

REFERENCES

1. M. Maeda, Y. Tanaka, M. Fukutoma, T. Asano, *Jpn. J. Appl. Phys.* **27**, 209 (1988).
2. Z. Z. Sheng and A. M. Hermann, *Nature* **332**, 55 (1988).
3. ———, *ibid.*, p. 138.
4. S. N. Putlin, E. V. Antipov, O. Chmaissem, M. Marezio, *ibid.* **362**, 226 (1993).
5. A. C. Larson and R. B. von Dreele, General Structure Analysis System, University of California (1985).

1 April 1994; accepted 18 May 1994

Crystal Structure of P22 Tailspike Protein: Interdigitated Subunits in a Thermostable Trimer

Stefan Steinbacher,* Robert Seckler, Stefan Miller, Boris Steipe, Robert Huber, Peter Reinemer*

The tailspike protein (TSP) of *Salmonella typhimurium* phage P22 is a part of the apparatus by which the phage attaches to the bacterial host and hydrolyzes the O antigen. It has served as a model system for genetic and biochemical analysis of protein folding. The x-ray structure of a shortened TSP (residues 109 to 666) was determined to a 2.0 angstrom resolution. Each subunit of the homotrimer contains a large parallel β helix. The interdigitation of the polypeptide chains at the carboxyl termini is important to protimer formation in the folding pathway and to thermostability of the mature protein.

The tailspike protein of *Salmonella typhimurium* phage P22 is a homotrimeric structural protein of 666 amino acid residues, which is noncovalently bound to the neck of the virus capsid and essential for phage adsorption to the bacterial host (1). It displays endorhamnosidase enzymatic activity, hydrolyzing the α -1,3-O-glycosidic linkage between rhamnose and galactose of *Salmonella* O-antigen polysaccharide (2).

The tailspike protein has served as a model system for the folding and assembly of large, multi-subunit proteins. The folding pathway comprises several consecutive steps, such as subunit folding and the formation of a protimer, in which the chains are stably associated but not fully folded, and a rate-limiting folding reaction from the protimer to native TSP (3). The native trimer is thermostable beyond 80°C, resistant to proteases, and not dissociated by SDS. However, even late intermediates in the folding pathway are thermolabile (4), and the folding efficiency decreases strongly with increasing temperature, both in vivo (5) and in vitro (6).

Folding of the TSP has been the subject of detailed genetic analysis, yielding a number of lethal mutations (7) and two types of point mutations affecting the folding efficiency at high temperature. (i) Temperature-sensitive folding (tsf) mutations (8) reduce the folding yield and enhance aggregation (5). There are at least 60 different tsf mutations distributed over the central third

of the TSP gene (8). (ii) Global suppressor mutations of the tsf phenotype increase the folding yield at high temperature and map to only a few sites in the gene (9). We deter-

mined the x-ray crystal structure of an NH₂-terminally shortened TSP at 2.0 Å resolution (10–12).

The shortened TSP crystallized in cubic space group P2₁3 with one monomer in the asymmetric unit. Thus, the three subunits of the homotrimer are related by threefold crystallographic symmetry. The structure was solved by multiple isomorphous replacement (MIR) techniques (Table 1) (12). The molecule is 133 Å in length and between 35 and 80 Å in diameter (Fig. 1A). Each monomer has the overall shape of a fish and is composed of six segments corresponding to the main body, the mouth, the dorsal fin, and the first, second, and third segments of the caudal fin, respectively (Fig. 1B). The secondary structure, which was analyzed with the program DSSP (13), is dominated by three parallel (A to C) and two antiparallel β sheets (D and E). In addition, the dorsal fin contains a strongly twisted antiparallel β sheet F. There are only five short α helices, α 1 to α 5. No disulfide bridges are present in the structure.

