



Solution Structure of Human p8 MTCP1 , a Cysteine-rich Protein Encoded by the MTCP1 Oncogene, Reveals a New a a a-Helical Assembly Motif

Philippe Barthe, Yin-Shan Yang, Laurent Chiche, François Hoh, Marie-Paule Strub, Laurent Guignard, Jean Soulier, Marc-Henri Stern, Herman van Tilbeurgh, Jean-Marc Lhoste, et al.

► To cite this version:

Philippe Barthe, Yin-Shan Yang, Laurent Chiche, François Hoh, Marie-Paule Strub, et al.. Solution Structure of Human p8 MTCP1 , a Cysteine-rich Protein Encoded by the MTCP1 Oncogene, Reveals a New a a a-Helical Assembly Motif. *Journal of Molecular Biology*, 1997, 274 (5), pp.801-815. 10.1006/jmbi.1997.1438 . hal-02359713

HAL Id: hal-02359713

<https://hal.science/hal-02359713>

Submitted on 14 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Solution Structure of Human p8^{MTCP1}, a Cysteine-rich Protein Encoded by the *MTCP1* Oncogene, Reveals a New α -Helical Assembly Motif

Philippe Barthe¹, Yin-Shan Yang¹, Laurent Chiche¹
 François Hoh¹, Marie-Paule Strub¹, Laurent Guignard¹, Jean Soulier²
 Marc-Henri Stern², Herman van Tilbeurgh¹, Jean-Marc Lhoste¹
 and Christian Roumestand^{1*}

¹Centre de Biochimie Structurale, CNRS-UMR 9955 INSERM-U414, Université de Montpellier I, Faculté de Pharmacie, 15 Avenue Charles Flahault, 34060 Montpellier Cedex, France

²Unité INSERM U462, Hôpital Saint-Louis, Paris 75475 France

MTCP1 (for Mature-T-Cell Proliferation) is the first gene unequivocally identified in the group of uncommon leukemias with a mature phenotype. The three-dimensional solution structure of the human p8^{MTCP1} protein encoded by the *MTCP1* oncogene was determined by homonuclear proton two-dimensional NMR methods at 600 MHz. After sequence specific assignments, a total of 931 distance restraints and 57 dihedral restraints were collected. The location of the three previously unassigned disulfide bridges was determined from preliminary DIANA structures, using a statistical analysis of intercystinyl distances. The solution structure of p8^{MTCP1} is presented as a set of 30 DIANA structures, further refined by restrained molecular dynamics using a simulated annealing protocol with the AMBER force field. The r.m.s.d. values with respect to the mean structure for the backbone and all heavy atoms for a family of 30 structures are 0.73(±0.28) and 1.17(±0.23) Å, when the structured core of the protein (residues 5 to 63) is considered. The solution structure of p8^{MTCP1} reveals an original scaffold consisting of three α helices, associated with a new cysteine motif. Two of the helices are covalently paired by two disulfide bridges, forming an α -hairpin which resembles an antiparallel coiled-coil. The third helix is oriented roughly parallel to the plane defined by the α -antiparallel motif and its axis forms an angle of $\approx 60^\circ$ with respect to the main axis of this motif.

© 1997 Academic Press Limited

Keywords: Leukemia; protein structure; nuclear magnetic resonance; cysteine motif; disulfide connectivities

*Corresponding author

Abbreviations used: 2D, two-dimensional; NOE, nuclear Overhauser enhancement; NOESY, nuclear Overhauser enhancement spectroscopy; DQF-COSY, double-quantum-filtered scalar-correlated spectroscopy; z-TOCSY, z-filtered total correlation spectroscopy; CSI, chemical shift index; r.m.s.d., root mean square deviation; PIR, Protein Identification Resource; Rop, mutant of repressor of primer from *Escherichia coli*; RP-HPLC, reverse phase high performance liquid chromatography; DTNB, 5,5'-dithiobis(2-nitrobenzoic acid); DTT, 1,4-dithiothreitol; IPTG, isopropyl-thio- β -D-galactopyranoside; PMSF, phenylmethylsulfonyl fluoride; (²H)TSP, 3-methylsilyl-[2,2,3,3-²H₄]propionate; TCR, T-cell receptor.

Introduction

T-cell lymphoproliferative diseases are often associated with recurrent chromosomal translocations involving T cell receptor genes (TCR) and genes that are thought to play a role in the pathogenesis of these diseases. These chromosomal translocations have been proposed to occur during the recombination events of the TCR genes, leading to genetic aberrations which deregulate proto-oncogenes now juxtaposed to the TCR genes. The characterization of numerous translocations found in immature (acute) T cell leukemias has led to the discovery of several genes which appear to be important in the malignant transformation process (for a review, see Rabbitts, 1994).

In contrast, little is known about the genetic aberrations associated with the heterogeneous group of uncommon leukemias having a mature phenotype. *MTCP1* (also called *c6.1B*; Fisch *et al.*, 1993; Stern *et al.*, 1993), located in the Xq28 chromosomal region, was the first gene to be unequivocally identified in this pathogenesis. It is involved in the translocation t(X;14)(q28;q11), a translocation currently associated with a rare subset of mature T-Cell proliferations (Dallapiccola *et al.*, 1984; Witzig *et al.*, 1986; Goyns *et al.*, 1993; Stern *et al.*, 1993; Madani *et al.*, 1996). It is also observed in ataxia telangiectasia patients where clonal proliferations of T lymphocytes bearing this t(X;14) translocation can be observed years before the emergence of leukemia (Taylor *et al.*, 1981; Canki *et al.*, 1983; Beatty *et al.*, 1986), indicating that this translocation is one of the early events of the malignant transformation (Stern *et al.*, 1989; Taylor *et al.*, 1992; Sherrington *et al.*, 1994; for a review see Taylor *et al.*, 1996).

Surprisingly, the two *MTCP1* major splicing forms, A and B, encode two entirely different proteins, a phenomenon seldomly observed in mammals. The longer transcript, B1, codes for a 107 residues (13 kDa) protein known as p13^{MTCP1} (Madani *et al.*, 1996). This protein shows high sequence similarity (40% identity) with p14^{TCL1} (Fu *et al.*, 1994), the product of the 14q32.1 oncogene *TCL1*. Since p13^{MTCP1} and p14^{TCL1} do not share homology with other known sequences, they may constitute a new family of oncogenic proteins. Type A transcripts code for a small 8 kDa (68 residues) protein, p8^{MTCP1} (Soulier *et al.*, 1994) recently localized in mitochondria (Madani *et al.*, 1995). The protein sequence shows a striking richness in cysteine residues (seven cysteine out of 68 amino acids), though it could not be classified among previously described cysteine motifs (Madani *et al.*, 1995). Computer analysis (Madani *et al.*, 1995) reveals sequence similarity with type-1 repeat fragments of thrombospondin proteins, also present in the β chain of the complement C8. Whereas the expression of p13^{MTCP1} is restricted to mature T-cell proliferation with t(X,14) translocations, the mitochondrial protein p8^{MTCP1} is expressed at low levels in most human tissues and is over-expressed in the proliferating T cells (Soulier *et al.*, 1994; Madani *et al.*, 1996). Thus, p8^{MTCP1} may be associated with a function common to many cell types. Supporting this assumption, the p8 protein homolog is highly conserved in mouse (Soulier *et al.*, 1994), indicating the evolutionary conservation of its potential function. Although its biological function is still unknown, the fact that p8^{MTCP1} is over-expressed in proliferating t(X,14) T-cells strongly suggests its participation in oncogenesis. We present here the first report on the structure determination of human p8^{MTCP1} by two-dimensional homonuclear NMR.

Results

Resonance assignments and secondary structure

The NMR sample was obtained by a method similar to that previously published by Soulier *et al.* (1994) (see Materials and Methods). Eleven residues were added (six residues at the N terminus and five at the C terminus end of the protein) in order to prevent enzymatic degradation and for sub-cloning facilities. Nearly complete sequence-specific ¹H resonance assignments for p8^{MTCP1} were obtained from the 2D homonuclear spectra following the sequential assignment procedure (Billeter *et al.*, 1982; Wagner & Wüthrich, 1982; Wider *et al.*, 1982; Wüthrich, 1986), using a combination of through-bond (DQF-COSY, TOCSY) and through-space (NOESY) connectivities. 2D spectra were usually recorded at 298 K, i.e. under conditions where optimal dispersion of amide ¹H resonances was observed. Residual overlap was resolved from analogous spectra recorded at 288 K and 308 K. Despite the size of the protein (79 residues) and its high helical content (*vide infra*), the dispersion of ¹H resonance in the 2D homonuclear spectra at 600 MHz was sufficient to enable identification of a continuous stretch of sequential NOEs over the whole sequence without the need for labelling and heteronuclear NMR experiments. However, the quality of the TOCSY spectra was rather poor, except for N-terminal (residues –5 to 4) and C-terminal (residues 64 to 73) residues. The sharp and intense lines observed in the TOCSY spectra, together with the strong $d_{\alpha N}$ sequential NOEs, strongly suggest that the N and C-terminal ends of the protein are fully flexible random coil polypeptide segments. Since the N and C-terminal additional residues appear in a disordered region and do not participate to the structured core of the protein, they will not be considered further in the discussion. Nevertheless, some spin systems were easily identified in the structured part of the protein (residues 5 to 63): the connectivities of all aromatic ring protons were straightforward and their connectivity to the C α H and C β H resonances was established using NOEs with the corresponding aromatic protons. Moreover, the valine, alanine, threonine and, to a lesser extent, the glycine spin systems were easily identified in the TOCSY spectra and constituted good starting points for the sequential assignment. The NOESY spectra in H₂O revealed a large number of sequential d_{NN} contacts which led to the recognition of three large stretches of adjacent residues (Figure 1). Additionally, the presence of strong sequential C α H–C $\delta\delta$ H NOEs in the three Xxx-Pro dipeptides unambiguously establishes the *trans* geometry of the peptide bonds involved. The resulting ¹H assignments are available as supplementary material.

