, Gln31, Arg59, Lys66 and Lys98.

The superposition of the mean NMR structure and the crystal structure in Fig. 5.2.1 shows, that besides the good structural coincidence for large parts of the polypeptide chain, some significant deviations occur in the outer loops containing residues 35–44 and 67–70. The RMSD comparison of the mean NMR structure and the crystal structure yields 1.27 Å for the backbone of the 72 residues that are well defined in the NMR structure and 1.39 Å if the heavy atoms of the 48 well-defined side-chains are added (Table 4.2). The following discussion is based on inspection of the structures with computer graphics, on the normalized displacements  $D_n$  and B-factor values  $\sqrt{B_n}$  reported in Fig. 5.2.2, and on the analysis of hydrogen bonds, dihedral angles and crystal contacts.

Except for residues 27–28, there is good similarity of the backbone fold from the second residue to the end of the second helix, and the normalized displacements between the mean NMR structure and the crystal structure correspond in magnitude to those between the mean NMR structure and the individual NMR conformers (Fig. 5.2.2). This is in spite of the large B-factors reported for residues 15–18 which form the second half of the first helix. Increased B-factors are also given for residues 4, 13–14, 19 and 28–29; the latter may explain only the apparent structural difference in the first 34 residues. The loop with residues 35–47 shows significant structural differences between the two structures, which cannot be fully explained by the low degree of structural definition for this fragment in the NMR structure. The two structures converge again for the first  $\beta$ -strand. The central two glycines of this strand, 52 and 53, adopt however a different conformation. Interestingly, the following residues 57–60 also show good structural coincidence in spite of both a large structural spread among the NMR conformers and increased B-factors. This coincidence is continued to residue 66. The fragment 67–70 shows sizeable differences combined with a low definition of



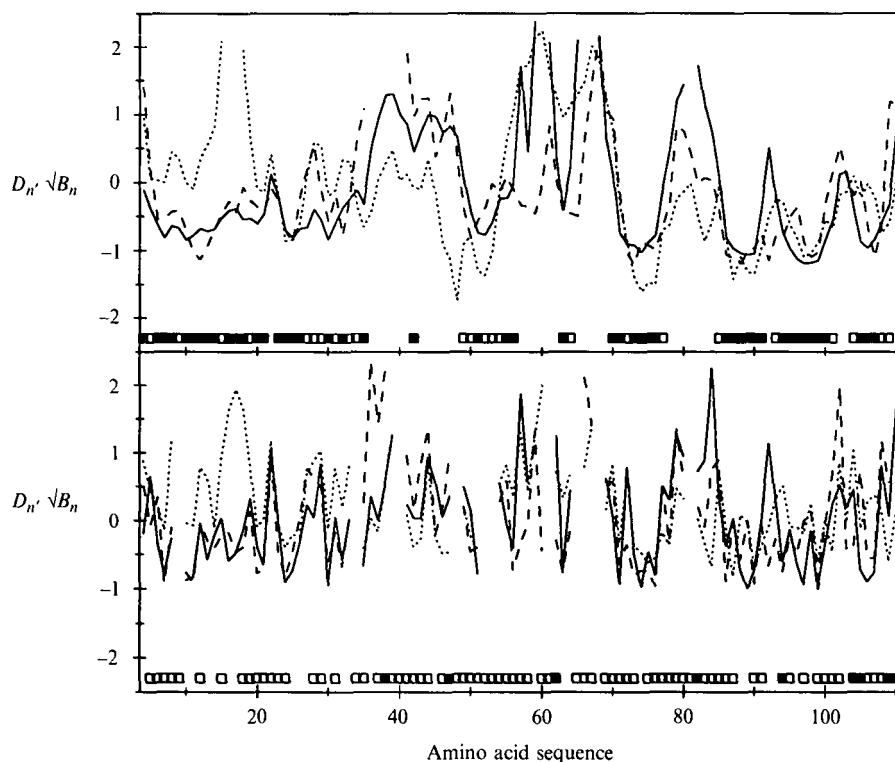


Fig. 5.2.2. Normalized displacements,  $D_n$  (equation 3a), and normalized square roots of the B-factors,  $\sqrt{B_n}$  (equation 3b), are plotted versus the sequence for barnase. See the caption to Fig. 5.1.3 for the information in the three panels and the meaning of the different line types and squares.

both the NMR and the crystal structure. The remaining part of the protein forms three well defined  $\beta$ -strands. Fragments of increased disorder within the NMR and the crystal structure, and between the two, correspond to the turns and loops connecting these strands.

Obvious differences in side-chain conformation in the fragments 4–34, 48–66 and 71–110, i.e. excluding the regions with large backbone differences between the two structures, are found for the following six residues. The backbone of Arg59 is not well defined in the NMR structure. Lys66 is at the beginning of the poorly defined loop 67–70 in both structures. Thr79 and the C-terminal residues Lys108 and Arg110 are in loops connecting  $\beta$ -strands, and are not well defined in the NMR structure due to their size and their location on the protein surface. The most interesting difference occurs for His102, which is reasonably well defined in both structures, and is involved in the enzymatic activity of barnase (see below).

As expected, most of the hydrogen bonding network involving the protein backbone coincides between the two structures. Exceptions were only found in the least well defined regions 35–47 and 67–70, for Gly53 in the centre of the first  $\beta$ -strand, and at the edges of the  $\beta$ -sheet, where donors or acceptors are sometimes replaced by equivalent groups of a sequentially neighbouring residue. The

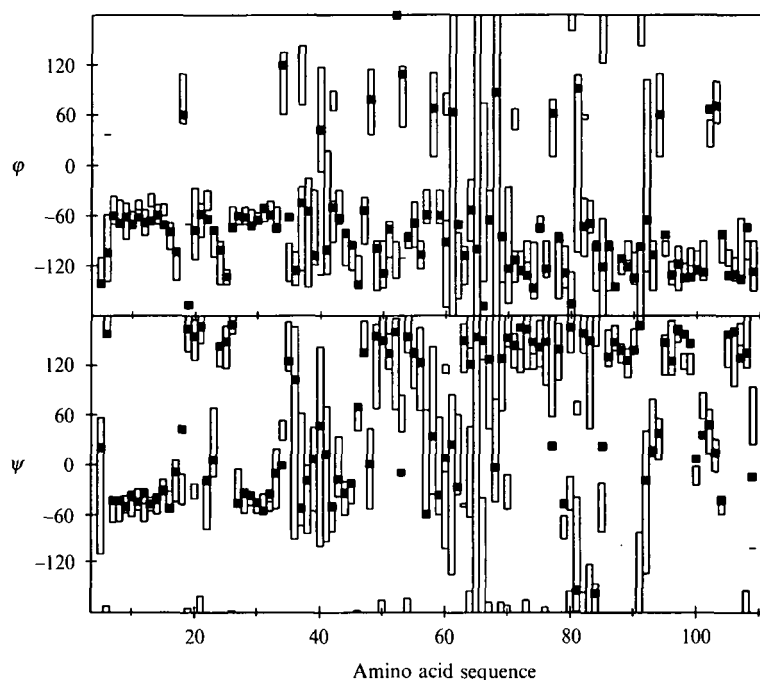


Fig. 5.2.3. Plot of the dihedral angles  $\phi$  and  $\psi$  versus the sequence for barnase. The open bars delineate the range of values adopted by the dihedral angles in the individual NMR conformers. The filled squares indicate the value of the corresponding dihedral angle in the X-ray structure.

different hydrogen-bonding pattern in the first  $\beta$ -strand is probably due to a hydrogen bond constraint used for the NMR structure calculation and derived from exchange data and secondary structure identification (Bycroft *et al.* 1990). This forces a hydrogen bond from the oxygen of Glu73 to the amide of Asp54, whereas the amide of Gly53 is bonded by this oxygen in the crystal structure.

The structural coincidence is also present on a very local level, as shown by the dihedral angles  $\phi$  and  $\psi$  (Fig. 5.2.3). All positive  $\phi$ -angles of the crystal structure were also found in the NMR structure (Bycroft *et al.* 1991). Differences for residues 18–19 and 34–35 correspond to rotations of the peptide planes. The different conformation of glycines 52 and 53 was already mentioned. Finally, differences were found to the  $\psi$ -angles of residues 77 and 85, which are located at a hinge position for the loop between the second and third  $\beta$ -strand. This loop makes most of its contacts with the ill-defined region 35–47. While almost all residues in the crystal structure are in contact with observed water molecules, only nine residues form inter-protein contacts (Fig. 5.2.2). About one third of these protein–protein contacts are between His102 or Gln104 and His18, and are thus related to the active site. Another third connects the three C-terminal residues to the residues 38 and 82, both of which are in more disordered regions.

Genuine structural differences were thus observed for the backbone of residues 35–47, the backbone of Gly53 and the side-chain of His102. The first difference affects an apparently flexible region and seems to be related to the numerous



Fig. 5.3.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of calbindin. The superposition was calculated for selected residues identified in Fig. 5.3.2.

crystal contacts of residue 38. The different hydrogen bonding pattern around Gly53 is probably due to a distance constraint enforcing a particular hydrogen bond (Bycroft *et al.* 1990). The different conformation of His102 between the two structures seems related to the crystal lattice contacts with His18, and to the large B-factors for this latter histidine. Possibly, His18 makes some contacts similar to those by the RNA substrate.

### 5.3 Calbindin

Calbindin is a 75 residue calcium-binding protein of the Calmodulin family (also called Vitamin D-dependent Ca-binding protein). Four helices (residues 2–16, 24–36, 45–54 and 62–75) connected by loops of seven or eight residues determine the fold of the bovine form of the protein, which is investigated here (Fig. 5.3.1). The loops between the first two helices and between the last two helices each bind one  $\text{Ca}^{2+}$  ion. This helix-loop-helix motif formed by both pairs of helices is common to many Ca-binding proteins. In the X-ray structure (Szebenyi & Moffat, 1986), the peptide bond to proline 43 was described to be in a *trans*-configuration; however, other studies have shown that an equilibrium between *cis* and *trans* exists (Chazin *et al.* 1989; Svensson *et al.* 1992). In the NMR structure study (Skelton *et al.* 1990; W. J. Chazin, personal communication), this proline was substituted by a glycine (and is thus also in a *trans*-configuration). In contrast to the crystal structure, the solution structure was determined in the absence of calcium.

The global RMSD values between the mean NMR structure and the 33 individual NMR conformers are 0.49 Å for the backbone of 58 residues, and 0.66 Å when the heavy atoms of 44 side-chains are added (see Fig. 5.3.2 for the

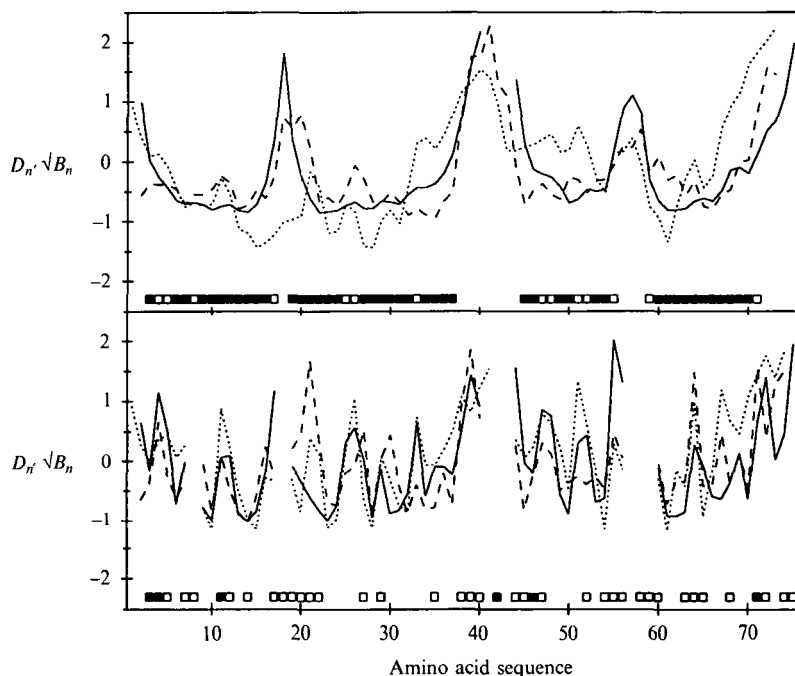


Fig. 5.3.2. Normalized displacements  $D_n$  (equation 3a) and normalized square roots of the B-factors  $\sqrt{B_n}$  (equation 3b) are plotted versus the sequence for calbindin. See the caption to Fig. 5.1.3 for the information in the three panels and the meaning of the different line types and squares.

residue selection). The local backbone precision of the NMR structure is given in Fig. 5.3.2 in terms of normalized displacements. The displacements are lowest for the four helices and highest for the loop connecting the two helix-loop-helix motifs (Fig. 5.3.2). The spread among the NMR conformers is also increased for the two Ca-binding loops, probably due to the absence of calcium in this structure.

The X-ray structure shows small B-factors for the backbone of the first helix-loop-helix motif including the Ca-binding loop (Fig. 5.3.2). The B-factors increase at the end of the second helix, are high for the loop connecting the two Ca-binding motifs, and remain above average for the third helix. After rather low values for the second Ca-binding loop and the first part of the last helix, they increase again for the C-terminus. The bottom of Fig. 5.3.2 shows only very few protein-protein contacts in the crystal lattice. However, Szebenyi & Moffat (1986) list additional intermolecular hydrogen bonds between the side-chain of Glu4 and the amide of Leu46, the side-chains of Glu5 and Asn56, the side-chains of Lys12 and Glu64, and the side-chains of Lys16 and Glu48.

The global RMSD values between the mean NMR structure and the crystal structure are 0.71 Å for the backbone of the 58 selected residues, and 1.10 Å when the heavy atoms of 44 selected side-chains are added (Table 4.2), indicating the close coincidence of the two structures. This similarity is strongest for the helices, followed by the Ca-binding loops (Figs 5.3.1 and 5.3.2). The largest difference

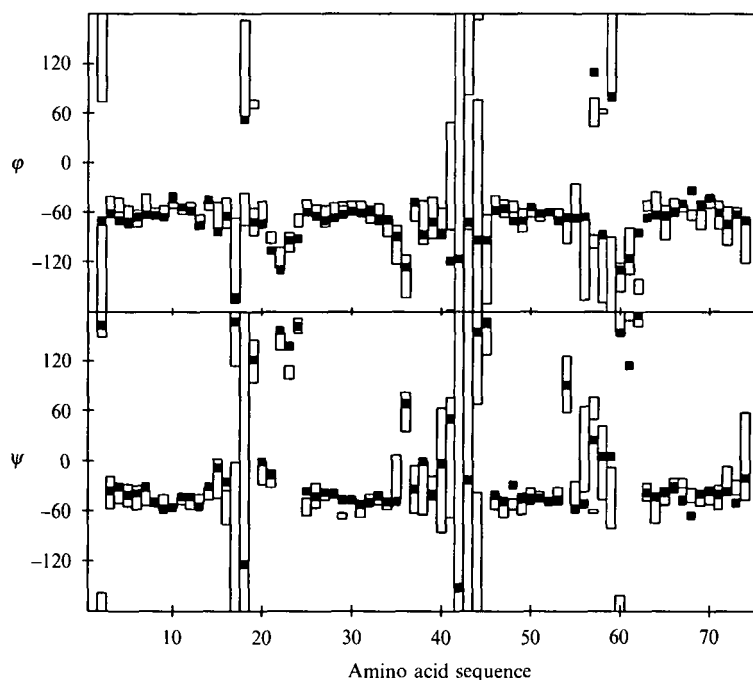


Fig. 5.3.3. Plot of the dihedral angles  $\phi$  and  $\psi$  versus the sequence for calbindin. The open bars delineate the range of values adopted by the dihedral angles in the individual NMR conformers. The filled squares indicate the value of the corresponding dihedral angle in the X-ray structure.

between the two structures is observed for the loop between the second and the third helix, where a structural heterogeneity is reported involving Pro43 (see above). In both structures, the two Ca-binding loops, residues 17–23 and 55–61, are connected by two hydrogen bonds between the amide and carbonyl groups of Leu23 and Val61.