The main body of each subunit is formed by a parallel β helix, comprising 13 complete turns (residues 143 to 540), in which parallel β strands are coiled into a large right-handed helix. The β helix is collapsed

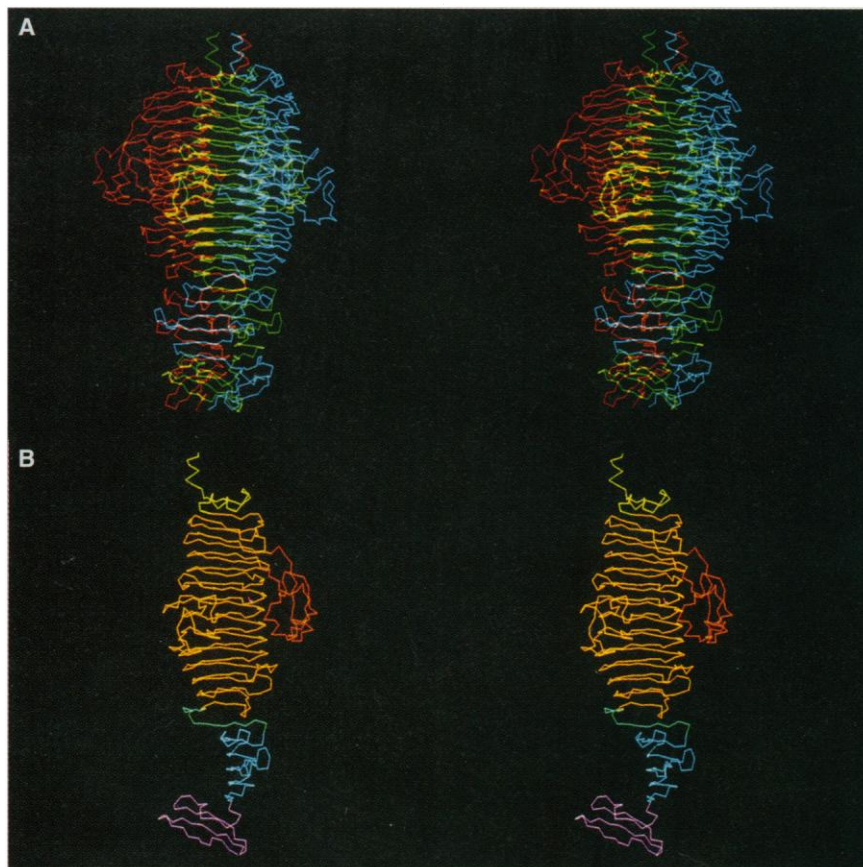


Fig. 1. Stereo view of the tertiary structure of TSP (C α traces). (A) The subunits of the trimer are shown in red, blue, and green with the NH₂-termini at the top. (B) Each subunit is composed of the mouth (yellow), the main body (orange), the dorsal fin (red), and the first (green), second (blue), and third (magenta) segments of the caudal fin.

S. Steinbacher, B. Steipe, R. Huber, P. Reinemer, Max-Planck-Institut für Biochemie, Abteilung Struktur-forschung, 82152 Martinsried, Germany.
R. Seckler and S. Miller, Universität Regensburg, Physikalische Biochemie, 93040 Regensburg, Germany.

*To whom correspondence should be addressed.

Table 1. Data collection and phasing statistics (24).

Parameter	Native I	Native II	HGCR	HGME	CMNP	UOSO	PTL4	CMPT
Total observations	62380	143312	51196	68103	66662	61404	87673	60060
Unique reflections	11402	35321	11466	11430	10950	11104	11444	11363
Completeness (%) [*]	94.6/78.9	88.2/57.9	96.4/82.6	94.7/90.9	90.7/67.8	92.1/86.3	94.6/93.3	94.3/90.1
$R_{\text{merge}}^{\dagger}$	0.109	0.132	0.124	0.113	0.176	0.128	0.106	0.110
$R_{\text{sym}}^{\ddagger}$	0.046	0.065	0.058	0.085	0.078	0.073	0.039	0.055
Binding sites			2	2 [#]	2	3	1	3
R_{iso} (25.0 to 3.0 Å) [§]			0.170	0.207	0.157	0.187	0.100	0.157
Phasing power			1.25	1.22	1.27	1.10	1.26	1.72