A total of 37 residues out of 79 have their C α H proton resonance shifted upfield from the random-coil values, indicating a three-dimensional fold rich

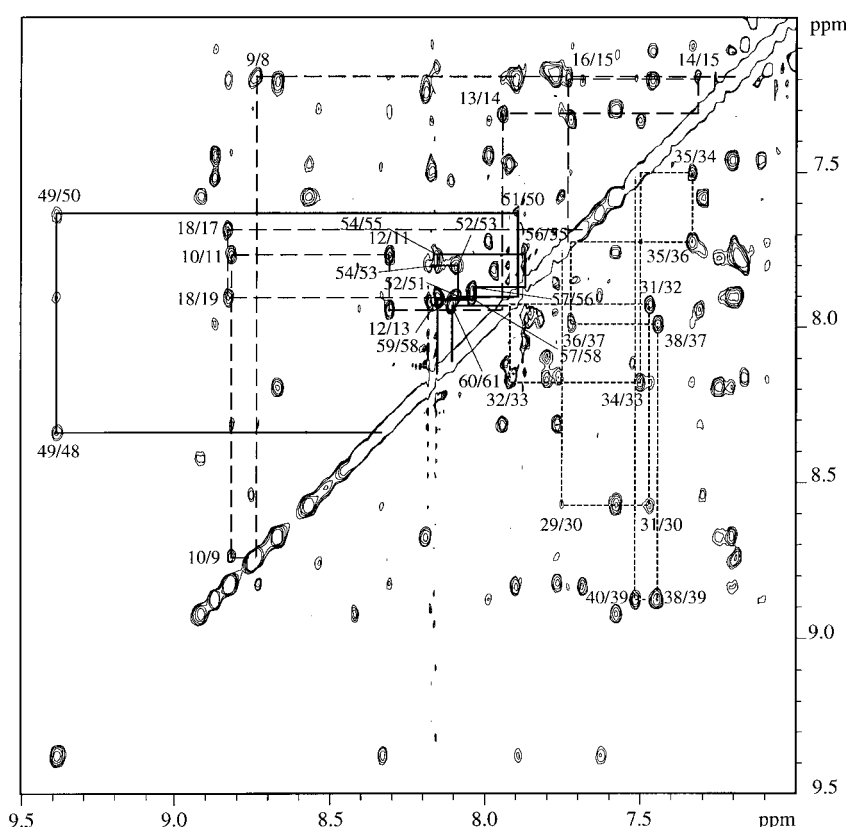


Figure 1. NOESY spectrum (amide protons region) of p8^{MTCPI} (600 MHz, 200 ms mixing time) recorded in H₂O at pH 6.5 and 298 K. The sequential d_{NN} connectivities are shown for the three helical regions: residues 8 to 19 (broken line); residues 29 to 40 (dotted line); residues 48 to 61 (continuous line).

in α -structures (Wishart *et al.*, 1991, 1992, 1995). This was supported by the presence of large stretches of medium range NOE contacts ($\leq i, i + 4$), as revealed by the diagonal plot (data unshown), and of weak $^3J_{\text{NH}\alpha}$ coupling constant values (see supplementary material), indicating that the main secondary elements are α -helices encompassing residues 8 to 20 (helix I), 29 to 40 (helix II) and 48 to 63 (helix III). The presence of $d_{\alpha\text{N}}(i, i + 2)$ connectivities at the extremities of helices I and II, with the concomitant disappearance $d_{\alpha\text{N}}(i, i + 4)$ connectivities, indicates that initial and final turns are distorted towards a 3_{10} conformation. Additionally, characteristic side-chain to backbone proton NOEs indicate that the first and third helices are stabilized by an N-capping box (Richardson & Richardson, 1988) involving residues Asp5 and Gln8, and Ser47 and Cys50, respectively. The perpendicular disposition of numerous long-range ($> i, i + 4$) NOE contacts with respect to helix I and helix II diagonal segments suggests that these two helices are orientated in an antiparallel fashion. In contrast, few NOEs were found between helix II and helix III.

Assignment of the disulfide bridges

A major problem confronted in the structure determination of p8^{MTCPI} was the assignment of the disulfide bridges. Prior to this work the cysteine residues involved in disulfide bondings

had not been identified, and as mentioned above, the p8^{MTCPI} sequence could not be related to any known cysteine motifs. Titration of the free thiol groups with DTNB according to Ellman's method (Riddles *et al.*, 1983) at different pH (see Materials and Methods) showed that only one cysteine exists in the reduced form and that the remaining cysteine residues were not involved in a zinc cluster. Consequently, the six cysteine residues are involved in three disulfide bonds. However, it was not possible to determine the disulfide bridge arrangement in p8^{MTCPI} by either biochemical means of a qualitative interpretation of the NOE data. Indeed, if the Cys17–Cys28 disulfide bond could be assigned on the basis of non ambiguous long-range NOE contacts between protons belonging to the two half-cystine residues, this approach failed for the determination of the two remaining disulfide bridges. This was due to the presence of the two vicinal hemi-cystine residues Cys38 and Cys39 which led to ambiguous long-range NOEs with Cys7 and Cys50. Thus, the strategy we used was to maximize the structural information content of the NMR data, so that the disulfide bridges could be assigned from a set of structures calculated without imposing any disulfide bonds.

A set of 20 DIANA structures (Güntert & Wüthrich, 1991) were calculated and analyzed in terms of convergence (see Materials and Methods) and the mean intercystinyl distances in this set of conformers were then determined (Table 1). Stereo-

Table 1. Distribution of distances (Å) between C^β and S^γ atoms of cysteine residues over 20 structures obtained after DIANA calculation

	Cys7	Cys12	Cys17	Cys28	Cys38	Cys39	Cys50
Cys7		9.5 ± 0.2 (12.3 ± 0.3)	14.1 ± 0.2 (14.3 ± 0.8)	14.2 ± 0.3 (14.4 ± 0.9)	4.8 ± 0.3 (3.2 ± 1.4)	5.0 ± 0.5 (3.2 ± 0.4)	5.9 ± 0.9 (5.8 ± 1.5)
Cys12	Excluded		9.5 ± 0.2 (10.02 ± 0.8)	11.9 ± 0.2 (11.0 ± 0.7)	12.1 ± 0.3 (12.9 ± 0.8)	13.8 ± 0.5 (14.0 ± 0.5)	13.7 ± 0.5 (16.0 ± 1.0)
Cys17	Excluded	Excluded		4.4 ± 0.3 (2.2 ± 1.7)	14.3 ± 0.2 (14.7 ± 0.4)	16.3 ± 0.4 (15.3 ± 1.1)	17.8 ± 0.6 (17.5 ± 1.5)
Cys28	Excluded	Excluded	Used		14.4 ± 0.4 (15.2 ± 0.6)	15.1 ± 0.4 (15.1 ± 0.8)	16.6 ± 0.7 (17.1 ± 1.0)
Cys38	Used	Excluded	Excluded	Excluded		5.3 ± 0.1 (5.9 ± 1.0)	8.9 ± 0.8 (8.4 ± 2.1)
Cys39	Possible	Excluded	Excluded	Excluded	Excluded		4.6 ± 0.8 (3.1 ± 1.5)
Cys50	Excluded	Excluded	Excluded	Excluded	Excluded	Used	

The average values and standard deviations are given above the diagonal (into brackets for S^γ). Structure calculations were performed using the DIANA program as described in Materials and Methods, without fixing any constraint between S atoms. The bold values are statistically the lowest distances and reflect disulfide bonds. Italic values would correspond to excluded disulfide arrangements, and the underlined value to a possible disulfide bond. These considerations are summarized below the diagonal.

specific assignments were available for all C^βH protons in the half-cystine residues, which gives a better precision on the spatial location of the sulfur atoms. We also measured the inter-C^β distances, independent of the χ_1 angle, to avoid incorrect disulfide assignments due to possible erroneous stereospecific assignments. As a result of this preliminary calculation, the cysteine residues appeared clustered in two distinct regions of the protein: Cys12, Cys17 and Cys28 are located close to the turn which joins helix I and helix II; Cys7, Cys38, Cys39 and Cys50 are situated in a region where the N terminus of helix I and helix III and the C terminus of helix II are in close proximity. These regions are far enough apart to avoid any disulfide pairing between half-cystine residues belonging to distinct clusters. In the first cluster, the Cys17–Cys28 disulfide bridge can be easily identified from NOE contacts between the two half-cystine residues: it follows that Cys12 is the reduced cysteine. In the second cluster, the disulfide bridge between the two vicinal half-cystine residues, Cys38 and Cys39, can be eliminated: the formation of such bridges is seldomly observed in natural proteins (Zhang & Snyder, 1989), because it requires the conversion of the Cys–Cys peptide bond from its usual *trans*-configuration into the energetically less favourable *cis*-configuration. Furthermore, it has been observed in peptide models that such cyclocystine loops manifest themselves in proton spectra in ways which were not observed for p8^{MTCP1} (Liff & Siddiqui, 1996). Discarding this sequential bridge eliminates the possibility of pairing between Cys7 and Cys50. This restrained the analysis to two possibilities: Cys38–Cys7, Cys39–Cys50 or Cys38–Cys50, Cys39–Cys7. If the distance analysis shows that the Cys39–Cys7 disulfide could occur, both C^β–C^β or S^γ–S^γ distances between Cys38 and Cys50 are too large to be compatible with a disulfide bridge. Therefore the correct disulfide pairing involves Cys38–Cys7 and Cys39–Cys50.