Other significant differences, i.e. differences exceeding the precision of both structures, are observed for the backbone of residues 20–28 and 59–62. These are most likely due to the presence of the calcium in the crystal structure, and are correlated to each other by the two hydrogen bonds from Leu23 to Val61. This difference is propagated into the second helix, where a hydrogen bond from Glu26 and Leu30 is observed only in the X-ray structure. In addition, a different hydrogen bond is formed by the side-chain of Asn21, which binds to the side-chain of Asp19 in the crystal structure and to the side-chain of Gln22 in the NMR structure. Finally, a significant difference in side-chain conformation is seen for Glu64, which forms a hydrogen bond to another protein molecule in the crystal (Szebenyi & Moffat, 1986).

Sizeable differences in backbone dihedral angles (Fig. 5.3.3) are all located in the three loops between the helices, and describe mostly rotations of peptide planes. For most dihedral angles, the range of observed values in the NMR conformers is very narrow, and the coincidence with the crystal structure is outstandingly good.

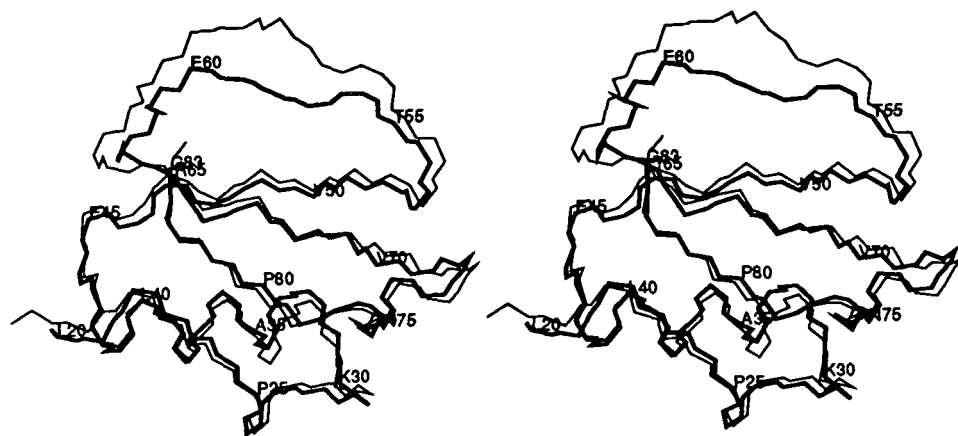


Fig. 5.4.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of chymotrypsin inhibitor 2. The superposition was calculated for selected residues identified in Fig. 5.4.2.

Thus, when accounting for the fact that calcium is present in the crystal structure but not in the NMR structure, the only significant difference is observed for the side-chain of Glu64, probably due to its intermolecular hydrogen bond.

#### 5.4 Chymotrypsin inhibitor 2

The chymotrypsin inhibitor 2 from barley seeds is an interesting member of the serine proteinase inhibitor family because it lacks the characteristic disulfide bridges of most other inhibitors. A crystal structure appeared in 1987 (McPhalen & James). In the same year a first NMR structure was reported (Clare *et al.* 1987*a*), and the two structures were compared (Clare *et al.* 1987*b*). Recently, a refined NMR structure was presented and compared to the crystal structure (Ludvigsen *et al.* 1991). The comparison given here is based on the latter NMR structure. The entire protein consists of 83 residues, however, neither the X-ray nor the NMR study could determine the conformation of the first 19 residues. In addition, both structures describe the loop with residues 54–62 as flexible. It is worth noting that the active site Met59–Glu60 is part of this loop. The core of the protein is formed by a  $\beta$ -sheet with both parallel and antiparallel relative strand orientations, covered on one side by a helix with residues 32–43 (Fig. 5.4.1). In the crystal structure the description of the  $\beta$ -sheet includes residues 22–24, 30, 47–53, 64–71, 75–77 and 80–82 (McPhalen & James, 1987). In the new NMR structure, the one residue long 'strand' 30 is extended to include residues 29–31. The strand 22–24 runs antiparallel to residues 80–82, and so does the second strand with respect to residues 76–75. The residues 75–77 and, in the NMR structure only, the residues 80–82 form an antiparallel sheet with the strand 64–71, which in turn runs parallel to the strand 47–53. The following discussion of similarities and differences of the two structures is based on global displacements and on square roots of B-factors (Fig. 5.4.2).

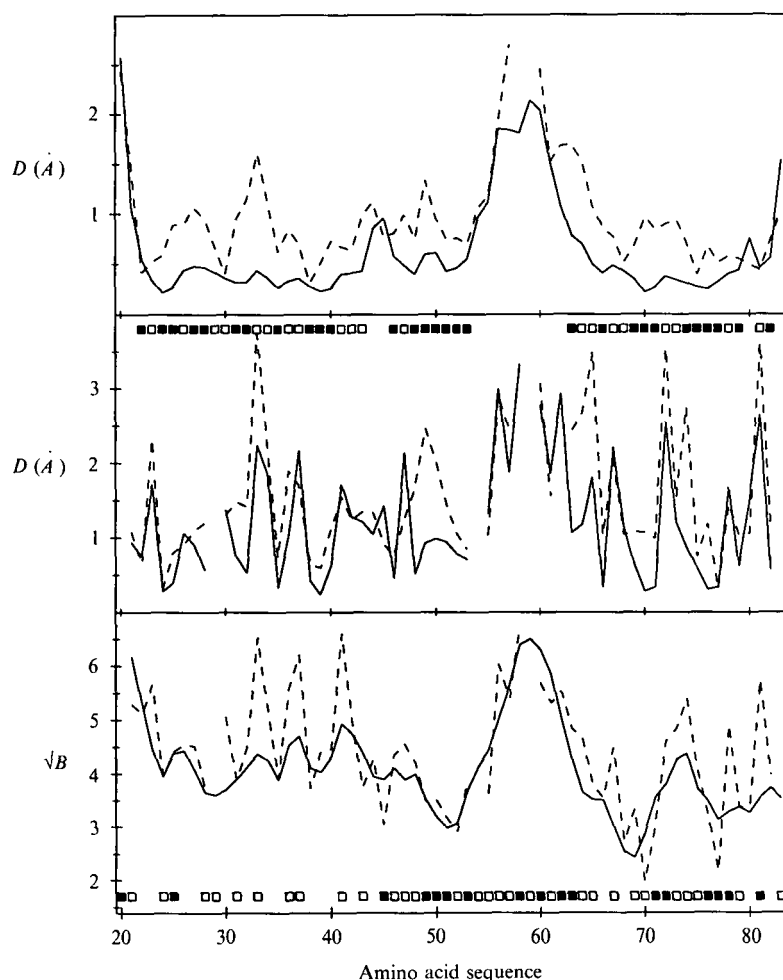


Fig. 5.4.2. Displacements and square roots of the B-factors are plotted versus the sequence for chymotrypsin inhibitor 2. See the caption to Fig. 5.1.2 for the information in the three panels and the meaning of the different line types and squares.

The NMR structure was based on 961 NOEs and 94 restraints on dihedral angles. The well defined parts include the backbone of residues 22–43, 46–53, 63–79 and 81–82, as well as 29 sidechains (see Fig. 5.4.2). The RMSD values of the 20 NMR conformers to their mean are 0.47 for the backbone and 0.59 for all well defined heavy atoms (Table 4.2). The global displacements for the backbone of the NMR structure show little variation, with peaks for residues 44–45 and the terminal residues, and of course for the flexible loop 54–62. Large side-chain displacements occur mostly in this loop.

A similar picture is shown in Fig. 5.4.2 for the B-factors of the X-ray structure. The pronounced peak for the flexible loop is flanked by particularly low B-factor values for the backbone of residues 49–53 and 68–69. Most of the protein–protein contacts in the crystal lattice affect residues 45, 49–50, 58, 60, 71, 76–78 and 81.

Fig. 5.4.2 shows that except for the less well defined backbone segments and a

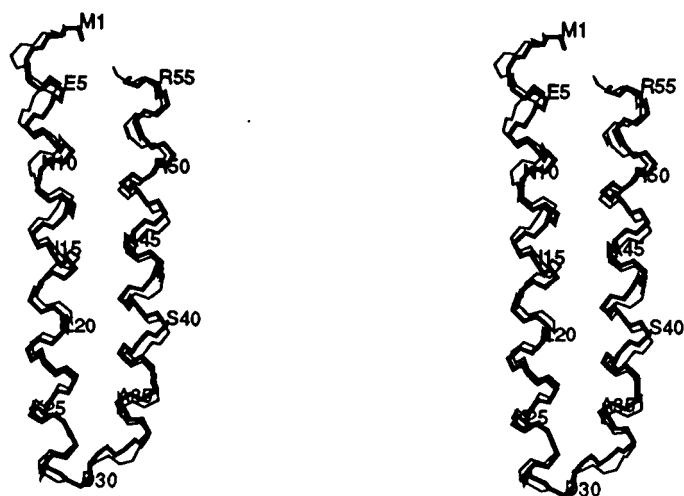


Fig. 5.5.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of the ColE1 rop protein. The superposition was calculated for selected residues identified in Fig. 5.5.2.

few residues (30, 38, 79–81), the deviation from the X-ray structure to the mean NMR structure exceeds that from the individual NMR conformers to the mean NMR structure by approximately a factor of two. This factor is also found for the RMSD values between the mean NMR structure and the X-ray structure (Table 4.2). Interestingly, the largest deviations between the two structures is sometimes within regular secondary structure elements, e.g. Glu33 in the helix, or Ile49 and Val70 in the last two  $\beta$ -strands. No obvious explanation is at hand for the first difference, except possibly the use of hydrogen bond constraints for the NMR structure: The hydrogen bond from Glu33 to Lys37 occurs in all NMR conformers, while it is absent in the X-ray structure. The other two differences may be explained by the numerous protein–protein contacts in the crystal lattice between Val50 and Asp71, and between Ile49 and residues 76–77. Contacts from residues 78 and 81 go to residues 58 and 60, and thus to the flexible loop. Other crystal contacts are between the glutamic acids 45 of two neighbouring molecules. These contacts may explain the most significant structural deviations for the side-chains, i.e. for Ile49 and in the loop 72–75, Glu33 again being an exception.

The agreement between the two structures of the backbone dihedral angles is very good, except for some residues in the loop 72–74, where crystal contacts are also likely to influence the local structure. As with the  $\alpha$ -amylase inhibitor tendamistat and with barnase, it is noteworthy that crystal lattice contacts again involve the active site with residues 58–60.

### 5.5 *ColE1 rop protein*

The ColE1 repressor of primer (*rop*), a 63 residue long protein, controls the replication of the plasmid pBR322. It consists of two antiparallel helices with residues 3–28 and 32–56, and exists as a homodimer forming a four-helix bundle



(Fig. 5.5.1). The X-ray structure (Banner *et al.* 1987) gives coordinates for the first 56 residues. The NMR structure reports coordinates for all 63 residues, but the C-terminal tail is not structured (Eberle *et al.* 1991). A brief comparison of the two structures is found in Eberle *et al.* (1991).

The extensive contact area between the two monomeric unit adds a new aspect to the evaluation of the NOEs for the determination of the NMR structure. In addition to the assignment of the corresponding cross peaks to individual protons, the NOEs needed to be attributed either to protons of the same monomer or to protons located on different subunits of the dimer. For the determination of the monomer conformation, 397 NOEs were observed. The dimer structure was determined with the help of 30 additional, intermolecular NOEs. Conformers of the dimers were obtained by docking two identical copies of the monomer. These were then subjected to molecular dynamics calculations without imposing symmetry constraints. For the final structure, the conformational difference among monomeric units of the same dimer is smaller than between monomeric units of different dimers. The selection of the best defined fragments of the NMR structure yields a RMSD value of 0.59 Å for the backbone of residues 5–14, 16–27, 32–50 and 52. As expected, the most disordered regions are the two chain termini and the loop connecting the two helices (Fig. 5.5.2). The addition of 18 side-chains located in the helices (see Fig. 5.5.2) increases the RMSD only marginally to 0.65 Å.

The distribution of local precision in the crystal structure as described by the B-factors follows the same pattern as in the NMR structure, i.e., the best defined regions are again the helices (Fig. 5.5.2). Difficulties in reproducing the unit cell parameters from crystal to crystal were explained by the formation of layers of Rop molecules whose relative position varies. Thus, intermolecular interactions in the crystal should occur mostly within the dimers.

Due to the simple tertiary fold, the structure comparison reveals only few differences. Many of the apparent backbone differences shown in Fig. 5.5.2 are located at the termini or in the loop connecting the two helices (residues 29–31). They thus correspond to the least well determined regions in both the NMR and the crystal structure. Significant differences are observed for residues 6–7, 32–33, 36, 43 and 46–47. Many of them correlate with crystal contacts between different dimer molecules involving residues 2–4, 7, 16, 20, 24, 32 and 35–36 (Fig. 5.5.2). In contrast to some other calculations of NMR structures (e.g. the two interleukins discussed later), the present structure determination made use of the Coulomb potential during the final steps. This might explain the salt bridge between the side-chains of Asp43 and Arg50 which is not found in the crystal structure. The consequence is a slight compression of the helix, which is responsible for the differences of the backbone of residues 43 and 46–47.