^{*}For native I and derivatives, 158–3.0/3.05–3.00 Å, and for native II, 158–2.0/2.03–2.00 Å. $\dagger R_{\text{merge}} = \sum_i \sum_h |I(h,i) - \langle I(h) \rangle| / \sum_i \sum_h I(h,i)$, where $I(h,i)$ is the intensity value of the i -th measurement of h and $\langle I(h) \rangle$ is the corresponding mean value of h for all i measurements of h ; the summation is over all measurements. $\ddagger R_{\text{sym}} = \sum_i (I_i - \langle I_i \rangle) / \sum_i I_i$, where I_i is the averaged value of point group-related reflections and $\langle I_i \rangle$ is the averaged value of a Bijvoet pair. $\S R_{\text{iso}} = \sum_i (F_{\text{PH}}^2 - F_i^2) / \sum_i (F_{\text{PH}}^2 + F_i^2)$, where F_{PH} and F_i are the derivative and the native structure-factor amplitudes, respectively. || F_i /residual: rms. mean heavy-atom contribution/rms. residual, defined as $[(F_{\text{PHC}}^2 - F_{\text{PH}}^2)/n]^{1/2}$ with the sum over all reflections, where F_{PHC} is the calculated structure factor and F_{PH} is the structure-factor amplitude of the heavy atom derivative, respectively. [#]Same as in CMNP.

to an approximately triangular cross section, with each face formed by one of the parallel β sheets A, B, and C (Fig. 2). The sheets consist of short strands of two to three, two to three, and three to six residues, respectively, but the hydrogen-bonding pattern extends into the connecting turns to generate an almost uniform β helix. Sheets A and B are connected by short turns, while turns and loops of more variable length connect sheets B and C and A and C, respectively. In the turn regions between β sheets B and C in turns 5 to 13 of the β helix, two residues make a sharp inward bend (residues 331 and 332 in Fig. 2) and form a long groove. In the turn region between β sheets A and C, loops with 5 to 25 residues extrude from turns 5 to 8 of the β helix. The insertions in turns 5, 7, and 8 pack together at the solvent-exposed side of the β helix. The loop inserted in turn 6 is involved in an intersubunit contact with the neighboring dorsal fin of the symmetry-related monomer. A complete β -helix turn comprises approximately 22 residues in the NH_2 -terminal and central parts and 16 residues in the last COOH-terminal turn. Concomitantly, its diameter decreases from 26 to 11 Å at the COOH-terminus.

In the trimer β sheets A and B of two adjacent β helices form a β sandwich. About 66 residues are involved in the intersubunit contacts between the β helices; 18 are charged and 18 are polar, generating a hydrophilic interface. The wide channel along the molecular threefold axis harbors about 80 ordered water molecules. In contrast, the β helix is mainly filled with hydrophobic residues and contains only 12 ordered water molecules, hydrogen-bridged to polar side chains or carbonyl groups. The hydrophobic side chains often form stacks of aliphatic and aromatic residues. Within these stacks, there are short repetitive elements composed of two Leu, two and three Ile, two and three Val, two Ala, four Gly, two and three Phe, and two Asn residues.

In the NH_2 -terminal mouth segment (residues 113 to 142), the β helix is extended by

one turn formed by the first strand of β sheet A and the seven-residue α helix $\alpha 2$, which lies almost parallel to the solvent-exposed face of the β helix. This motif resembles the α - β turn type found recently as a repetitive element in porcine ribonuclease inhibitor (14). In TSP, another seven-residue α helix, $\alpha 1$, precedes the α - β turn and forms a three-helix bundle in the trimer, burying hydrophobic residues in the interface.

Recently, smaller β helical structures have been found in alkaline protease (AP) from *Pseudomonas aeruginosa* (15) and pectate lyase (PelC) from *Erwinia chrysanthemi* (16). Both the number and length of the β -helix turns vary greatly between these proteins. Also, the core of the β helices contains calcium in AP, stacks of polar side chains in PelC, and hydrophobic residues in TSP. A common functional property of these molecules is not apparent except for the binding of carbohydrate in PelC and TSP. Both Baumann *et al.* (15) and Yoder *et al.* (16) suggest that the β helix might participate in translocation across a membrane. Because TSP is not secreted into the extracellular space as is PelC or AP, and because the mature trimer is very stable and

the protein shows a low folding efficiency, it is unlikely that the β helix is associated with simple unfolding and refolding. The dorsal fin (residues 197 to 259) is inserted into the β helix between β strands $\beta B3$ and $\beta C3$ and folded into a globular domain with mainly irregular structure; three very short α helices, $\alpha 3$, $\alpha 4$, and $\alpha 5$ of six, five, and four residues, respectively; and a strongly twisted, triple-stranded antiparallel β sheet F of topology 2-1-3 and only four regular hydrogen bonds. The dorsal fin domain is, mediated by hydrophobic residues, in close contact with six turns connecting β sheets B and C of the β helix.