Structure calculation and refinement

The final structure calculation steps included the disulfide bridges which were thus treated as additional distance restraints (four restraints per bridge). Inclusion of the disulfide restraints improved the convergence of the calculation. Forcing the alternative disulfide bridges (Cys38–Cys50, Cys39–Cys7 and Cys38–Cys39, Cys7–Cys50) led to numerous distance or dihedral violations confirming the disulfide pairing described in the previous section. A total of 988 structural (distance and dihedral) constraints out of the 1038 collected were meaningful and taken into account for distance-geometry calculations (see Materials and Methods). The remaining constraints were found to be irrelevant for the calculation, either because they correspond to fixed interproton distances or because they correspond to distances that cannot be violated by any conformation of the polypeptidic chain. Of the final 800 structures, 50 had a final target function smaller than 5 Å² and satisfied all experimental constraints with no NOE distance constraint violation exceeding 0.88 Å and no dihedral angle violation exceeding 9.9°. To speed up the calculations, only a subset comprising the 30 structures with the best target value (<3.95 Å²) and with no NOE distance violation exceeding 0.76 Å and no dihedral angle violation exceeding 6.3° was chosen for further refinement. The survey of the structural statistics and of the residual violations of experimental constraints for this 30 conformers is shown in Table 2. Discarding residues 1 to 4 and 64 to 68, for which no long-range NOEs were available, the low average pairwise r.m.s.d. values (1.20 ± 0.28 Å) indicate that all conformers converge to the same overall fold. We thus consider them representative for the conformational space consistent with the data.

These 30 DIANA conformers with lowest target function value were further refined by simulated annealing and restrained energy minimization

Table 2. Experimental constraints and refinement statistics of the 30 conformers representing the solution structure of p8^{MTCPI} before and after restrained energy minimization

Distance constraints		
Intraresidue	203	
Sequential	278	
Medium-range	292	
Long-range	158	
Disulfide bonds	3	
Constraints per residue	24	
Dihedral constraints		
PHI (Φ)	35	
CHI1 (χ_1)	22	
Parameter	DIANA	DIANA + AMBER
Target function (\AA^2)	3.06 \pm 0.52	
Upper limit violations		
Number >0.2 \AA	6.77 \pm 2.71	5.67 \pm 1.58
Sum of violations (\AA)	9.17 \pm 1.06	8.13 \pm 0.64
Maximum violation (\AA)	0.43 \pm 0.12	0.30 \pm 0.02
Dihedral angle violations		
Number >5°	0	0.90 \pm 0.88
Sum of violations (°)	9.54 \pm 4.31	12.90 \pm 4.51
Maximum violation (°)	3.46 \pm 1.23	5.43 \pm 0.91
Van der Waals violations		
Number >0.2 \AA	3.57 \pm 1.85	
Sum of violations (\AA)	6.30 \pm 1.20	
Maximum violation (\AA)	0.31 \pm 0.09	
AMBER energies (Kcal mol ⁻¹)		
Bond energy		23.4 \pm 1.2
Valence angle energy		117.9 \pm 4.1
Van der Waals energy		-418.7 \pm 12.2
Electrostatic energy		-1896.2 \pm 21.8
Constraint energy		48.9 \pm 2.8
Total non-bonding energy		-1508.3 \pm 20.4
Total energy		-1029.7 \pm 17.2
r.m.s.d. values (\AA)		
Residues		
1-68 (all) BA ^a /HA ^b	3.16 \pm 0.84/3.76 \pm 0.75	2.32 \pm 0.38/2.64 \pm 0.37
5-63 BA ^a /HA ^b	1.20 \pm 0.28/2.01 \pm 0.29	0.73 \pm 0.28/1.17 \pm 0.23
8-40 BA ^a /HA ^b	0.59 \pm 0.13/1.43 \pm 0.15	0.23 \pm 0.11/0.69 \pm 0.10
48-63 BA ^a /HA ^b	1.33 \pm 0.54/2.33 \pm 0.51	0.60 \pm 0.20/1.32 \pm 0.25

^a Backbone atoms.^b All heavy atoms.

using the AMBER forcefield (Pearlman *et al.*, 1995). The results of the refinement are collected in Table 2. As expected the refinement resulted in a substantial improvement of the van der Waals energy and of the electrostatic contribution to the energy. Importantly, the decrease in the non-bonding energy was achieved without significant increase in the residual violations of the experimental constraints as compared with the DIANA conformers. The average global (residues 5 to 63) r.m.s.d. backbone and heavy atoms goes from 1.20(\pm 0.28) and 2.01(\pm 0.29) to 0.73(\pm 0.28) and 1.17(\pm 0.23), respectively, after refinement. In conjunction with the high number of NOEs per residue (Figure 2), this indicates a good sampling of the conformational space. Perhaps the most important result of the refinement was the marked improvement in the geometry of most hydrogen bonds in the individual structures as a result of the electrostatic term in the forcefield. Altogether, the refined structures provide a better estimate of the conformational space in agreement with the experimental constraints than that obtained from DIANA alone.

The energy-minimized average structure calculated from the above family of 30 refined structures was used to check the quality of the structure in a Ramachandran plot through the program PROCHECK (Laskowski *et al.*, 1993) (data not shown). The protein contains 61 "meaningful" residues (i.e. non-glycine, non-proline, and non-terminal residues) for the Ramachandran plot. In the energy-minimized average structure, 54 residues (i.e. 88.5%) fall in the most favored regions of the Ramachandran plot; six residues (i.e. 9.8%) are in the additional allowed regions; only one residue (0.3%), Lys64, located in the disordered C-terminal end of the protein, falls in the generously allowed and no residue falls in the disallowed region.

Description of the NMR solution structure of p8^{MTCPI}

The main structural motif of p8^{MTCPI} consists of two antiparallel helices spanning residues 8 to 20 (helix I) and 29 to 40 (helix II), strapped in a well-determined orientation by the two disulfides 7-38

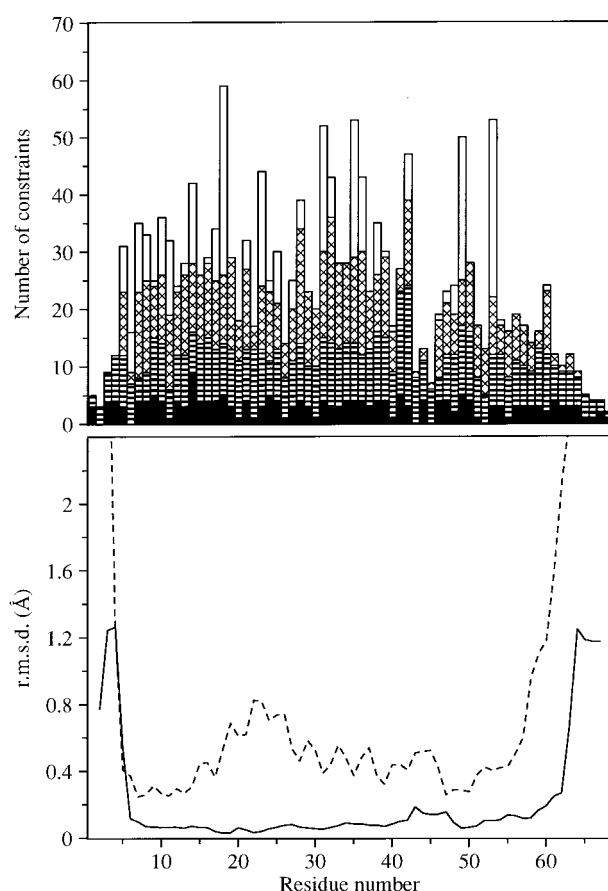


Figure 2. (Top) Plot as a function of the amino acid sequence of the number of NOE constraints used in the final structure calculation of p8^{MTCP1}. NOE categories are shown as follows: intraresidue, black; sequential, horizontal lines; medium-range, cross-hatching; long-range, open. (Bottom) Plots *versus* the amino acid sequence of the mean of the global r.m.s. differences from the energy-minimized average structure for the backbone atoms superimposed over the peptidic segment 5 to 63 (broken line), and the mean local r.m.s. deviation for the backbone superposition of all tripeptide segments along the sequence relative to the mean NMR structure (continuous line). The r.m.s.d. values for the tripeptide segments are plotted at the position of the central residue.

and 17–28 (Figure 3). Such an antiparallel arrangement forms an α -helical hairpin which resembles a two-stranded α -helical coiled-coil. The angle, however, between the axes of the two cylinders representing helices I and II is about 5°, which is significantly smaller than that observed for a typical super-coil in which hydrophobic contact surface between the two helices is maximized (Schultz & Schirmer, 1979). The two disulfide bridges are well-defined and adopt the common left-handed spiral conformation (Richardson, 1981) in all the calculated structures (Figure 4). According to Srinivasan *et al.* (1990), the conformation of the disulfide bond Cys7–Cys38 is of type 2, that of the Cys17–Cys28 is of type 3. The two antiparallel

helices are linked by two interlocking turns. The first one is a non-classical turn characterized by a $i, i+5$ hydrogen bond (23 \rightarrow 18), whereas the second one could be regarded as a 3_{10} helix turn with a characteristic $i, i+3$ hydrogen bond (24 \rightarrow 21, 27 \rightarrow 24). A relatively well-defined loop (from residues Gln41 to Arg46) connects helix II to helix III, spanning residues 48 to 63. The third disulfide bridge 39–50 links the top of helix III to the tip of helix II. This bridge adopts the rather unusual right-handed conformation in all the refined structures (Figure 4), and does not fit to any of the conformations described by Srinivasan *et al.* (1990). Note that the three disulfide bridges have the same geometry in all of the 30 refined protein conformers, reflecting the accuracy of their definition by the NMR restraints. The N-terminal end of helix III is positioned near the C-terminal end of helix II. Helix III is contained in a plane approximately parallel to the plane defined by the antiparallel helical motif. Its axis forms an angle of $57(\pm 10)^\circ$ with respect to the main axis of the antiparallel helical motif. The N (residues 1 to 4) and C termini (residues 64 to 68) appear essentially mobile and disordered.

Analysis of the local backbone displacements (Wagner *et al.*, 1987) (Figure 2) and of the order parameter S (Hyberts *et al.*, 1992) calculated for the 30 energy-refined conformers (supplementary material) indicates that all three helices of p8^{MTCP1}, as well as the two interlocking turns joining helix I and helix II, are well-defined. The N and C termini, and to a lesser extent the loop between helix II and III, are less well-defined, primarily due to a lack of long and medium range NOEs, presumably a consequence of increased flexibility. Rather than the local r.m.s.d., where stretches of three residues are aligned to show the differences in local structures, the calculation of the order parameter S for dihedral angles allows us to analyze the precision of the local structures, since this parameter is entirely independent of the alignment procedure. A value very close to 1 generally observed for most ϕ and ψ angles in the three helices reflects their high definition, even though no dihedral constraints are available for all ϕ angles. Similarly, the values of S observed for χ_1 angles indicate that the side-chain position for most residues in the helices is well defined, despite the fact that stereospecific assignment of the H^β protons was available only for 33% of the residues. This suggests that the well-defined χ_1 for most of the side-chains in the helices is a consequence of the protein packing rather than the direct experimental data.