The largest apparent differences of side-chains between the two structures are observed for exposed residues in the monomer, namely residues 7, 10–11, 13, 16, 18 and 25 on the first helix, residues 36, 39, 43, 47 and 50 on the second helix, and residues 30 and 32 in the loop (Fig. 5.5.2). With the exception of two residues, these side-chains are also the ones with the largest disorder in the two structures

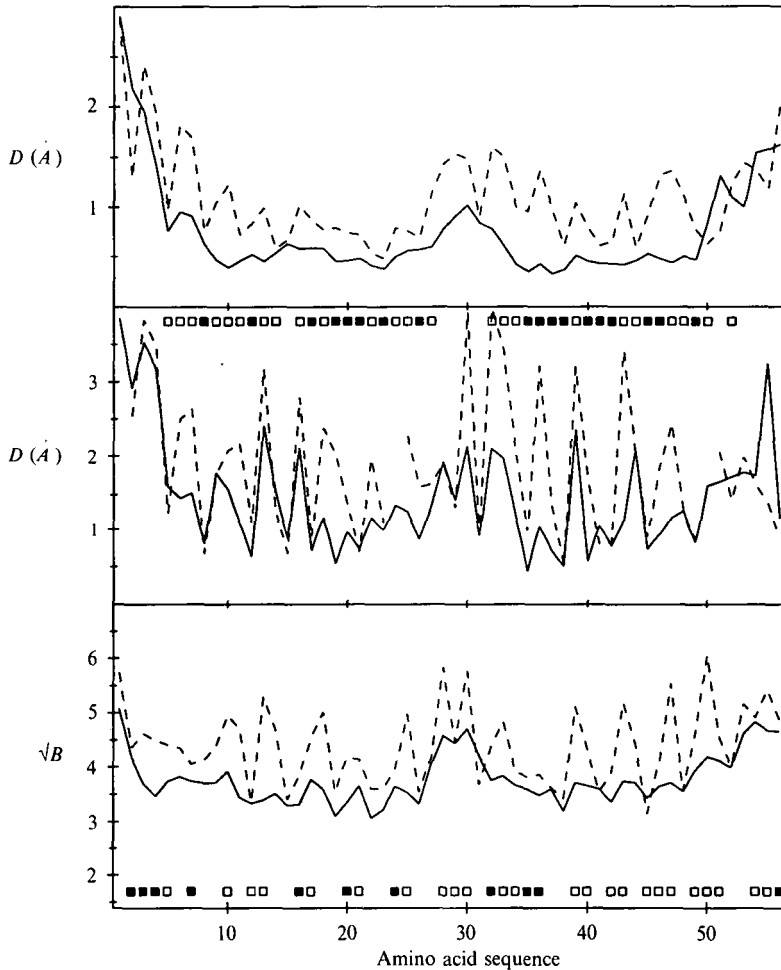


Fig. 5.5.2. Displacements and square roots of the B-factors are plotted versus the sequence for the ColE1 rop protein. See the caption to Fig. 5.1.2 for the information in the three panels and the meaning of the different line types and squares.

(Figs. 5.5.2). The exceptions are Thr7 and Asp36 which form inter-dimer hydrogen bonds.

Due to the helical character, the distribution of backbone torsion angles is very uniform, and similar in both structures. A genuine difference concerns the  $\chi^1$ -angle (not shown) of Val19. The two methyl groups, which are located in the interior of the monomeric unit, are rotated by about  $120^\circ$ . This side-chain is among the best defined in both structures. Other differences in  $\chi^1$  are observed for the surface residues 30, 40 and 50, all of which are not particularly well defined in the two structures.

### 5.6 Interleukin $1\beta$

With 153 residues, interleukin  $1\beta$  is the largest protein presented in this comparison study. As a member of the cytokine family, it is produced mainly by

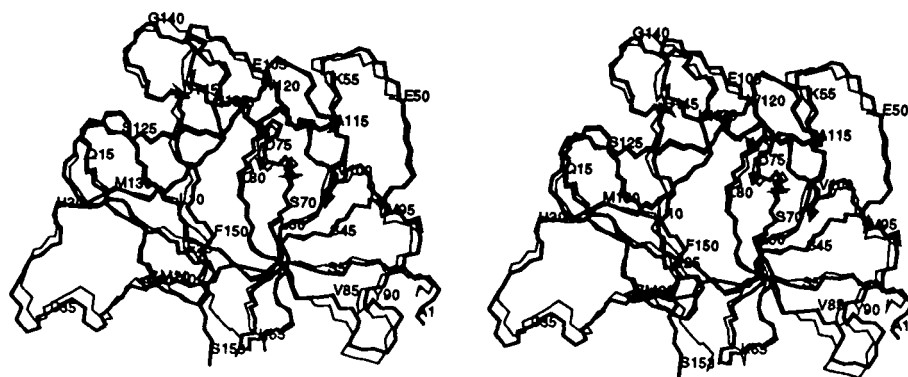


Fig. 5.6.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of interleukin  $1\beta$ . The superposition was calculated for selected residues identified in Fig. 5.6.2.

phagocytes in response to infections. Several research groups have determined the structure of this protein independently: three X-ray structures (Finzel *et al.* 1989; Priestle *et al.* 1989; Veerapandian *et al.* 1992) and one NMR structure (Clare *et al.* 1990*a*; 1991) have appeared at nearly the same time. A detailed comparison between the NMR and the X-ray structure was given by Clare & Gronenborn (1991*b*). The structure of interleukin  $1\beta$  (Fig. 5.6.1) consists of 12  $\beta$ -strands (residues 5–12, 17–22, 25–30, 40–52, 55–62, 67–74, 77–85, 99–106, 109–114, 120–125, 130–135 and 143–151) and a  $3_{10}$  helix (residues 33–39). All pairs of strands are oriented in an antiparallel fashion, and the peptide bond before Pro90 is in a *cis* conformation.

The NMR structure was derived from 2780 distance constraints and 366 dihedral angle constraints. The RMSD values for the complete backbone and for all heavy atoms, 0.48 Å and 0.85 Å, are the smallest ones in Table 4.2. After the automatic selection of backbone fragments with 118 residues and of 67 side-chains (see Fig. 5.6.2) these RMSD values drop to 0.26 Å and 0.29 Å, respectively. The NMR structure is thus particularly well defined. The normalized curve for the global displacements between the NMR conformers and their mean structure (Fig. 5.6.2) correlates well with the distribution of the  $\beta$ -strands.

From the three crystal structures, the one by Priestle *et al.* (1989) was chosen here, because coordinates for all atoms are included in this structure, and no second conformation of any residue is given. The lack of coordinates for the first two residues in the other two crystal structures, and the presence of two conformations for residue 64 in the structure by Finzel *et al.* (1989) indicates some flexibility in the conformation of these residues. Otherwise, the three structures are very similar with backbone RMSD differences for residues 3–62 and 66–153 between 0.31 and 0.36 Å. The normalized curve for the square roots of the B-factors shows again a strong correlation with the regular secondary structure elements (Fig. 5.6.2).

With 0.64 Å and 0.74 Å, the RMSD values between the NMR and the X-ray structure for the selected regions of the protein are almost 2.5 times larger than the

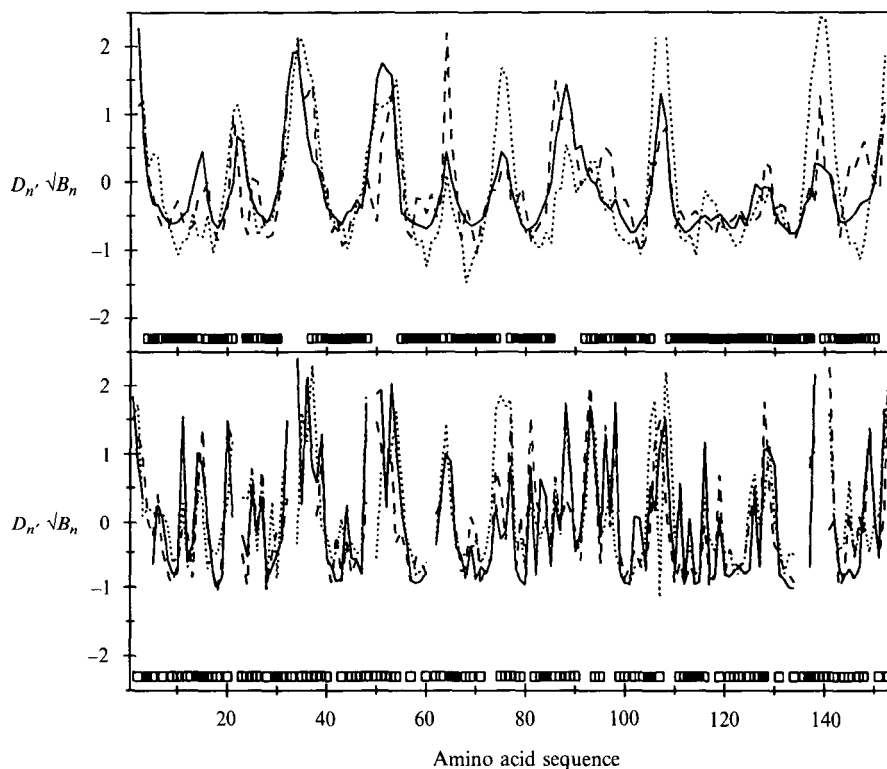


Fig. 5.6.2. Normalized displacements  $D_n$  (equation 3 *a*) and normalized square roots of the B-factors  $\sqrt{B_n}$  (equation 3 *b*) are plotted versus the sequence for interleukin 1 $\beta$ . See the caption to Fig. 5.1.3 for the information in the three panels and the meaning of the different line types and squares.

corresponding RMSD values describing the NMR structure. The superposition in Fig. 5.6.1 shows that significant deviations of the two backbones occur for terminal residues and for loops located on the periphery of the protein. The normalized displacements and B-factors of Fig. 5.6.2 show that the variation of the B-factors and of the NMR displacements coincide, and follow the pattern of the secondary structure. Differences in the backbone conformation that are not explained by lower precision in one or both of the two structures occur for the following residues. Glu64 can adopt a second conformation in the crystal environment (see above). The carbonyl oxygen of Asp86 hydrogen bonds to the amide of Asn89 in the crystal structure and to the amide of Tyr90 in the NMR structure. Thr147 makes two backbone hydrogen bonds to Arg11 in the crystal structure, whereas in the NMR structure the side-chain OH of Thr147 binds to the oxygen of Arg11.

The largest local difference as identified in a plot of backbone dihedral angles (not shown) is encountered for residues 106–108. In the crystal structure, the  $\psi$ -angles for residues 106–107 and the  $\phi$ -angles for residues 107–108 are all about  $-90^\circ$ . The NMR structure, however, has values of all these angles of about  $+60^\circ$ . Residues 106–108 are located on the surface. They are thus not among the best

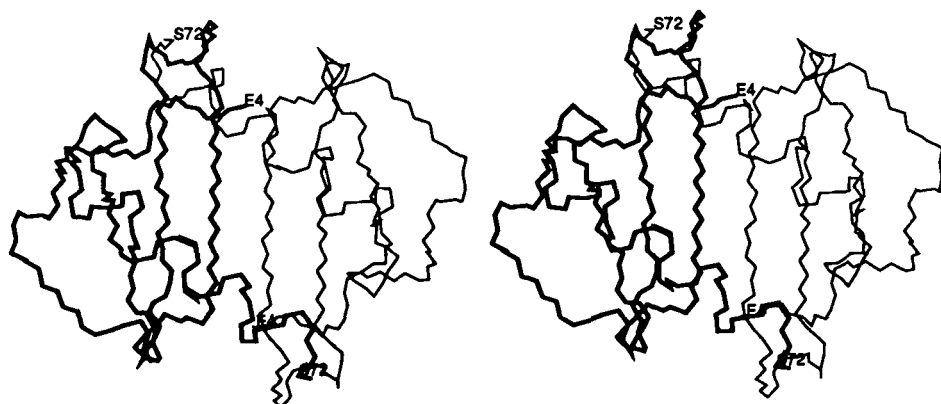


Fig. 5.7.1. Structure of the dimer of interleukin 8. The crystal structure is drawn with a heavy line for one monomeric unit and a thin line for the other.

defined parts in the NMR structure, and the B-factors of the crystal structure are the largest except for the terminal residues.

The following side-chains show significant differences of their conformation in the two structures (Fig. 5.6.2). Glu15 is located on the surface. In the NMR structure its conformation is not very well defined but tends to point away from the protein. In the X-ray structure, residues 14 and 15 are in contact with another protein molecule, thus the side-chain of Glu15 might be pushed to the side. Gln81, Lys92 and Glu128 are also surface residues. In the crystal structure they interact with other side-chains forming hydrogen bonds or salt bridges: Gln81–Lys74, Lys92–Tyr90 and Glu128–His30. In addition, Glu128 forms an intermolecular salt bridge to Lys65. The intramolecular surface interactions are not likely to be detected by NMR. Finally, the observed difference for Asn107 is due to the different backbone conformations; in fact the side-chain tries to compensate this difference by placing its peripheral group at similar positions in the two structures.

### 5.7 Interleukin 8

Interleukin 8, another cytokine, is involved in the response to inflammatory stimuli. It forms a homodimer by joining  $\beta$ -sheets from two monomeric units to a larger sheet (Fig. 5.7.1). Each subunit consists of 72 residues. These  $\beta$ -strands with residues 23–30, 36–43 and 48–51 are connected by short loops (Fig. 5.7.2). The first strands of the two subunits run antiparallel to each other, thus connecting the two  $\beta$ -sheets to form a six-stranded, fully antiparallel sheet in the dimer. The helices of the two units (residues 56–72) run antiparallel to each other on the same side of the  $\beta$ -sheet at an angle of about  $60^\circ$  to the  $\beta$ -strands. As for the ColE1 rop protein, the discrimination between intra- and inter-subunit NOEs presented an additional challenge to the determination of the NMR structure (Clare *et al.* 1990b). An interesting aspect of the determination of the X-ray structure was the use of the NMR structure as starting model to solve the phase

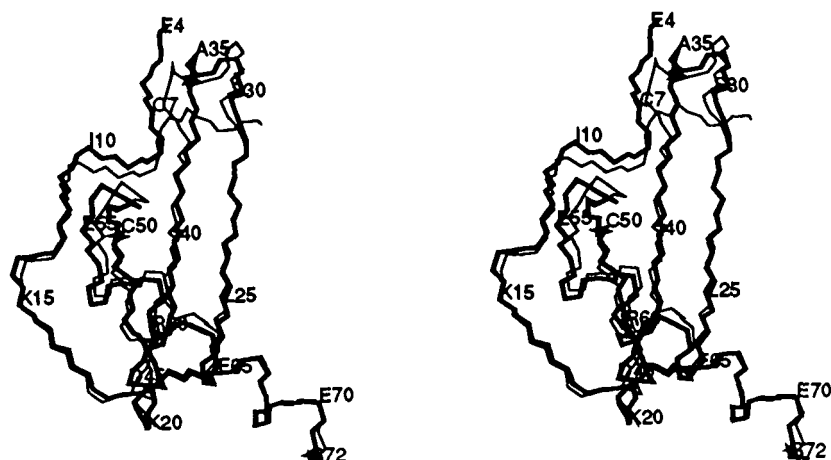


Fig. 5.7.2. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of interleukin 8 (only one monomeric unit). The superposition was calculated for selected residues identified in Fig. 5.7.3.

problem (Baldwin *et al.* 1991). Comparisons of the two structures were presented in Baldwin *et al.* (1991) and in Clore & Gronenborn (1991*c*). The NMR structure includes coordinates for residues 2–72, while the X-ray reports electron density, and thus coordinates, only for residues 4–72.

For the NMR structure, a total of 2242 constraints were used. Each subunit was defined by 847 distance constraints from NOEs, 52 constraints for hydrogen bonds and 181 dihedral angle constraints. In addition, 70 distance constraints from NOEs and 12 hydrogen bond constraints connect the two subunits (note that two constraints enforce one hydrogen bond). The automatic selection of well defined parts of the NMR structure (see Chapter 2) eliminates the backbone of residues 4–7, 9–11, 35–39, 45–46, 52–54 and 70–72. The variations of the global backbone displacements for the NMR conformers relative to their mean structure are, however, very small (Fig. 5.7.3), and exceed 0.4 Å only for a few residues at the chain termini. Similarly, nearly all global displacements for the side-chains are smaller than 2.0 Å, but the selection procedure described only the side-chains of 27 residues with displacements smaller than 0.5 Å as well defined (Fig. 5.7.3).

In the crystal structure, larger values for the B-factors of the backbone are again only given for the terminal residues (Fig. 5.7.3). The best-characterized parts include the three  $\beta$ -strands and the N-terminal half of the helix.