In contrast to the main body, where the subunits form independently folded domains, the three polypeptide chains merge into a single common domain in the caudal fin, which is composed of three segments. In the first segment (residues 541 to 555), a wide loop winds around the molecular triad and merges with the strands of the symmetry mates. The second (residues 556 to 619) and third (residues 620 to 666) segments of the caudal fin are formed, respectively, by antiparallel β sheets D and E, composed of five and three strands, which are two to six and six residues long (Fig. 1B). The first caudal fin segments of the symmetry mates extend β sheet D toward the NH_2 -terminus by two parallel strands of four and two residues, respectively. In the trimer, the β sheets D and E pack against their symmetry mates. The strands of the slightly twisted β sheets D lie almost perpendicular to the molecular

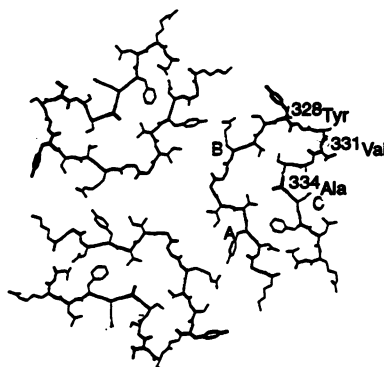


Fig. 2. Cross section of the parallel β helices (residues 317 to 339) in the trimer. Amino acids Val³³¹ and Ala³³⁴ at the global suppressor sites and Tyr³²⁸ are shown in bold (see also Figs. 3 and 4). The β sheets A and C form a β sandwich, and β sheets A and B enclose an angle of approximately 120°.

Fig. 3. Sites of lethal mutations (red) are clustered in the caudal fin, whereas tsf mutations (blue) cluster in the main body and in the dorsal fin. Sites of global suppressors of the tsf phenotype (white) are located in the central part of the β helix.



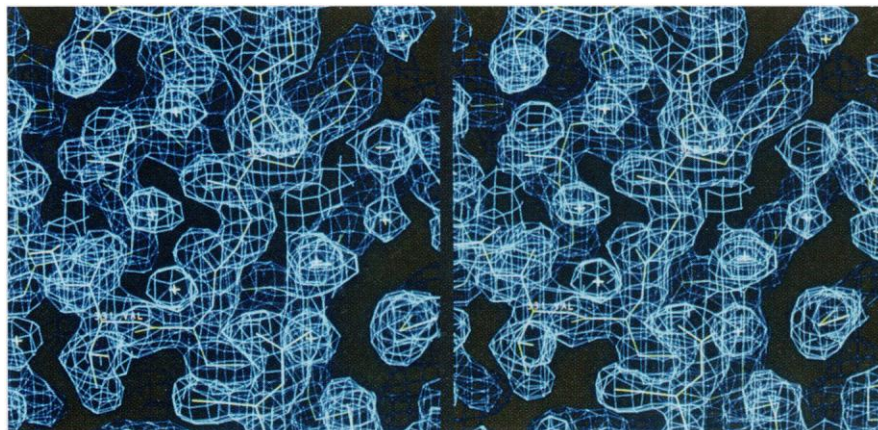


Fig. 4. Stereo view of a $2F_o - F_c$ map contoured at 1.5σ above background around Tyr³²⁸ and Val³³¹, which are outliers in the Ramachandran plot (5 non-Gly outliers in total).

triad and form a trigonal prism. The strongly twisted β sheets E are rotated by about 90° around the molecular threefold axis with respect to β sheets D and tilted toward the axis by about 30° . Each β sheet E forms two six-stranded β barrels by its three strands associated in their upper part with the preceding and in their lower part with the succeeding symmetry mates, respectively, thus forming a three-blade propeller.

Obviously, the strongest intersubunit contacts of the trimer are formed in the caudal fin by tightly interconnected β sheets such that the secondary structure involves all three subunits. The native tertiary structure cannot be adopted in the absence of the symmetry mates. This condition helps to explain why monomeric folding intermediates are only marginally stable (4, 17). Thus, the carboxyl-terminal part of the structure is likely a clue to the formation of a protimer folding intermediate with associated, but not fully folded, polypeptide chains.