Calculation of the global r.m.s.d. values for the different atom selections (Table 2) shows that the location of helix III is globally significantly less defined when compared with the first two helices. Indeed, when enforcing the alignment of the 30 refined structures on the α -antiparallel motif (residues 8 to 40) for the calculation of the average structure, the r.m.s.d. value rises to $1.74(\pm 0.96)$ Å in the third helix (residues 48 to 63). In contrast,

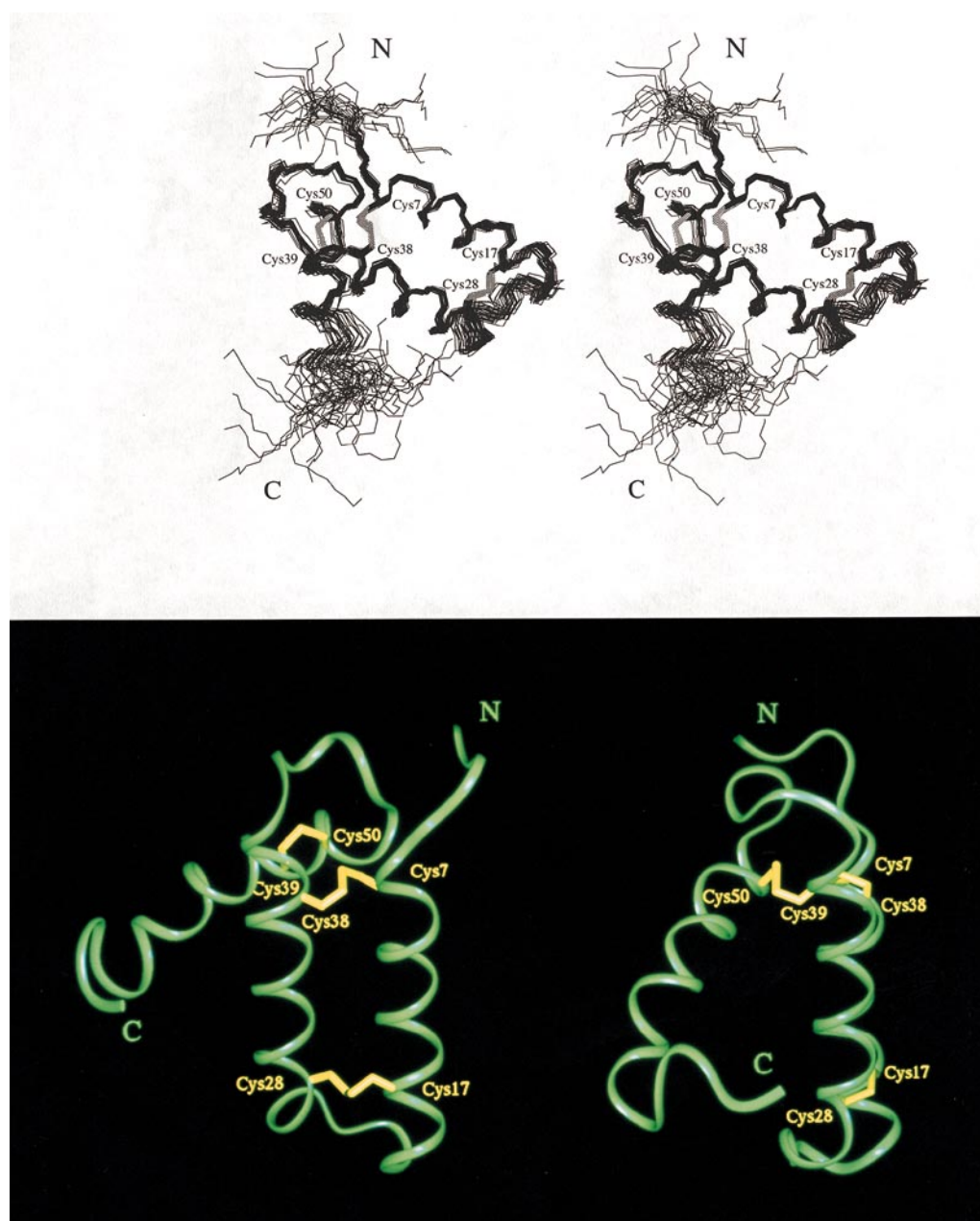


Figure 3. The refined structure of p8^{MTCPI}. (Top) Stereoview of the 30 best structures of p8^{MTCPI}, superimposed over the backbone heavy atoms N, C α and C for residues 5 to 63. Only backbone atoms are shown, except for the three disulfide bridges (7–38, 17–28, and 39–50) (shaded grey). Cysteine residues are labelled. (Bottom) Ribbon plots of the structure of p8^{MTCPI} which is closest to the average, showing the backbone (in green) and disulfide bonds (in yellow). The two views are related by a $\approx 90^\circ$ rotation about the vertical axis.

enforcing the fit on the third helix alone leads to a r.m.s.d. value of $2.10(\pm 0.96)$ Å on the double-helix motif. Technically, the increased global disorder results from the very few long range NOEs observed between helix III and the α -hairpin motif.

Inspection of the dihedral angle distribution and of the hydrogen bonds reveals that Asp5, and Ser47 form N-caps (Richardson & Richardson, 1988), as previously suspected by the inspection of the NOESY maps. At the N terminus of helix III, the side-chain hydroxyl oxygen atom of Ser47 H-bonds with the backbone amide proton of Cys50, as it is usually found for classical N-caps. A more

remarkable conformation is observed for the side-chain of Asp5: on carboxyl oxygen atom forms a hydrogen bond with the backbone amide proton of Gln8, capping the N terminus of helix I, and the other carboxyl oxygen forms a hydrogen bond with the backbone amide proton of Val48, thus participating to the N-capping of helix III.

Discussion

α -Helical coiled-coils represent a wide-spread assembly motif (Kohn *et al.*, 1997). A coiled-coil is composed of two, three or four right-handed

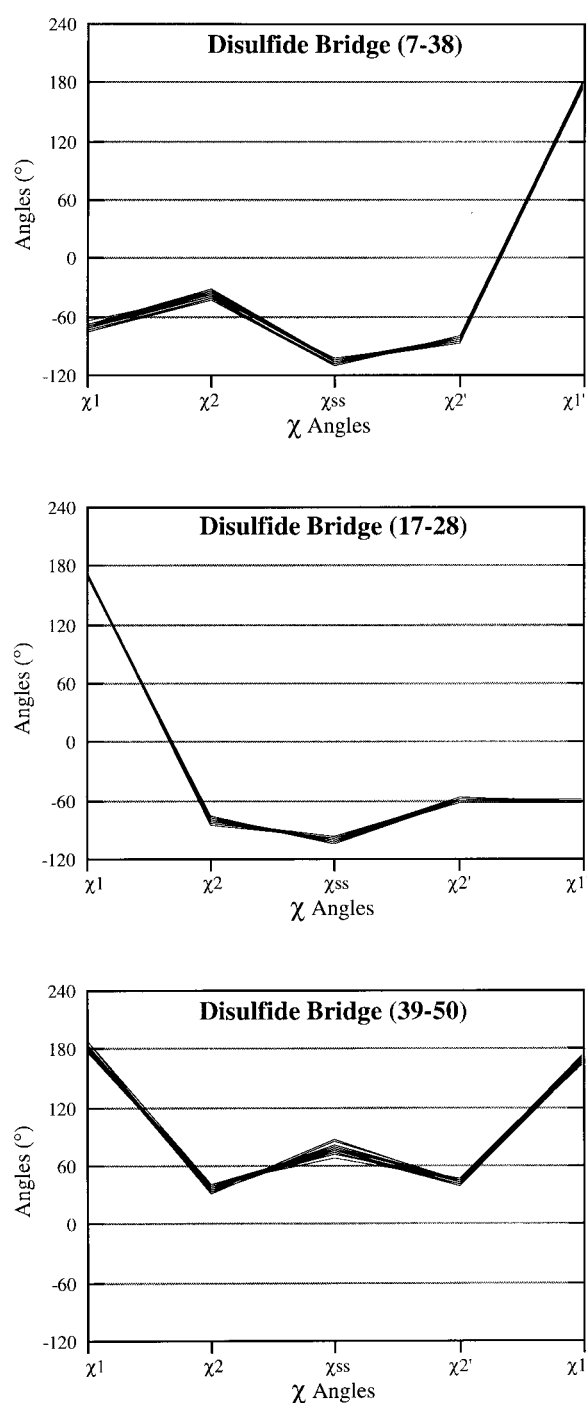


Figure 4. Disulfide bridge geometries for the 30 refined conformers of p8^{MTCPI}. χ_1 and χ_2 correspond to the first numbered cysteine residue in the disulfide bridge, χ_1' and χ_2' to the second one.

amphipathic α -helices wrapped around each other in a left-handed supercoil with a crossing angle of approximately 20° between helices (Schultz & Schirmer, 1979) such that their hydrophobic surfaces are in continuous contact to form, respectively, dimeric, trimeric or tetrameric coiled-coils. The α -helical assembly motif observed for helices I

and II in p8^{MTCPI} represents an original alternative to the stabilization of two antiparallel α -helices. Indeed, a disulfide linkage between α -helical structures (Pullman & Pullman, 1974) is scarcely observed and may involve unusual stereochemical constraints. Other examples of naturally occurring disulfide linkages between α -helical structures are to be found in small toxins such as the plant α -hordothionin from Korean barley (Rao *et al.*, 1994), the *Escherichia coli* heat-stable enterotoxin b (Sukumar *et al.*, 1995) and the recently published neurotoxin B-IV from the marine worm *Cerebratulus lacteus* (Barnham *et al.*, 1997). All these small proteins represent an α -helical hairpin structure, with the branches of the hairpin linked by two (enterotoxin b) or four (α -hordothionin, neurotoxin B-IV) disulfide bonds. In all these cases, the helical geometries are deeply distorted, except for the α -helical hairpin region between the two proximal disulfide bridges Cys16–Cys48 and Cys23–Cys41 in enterotoxin b where the two helices adopt a geometry close to that found in coiled-coil structures. p8^{MTCPI} represents the first example of a natural protein containing an α -antiparallel helical hairpin motif stabilized by two disulfide bridges, without any distortion of either the helices or the disulfide geometries. However, the steric constraints induced by the two disulfide bridges impose an angle of about 5° between the axes of helix I and helix II. This is significantly smaller than the typical value of 20° observed in a super-coil. Disulfide bridges are likely to stabilize this α -hairpin structure, even though the role of individual bonds is difficult to access. Zhou *et al.* (1993) reported that disulfides at the interface of parallel synthetic coiled-coils had a strong influence on their stability, with either stabilizing or destabilizing effects, depending on their location.