The RMSD value of 2.0 Å for the complete backbone between the mean NMR structure and the X-ray structure indicates the presence of some significant structural differences. For the backbone selected as well defined in the NMR structure this RMSD is reduced to 0.95 Å, which is still significantly larger than the 0.24 Å from the individual NMR conformers to their mean structure. In addition, the displacements between the NMR and the X-ray structure are all larger than those between the NMR conformers and their mean structure. The major backbone deviations between the two structures involve residues 4–10, 31–33, 35–36, 44, 53–54 and 71–72. Except for two terminal residues on each chain

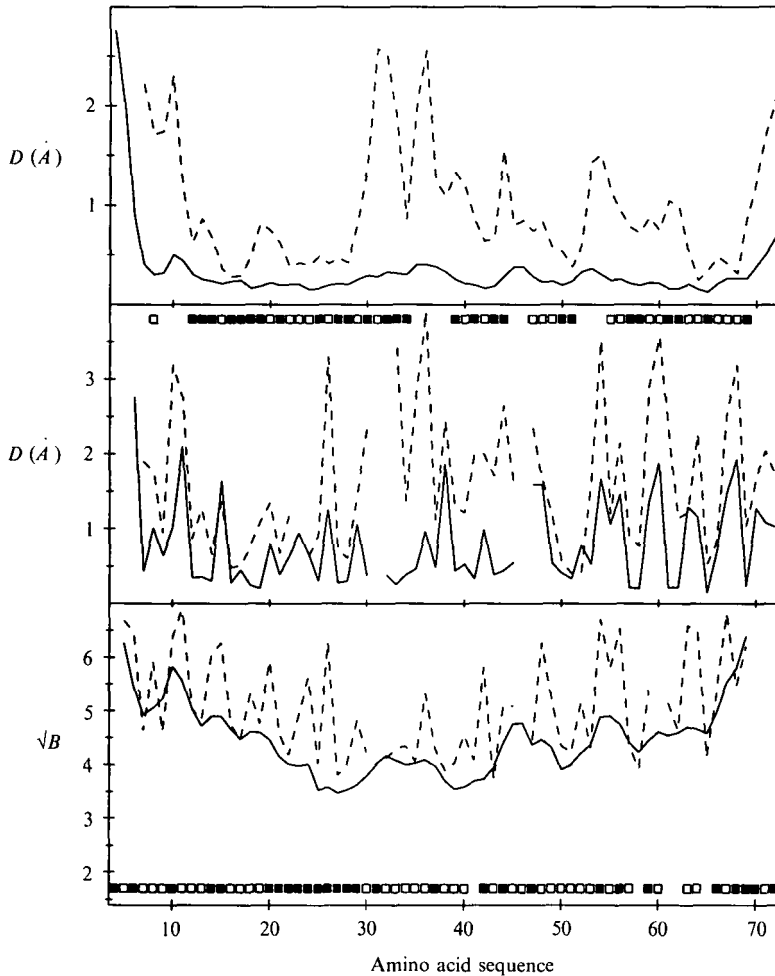


Fig. 5.7.3. Displacements and square roots of the B-factors are plotted versus the sequence for interleukin 8. See the caption to Fig. 5.1.2 for the information in the three panels and the meaning of the different line types and squares.

end, these differences exceed the variations attributed to both the NMR and the X-ray structure due to their precision. The N-terminus and residues 31–36 are located next to each other at one edge of the  $\beta$ -sheet and connected by the disulfide bridge 7–34, while the other residues are at different, exposed corners of the protein. Lattice contacts to other dimers may explain the structural variations of residues 6, 10, 44 and 54. The conformational difference of the loop 31–35 might be relevant to the binding of the protein to its receptor, which was proposed to involve His33 (Baldwin *et al.* 1991).

Differences in side-chain conformation are observed for residues 10–11, 23, 26, 32–34, 54, 60 and 68 (Fig. 5.7.3). Residues 10–11, 54, 60 and 68 have large side-chain B-factors, and are also found to be disordered in the NMR structure (residues 11, 60 and 68) or contact other protein molecules in the crystal lattice (residues 10 and 54). Residues 23 and 26 are in the first  $\beta$ -strand and thus in the

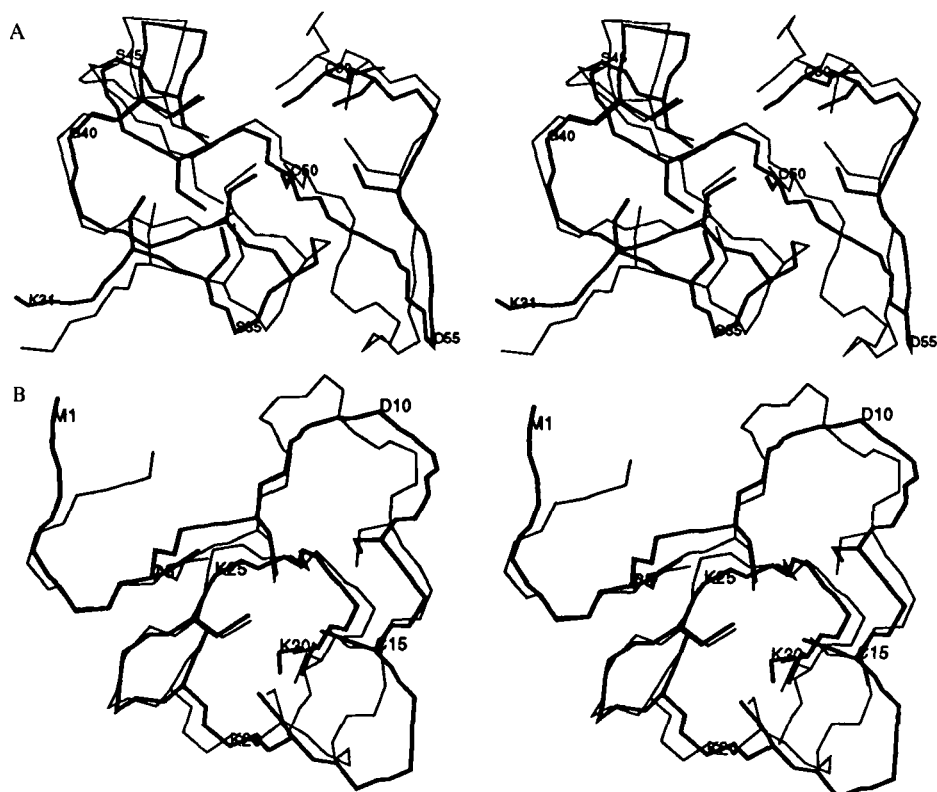


Fig. 5.8.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of rat metallothionein. The superposition was calculated for selected residues identified in Fig. 5.8.2. A:  $\alpha$ -domain. B:  $\beta$ -domain.

interface between the two monomeric units. In the NMR structure, the  $\chi^1$ -angle is well defined, but different from the one in the crystal structure by about  $120^\circ$ . As a result, the side-chain of Lys23 on one monomeric unit interacts in the NMR structure with the side-chain of Glu29 of the other monomeric unit. The two arginines 26 of the dimer are located next to each other in the middle of the dimer-forming  $\beta$ -strands. In the crystal structure, they point away from each other as expected from two positively charged groups. Possibly due to the absence of an electrostatic potential in the calculation of the NMR structure, this repulsion effect is not observed in this structure. The remaining differences cluster around His33 and are further discussed below.

Five hydrogen bonds are present in 70% or more of the NMR conformers, but not in the crystal structure. Three are between the backbone atoms of residues separated by four positions along the sequence, and in each case several of the five residues involved are in a less well defined region (carbonyls of residues 43, 67 and 68 to amides of 47, 71 and 72, respectively). They might be due to hydrogen bond constraints imposed at the end of regular secondary structure elements. The hydrogen bond from backbone of Asp44 to the side-chain of Ser45 is exposed and located on one corner of the protein. An interesting difference between the



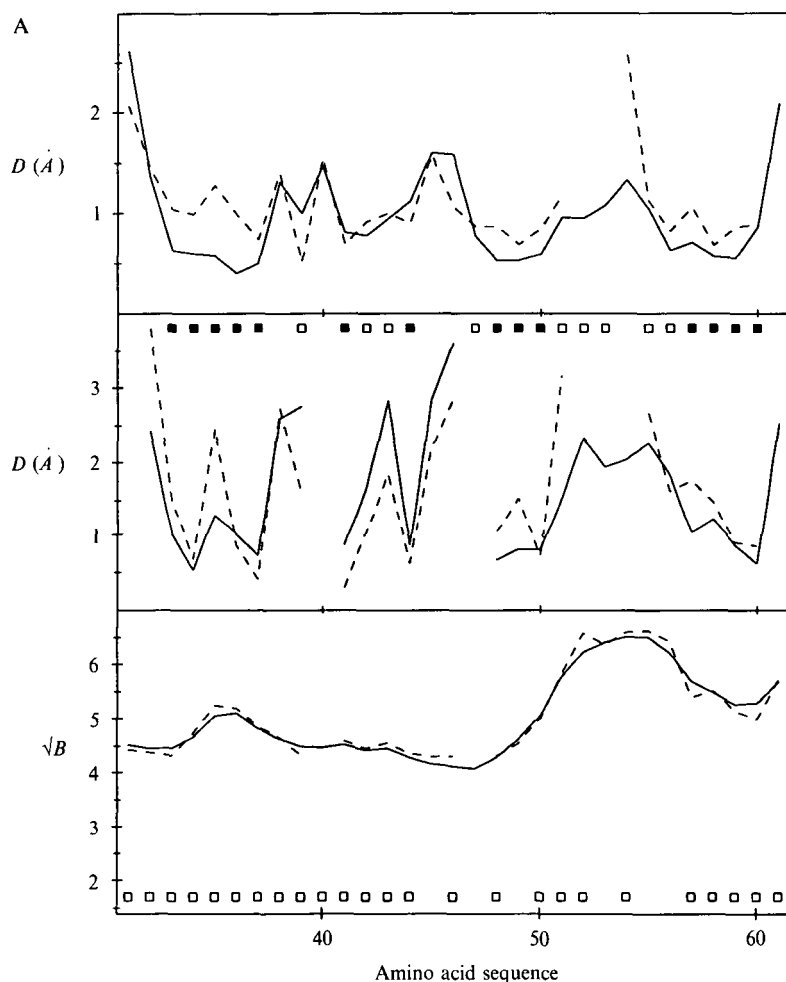
solution and crystal structure is observed for the aforementioned His33. The conformation of this side-chain is well defined in both structures, but forms different hydrogen bonds in the NMR structure (to the backbone amide of Gln8) and in the crystal structure (to the carbonyl of Glu29) as pointed out by Baldwin *et al.* (1991). Unusually low  $pK_a$  values of 3.7 and 4.9 were measured for His18 and His33, respectively (Clare *et al.* 1990b). These can be explained by the environment of the imidazole ring in the NMR structure but not in the crystal structure.

Baldwin *et al.* (1991) and Clare & Gronenborn (1991c) pointed out that a major structural difference is observed when the dimers are superimposed. In particular, the distances between the helices decreased from 14 Å–12 Å when going from the NMR to the X-ray structure. The authors note that the crystal structure of the dimer is inconsistent with the NMR constraints. They propose a model in which the NMR structure represents an 'open' form for receptor binding and the X-ray structure a 'closed' form (Clare & Gronenborn, 1991c). This would imply that both structures describe naturally occurring conformational states of the protein.

### 5.8 Metallothionein

Metallothionein from rat is a small protein with 61 residues that can bind up to seven metal ions. It consists of two domains with little contact between them. The  $\alpha$ -domain is formed by residues 31–61 and the  $\beta$ -domain by residues 1–30 with two lysines 30 and 31 forming an extended linking region. The  $\beta$ -domain binds three metal ions, which are, in the case of the crystal structure (Robbins *et al.* 1991), one cadmium and two zinc atoms, and in the NMR structure (Schultze *et al.* 1988) three cadmium atoms. The  $\alpha$ -domain binds four cadmium ions in both structures. Of particular interest in the present context is the history of the structure determination of rat metallothionein. The first crystal structure (Furey *et al.* 1986) showed significant differences to the solution structure (Schultze *et al.* 1988), in particular with respect to the metal co-ordination. Confidence in the NMR structure was gained due to its close similarity to the solution structure of the homologous metallothionein from rabbit (Arseniev *et al.* 1988). Re-investigation of the X-ray structure yielded a revised protein architecture which closely corresponds to the NMR structure (Robbins *et al.* 1991). A comparison of the two most recent structures is given in Braun *et al.* (1992). Almost no experimental data is provided by NMR describing the relative position and orientation of the two domains; therefore separate structure comparisons are given for the  $\alpha$ - and the  $\beta$ -domain. Fig. 5.8.1, A and B, shows superpositions of the backbones and the cysteines of the NMR and the X-ray structures. The cysteine residues, which co-ordinate the metal ions, are added because they are among the best defined parts of the protein in the NMR structure. Regular secondary structure elements, which usually form the well defined parts of a protein, are not present in metallothionein.

This lack of regular secondary structure explains the limited quality of the NMR structure. For example, no hydrogen bond is found consistently in all 10



For legend see facing page.

conformers representing the NMR structure. With the selection procedure of Chapter 2 applied to the  $\beta$ -domain, a RMSD from the NMR conformers to their mean of  $0.86 \text{ \AA}$  is obtained for the backbone of residues 3–8, 13–15 and 18–29 (Table 4.2). Adding 15 side-chains yields a value of  $1.05 \text{ \AA}$ . The corresponding numbers for the  $\alpha$ -domain are  $0.83 \text{ \AA}$  for the backbone of residues 33–37, 39, 41–44, 47–53 and 55–60, and  $0.87 \text{ \AA}$  when 14 side-chains are added. All 19 cysteines are among the best defined regions of the protein, while the chain termini and the loops between cysteine residues exhibit an increased structural spread (Fig. 5.8.2).

The curves for the square roots of the B-factors are very similar for the backbone and the side-chains (Fig. 5.8.2). These curves do not correlate with the distribution of the cysteines, instead large B-factors occur for residues 1–4 and 51–56. Increased values are also observed for residues 35–36, the latter being a cysteine, and for the side-chains of residues 18, 20 and 22.