Two point mutations, Gln $\rightarrow$ Tyr⁴⁸⁹ (18) and Gly $\rightarrow$ Arg⁵⁰⁵ (7), defective in endorhamnosidase activity must affect the O-antigen hydrolysis indirectly because Gln⁴⁸⁹ points into the interior of the β helix and Gly⁵⁰⁵ is located in a turn near the subunit interface. All sites where a single amino acid exchange leads to the tsf phenotype are located between residues 163 and 493 (19). A large proportion (10 out of 32) of the independent tsf sites is located in the dorsal fin domain (Fig. 3). This position indicates the importance of the site for subunit folding and suggests that it might function as a folding nucleus. Of 32 independent tsf sites, 24 are exposed to solvent, with about 15 sites in surface turns or loops. Only three of the affected residues point into the interior of the β helix, and another four tsf sites are at the subunit interface. Lethal point mutations clearly cluster in the caudal fin domain (9 out of 12), suggesting that the main body is significantly more tolerant to

destabilizing mutations than is the caudal fin.

The global suppressor sites at residues 331 and 334 (Figs. 2 and 3) are located in the sixth turn of the β helix. The Val³³¹ residue is partly solvent-exposed and located in the loop between β strands $\beta B6$ and $\beta C6$, whereas Ala³³⁴ is the first residue of β strand $\beta C6$ pointing into the interior of the β helix. Interestingly, Val³³¹ as well as Tyr³²⁸, which is the last residue of β strand $\beta B6$, are outliers in the Ramachandran plot (five non-Gly outliers in total) with $\phi = -118^\circ$, $\psi = -141^\circ$ and $\phi = 64^\circ$, $\psi = -143^\circ$, respectively, clearly defined by electron density (Fig. 4). Alanine or Gly acts as a tsf suppressor at site 331 (9, 20) and stabilizes TSP and its folding intermediates (4) probably because it releases backbone strain. At site 334, the substitution of Val or Ile for Ala might improve the stacking of hydrophobic side chains in the interior of the β helix. In summary, the unusual x-ray structure of TSP is compatible with the available genetic and biochemical data.