Even if the two disulfide bridges play an important role in conferring conformational stability to the α -antiparallel helical motif found in p8^{MTCPI}, the contribution of interchain hydrophobic interactions cannot be neglected. Interestingly, as observed in typical α -helical coiled coils, helix I and helix II are markedly amphipathic, whereas helix III is essentially hydrophilic (Figure 5). The residues at the interface of helix I and helix II, mostly hydrophobic, together with hydrophobic residues located in the loop connecting helix II and III and in the first turn of helix III form a globular hydrophobic core containing the two disulfide bridges 7–38 and 39–50 (Figure 6). Moreover, regular heptad motifs (denoted as *a b c d e f g*; McLachlan & Stewart, 1975) are observed in helix I and II (Figure 5). The residues at position *a* and *d*, which typically form the hydrophobic core in α -helical coiled-coils, are mainly hydrophobic residues or half-cystine residues. Thus, disulfide bridging substitutes to hydrophobic interactions to stabilize the α -helical assembly. This additional stabilization by disulfide bridging is probably necessary due to the relative shortness of the two helices (only two heptads), which is hardly compa-

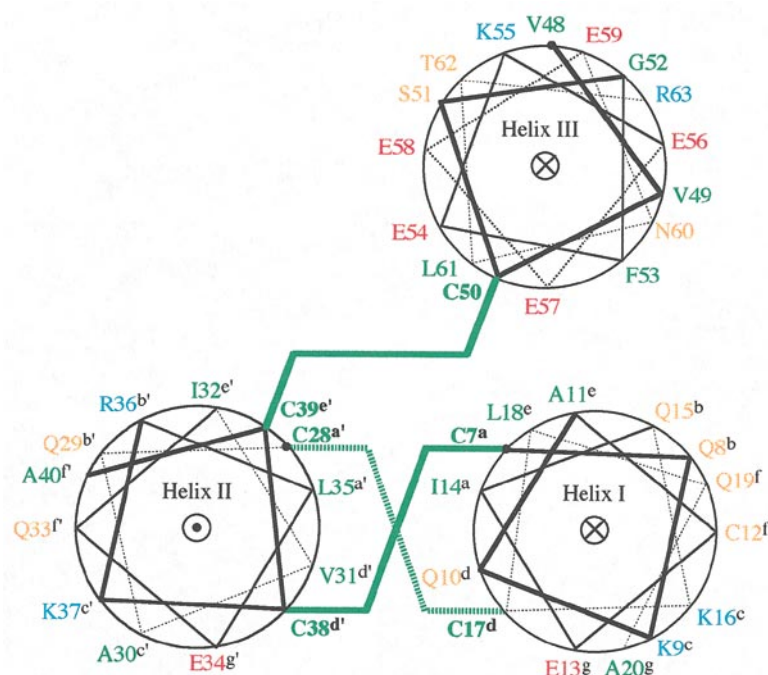


Figure 5. "Wheel diagram" of the three helices of $p8^{MTCP1}$. For the sake of clarity, helix III is presented in the same plane as helix I and helix II. Similarly, the three helices are represented as regular α -helices, whereas the pitch of helix I and II has an intermediate value between α -helix and coiled-coil. Hydrophobic residues and disulfide bridges are labelled in green, polar residues in orange, positively charged residues in blue and negatively charged residues in red. The letters in exposit indicate the heptad repeats.

tible with a self-stabilized coiled-coil assembly (about four heptads at least, generally) involving mainly hydrophobic interactions.

The disulfide disposition in the α -helical hairpin found in $p8^{MTCP1}$ is markedly different from that of neurotoxin B-IV. The half-cystine residues at positions 16, 23, 41 and 48, delimiting the regular helical region of the hairpin in neurotoxin B-IV, are located at $d(d')$ positions of the heptad, and are involved in $d-d'$ disulfide bridges. Thereby, the hydrophobic bonds lay in the same face of the hairpin and are solvent exposed. In $p8^{MTCP1}$, the half-cystine residues Cys7, Cys17 (helix I) and

Cys28, Cys38 (helix II) are located at position a and d (a' , d') on each helix, respectively. This lead to the formation of $a-d'$ ($a'-d$) disulfide bridges, which appears as a more favorable geometry, the hydrophobic bridges being buried in the hydrophobic interface. This different disposition arises from the different cysteine motifs observed in $p8^{MTCP1}$ and neurotoxin B-IV: considering only the previously quoted half-cystine residues, the latter exhibits a -Cys- X_7 -Cys- X_m -Cys- X_7 -Cys- cysteine motif, whereas $p8^{MTCP1}$ is characterized by a -Cys- X_9 -Cys- X_m -Cys- X_9 -Cys- motif. Note that in these two motifs, the half-cystine residues are regularly

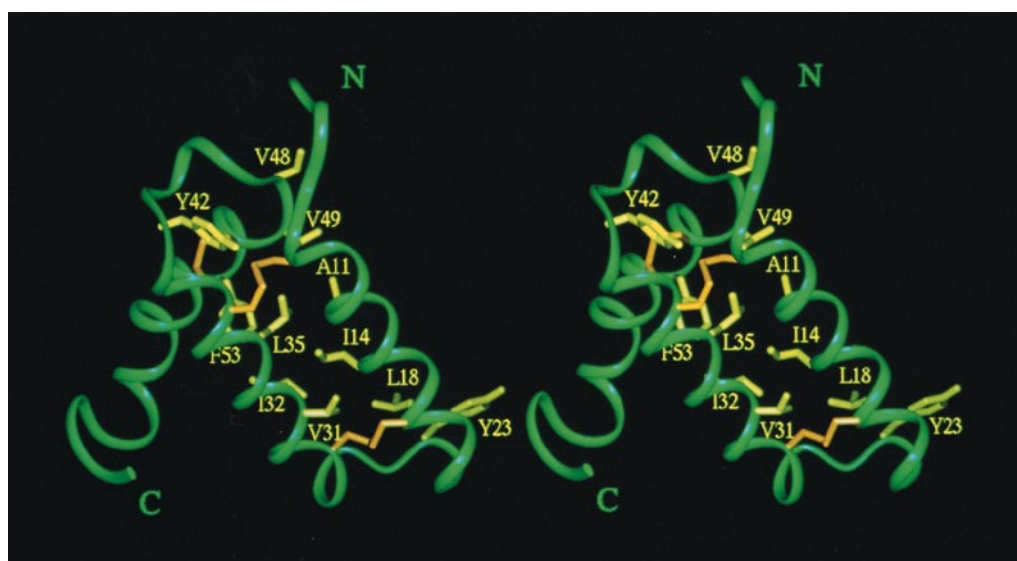


Figure 6. Stereoview (ribbon representation) of the hydrophobic core of $p8^{MTCP1}$. The backbone and the side-chain heavy atoms of the residues forming the core of the molecule are shown. Side-chains are identified with the one-letter amino acid code and the sequence numbers. The backbone is colored in green, the side-chain heavy atoms in yellow. The disulfide bridges are shown in orange.

spaced, which seems compatible with the regular geometry of the helices. Indeed, enterotoxin b, where at least one helix of the α -hairpin is highly distorted, exhibits a -Cys-X₁₀-Cys-X_m-Cys-X₁₁-Cys-motif, and the distorted regions of the α -hairpin found in neurotoxin B-IV correspond to regions where disulfide bonds connect non-equally spaced half-cystine residues.

Charged residues are generally located at the *e* and *g* positions of the heptad repeat: they flank the hydrophobic interface and often participate in interhelical *e-e'* or *g-g'* (*e-g'* or *g-e'* in α -parallel coiled-coils) electrostatic interactions, thereby improving the stability of the helical assembly motif. This is however not the case for p8^{MTCP1}. At positions *g* and *g'*, which are located in the solvent-exposed face of the α -hairpin, residues with the same charge (Glu13 and Glu34) are found so that no favorable interaction can occur. Examination of the refined structures reveals that Glu13 is involved in an intrahelical salt-bridge with either Lys9 or Lys 16 (31% and 35% of the computed structures, respectively). Positions *e* and *e'* are occupied by hydrophobic residues (Ala 11 and Ile 32, respectively) on the opposite face of the α -antiparallel helical motif and involved in hydrophobic contacts with helix III (residues Val49 and Phe53). Presumably due to the presence of the two disulfide bridges, additional interhelical *e-e'* or *g-g'* contacts are not needed to improve the stability of the α -helical assembly. Note that the packing of the initial turns of helix III against the α -antiparallel helical motif is further stabilized by a salt-bridge involving Arg36 (helix II) and either Glu56 or Glu57 (helix III).

It has been predicted that the formation of a coiled-coil is primarily dependent on the presence of these heptad repeat sequences and that discontinuities in heptad periodicity induce changes in the direction of the polypeptide chain in coiled-coils and four- α -helix bundles, including formation of α - α hairpin bends (Cohen & Parry, 1986; Banner *et al.*, 1987). We could speculate that the regular heptad repeats of p8^{MTCP1} probably participate in the earlier events of protein folding, contributing to the correct positioning of helices I and II prior to disulfide bridge formation. Supporting this assumption, the disulfide bridges form spontaneously during protein purification and only one protein species is obtained from the *E. coli* culture.