Differences between the NMR and the crystal structure, that are in regions

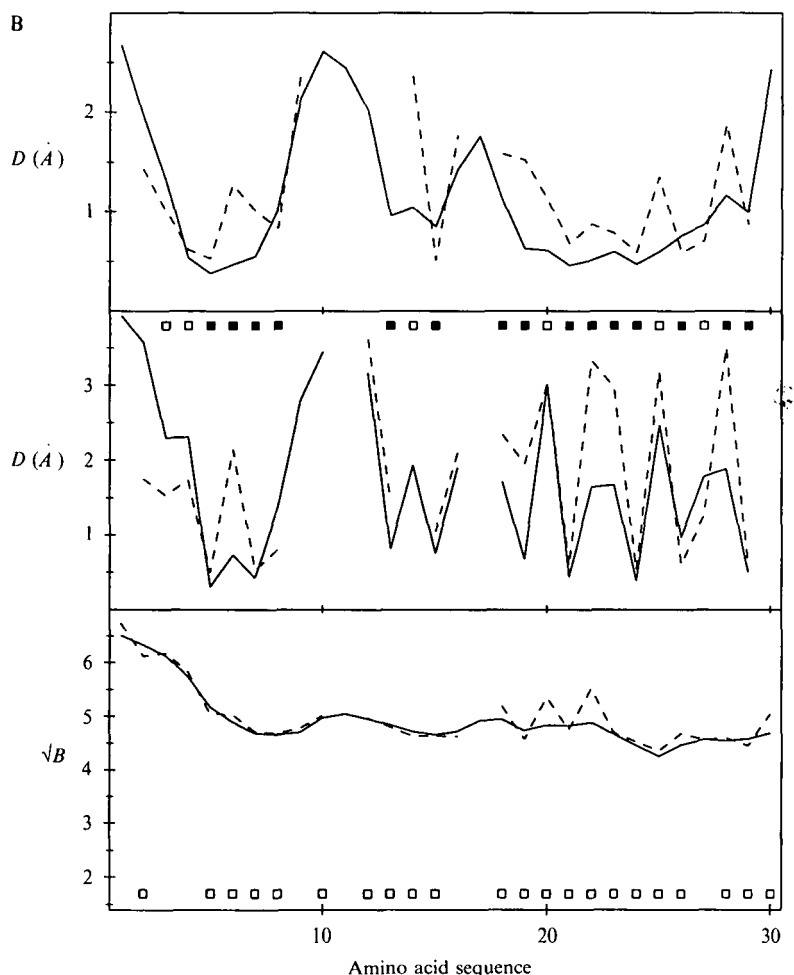


Fig. 5.8.2. Displacements and square roots of the B-factors are plotted versus the sequence for rat metallothionein. See the caption to Fig. 5.1.2 for the information in the three panels and the meaning of the different line types and squares. A:  $\alpha$ -domain. B:  $\beta$ -domain.

where both structures are well defined, involve the backbone and side-chain of Ser6, the side-chain of Thr9, Cys13 and Ser14, Gly17, Cys19, the side-chains of Lys22 and Gln23, Lys25 and Ser28 in the  $\beta$ -domain. In the  $\alpha$ -domain, the backbone of the loop with residues 34–37 is slightly shifted between the two structures.

It is interesting to note that the domain with more differences between the solution and the crystal structure, the  $\beta$ -domain, is also the one with different ions in the two structures, i.e. one cadmium and two zinc atoms in the crystal structure and three cadmium in the solution structure. Since the zinc atoms bind to cysteines 7, 15 and 24 in the X-ray structure, the presence of the different atoms may explain some of the local variations in the  $\beta$ -domain. Another explanation is again offered by the crystal contacts which involve residues 9–10, 14–15, 18, 21–24, 30–31, 40, and 45–47 (Robbins & Stout, 1991). One should however also

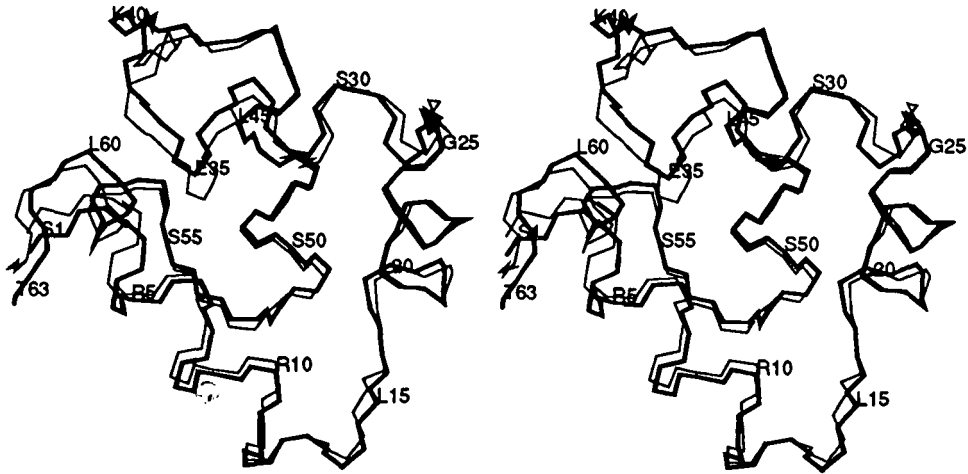


Fig. 5.9.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of the 434 repressor. The superposition was calculated for selected residues identified in Fig. 5.9.2.

keep in mind that, mostly due to the lack of regular secondary structure, the NMR structure is not of very high quality, so that differences may also be due to the limited precision of the solution structure.

### 5.9 434 repressor(1–69)

The N-terminal domain of the 434 repressor (called ‘434 repressor(1–69)’ due to the residues forming this domain) is a typical representative of the DNA-binding proteins with a helix-turn-helix motif. A wealth of biochemical and genetic data is available for the free and the DNA-bound protein. X-ray structures of both the free 434 repressor(1–69) (Mondragón *et al.* 1989) and of the complexes with two DNA operator sequences (Anderson *et al.* 1987; Aggarwal *et al.* 1988) are available. Very recently, the NMR structure of this protein was presented together with a detailed comparison of the solution and the crystal structure of the free protein (Neri *et al.* 1992). Since very similar comparison tools were used in this publication for the structure comparisons, the following section will only summarize these results. The 434 repressor(1–69) (Fig. 5.9.1) consists of five helices (residues 2–13, 17–24, 28–35, 45–52 and 56–60) and a flexible C-terminus. In fact, both structures define a unique conformation only for the segment 1–63. The helix-turn-helix motif responsible for the specific DNA-binding consists of residues 17–35.

A particular feature of the NMR structure is the large number of coupling constants, both homonuclear and heteronuclear (Neri *et al.* 1992). 103 coupling constants were obtained, 46 of which determine the side-chain torsion angle  $\chi^1$  for the 69 residues of this protein. The RMSD values of about 0.4 Å, calculated for the backbone of residues 3–33, 35 and 43–59 and for the addition of 19 side-chains (see Fig. 5.9.2), indicate a well defined structure (Table 4.2). Lower precision is

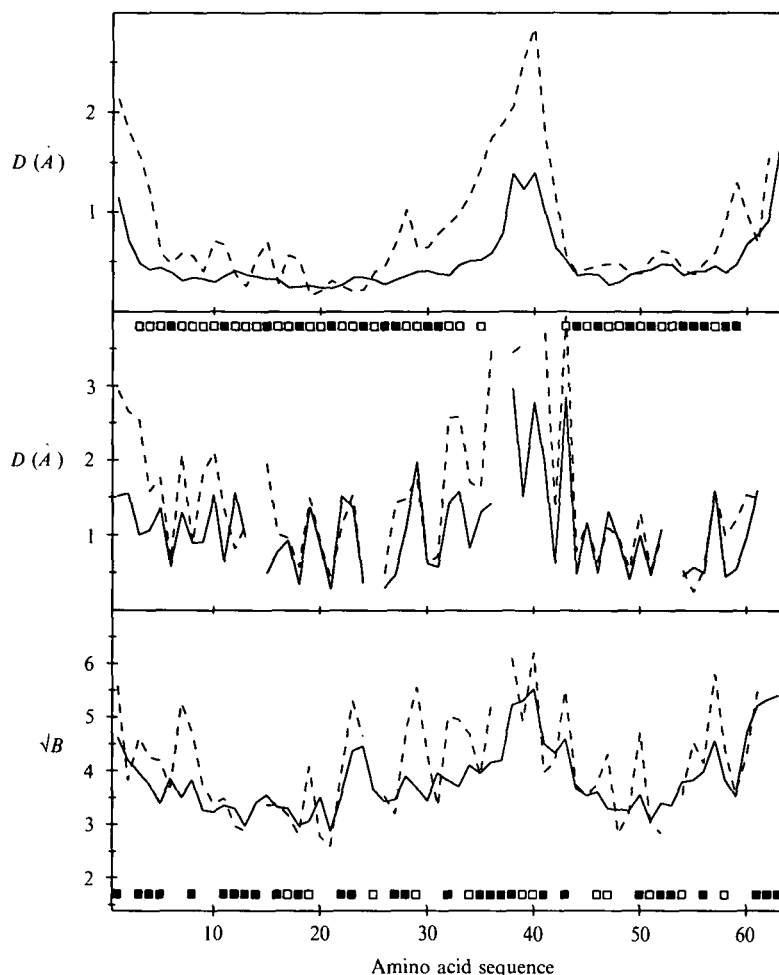


Fig. 5.9.2. Displacements and square roots of their B-factors are plotted versus the sequence for the 434 repressor. See the caption to Fig. 5.1.2 for the information in the three panels and the meaning of the different line types and squares.

achieved for two residues on each chain end and, more importantly, for the loop 36–42 (Fig. 5.9.2).

In the X-ray structure, the same residues that exhibit a larger structural spread in the NMR structure also show increased B-factors (Fig. 5.9.2). A feature pointed out by Neri *et al.* (1992) is the very short crystal contacts obtained from the coordinates. Five distances between heavy atoms of different protein molecules are shorter than 2 Å, the shortest being 1.57 Å between the side-chain oxygen atoms of Ser3 and Thr63.

For large parts of the protein, the displacements between the X-ray and the NMR structure are similar to the displacements among the NMR conformers (Fig. 5.9.2). The DNA-binding helix-turn-helix motif (residues 17–36) is among those parts with good structural coincidence between the solution and the crystal form. Differences in backbone structure within the well defined parts are observed

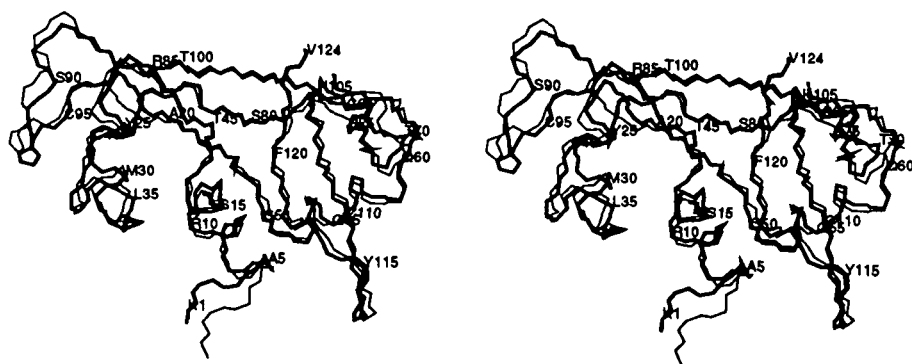


Fig. 5.10.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of ribonuclease. The superposition was calculated for selected residues which are identified in Fig. 5.10.2.

for several N-terminal residues, for the residues flanking the less well defined region 38–40, and for the residues 58–60 of the last helix. In these regions, the following residues show a different conformation of the side-chain. Lys7 has large B-factors. Lys9 and Arg10 form hydrogen bonds or salt bridges to Gln12 and Glu35, respectively, only in the crystal structure. NMR is unlikely to detect these surface interactions.

Neri *et al.* (1992) explained the genuine differences involving the N-terminal residues, the residues just before and just after the flexible loop 38–40, and the residues of the last helix by the presence of short protein–protein contacts in the crystal lattice. Contacts shorter than 3 Å are observed for residues 1, 3–5, 8, 11–14, 16, 18, 22–23, 27–28, 32, 35–38, 41, 43, 50, 52–53, 56 and 61–63. According to this hypothesis, the first helix would be pressed against the rest of the protein, which would mainly affect its free N-terminal part. A consequence of this could be the different hydrogen bonds from the amide of Gly14 which go to Arg10 in the NMR structure and to Ile11 in the crystal structure. Another difference in the hydrogen bonding involves residues 27–30, since the interaction Thr27–Ser30 was observed only in the NMR structure. The most disordered region in the NMR structure, residues 38–40, might be stabilized in the crystal structure by lattice contacts. Finally, it appears that the last helix is slightly shifted due to the contacts at both ends.

### 5.10 Ribonuclease

Bovine ribonuclease is one of the best studied proteins by X-ray diffraction on single crystals (for a review see Wlodawer, 1985). It was the third enzyme whose crystal structure was solved. The fold of bovine ribonuclease (Fig. 5.10.1) combines three helices (residues 3–13, 24–34 and 50–60) with two antiparallel  $\beta$ -sheets (first sheet: residues 41–48, 79–91 and 94–104, with a bulge for residues 88–89; second sheet: residues 61–64, 71–75, 105–113 and 114–124, with residue 120 not properly belonging to the sheet; see Protein Data Bank entry 7RSA describing the crystal structure of Wlodawer *et al.* (1988)).

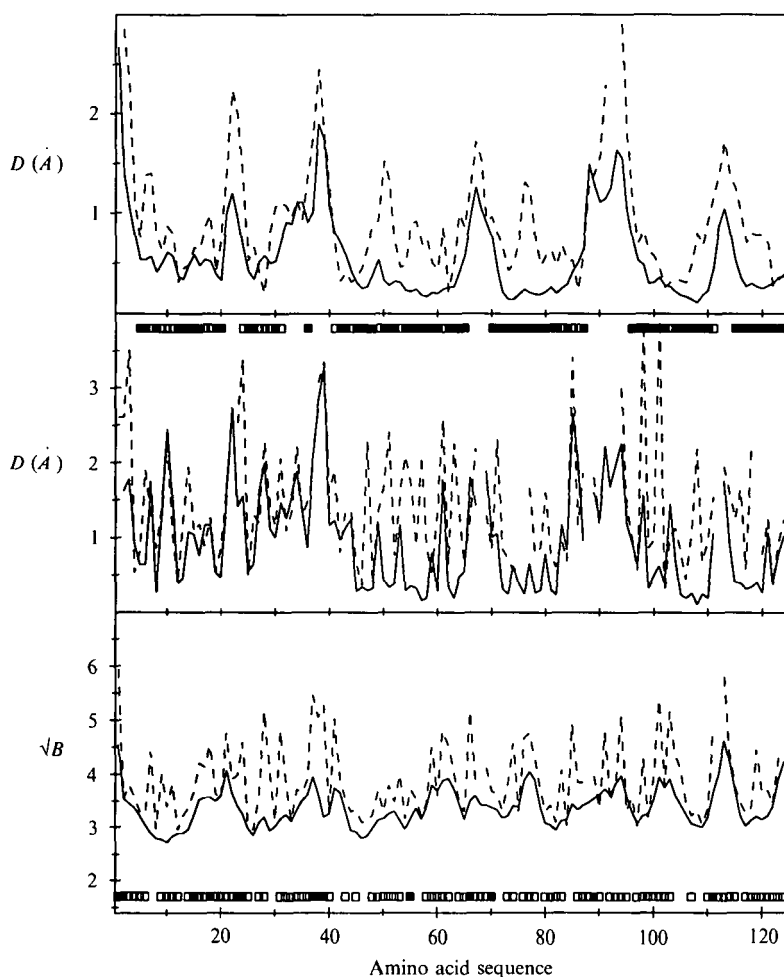


Fig. 5.10.2. Displacements and square roots of the B-factors are plotted versus the sequence for ribonuclease. See the caption to Fig. 5.1.2 for the information in the three panels and the meaning of the different line types and squares.

The NMR structure (Rico *et al.* 1991) was determined under conditions where the enzyme was saturated with phosphate. The automatic selection of well defined regions in the NMR structure excludes only residues belonging to surface loops (see Fig. 5.10.2). The RMSD values for the selected residues are 0.44 Å for the backbone, and 0.52 Å when 72 side-chains are added (Table 4.2). The helix 50–60, the  $\beta$ -strands 61–64, 71–75 and parts of 105–113 are the best defined parts of the solution structure (Fig. 5.10.2); they build one corner of the molecule (Fig. 5.10.1). The publication of the NMR structure includes also a comparison to different crystal structures of this protein (Rico *et al.* 1991).