REFERENCES AND NOTES

- R. T. Sauer, W. Krovatin, A. R. Poteete, P. B. Berget, *Biochemistry* **21**, 5811 (1982); D. P. Goldenberg, P. B. Berget, J. King, *J. Biol. Chem.* **257**, 7864 (1982); J. V. Israel, T. F. Anderson, M. Levine, *Proc. Natl. Acad. Sci. U.S.A.* **57**, 284 (1967).
- S. Iwashita and S. Kanegasaki, *Eur. J. Biochem.* **65**, 87 (1976).
- D. P. Goldenberg and J. King, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3403 (1982); A. Fuchs, C. Seiderer, R. Seckler, *Biochemistry* **30**, 6598 (1991).
- M. Danner and R. Seckler, *Protein Sci.* **2**, 1869 (1993).
- C. A. Haase-Pettingell and J. King, *J. Biol. Chem.* **263**, 4977 (1988).
- R. Brunschier, M. Danner, R. Seckler, *ibid.* **268**, 2767 (1993); A. Mitraki, M. Danner, J. King, R. Seckler, *ibid.*, p. 20071.
- J. J. Schwarz and P. B. Berget, *ibid.* **264**, 20112 (1989).
- D. P. Goldenberg and J. King, *J. Mol. Biol.* **145**, 633 (1981); D. H. Smith and J. King, *ibid.*, p. 653.
- A. Mitraki, B. Fane, C. Haase-Pettingell, J. Sturtevant, J. King, *Science* **253**, 54 (1991); A. Mitraki and J. King, *FEBS Lett.* **307**, 20 (1992).
- M. Danner, A. Fuchs, S. Miller, R. Seckler, *Eur. J. Biochem.* **215**, 653 (1993); unpublished results. Shortened TSP was expressed in *Escherichia coli* and contains an initial Met and, in comparison with the published sequence (1), a single substitution of Gly for Ser⁵¹³, which was also present in the parent wild-type DNA. The fragment lacks 108 amino-terminal residues that are probably involved in head binding (7, 11); maintains spectral properties, enzymatic activity, and SDS resistance of intact TSP; and resembles the complete protein in its folding pathway and temperature sensitivity of folding.
- B.-L. Chen and J. King, *Biochemistry* **30**, 6260 (1991); P. A. Maurides, J. J. Schwarz, P. B. Berget, *Genetics* **125**, 673 (1990).
- For crystallization, hanging droplets were made from 5 μ l of a protein solution (10 mg/ml) in 10 mM Hepes (pH 7.0) and 3 μ l of precipitating buffer containing 1.5 M $(\text{NH}_4)_2\text{SO}_4$ in 0.1 M sodium phosphate (pH 10.0). Crystallization was achieved at 4°C by vapor diffusion against 1.0 M $(\text{NH}_4)_2\text{SO}_4$ in 0.1 M sodium phosphate (pH 10.0). Cube-shaped crystals of approximately 0.35 mm were apparent after 3 days. They were either harvested in 1.0 M $(\text{NH}_4)_2\text{SO}_4$ in 0.1 M sodium phosphate (pH 10.0) for the collection of native data sets or changed to an appropriate stabilizing solvent to facilitate the formation of heavy atom derivatives (Table 1). The cubic crystals diffracted to 2.0 $\text{\AA}$ resolution and belonged to space group $P2_13$ with lattice constant $a = 120.90 \text{ \AA}$. All x-ray measurements were done on a FAST television area detector mounted on a Rigaku rotating anode generator operated at 5.4 kW ($\lambda = \text{CuK}\alpha = 1.5418 \text{ \AA}$). The x-ray intensities were evaluated with the MADNES system and were scaled, corrected for absorption effects, and averaged with ABCOR and PROTEIN (21). Heavy atom derivatives were prepared by soaking under the conditions given in Table 1 and were analyzed by difference Patterson methods and cross-phased difference Fourier maps with PROTEIN. Heavy atom sites and the phases were refined with PROTEIN. A first 3 $\text{\AA}$ MIR map showed well-defined secondary structure elements. A model of the polypeptide chain was built with FRODO (22) on an ESV-30 Graphic system workstation (Evans & Sutherland, Salt Lake City, UT). The model was subsequently completed and refined by energy-restrained crystallographic refinement with X-PLOR, and the parameters were derived by Engh and Huber (23). After each round of refinement, a phase-combined Fourier map was calculated and the model was checked and rebuilt; the resolution was extended from 3.0 (data set native I) to 2.0 $\text{\AA}$ (data set native II) during the refinement procedure. Ordered water molecules with $B < 52 \text{ \AA}^2$ were identified in $2F_o - F_c$ maps contoured at 1.5σ . The coordinates have been submitted to the Brookhaven Protein Data Bank.
- W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983); β -sheet strands are designated as $\beta A1$, $\beta A2$, ... for sheet A; $\beta B1$, $\beta B2$, ... for sheet B; and so forth.
- B. Kobe and J. Deisenhofer, *Nature* **366**, 751 (1993).
- U. Baumann, S. Wu, K. M. Flaherty, D. B. McKay, *EMBO J.* **12**, 3357 (1993).
- M. D. Yoder, N. T. Keen, F. Jurnak, *Science* **260**, 1503 (1993).
- D. P. Goldenberg, D. H. Smith, J. King, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7060 (1983).
- P. B. Berget and A. R. Poteete, *J. Virol.* **34**, 234 (1980).
- R. Villafane and J. King, *J. Mol. Biol.* **201**, 607 (1988).
- B. Fane, R. Villafane, A. Mitraki, J. King, *J. Biol. Chem.* **266**, 11640 (1991); S. C. Lee, H. Koh, M.-H. Yu, *ibid.*, p. 23191.
- A. Messerschmidt and J. W. Pflugrath, *J. Appl. Crystallogr.* **20**, 306 (1987); A. Messerschmidt, M. Schneider, R. Huber, *ibid.* **23**, 436 (1990); W. Steigemann, in *Crystallographic Computing*, D. Moras, A. D. Podjarny, J. C. Thiery, Eds. (Oxford Univ. Press, Oxford, 1991), vol. 5, 115-125.
- T. A. Jones, *J. Appl. Crystallogr.* **11**, 268 (1978).
- A. T. Brünger, *X-PLOR, Version 3.1: A System for Crystallography and NMR* (Yale Univ. Press, New