The three-dimensional structure of p8^{MTCP1} presents a hitherto undescribed folding topology for an α -protein, associated with a new cysteine motif. Thereby, p8^{MTCP1} bears no resemblance to any potentially oncogenic protein previously described, nor does it belong to a well-known functional protein family. The distance matrix algorithm of Holm & Sander (1993) implemented in the program DALI did not find any significant hit except for the coiled-coil of Rop (ColEI repressor of primer; Vlassi *et al.*, 1994; Banner *et al.*, 1987) and for a number of helical regions of different proteins (data not shown), but none of these structures exhi-

bit disulfide bridging between helices. These results only reflect the fact that geometry of the α -hairpin in p8^{MTCP1} resembles a common α -antiparallel coiled-coil, except for the value of the crossing-angle. Finally, little information on the functional role of p8^{MTCP1} can be extracted from these 3D structure homologies, since coiled-coils are widespread scaffolds known to support a variety of different activities (Kohn *et al.*, 1997).

Besides, p8^{MTCP1} shares 40% sequence homology (22% identity), but a different cysteine motif, with a segment of type-1 repeat of thrombospondin proteins, a repeat which is also present in the beta chain of human complement C8 protein (Soulier *et al.*, 1994). The significance of this high sequence homology remains unknown, but it is unlikely that the small mitochondrial protein p8^{MTCP1} shares a common function with such big extracellular mosaic proteins. This may indicate a fortuitous genetic exchange between ancestor genes.

Concluding Remarks

The solution structure of the human p8^{MTCP1} protein, from the MTCP1 oncogene, has been obtained from homonuclear NMR methods. Both the r.m.s.d. and angular order parameters indicate that the structure is resolved with a good accuracy, representing a significant achievement for a non-enriched protein. The new cysteine motif presented by this protein has been found associated with an original alternative for the stabilization of a coiled-coil-like structure: two antiparallel α -helices are tightly strapped together by two disulfide bridges at each end. Additional inter-helix hydrophobic contacts impart stability to this scaffold, but we think that these interactions are more likely to play an important role in the earlier events of the protein folding. Thus, the finding of this new α -helical motif should provide a new insight into the folding pathway of α -helical proteins. Moreover, α -helical motifs have been extensively studied as prime candidates for attempts at *de novo* design of proteins with novel properties (DeGrado *et al.*, 1989; Cohen & Parry, 1990; Hecht *et al.*, 1990; Morii *et al.*, 1991; Kamtekar *et al.*, 1993). To this aim, the α -antiparallel helix scaffold present in the 3D structure of p8^{MTCP1} should constitute a very useful tool: when compared to classical α -helical bundles, the disulfide bonds should provide most of the stabilization energy, leaving a larger part of the protein structure available for mutations. The coordinates of the structure of p8^{MTCP1} are being deposited to the Protein Data Bank (Brookhaven National laboratory) (code access for the average structure: lhp8; code access for the 30 conformers: 2hp8).

Materials and Methods

Sample preparation

The cDNA encoding p8^{MTCP1} was subcloned into the EcoRI site of the pGEX2 T vector (Pharmacia) leading to

an in-frame fusion with the GST gene (Soulier *et al.*, 1994).

BL21 *E. coli* transformed with the recombinant plasmid were used to inoculate 400 ml of medium (Tryptone 20 g/l, Yeast extract 10 g/l, sodium citrate 5 g/l, KH₂PO₄ 5 g/l adjusted to pH 7.0) containing 100 µg/ml ampicillin. After overnight growth at 37°C, the culture was used to inoculate four liters of the same medium supplemented with ampicillin (100 µg/ml), MgSO₄ (10 mM), glucose (5 g/l), biotin (1 mg/l), thiamine (10 mg/l), nicotinamide (10 mg/l) and grown at 37°C for three hours. Expression was induced for three to four hours by addition of 0.5 mM IPTG. The cells were harvested by centrifugation for 20 minutes at 8000 rpm and stored at -80°C.

Cells, 25 g wet weight, stored at -80°C were thawed and resuspended in 100 ml of ice cold 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM EDTA, 1 mM DTT, 5 mM benzamidine, 0.1 g/l PMSF and 0.5 mg/ml lysozyme. The suspension was stirred on ice for 30 minutes and then probe sonicated for four minutes with 0.2 second bursts at 160 W. The particulate material was removed by centrifugation at 50,000 g for two hours and 4°C. PEI was added to the supernatant to a final concentration of 0.3% (w/v) and the solution was stirred on ice for an additional ten minutes. The precipitate was removed by centrifugation at 50,000 g for 30 minutes and the supernatant was applied onto a glutathione-Sepharose®-6B (Pharmacia Biotech) column (5 cm × 1 cm) equilibrated with 50 mM Tris-HCl (pH 6.5), 20 mM NaCl. The column was washed with 500 ml of the same buffer containing 2 M NaCl and then re-equilibrated with the above buffer. GST cleavage was carried out on the column in 40 ml of buffer containing 20 µl of human thrombin (100 units/ml), at room temperature, overnight with gentle shaking. p8^{MTCP1} was eluted in 10 ml fractions. The protein containing fractions were identified by 15% (w/v) SDS-PAGE and pooled. p8^{MTCP1} was loaded onto a SP-Sepharose® FF (Pharmacia Biotech) column (1.6 cm × 10 cm) equilibrated in 50 mM Tris-HCl (pH 6.5), 20 mM NaCl. The column was developed with a 250 ml linear gradient from 0.02 to 2 M NaCl in the above buffer and 1 ml fractions were collected at a flow rate of 1 ml/min. The fractions containing the protein, as shown by SDS-PAGE, were pooled and concentrated using a Centriprep® 3 concentrator (Amicon®). p8^{MTCP1} was further purified by gel filtration on Sephacryl® HR 100 (Pharmacia Biotech) column (1.6 cm × 100 cm) equilibrated with 50 mM sodium phosphate buffer (pH 6.5). Fractions (1 ml) were collected at a flow rate of 0.2 ml/min. The protein was eluted as a single peak and the corresponding fractions were pooled and concentrated.

Thiol titration

The titration of thiol groups in the protein was done according to Ellman's method (Riddles *et al.*, 1983) by incubating (five minutes) an aliquot of 10 µl from the NMR sample in a freshly prepared solution of DTNB. As an excess of deuterated DTT was present in solution, it was not possible to accurately couple the quantity released TNB to the number of free cysteines in the protein. We therefore purified the modified protein by RP-HPLC and analyzed it by mass spectroscopy. The electron spray spectrum gives a main peak at 9065.5 Da, indicating that the protein reacts with one equivalent of DTNB (expected mass: 9067). We obtained a similar result when the aliquot from the

NMR sample was RP-HPLC purified prior to modification with DTNB: due to the very low pH of the eluant buffer (H₂O-AcN-TFA, pH 2), this clearly demonstrates that the cystinyl residues are not involved in a zinc cluster.

NMR measurements

All NMR experiments, i.e. DQF-COSY (Rance *et al.*, 1983), z-TOCSY (Braunschweiler & Ernst, 1983; Davis & Bax, 1985; Rance, 1987), NOESY (Jeener *et al.*, 1979; Kumar *et al.*, 1980) were carried out at 600 MHz, on a Bruker AMX600 spectrometer equipped with a z-gradient ¹H-¹³C-¹⁵N triple resonance probe. All samples were 2 mM in protein, with approximately 1.5 equivalents of (²H)DTT, the pH (pD) was adjusted to 6.5 in order to prevent protein degradation. The pD values are uncorrected for isotopic effects. Samples dissolved in H₂O contained 10% (v/v) ²H₂O for the lock. Data were acquired on samples maintained at 25°C or 35°C, and chemical shifts are reported relative to (²H)TSP. Deuterated reagents ²H₂O, ²HCl, NaO²H, (²H)DTT, (²H)TSP were from EURISOTOP.

For NOESY experiments, two mixing times of 100 and 200 ms were used. z-TOCSY experiments were carried out using the TOWNY isotropic transfer sequence (Kadkhodaei *et al.*, 1993) and a mixing time of 60 ms. In all 2D experiments, quadrature detection in the indirectly observed dimension was obtained with States-TPPI (Marion *et al.*, 1989a). Solvent suppression in z-TOCSY and NOESY experiments was carried out using the WATERGATE method (Piotto *et al.*, 1992) in association with water-flip-back pulses (Dalhuin *et al.*, 1996; Lippens *et al.*, 1995). For DQF-COSY, water suppression was obtained with a low-power irradiation of the solvent signal during the relaxation delay. In addition, a solvent-suppression filter was applied to the time-domain data (Marion *et al.*, 1989b). The spectral width used in both dimensions was 7800 Hz; 4096 complex data points in *t*₂ and 512 experiments in *t*₁ were usually acquired, except for DQF-COSY where 1024 experiments were acquired in *t*₁. Typically, 128 scans (NOESY, DQF-COSY) or 96 scans (z-TOCSY) were acquired per increment. The data sets were processed using UXNMR software. Alternatively, NMR spectra were also treated with GIFA (Pons *et al.*, 1996) software utilities. Prior to Fourier transformation the raw data were multiplied by a squared cosine window function in *t*₂ and a shifted (π/4) sine bell window function in *t*₁ and were zero-filled to the next power of 2 leading to a resolution of 1.9 Hz in the *f*₂ dimension and 7.62 Hz (NOESY, z-TOCSY) or 3.8 Hz (DQF-COSY) in the *f*₁ dimension. Residual baseline distortions in *f*₂ were removed with a fifth-order polynomial baseline correction, and a linear baseline correction was applied in the *f*₁ dimension.