The crystal structure used here is the best-resolved crystal structure of this study with a resolution of 1.26 Å and an R-factor of 0.15 (Wlodawer *et al.* 1988). Another high-resolution crystal structure was reported earlier by Borkakoti *et al.* (1982; resolution: 1.45 Å); differences between these two crystal structures reveal

additional insight into apparent differences between the crystal and the solution structure. In contrast to the solution structure, which was saturated with phosphate, the crystal structure is free of phosphate. Particular features of this crystal structure are the large number of water molecules (188) and the description of two alternative conformations for a total of 13 side-chains. Only two of these side-chains, Ser50 and Lys61, are well defined in the NMR structure. All B-factors of the backbone of this crystal structure are smaller than 25 Å<sup>2</sup> (Fig. 5.10.2).

For the residues that are well defined in the NMR structure, the RMSD values between the NMR and the X-ray structure amount to 0.76 Å for the backbone and 1.25 Å when the 72 selected side-chains are added. With the exception of Gln101, all side-chains with large B-factors are disordered in the NMR structure.

With the above precision of the individual structures in mind, the most significant differences are observed for the backbone preceding the helix 50–60, for the backbone of residues 76–77 and for the side-chains of Tyr76, Gln101 and His119. The orientations of the peptide planes of the surface residues 48 and 49 are not well characterized in the NMR structure (Rico *et al.* 1991), and thus result in some differences relative to the crystal structure. For the helix 50–60, particularly small displacements describing the NMR structure are obtained (Fig. 5.10.2). The displacements from the NMR structure to the X-ray structure, which are between 0.4 and 0.8 Å, correspond to a small translation of this helix along its axis.

In solution, Tyr76 forms a hydrogen bond to the backbone of Lys61, which is not present in the X-ray structure. This interaction may be correlated with the difference in the backbone structure of residues 76 and 77 or to the observation of two conformations for Ser77 in the crystal structure. The side-chain of Gln101 forms hydrogen bonds to Ala20 and to Thr82 in all NMR conformers, but not in the X-ray structure. The position of its amide group in the NMR structure is replaced by water molecules in the X-ray structure, which also form the above mentioned hydrogen bonds. A conformation similar to the NMR structure was reported for earlier crystal structures and was also in the starting model for the X-ray refinement. Attempts to model the side-chain according to this older conformation failed, even when allowing for two conformations (Wlodawer *et al.* 1988). The authors of the crystal structure express however their belief that the conformation of Gln101 depends on the conditions of crystallization. Finally, the side-chain of His119 occupies very different space due to a rotation of  $\chi^1$  of about 180°. A similar explanation as for Gln101 is proposed by Rico *et al.* (1991) for this difference. They observed an NOE from His119 to Val118 under the experimental conditions for which the NOE data set was collected. At higher pH, an additional NOE to Asp121 was measured. Since all these residues are part of a  $\beta$ -strand, the presence of both NOEs can only be explained by a dynamic equilibrium in which the histidine ring would move between the two distinct conformations given in the NMR and in the X-ray structure. The crystallization conditions were indeed at higher pH than the NMR experiments (Wlodawer *et al.* 1988; Rico *et al.* 1991). Moreover, the crystal structure of Borkakoti *et al.* (1982) reported both conformations of His119 as alternatives for which electron density is present.



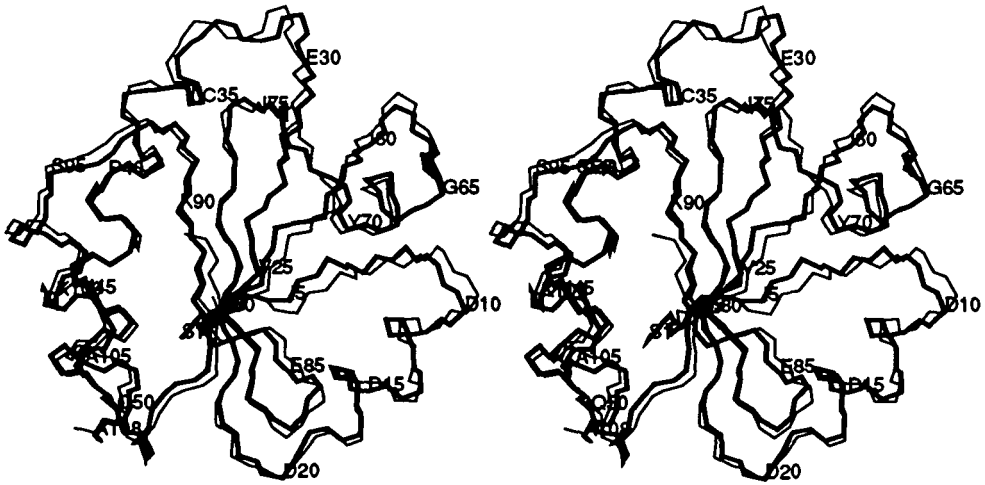


Fig. 5.11.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of *E. coli* thioredoxin. The superposition was calculated for selected residues identified in Fig. 5.11.2.

### 5.11 Thioredoxin

The small enzyme thioredoxin from *E. coli* (108 amino acids) provides a redox couple due to the presence of the disulfide bridge closing the loop Cys32–Gly–Pro–Cys35, which can reversibly be oxidized. The following comparison of *E. coli* thioredoxin is the only one of this study for two non-identical structures, since the protein could be crystallized only in its oxidized form, i.e. in the presence of a disulfide bridge between Cys32 and Cys35 (Katti *et al.* 1990), while the NMR structure was obtained for the reduced form (Dyson *et al.* 1990). A central five-stranded  $\beta$ -sheet (residues 4–7, 53–59, 22–29, 76–81 and 88–91) is surrounded by three helices (residues 11–18, 34–49 and 95–108), and another helical structure consists of residues 64–70 (Fig. 5.11.1).

In solution, a well-defined structure was obtained for the backbone of 76 residues (see Fig. 5.11.2). In contrast to other structures, this does not include all residues in the regular secondary elements. All helices and the first  $\beta$ -strand contain residues with an increased structural spread of the backbone. The RMSD for the selected backbone fragments between the NMR conformers and their mean structure is 0.37 Å; this RMSD increases slightly to 0.43 Å when 39 side-chains are added (Table 4.2). Of interest is the absence of most of the active site tetrapeptide 32–35 from the well defined fragments. It should be noted that with the exception of only ten residues (1–3, 8–9, 20, 62–64, 108) the global backbone displacements among the NMR conformers are below 0.7 Å (Fig. 5.11.2).

The crystal structure consists of two molecules per asymmetric unit with different contacts in the crystal lattice. For the molecule considered mostly in this comparison (molecule A in Katti *et al.* 1990), the first helix and the following loop (residues 11–21) are in contact with other protein molecules in the crystal lattice, while the second molecule shows more intermolecular contacts from its

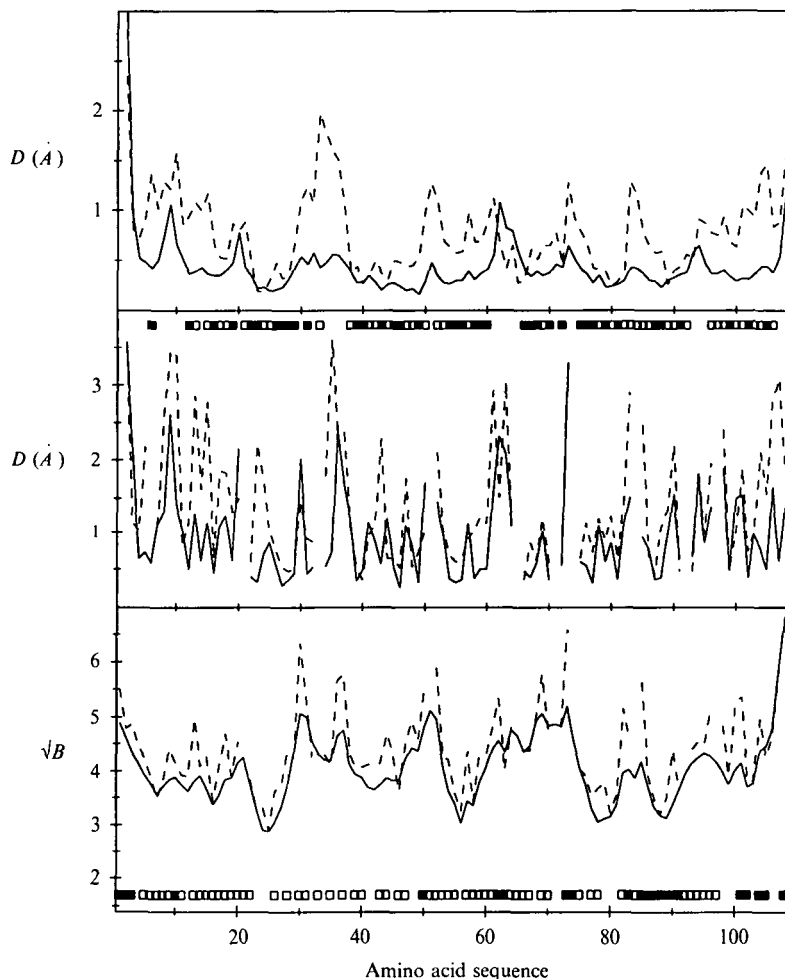


Fig. 5.11.2. Displacements and square roots of the B-factors are plotted versus the sequence for *E. coli* thioredoxin. See the caption to Fig. 5.1.2 for the information in the three panels and the meaning of the different line types and squares.

C-terminal helix. In all molecules the last  $\beta$ -strand runs in an antiparallel fashion to the same strand of another molecule, thus extending the  $\beta$ -sheet. In addition, copper ions connect the first two residues of one molecule with residue 10 of another. The backbone RMSD between the mean NMR structure and the X-ray structure is 0.66 Å, but after excluding residues 10–14, 17–21, 49–52, 82–86 and 104–108, the two forms become very similar with a backbone RMSD of 0.24 Å. Two conformations are given for the side-chains of His6, Ile23 (only molecule A), Asp47, Glu48 (only molecule A), Leu58 and Leu80.

The global displacements in Fig. 5.11.2 between the mean NMR structure and the X-ray structure indicate the following differences. The structural difference for residues 5–15 and 93–107 is paralleled by different conformations of the two molecules in the asymmetric unit of the crystal. Furthermore, residues 1, 2 and 10 are affected by the presence of copper ions in the crystal. The differences in the

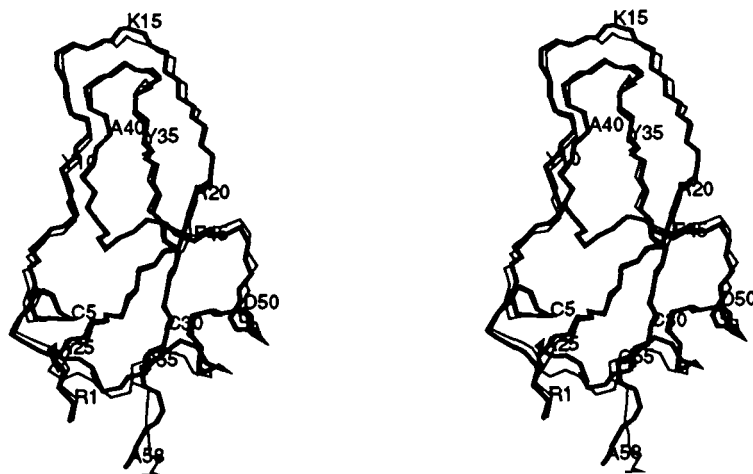


Fig. 5.12.1. Stereo view with the backbones of the mean NMR structure (thick line) and of the X-ray structure (thin line) of BPTI. The superposition was calculated for selected residues identified in Fig. 5.12.2.

loop preceding the last  $\beta$ -strand may be explained by the extension of the  $\beta$ -sheet to another protein molecule in the crystal structure (see above). The backbone differences of residues 50–60 might be a consequence of the  $\beta$ -sheet interactions with residues 4–7 and the distortions in this latter fragment. The conformational changes of residues 30–37 between the X-ray and the NMR structure are due to the presence of the disulfide bridge Cys32–Cys35 in the X-ray structure only. The different  $\chi^1$ -angle for Cys35 increases the sulphur–sulphur distance from 2 Å in the closed bridge to about 7 Å in the NMR structure (Dyson *et al.* 1990). The apparently different side-chain conformations of His6 and Ile23 (Fig. 5.11.2) are caused by the use of only one of two alternate conformations of these side-chains in the crystal structure. Finally, the  $\chi^1$ -angle of the side-chain of Asp43 is also rotated by about 120°. The close distance of 4 Å of one side-chain oxygen to a copper ion in the crystal structure offers a possible explanation.

### 5.12 Trypsin inhibitor (BPTI)

The final protein discussed in this study is the bovine pancreatic trypsin inhibitor, BPTI. This protein has a long history of NMR applications to proteins. In fact, it was the first protein for which almost complete resonance assignments were obtained using the systematic approach of the sequential resonance assignment method (Wagner & Wüthrich, 1982). A first NMR structure of BPTI was presented in 1987 (Wagner *et al.* 1987); several structure determinations followed and have often served for comparisons of algorithms or for tests of assignment methods. The first X-ray structure dates from 1975 (Deisenhofer & Steigemann, 1975), and two additional crystal forms of the protein followed (Wlodawer *et al.* 1984; 1987*a*) and have been compared to the first form (Wlodawer *et al.* 1987*b*). Very recently, a new NMR structure has appeared and it was also compared to all three crystal structures (Berndt *et al.* 1992). Similar tools to the ones used in the

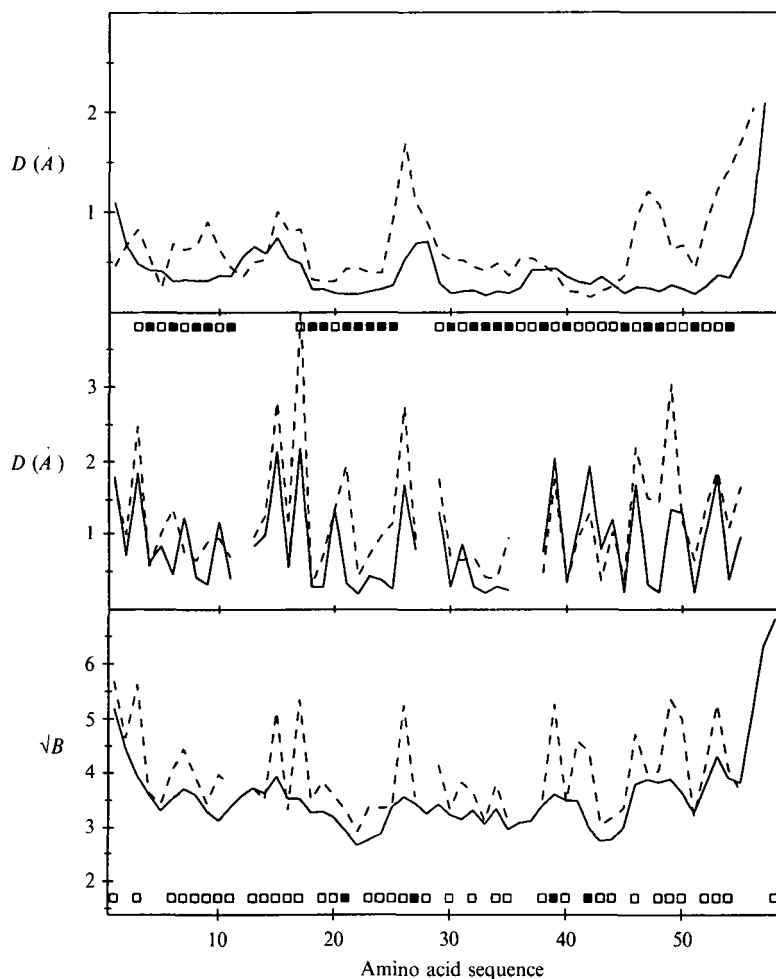


Fig. 5.12.2. Displacements and square roots of the B-factors are plotted versus the sequence for BPTI. See the caption to Fig. 5.1.2 for the information in the three panels and the meaning of the different line types and squares.

present study were applied in this BPTI comparison. Therefore, the following section is a short summary of the new findings, based on the crystal structure deposited as entry 5PTI in the Protein Data Bank, but the X-ray entries 4PTI and 6PTI of the Protein Data Bank are also consulted.

BPTI consists of a short  $3_{10}$ -helix (residues 3–7), an antiparallel  $\beta$ -sheet (residues 18–24, 29–35 and 44), an  $\alpha$ -helix (residues 47–56) and connecting loops (Fig. 5.12.1). The automatic selection procedure of Chapter 2 yields for the backbone of residues 3–11, 17–25, 29–54 a RMSD of 0.32 Å between the NMR conformers and their mean structure, and of 0.34 Å if 24 side-chains are added (Table 4.2). The backbone of both the NMR and the X-ray structures is very well defined except for residues 1 and 56–58 (Figs. 5.12.2).

Apparent differences between these structures are observed for the backbone of residues 6–9, 15–17, 25–27, 46–48 and 52–55, and, in addition to these backbone

fragments, for the side-chain of Glu49 (Fig. 5.12.2). The latter difference may be related to the large B-factors reported for this side-chain (Fig. 5.12.2). No explanation is available for the backbone difference observed for residues 6–9, except that residues 8 and 9 are prolines and that this region is connected to another region with structural differences around Cys55 via the disulfide bridge from Cys5. The fragments 15–17 and 46–48 coincide with structural differences between the two crystal structures 4PTI and 5PTI (Wlodawer *et al.* 1984). The side-chain of Arg17 hydrogen bonds to the backbone of Lys26 and Ala58 from other protein molecules in the structure 4PTI but not in the structure 5PTI. Four intermolecular hydrogen bonds located on one side of the loop 25–27 may induce a shift of this loop in the crystal structure (Berndt *et al.* 1992). The Protein Data Bank entry 5PTI reports an atom named 'UNK' that 'is probably a potassium ion in a partially occupied site'. This atom is only 2.4 Å from the sulphur of Met52 (in one of two reported conformations of this side-chain). It is also in van der Waals contact to Gly56, e.g. the distance to one of the  $\alpha$ -protons is only 3.1 Å. The addition of an ion at this location inside the protein could influence the position of the C-terminal helix 47–56. On the other hand, no such atom was reported in the other crystal structures of BPTI, and no significant difference of the backbone structure 52–55 is observed between 4PTI and 5PTI.

## 6. CONCLUSIONS

The first conclusion drawn from the detailed comparisons given in the previous chapter is the close coincidence of the structures obtained by X-ray diffraction on single crystals and by NMR on protein solutions. This similarity often extends to the local precision, that is, less well defined fragments in one structure often are also disordered in the other structure. Furthermore, many apparent differences between structures determined by NMR and X-ray are not significant once this precision of the structural elements is taken into account.

A possible source for true structural differences are the assumptions and interpretations made during the determination of the structures by the NMR and by X-ray techniques. For example, the explicit specification of hydrogen bond constraints in the input for the distance geometry calculations of NMR structures may cause distortions, in particular in peripheral residues of regular secondary structure elements. Some local differences between NMR and X-ray structures of barnase, chymotrypsin inhibitor 2 and interleukin 8 might be due to the use of these hydrogen bond constraints. Observed discrepancies may also be judged differently after searching the electron density map of a crystal structure for a possible second conformation of the protein fragment in question. Several crystal structures described in this study (barnase, ribonuclease and BPTI) report multiple conformations for side chains. A similar situation is encountered when different conformations are observed among protein molecules in the same crystal (e.g. thioredoxin), or when the protein structure changes due to different crystallization procedure or to different crystal forms (e.g. BPTI).

Another property of structures determined by NMR is the decrease in the

number of NOEs per residue on the surface with respect to the core of the protein. Since the selection of best defined side-chains in the structure comparisons of the previous chapter was based exclusively on the NMR structures, most surface side-chains were eliminated from the detailed discussion. It is difficult to assess the contribution of individual residues to the protein surface since the calculated surface accessibilities of individual residues strongly depend on the side-chain conformations. However, it is known that charged side-chains are usually located on the protein surface. In the NMR structures, the normalized displacement,  $D_n$  (equation 3a), is larger than zero for 82% of the charged residues. Thus, four fifths of the charged side-chains are more disordered than the average side-chain of the protein. The corresponding value for the normalized square root of the B-factors of X-ray structures,  $\sqrt{B_n}$  (equation 3b), is 79%. This indicates that the structural variations of surface side-chains in NMR structures are mainly due to genuine properties of the protein surface rather than to the lack of experimental data.

Several structural differences were found to originate from the different environments found in the crystal and in solution. In the single crystal form of a protein, distortions may result from the presence of various small compounds in the crystal lattice (possible examples are the copper ions in thioredoxin or the atom labelled 'UNK' in BPTI), from the freezing of single conformations of protein fragments that are flexible in other environments (an indication for this phenomenon is given by the occurrence of structural differences in single crystals of the same protein, see above), or from short contacts between neighboring proteins in the crystal (e.g. barnase, calbindin, interleukin  $1\beta$ , the 434 repressor(1-69) and others). Particularly interesting examples of crystal contacts are the potential substrate mimics of parts of one protein molecule which contact the active sites of other protein molecules. Crystal contacts involving the active site are observed, for example, for the  $\alpha$ -amylase inhibitor, barnase and chymotrypsin inhibitor 2.

Experimental conditions, like the choice of the pH value or of the salt concentration, may of course also influence the protein structure, although these primarily ionize the side-chains and the chain termini. In ribonuclease, different conformations were observed by NMR for the active site histidine 119 upon changing the pH. A condition-dependent mechanism was postulated for the interleukin 8 dimer manifesting an 'open' form of the enzyme in the NMR structure and a 'closed' form in the X-ray structure.

The goal of this study was to describe the effects of the two methods on structures determined using identical proteins. Yet a few selected proteins exhibit minor differences in the chemical structure: In calbindin a point mutation prevents a *cis-trans* isomerization in the NMR structure. The NMR and the X-ray structure of metallothionein were determined with different metal ligands. Finally, a disulfide bridge was present in the X-ray structure, but not in the NMR structure of thioredoxin. These few examples show that even very small chemical changes induce structural modifications that often exceed the differences observed between the NMR and X-ray structures of identical proteins.

For a few differences no explanation could be given in the comparisons of the previous chapter (e.g. Ile53 of the  $\alpha$ -amylase inhibitor or Val19 of the ColE1 rop protein), however they affect only a small number of atoms in the protein. Moreover, it is not excluded that there still is a simple explanation for these discrepancies. The existence of large scale differences between NMR and X-ray structures is therefore very unlikely. Major structural differences between the NMR and the X-ray structure have been published for two proteins. The first case, which originated from different metal coordination determined for metallothionein (Furey *et al.* 1986; Schultze *et al.* 1988) has been resolved by the publication of a revised crystal structure (Robbins *et al.* 1991). No answer is yet given for the apparent differences for the *E. coli* HPr protein (Klevit & Waygood, 1986; El-Kabbani *et al.* 1987). The recent publication of assignment corrections for the NMR structure (Hammen *et al.* 1991; van Nuland *et al.* 1992), and of a crystal structure for the homologous protein HPr from *Bacillus subtilis* (Herzberg *et al.* 1992) are possibly first steps in clarifying this remaining discrepancy.

## 7. ACKNOWLEDGEMENTS

The author thanks Drs K. D. Berndt, J. H. Bushweller, P. Güntert and K. Wüthrich for fruitful discussion, Drs K. D. Berndt and J. H. Bushweller for reading the manuscript, Drs W. J. Chazin, G. M. Clore, G. G. Dodson, H. Eklund, A. Pastore, F. M. Poulsen, M. Rico, D. Tsernoglou, A. Wlodawer for providing structure coordinates prior to availability from the Protein Data Bank, and the Protein Data Bank for making coordinates of protein structures available. This work was supported by the Schweizerischer Nationalfonds (projects 31.25174.88 and 31.32033.91).

## 8. REFERENCES

- ABOLA, E. E., BERNSTEIN, F. C., BRYANT, S. H., KOETZLE, T. F. & WENG, J. (1987). Protein Data Bank, in: Crystallographic Databases—Information Content, Software Systems, Scientific Applications (ed. F. H. Allen, G. Bergerhoff and R. Sievers), Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, pp. 107–132.
- AGGARWAL, A. K., RODGERS, D. W., DROTTAR, M., PTASHNE, M. & HARRISON, S. C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science, N.Y.* **242**, 899–907.
- ANDERSON, J., PTASHNE, M. & HARRISON, S. C. (1987). Structure of the repressor-operator complex of bacteriophage 434. *Nature, Lond.* **326**, 846–852.
- ARSENIEV, A., SCHULTZE, P., WÖRGÖTTER, E., BRAUN, W., WAGNER, G., VASÁK, M., KÁGI, J. H. R. & WÜTHRICH, K. (1988). Three-dimensional structure of rabbit liver [Cd<sub>7</sub>]metallothionein-2a in aqueous solution determined by nuclear magnetic resonance. *J. molec. Biol.* **201**, 637–657.
- BALDWIN, E. T., WEBER, I. T., CHARLES, R. ST., XUAN, J.-C., APPELLA, E., YAMADA, M., MATSUSHIMA, K., EDWARDS, B. F. P., CLORE, G. M., GRONENBORN, A. M. & WLODAWER, A. (1991). Crystal structure of interleukin-8: Symbiosis of NMR and crystallography. *Proc. natn. Acad. Sci. U.S.A.* **88**, 502–506.

- BANNER, D. W., KOKKINIDIS, M. & TSERNOGLOU, D. (1987). Structure of the ColE1 Rep protein at 1.7 Å resolution. *J. molec. Biol.* **196**, 657–675.
- BERNDT, K. D., GÜNTERT, P., ORBONS, L. P. M. & WÜTHRICH, K. (1992). Determination of a high-quality NMR solution structure of the bovine pancreatic trypsin inhibitor (BPTI) and comparison with three crystal structures. *J. molec. Biol.* (In the press).
- BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., JR., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMAMOUCHI, T. & TASUMI, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. molec. Biol.* **112**, 535–542.
- BILLETER, M., KLINE, A. D., BRAUN, W., HUBER, R. & WÜTHRICH, K. (1989). Comparison of the high-resolution structures of the  $\alpha$ -amylase inhibitor tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals. *J. molec. Biol.* **206**, 677–687.
- BILLETER, M., SCHAUMANN, TH., BRAUN, W. & WÜTHRICH, K. (1990). Restrained energy refinement with two different algorithms and force fields of the structure of the  $\alpha$ -amylase inhibitor tendamistat determined by NMR in solution. *Biopolymers* **29**, 695–706.
- BORKAKOTI, N., MOSS, D. S. & PALMER, R. A. (1982). Ribonuclease-A: least-squares refinement of the structure at 1.45 Å resolution. *Acta crystallogr. B* **38**, 2210–2217.
- BRÄNDEN, C.-I. & JONES, T. A. (1990). Between objectivity and subjectivity. *Nature, Lond.* **343**, 687–689.
- BRAUN, W. & GÖ, N. (1985). Calculation of protein conformations by proton–proton distance constraints. A new efficient algorithm. *J. molec. Biol.* **186**, 611–626.
- BRAUN, W., VASÁK, M., ROBBINS, A. H., STOUT, C. D., WAGNER, W., KÄGI, J. H. R. & WÜTHRICH, K. (1992). Comparison of the NMR solution structure and the X-ray crystal structure of rat metallothionein-2. *Proc. natn. Acad. Sci. U.S.A.* (In the press).
- BRÜNGER, A. T. (1990). X-PLOR 2.1 Manual. Yale University, New Havey CT 06511.
- BUERGER, M. J. (1980) *Crystal-structure Analysis*. Robert E. Krieger Publishing Company. New York: Huntington.
- BYCROFT, M., SHEPPARD, R. N., LAU, F. T.-K. & FERSHT, A. R. (1990). Sequential assignment of the <sup>1</sup>H nuclear magnetic resonance spectrum of barnase. *Biochemistry* **29**, 7425–7432.
- BYCROFT, M., LUDVIGSEN, S., FERSHT, A. R. & POULSEN, F. M. (1991). Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry* **30**, 8697–8701.
- CHAZIN, W. J., KÖRDEL, J., DRAKENBERG, T., THULIN, E., BRODIN, P., GRUNDSTRÖM, T. & FORSÉN, S. (1989). Proline isomerism leads to multiple folded conformations of calbindin D<sub>9k</sub>: direct evidence from two-dimensional <sup>1</sup>H NMR spectroscopy. *Proc. natn. Acad. Sci., U.S.A.* **86**, 2195–2198.
- CLORE, G. M. & GRONENBORN, A. M. (1991a). Applications of three- and four-dimensional hetero-nuclear NMR spectroscopy to protein structure determination. *Prog. NMR Spect.* **23**, 43–92.
- CLORE, G. M. & GRONENBORN, A. M. (1991b). Comparison of the solution nuclear magnetic resonance and X-ray crystal structures of human recombinant interleukin-1 $\beta$ . *J. molec. Biol.* **221**, 47–53.
- CLORE, G. M. & GRONENBORN, A. M. (1991c). Comparison of the solution nuclear magnetic resonance and crystal structures of interleukin-8. Possible implications for the mechanism of receptor binding. *J. molec. Biol.* **217**, 611–620.