NH-C^αH and C^αH-C^βH coupling constants were measured in the DQF-COSY spectra (2 Hz of digital resolution). In order to compensate for the over-estimation of the ³J_{NH-H_α} due to the broad line width of the anti-phase multiplets, we used the method previously described by Ludvigsen *et al.* (1991), thanks to a macro-program implemented in the GIFA software. Spin systems determination and sequential assignment of the ¹H resonances were computer assisted using the in-house CINDY software (Padilla, to be published), operating on a Silicon Graphics O₂ workstation.

Molecular modeling calculations

All calculations were carried out on a Hewlett-Packard HP735 workstation. The volumes of cross-peaks were integrated in the 100 ms NOESY spectra and were classed as strong, medium and weak, and converted to the corresponding distance constraints of 1.8 to 2.8 Å, 2.4 to 3.6 Å, and 3.5 to 4.8 Å, respectively. Additionally, some cross-peaks which were present only in the 200 ms 2D NOESY maps were converted to distance constraints of 5.5 Å (no lower limit). A total of 35 restraints on the ϕ dihedral angles were obtained from the coupling constants $^3J_{\text{NH-H}\alpha}$ using the empirically calibrated Karplus relation (Karplus, 1963; Pardi *et al.*, 1984). 31 angles corresponding to coupling constraints smaller than 6 Hz were restrained in the $-60(\pm 20)^\circ$ interval. Four angles corresponding to coupling constants greater than or equal to 9 Hz were restrained to $-120(\pm 25)^\circ$. Moreover, 22 χ_1 angles were obtained from the analysis of the $^3J_{\text{H}\alpha\text{-H}\beta}$ coupling constants and intra-residue NOEs (Hyberts *et al.*, 1987). The limits of the intervals were set to $\pm 45^\circ$.

3D structures were deduced from the experimental distance and angle restraints using first the variable target function program DIANA (Güntert & Wüthrich, 1991). Each run started from 1000 randomized conformers. When no stereospecific assignment was possible, pseudoatoms were defined and corrections added as described by Wüthrich *et al.* (1983). Fifty structures were generated from a preliminary set of distance and angle constraints. The analysis of the structures allowed us to resolve ambiguities in the assignment of NOE cross-peak arising from chemical shift degeneracy and thus the derivation of additional interproton restraints. To this aim, we used a routine of the CINDY software which allows the calculation of the NOESY 2D map from the calculated structures: the superimposition of the calculated and experimental maps gives an easy analyzable graphic representation of both the qualitative and semi-quantitative information from the NOE cross-peaks. This procedure was iterated several times by calculating at each step a new set of structures with an improved list of restraints. A set of 50 structures was generated with the final list of restraints. This list (from which values redundant with the covalent geometry have been eliminated by DIANA) consisted of 201 intra-residues, 290 sequential, 292 medium-range ($i-j < 5$), and 157 long-range upper bound restraints, 721 lower bound restraints, and 35 backbone and 22 side-chain dihedral restraints. No additional restraints were used for hydrogen bonds. Among the 50 preliminary structures generated by DIANA, 20 presented a value of the target function smaller than 3.8 Å and no distance violation larger than 0.62 Å. The ϕ and χ_1 angles showed no violation greater than 10.4° . The low value of the target function obtained from a high number of constraints indicates the consistency of the data set and corresponds to structures of good quality. In these preliminary structures, no disulfide bridge was enforced, and the calculations were done using cystinyl residues with deprotonated sulfur atoms. In this case the van der Waals radii for the cystinyl C $^\beta$ and S $^\gamma$ atoms are reduced when calculating their mutual quadratic repulsive interaction, allowing all possible disulfide interactions to occur. These structures were then used to determine the location of the disulfide bridges by a statistical analysis of the distances between the sulfur and/or the C $^\beta$ of the cystinyl residues. The final DIANA structure calculation included the usual distance constraints to enforce the three disulfide bridges, i.e.

ranges of 2.0 to 2.1 Å for $d(\text{S}^\gamma, \text{S}^\gamma)$, 3.0 to 3.1 Å for $d(\text{C}^\beta, \text{S}^\gamma)$ and 3.5 to 4.5 Å for $d(\text{C}^\beta, \text{C}^\beta)$. The complete list of experimental restraints is available upon request from the author.

The 30 best structures (based on the final target penalty function values) were further submitted to molecular mechanics energy refinement with the SANDER module of AMBER 4.1 (Pearlman *et al.*, 1995) using the 1994 force field (Cornell *et al.*, 1995). To reduce well-known artifacts arising from *in vacuo* simulations (charged and polar side-chains tend to fold back toward the protein to find hydrogen bonding partners), the net charge of ionised side-chains was reduced to ± 0.2 using a modified residue database, and a distance-dependent dielectric constant was used ($\epsilon = r$, Guenot & Kollman, 1992). During the molecular dynamics runs, the covalent bond lengths were kept constant by applying the SHAKE algorithm (van Gunsteren & Berendsen, 1977) allowing a 2 fs time step to be used. The non-bonded pair list was updated every 50 steps and the temperature was regulated by coupling the system to a heat bath with a coupling constant of 0.2 ps. Pseudo-energy terms taking into account the NMR inter-proton distance restraints were defined as follows *via* four threshold distance values: r_1 , r_2 , r_3 and r_4 . In all cases r_1 and r_2 were set to 1.3 and 1.8 Å, respectively. r_3 was taken as the upper boundary used in DIANA calculations, and r_4 was chosen as $r_3 + 0.5$ Å. For an observed distance lying between r_2 and r_3 , no restraint was applied. Between r_1 and r_2 or between r_3 and r_4 , parabolic restraints were applied. Outside the r_1 to r_4 range, the restraints were linear with slopes identical at parabolic slopes at points r_1 and r_4 . When no stereospecific assignment could be achieved for the methyl or methylene protons, an $\langle r^{-6} \rangle^{-1/6}$ averaging scheme was used instead of pseudoatoms. 5000 cycles of restrained energy minimisation were first carried out followed by a 50 ps long simulated annealing procedure in which the temperature was raised to 900 K for 30 ps then gradually lowered to 300 K. During this stage, the force constants for the NMR distance constraints and for the angular constraints were gradually increased from 3.2 to 32 kcal mol $^{-1}$ Å $^{-2}$ and 0.5 to 50 kcal mol $^{-1}$ rad $^{-2}$, respectively. A final restrained minimisation led to the refined structures discussed below.

The structures were displayed and analyzed on a Silicon Graphics O $_2$ station using the INSIGHT program (version 2.3.0, Biosym Technologies, San Diego).

Acknowledgments

We thank Dr F. Penin (Laboratoire de Conformation des Protéines, Lyon) for helpful discussions, Dr J. P. Lecaer (Laboratoire de Neurobiologie, ESPCI-CNRS URA 2054, Paris) for mass spectra analysis of the proteins, and Dr R. Ghirlando (C.B.S., Montpellier) for critical reading of the manuscript. This work was supported by grants from the Ligue Nationale contre le Cancer and from the Association pour la Recherche sur le Cancer.

References

- Banner, D. W., Kokkinidis, M. & Tsernoglou, D. (1987). Structure of the ColE1 Rop protein at 1.7 Å resolution. *J. Mol. Biol.* **196**, 657–675.

- Barnham, K., Dyke, T. R., Kem, W. & Norton, S. N. (1997). Structure of neurotoxin B-IV from the marine worm *Cerebratulus lacteus*: a helical hairpin cross-linked by disulfide bridges. *J. Mol. Biol.* **268**, 886–902.
- Beatty, D. W., Arens, L. J. & Nelson, M. M. (1986). Ataxia-telangiectasia. X,14 translocation, progressive deterioration of lymphocyte numbers and function and abnormal in vitro immunoglobulin production. *S. Afr. Med. J.* **69**, 115–118.
- Billeter, M., Braun, W. & Wüthrich, K. (1982). Sequential resonance assignments in protein ¹H nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances and statistical analysis of proton-proton distances in single crystal protein conformations. *J. Mol. Biol.* **155**, 321–346.
- Braunschweiler, J. & Ernst, R. R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521–528.
- Canki, N., Tivadar, I., Zupancic, N. & Devebec, M. (1983). Citogenetska studija sedmih bolnic z atak-sijo-teleangiiektazijo. *Zdrav. Vestn.* **52**, 567–570.
- Cohen, C. & Parry, D. A. D. (1986). α -Helical coiled coils: a widespread motif in proteins. *Trends Biochem. Sci.* **11**, 245–248.
- Cohen, C. & Parry, D. A. D. (1990). α -Helical coiled-coils and bundles: how to design an α -helical protein. *Proteins: Struct. Funct. Genet.* **7**, 1–15.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr, Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **117**, 5179–5197.
- Dalhuin, C., Wieruszeski, J. M. & Lippens, G. (1996). An improved homonuclear TOCSY experiment with minimal water saturation. *J. Magn. Reson. B*, **111**, 168–170.
- Dallapiccola, B., Alimena, G., Chessa, L., Gastaldi, R., De Rossi, G., Semenzato, G., Quinti, I. & Pandolfi, F. (1984). Chromosome studies in patients with T-CLL chronic lymphocytic leukemia and expansions of granular lymphocytes. *Int. J. Cancer*, **34**, 171–176.
- Davis, D. G. & Bax, A. (1985). Assignment of complex ¹H NMR spectra with two-dimensional homonuclear Hartmann-Hahn spectroscopy. *J. Am. Chem. Soc.* **107**, 2820–2821.
- DeGrado, W. F., Wasserman, Z. R. & Lear, J. D. (1989). Protein design, a minimalist approach. *Science*, **243**, 622–628.
- Fisch, P., Forster, A., Sherrington, P. D., Dyer, M. J. & Rabbit, T. H. (1993). The chromosomal translocation t(X;14)(q28;q11) in T-cell pro-lymphocytic leukemia breaks within one gene and activates another. *Oncogene*, **8**, 3271–3276.
- Fu, T.-B., Virgilio, L., Narducci, M. G., Facchiano, A., Russo, G. & Croce, C. M. (1994). Characterization and localization of the TCL-1 oncogene product. *Cancer Res.* **54**, 6297.
- Goyns, M. H., Hammond, D. W., Harrison, C. J., Menasce, L. P., Ross, F. M. & Hancock, B. W. (1993). Structural abnormalities of the X chromosome in non-Hodgkin's lymphoma. *Leukemia*, **7**, 848–852.
- Guenot, J. M. & Kollman, P. A. (1992). Molecular dynamics studies of a DNA-binding protein. *Protein Sci.* **1**, 1185–1205.
- Güntert, P. & Wüthrich, K. (1991). Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J. Biomol. NMR*, **1**, 447–456.
- Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. (1990). De novo design, expression and characterization of Felix: a four-helix bundle protein of native-like sequence. *Science*, **249**, 884–891.
- Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
- Hyberts, S. G., Märki, W. & Wagner, G. (1987). Stereospecific assignments of side-chain protons and characterization of torsion angles in eglin c. *Eur. J. Biochem.* **164**, 625–635.
- Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. (1992). The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751.
- Jeener, J., Meier, B. H., Bachman, P. & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4552.
- Kadkhodaei, M., Hwang, T.-L., Tang, J. & Shaka, A. J. (1993). A simple windowless mixing sequence to suppress cross relaxation in TOCSY experiments. *J. Magn. Reson. A*, **105**, 393–399.
- Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. (1993). Protein design by binary patterning of polar and nonpolar amino acids. *Science*, **262**, 1680–1685.
- Karplus, M. (1963). Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* **85**, 2870–2871.
- Kohn, W. D., Mant, C. T. & Hodges, R. S. (1997). α -Helical protein assembly motifs. *J. Biol. Chem.* **272**, 2583–2586.
- Kumar, A., Ernst, R. R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
- Liff, M. I. & Siddiqui, S. S. (1996). NMR evidence of formation of cyclocystine loops in peptide models of the high sulfur proteins from wool. *Int. J. Biol. Macromol.* **19**, 139–143.
- Lippens, G., Dalhuin, C. & Wieruszeski, J. M. (1995). Use of a water flip-back pulse in the homonuclear NOESY experiment. *J. Biol. NMR*, **5**, 327–331.
- Ludvigsen, S., Andersen, K. V. & Poulsen, F. M. (1991). Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of ϕ angles. *J. Mol. Biol.* **217**, 731–736.
- Madani, A., Soulier, J., Schmid, M., Plichtova, R., Lermé, F., Gateau-Roesch, O., Garnier, J. P., Pla, M., Sigaux, F. & Stern, M.-H. (1995). The 8kD product of the putative oncogene MTCP-1 is a mitochondrial protein. *Oncogene*, **10**, 2259–2262.
- Madani, A., Choukroun, V., Soulier, J., Cacheux, V., Claisse, J. F., Valensi, F., Daliphard, S., Cazin, B., Levy, V., Leblond, V., Daniel, M. T., Sigaux, F. &

- Stern, M. H. (1996). Expression of P13^{MTCP1} is restricted to T-cell proliferations with t(X;14) translocations. *Blood*, **87**, 1923–1927.
- Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989a). Rapid recording of 2D NMR data without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399.
- Marion, D., Ikura, M. & Bax, A. (1989b). Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data. *J. Magn. Reson.* **84**, 425–430.
- McLachlan, A. D. & Stewart, M. (1975). Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J. Mol. Biol.* **98**, 293–304.
- Morii, H., Ichimura, K. & Uedaira, H. (1991). Asymmetric inclusion by the novo designed proteins: fluorescence probe studies on amphiphilic α -helix bundles. *Proteins: Struct. Funct. Genet.* **11**, 133–141.
- Pardi, A., Billeter, M. & Wüthrich, K. (1984). Calibration of the angular dependence of the amide proton- C^α proton coupling constants, $^3J_{NH^\alpha}$, in a globular protein. *J. Mol. Biol.* **180**, 741–751.
- Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III, Fergusson, D. M., Seibel, G. L., Chandra Singh, U., Weiner, P. K. & Kollman, P. A. (1995). *AMBER 4.1*, University of California, San Francisco.
- Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single quantum spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
- Pons, J. L., Malliavin, T. E. & Delsuc, M. A. (1996). Gifa V.4: a complete package for NMR data set processing. *J. Biol. NMR*, **8**, 445–452.
- Pullman, B. & Pullman, A. (1974). Molecular orbital calculations of the conformation of amino acid residues of proteins. *Advan. Protein Chem.* **28**, 347–526.
- Rabbitts, T. H. (1994). Chromosomal translocation in human cancer. *Nature*, **372**, 143–149.
- Rance, M. (1987). Improved techniques for homonuclear rotating-frame and isotropic mixing experiments. *J. Magn. Reson.* **74**, 557–564.
- Rance, M., Sorensen, O. W., Bodenhausen, G., Ernst, R. R. & Wüthrich, K. (1983). Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479–495.
- Rao, A. G., Hassan, M. & Hempel, J. C. (1994). Structure-function validation of high lysine analogs of α -hordothionin designed by protein modeling. *Protein Eng.* **7**, 1485–1493.
- Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–330.
- Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the end of α helices. *Science*, **240**, 1648–1652.
- Riddles, P. W., Blakely, R. L. & Zerner, B. (1983). Reassessment of Ellman's reagent. *Methods Enzymol.* **91**, 49–60.
- Sherrington, P. D., Fish, P., Taylor, A. M. R. & Rabbitts, T. H. (1994). Clonal evolution of malignant and non-malignant T cells carrying t(14;14) and t(X;14) in patients with ataxia telangiectasia. *Oncogene*, **9**, 2377–2381.
- Shultz, G. E. & Schirmer, R. H. (1979). *Principles of Protein Structure*, pp. 79–81, Springer-Verlag, New York.
- Soulier, J., Madani, A., Cacheux, V., Rosenzweig, M., Sigaux, F. & Stern, M.-H. (1994). The MTCP-1/c6.1B gene encodes for a cytoplasmic 8 kDa protein over-expressed in T cell leukemia bearing a t(X;14) translocation. *Oncogene*, **9**, 3565–3570.
- Srinivasan, N., Sowdhamini, R., Ramakrishnan, C. & Balaram, P. (1990). Conformations of disulfide bridges in proteins. *Int. J. Peptide Protein Res.* **36**, 147–155.
- Stern, M. H., Theodorou, I., Aurias, A., Maier-Redelsperger, M., Debre, M., Debre, P. & Griscelli, C. (1989). T-cell nonmalignant clonal proliferation in ataxia telangiectasia: a cytological, immunological, and molecular characterization. *Blood*, **73**, 1285–1290.
- Stern, M. H., Soulier, J., Rosenzweig, M., Nakahara, K., Canki-Klain, N., Aurias, A., Sigaux, F. & Kirsch, I. R. (1993). MTCP-1: a novel gene of the human chromosome Xq28 translocated to the T cell receptor α/δ locus in mature T Cell proliferations. *Oncogene*, **8**, 2475–2483.
- Sukumar, M., Rizo, J., Wall, M., Dreyfus, L. A., Kupersztoch, Y. M. & Gierasch, L. M. (1995). The structure of *Escherichia coli* heat-stable enterotoxin b by nuclear magnetic resonance and circular dichroism. *Protein Sci.* **4**, 1718–1729.
- Taylor, A. M. R., Oxford, J. M. & Metcalfe, J. A. (1981). Spontaneous cytogenetic abnormalities in lymphocytes from thirteen patients with ataxia telangiectasia. *Int. J. Cancer*, **27**, 311–319.
- Taylor, A. M. R., Lowe, P. A., Stacey, M., Thick, J., Campbell, L., Beatty, D., Biggs, P. & Formstone, C. J. (1992). Development of T-cell leukemia in an ataxia telangiectasia patient following clonal selection in t(X;14)-containing lymphocytes. *Leukemia*, **6**, 961–966.
- Taylor, A. M. R., Metcalfe, J. A., Thick, J. & Mak, Y. F. (1996). Leukemia and lymphoma in ataxia telangiectasia. *Blood*, **87**, 423–438.
- van Gunsteren, W. F. & Berendsen, H. J. C. (1977). Algorithms for macromolecular dynamics and constraints dynamics. *Mol. Phys.* **34**, 1311–1327.
- Vlassi, M., Steif, C., Weber, P., Tsernoglou, D., Wilson, K. S., Hinz, H.-J. & Kokkinidis, M. (1994). Restored heptad pattern continuity does not alter the folding of a four- α -helix bundle. *Struct. Biol.* **1**, 706–716.
- Wagner, G. & Wüthrich, K. (1982). Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J. Mol. Biol.* **155**, 347–366.
- Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N. & Wüthrich, K. (1987). Proteins structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611–639.
- Wider, G., Lee, K. H. & Wüthrich, K. (1982). Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Glucagon bound to perdeuterated dodecylphosphocholine micelles. *J. Mol. Biol.* **155**, 367–388.
- Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* **222**, 311–333.
- Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry*, **31**, 1647–1651.

- Wishart, D. S., Bigam, C. G., Yao, J., Abilgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995). ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.
- Witzig, T. E., Phylly, R. L., Li, C.-Y., Homburger, H. A., Dewald, G. W. & Handwerger, B. S. (1986). T-cell chronic lymphocytic leukemia with a helper/inducer membrane phenotype: a distinct clinicopathologic subtype with a poor prognosis. *Am. J. Hematol.* **21**, 139–155.
- Wüthrich, K. (1986). *NMR of proteins and nucleic acids*, Wiley, New York.
- Wüthrich, K., Billeter, M. & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.
- Zhang, R. & Snyder, G. (1989). Dependence of formation of small disulfide loops in two-cysteine peptides on the number and types of intervening amino-acids. *J. Biol. Chem.* **264**, 18472–18479.
- Zhou, N. E., Kay, C. M. & Hodges, R. S. (1993). Disulfide bond contribution to protein stability: pos-

itional effects of substitution in the hydrophobic core of the two-stranded α -helical coiled-coil. *Biochemistry*, **32**, 3178–3187.

Edited by P. E. Wright

(Received 13 June 1997; received in revised form 15 September 1997; accepted 18 September 1997)



<http://www.hbuk.co.uk/jmb>

Supplementary material comprising a Table and two Figures is available from JMB Online.