- CLORE, G. M., GRONENBORN, A. M., KJAER, M. & POULSEN, F. M. (1987a). The determination of the three-dimensional structure of the barley serine inhibitor 2 by nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Protein Engineering*, **1**, 305–311.
- CLORE, G. M., GRONENBORN, A. M., JAMES, M. N. G., KJAER, M., MCPHALEN, C. A. & POULSEN, F. M. (1987b). Comparison of the solution and X-ray structures of barley serine protease inhibitor 2. *Protein Engineering*, **1**, 313–318.
- CLORE, G. M., DRISCOLL, P. C., WINGFIELD, P. T. & GRONENBORN, A. M. (1990a). Low resolution structure of interleukin-1 $\beta$  in solution derived from  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear three-dimensional nuclear magnetic resonance spectroscopy. *J. molec. Biol.* **214**, 811–817.
- CLORE, G. M., APPELLA, E., YAMADA, M., MATSUSHIMA, K. & GRONENBORN, A. M. (1990b). Three-dimensional structure of interleukin 8 in solution. *Biochemistry* **29**, 1689–1696.
- CLORE, G. M., WINGFIELD, P. T. & GRONENBORN, A. M. (1991). High-resolution three-dimensional structure of interleukin 1 $\beta$  in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy. *Biochemistry* **30**, 2315–2323.
- DEISENHOFER, J. & STEIGEMANN, W. (1975). Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta crystallogr. B* **31**, 238–250.
- DYSON, H. J., GIPPERT, G. P., CASE, D. A., HOLMGREN, A. & WRIGHT, P. E. (1990). Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **29**, 4129–4136.
- EBERLE, W., PASTORE, A., SANDER, C. & RÖSCH, P. (1991). The structure of ColE1 rop in solution. *J. Biomol. NMR* **1**, 71–82.
- EL-KABBANI, O. A. L., WAYGOOD, E. B. & DELBAERE, L. T. J. (1987). Tertiary structure of histidine-containing protein of the phosphoenolpyruvate:sugar phosphotransferase system of *Escherichia coli*. *J. biol. Chem.* **262**, 12926–12929.
- FERRIN, T. E., HUANG, C. C., JARVIS, L. E. & LANGRIDGE, R. (1988). The MIDAS display system. *J. molec. Graph.* **6**, 13–27.
- FINZEL, B. C., CLANCY, L. L., HOLLAND, D. R., MUCHMORE, S. W., WATENPAUGH, K. D. & EINSPAHR, H. M. (1989). Crystal structure of recombinant human interleukin-1 $\beta$  at 2.0 Å resolution. *J. molec. Biol.* **209**, 779–791.
- FUREY, W. F., ROBBINS, A. H., CLANCY, L. L., WINGE, D. R., WANG, B. C. & STOUT, C. D. (1986). Crystal structure of Cd, Zn metallothionein. *Science, N.Y.* **231**, 704–710.
- GLUSKER, J. P. & TRUEBLOOD, K. N. (1985). *Crystal Structure Analysis*. New York: Oxford University Press.
- GÜNTERT, P., BRAUN, W., BILLETER, M. & WÜTHRICH, K. (1989). Automated stereospecific  $^1\text{H}$  NMR assignments and their impact on the precision of protein structure determinations in solution. *J. Am. chem. Soc.* **111**, 3997–4004.
- GÜNTERT, P., BRAUN, W. & WÜTHRICH, K. (1991). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. molec. Biol.* **217**, 517–530.
- HAMMEN, P. K., WAYGOOD, E. B. & KLEVIT, R. E. (1991). Reexamination of the secondary and tertiary structure of histidine-containing protein from *Escherichia coli* by homonuclear and heteronuclear NMR spectroscopy. *Biochemistry* **30**, 11842–11850.

- HAVEL, T. F. (1990). The sampling properties of some distance geometry algorithms applied to unconstrained polypeptide chains: A study of 1830 independently computed conformations. *Biopolymers* **29**, 1565–1585.
- HAVEL, T. F. & WÜTHRICH, K. (1984). A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular  $^1\text{H}$ – $^1\text{H}$  proximities in solution. *Bull. Math. Biol.* **46**, 673–698.
- HERZBERG, O., REDDY, P., SUTRINA, S., SAIER, M. H. JR., REIZER, J. & KAPADIA, G. (1992). Structure of the histidine-containing phosphocarrier protein HPr from *Bacillus subtilis* at 2.0 Å resolution. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2599–2503.
- HOBOHM, U., SCHARF, M., SCHNEIDER, R. & SANDER, C. (1992). Selection of representative protein data sets. *Protein Science* **1**, 409–417.
- HYNES, T. R., RANDAL, M., KENNEDY, L. A., EIGENBROT, C. & KOSSIAKOFF, A. A. (1990). X-ray crystal structure of the protease inhibitor domain of Alzheimer's amyloid  $\beta$ -protein precursor. *Biochemistry* **29**, 10018–10022.
- KAPTEIN, R., ZUIDERWEG, E. R. P., SCHEEK, R. M., BOELENS, R. & VAN GUNSTEREN, W. F. (1985). A protein structure from nuclear magnetic resonance data. *lac* repressor headpiece. *J. molec. Biol.* **182**, 179–182.
- KATTI, S. K., LEMASTER, D. M. & EKLUND, H. (1990). Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J. molec. Biol.* **212**, 167–184.
- KENDREW, J. C., DICKERSON, R. E., STRANDBERG, B. E., HART, R. G., DAVIES, D. R., PHILIPS, D. C. & SHORE, V. C. (1960). Structure of myoglobin. *Nature, Lond.* **185**, 422–427.
- KLEVIT, R. E. & WAYGOOD, E. B. (1986). Two-dimensional  $^1\text{H}$  NMR studies of histidine-containing protein from *Escherichia coli*. 3. Secondary and tertiary structure as determined by NMR. *Biochemistry* **25**, 7774–7781.
- KLINE, A. D., BRAUN, W. & WÜTHRICH, K. (1986). Studies by  $^1\text{H}$  nuclear magnetic resonance and distance geometry of the  $\alpha$ -amylase inhibitor tendamistat. *J. molec. Biol.* **189**, 377–382.
- KLINE, A. D., BRAUN, W. & WÜTHRICH, K. (1988). Determination of the complete three-dimensional structure of the  $\alpha$ -amylase inhibitor tendamistat in aqueous solution by nuclear magnetic resonance and distance geometry. *J. molec. Biol.* **204**, 675–724.
- KUSZEWSKI, J., NILGES, M. & BRÜNGER, A. T. (1992). Sampling and efficiency of metric matrix distance geometry: A novel partial metrization algorithm. *J. Biomol. NMR* **2**, 33–56.
- LUDVIGSEN, S., SHEN, H., KJAER, M., MADSEN, J. CHR. & POULSEN, F. M. (1991). Refinement of the three-dimensional solution structure of barley serine proteinase inhibitor 2 and comparison with the structures in crystals. *J. molec. Biol.* **222**, 621–635.
- MAUGUEN, Y., HARTLEY, R. W., DODSON, E. J., DODSON, G. G., BRICOGNE, G., CHOTHIA, C. & JACK, A. (1982). Molecular structure of a new family of ribonucleases. *Nature, Lond.* **297**, 162–164.
- MCLACHLAN, A. D. (1979). Gene duplication in the structural evolution of chymotrypsin. Appendix: Least Square fitting of two structures. *J. molec. Biol.* **128**, 49–79.
- MCPHALEN, C. A. & JAMES, M. N. G. (1987). Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry* **26**, 261–269.
- MONDRAGÓN, A., SUBBIAH, S., ALMO, S. C., DROTTAR, M. & HARRISON, S. C. (1989).

- Structure of the amino-terminal domain of phage 434 repressor at 2.0 Å resolution. *J. molec. Biol.* **205**, 189–200.
- NERI, D., BILLETER, M. & WÜTHRICH, K. (1992). Determination of the nuclear magnetic resonance solution structure of the DNA-binding domain (residues 1 to 69) of the 434 repressor and comparison with the X-ray crystal structure. *J. molec. Biol.* **223**, 743–767.
- PERUTZ, M. F. (1964). The hemoglobin molecule. *Scientific American*, November, 64–76.
- PFLUGRATH, J. W., WIEGAND, G. & HUBER, R. (1986). Crystal structure determination, refinement and the molecular model of the  $\alpha$ -amylase inhibitor Hoe-467A. *J. molec. Biol.* **189**, 383–386.
- PRIESTLE, J. P., SCHÄR, H.-P. & GRÜTTER, M. G. (1989). Crystallographic refinement of interleukin 1 $\beta$  at 2.0 Å resolution. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9667–9671.
- RICO, M., SANTORO, J., GONZÁLEZ, C., BRUIX, M., NEIRA, J. L., NIETO, J. L. & HERRANZ, J. (1991). 3D structure of bovine pancreatic ribonuclease A in aqueous solution: An approach to tertiary structure determination from a small basis of <sup>1</sup>H NMR NOE correlations. *J. Biomol. NMR* **1**, 283–298.
- ROBBINS, A. H. & STOUT, C. D. (1991). X-ray structure of metallothionein. In *Metallobiochemistry, Part B, Metallothionein and Related Molecules* (ed. J. F. Riordan and B. L. Vallee). *Methods in Enzymology* **205**, 485–502.
- ROBBINS, A. H., MCREE, D. E., WILLIAMSON, M., COLLETT, S. A., XUONG, N. H., FUREY, W. F., WANG, B. C. & STOUT, C. D. (1991). Refined crystal structure of Cd, Zn metallothionein at 2.0 Å resolution. *J. molec. Biol.* **221**, 1269–1293.
- SCHULTZE, P., WÖRGÖTTER, E., BRAUN, W., WAGNER, G., BASÁK, M., KÄGI, J. H. R. & WÜTHRICH, K. (1988). Conformation of [Cd<sub>7</sub>]-metallothionein-2 from rat liver in aqueous solution determined by nuclear magnetic resonance spectroscopy. *J. molec. Biol.* **203**, 251–268.
- SINGH, U. C., WEINER, P. K., CALDWELL, J. W. & KOLLMAN, P. A. (1986). AMBER 3.0, University of California, San Francisco CA 94143.
- SKELTON, N. J., FORSÉN, S. & CHAZIN, W. J. (1990). <sup>1</sup>H NMR resonance assignments, secondary structure, and global fold of apo bovine calbindin D<sub>9k</sub>. *Biochemistry* **29**, 5752–5761.
- SVENSSON, L. A., THULIN, E. & FORSÉN, S. (1992). Proline *cis-trans* isomers in calbindin D<sub>9k</sub> observed by X-ray crystallography. *J. molec. Biol.* **223**, 601–606.
- SZEBENYI, D. M. E. & MOFFAT, K. (1986). The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine. *J. biol. Chem.* **261**, 8761–8777.
- TORDA, A. E., SCHEEK, R. M. & VAN GUNSTEREN, W. F. (1990). Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. *J. molec. Biol.* **214**, 223–235.
- VAN GUNSTEREN, W. F. & BERENDSEN, H. J. C. (1987). Groningen molecular simulation (GROMOS) library manual, Biomos. The Netherlands: Groningen.
- VAN NULAND, N. A. J., VAN DIJK, A. A., DIJKSTRA, K., VAN HOESEL, F. H. J., SCHEEK, R. M. & ROBILLARD, G. T. (1992). Three-dimensional <sup>15</sup>N-<sup>1</sup>H-<sup>1</sup>H and <sup>15</sup>N-<sup>13</sup>C-<sup>1</sup>H nuclear magnetic resonance studies of HPr a central component of the phosphoenolpyruvate-dependent phosphotransferase system of *Escherichia coli*. *Eur. J. Biochem.* **203**, 483–491.
- VEERAPANDIAN, B., GILLILAND, G. L., RAAG, R., SVENSSON, A. L., MASUI, Y., HIRAI, Y.

- & POULOS, T. L. (1992). Functional implications of interleukin- $1\beta$  based on the three-dimensional structure. *Proteins* **12**, 10–23.
- WAGNER, G. & WÜTHRICH, K. (1982). Sequential resonance assignments in protein  $^1\text{H}$  nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J. molec. Biol.* **155**, 347–366.
- WAGNER, G., BRAUN, W., HAVEL, T. F., SCHAUMANN, TH. M., GO, N. & WÜTHRICH, K. (1987). Protein structures in solution by nuclear magnetic resonance and distance geometry. *J. molec. Biol.* **196**, 611–639.
- WILLIAMSON, M. P., HAVEL, T. F. & WÜTHRICH, K. (1985). Solution conformation of proteinase inhibitor IIA from bull seminal plasma by  $^1\text{H}$  nuclear magnetic resonance and distance geometry. *J. molec. Biol.* **182**, 295–315.
- WLODAWER, M. P. (1985). Structure of bovine pancreatic ribonuclease. *Biol. Macromol. Assem.* **2**, 394–439.
- WLODAWER, A., WALTER, J., HUBER, H. & SJÖLIN, L. (1984). Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. *J. molec. Biol.* **180**, 301–329.
- WLODAWER, A., NACHMAN, J., GILLILAND, G. L., GALLAGHER, W. & WOODWARD, C. (1987a). Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J. molec. Biol.* **198**, 469–480.
- WLODAWER, A., DEISENHOFER, J. & HUBER, H. (1987b). Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J. molec. Biol.* **193**, 145–156.
- WLODAWER, A., SVENSSON, L. A., SJÖLIN, L. & GILLILAND, G. L. (1988). Structure of phosphate-free ribonuclease A refined at 1.26 Å. *Biochemistry* **27**, 2705–2717.
- WÜTHRICH, K. (1986). *NMR of Proteins and Nucleic Acids*. New York: Wiley.
- WÜTHRICH, K. (1989). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science, N.Y.* **243**, 45–50.
- ZUIDERWEG, E. R. P., BILLETER, M., BOELENS, R., SCHEEK, R. M., WÜTHRICH, K. & KAPTEIN, R. (1984). Spatial arrangement of the three  $\alpha$  helices in the solution conformation of *E. coli lac* repressor DNA-binding domain. *FEBS Lett.* **174**, 243–247.

### Note added in proof

The two remaining nuclear encoded subunits (the 33rd and 34th) of bovine complex I have been sequenced recently and are known as B14.5a and B14.5b, in accord with the nomenclature used for its subunits. [Arizmendi *et al.* (1992) *FEBS Lett* in press]. Therefore, the 34 completed nuclear encoded subunits contain 5,837 amino acids and have a corresponding combined molecular mass of 668,364. Together with the 7 subunits encoded in mt-DNA, the calculated mass of the 7,955 amino acids in the complex is 906,826.

## Appendix. Crystallographic parameters

Protein	Space group	Proteins per cell	Unit cell dimensions			Unit cell angles		
$\alpha$ -Amylase inhibitor	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	4	61.6	40.6	26.8	90.0	90.0	90.0
Barnase	P <sub>3</sub> <sub>2</sub>	9	59.53	59.53	81.67	90.0	90.0	120.0
Ca-binding protein	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2	4	56.2	42.6	29.2	90.0	90.0	90.0
Chymotrypsin inhib. 2	P6 <sub>2</sub> 2	12	69.015	69.015	52.890	90.0	90.0	120.0
ColE1 rop protein	C2	2*2	60.4	40.4	27.7	90.0	105.8	90.0
Interleukin 1 $\beta$	P <sub>4</sub> <sub>3</sub>	4	54.86	54.86	77.02	90.0	90.0	90.0
Interleukin 8	P <sub>3</sub> <sub>1</sub> 2 <sub>1</sub>	6	40.3	40.3	90.1	90.0	90.0	120.0
Metallothionein	P <sub>4</sub> <sub>3</sub> 2 <sub>1</sub> 2	8	30.9	30.9	120.4	90.0	90.0	90.0
434 repressor(1-69)	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	4	32.8	37.5	44.6	90.0	90.0	90.0
Ribonuclease A	P <sub>2</sub> <sub>1</sub>	4	30.18	38.40	53.32	90.0	105.85	90.0
Thioredoxin	C2	4*2	89.50	51.06	60.45	90.0	113.5	90.0
Trypsin inhibitor*	P <sub>2</sub> <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	4	74.1	23.4	28.9	90.0	90.0	90.0

\* Parameters for crystal form II of trypsin inhibitor (Wlodawer *et al.* 1984; Protein Data Bank entry 5PTI).