24. The native I data set was used during the initial model building and in the first stages of refinement, and the native II data set was used for final refinement. The mean figure of merit was 0.66 (25.0 to 3.0 Å). The current model has been refined to a crystallographic *R* factor of 18.4% with the use of data from 8.0 to 2.0 Å resolution with $||\sigma| > 2.5$ (34,670 unique reflections) and comprises 4115 nonhydrogen protein atoms maintaining strict geometry with deviations from ideal values of bond lengths and angles of 0.010

romercuri-4-nitrophenol (1.5 hours), 1 M Na₂SO₄, 0.1 M K₂HPO₄-NaOH (pH 10.0); UOSO—10 mM UO₂SO₄ (2 hours), 1 M Na₂SO₄, 0.1 M NaAc-HAc (pH 5.0); PTL4—2 mM K₂PtCl₆ (6 hours), 1 M Na₂SO₄, 0.1 M K₂HPO₄-NaOH (pH 10.0); CMPT—2 mM K₂PtCl₆, 1 mM 2-chloromercuro-4-nitrophenol (3 hours), 1 M Na₂SO₄, 0.1 M K₂HPO₄-NaOH (pH 10.0).

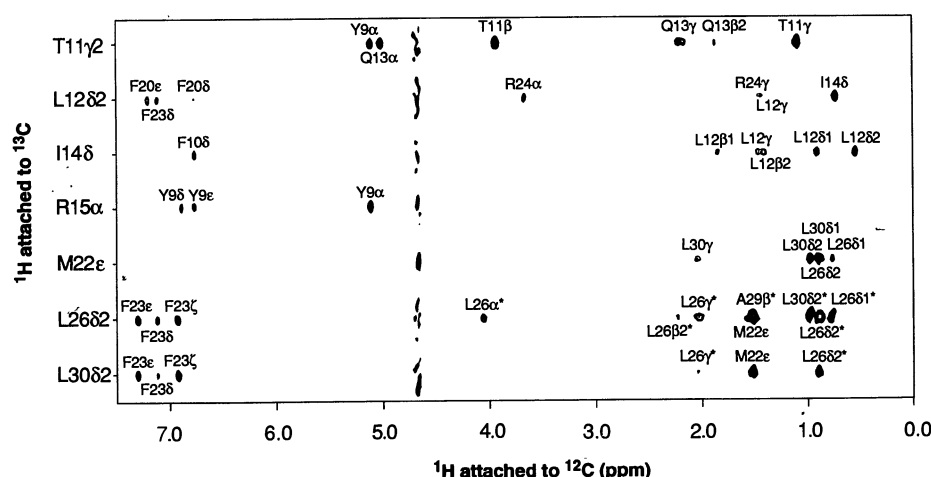
11 March 1994; accepted 26 May 1994

G. Marius Clore,* James G. Omichinski, Kazuyasu Sakaguchi,
Nicola Zambrano, Hiroshi Sakamoto, Ettore Appella,
Angela M. Gronenborn*

The p53 protein is a sequence-specific transcriptional activator that has a key function in tumor suppression (1). Inactivation of its tumor suppressor activity, either through mutation or by association with viral or cellular proteins, contributes to the development of as many as 50% of human cancers (1, 2). The p53 protein is believed to exert its tumor suppressor activity by stimulating the transcription of the p21 gene product that in turn inhibits cyclin-dependent kinase-4, thereby blocking cell division (3). The p53 protein is composed of four domains: an NH₂-terminal transactivation domain, a central DNA binding domain, an oligomerization domain, and a basic COOH-terminal nuclear localization domain (4). Although most mutations found in human cancers are located within the DNA binding domain (2), several observations suggest that the oligo-

oligomerization domain) is still capable of tumor suppression. (ii) Many mutants of p53 exert their effects through a negative-dominant mechanism whereby heterotetramers of wild-type and mutant p53 no longer bind DNA sequence specifically or bind with much reduced affinity. (iii) The transforming activity of mutants within the DNA binding domain of p53 can be abolished by making them incapable of forming oligomers as long as wild-type p53 is available from a non-mutant allele. These data suggest that inhibition of p53 oligomerization may be a useful therapeutic avenue for cancer chemotherapy. It has recently been established that a 42-amino acid peptide comprising residues 319 to 360 of p53 constitutes the minimal unit required for tetramerization (6). Here, we present the three-dimensional (3D) solution structure of the tetramerization domain (residues 319 to 360) of p53 by means of multidimensional heteronuclear NMR spectroscopy.

Because of the complexity of the system and the necessity to distinguish intra- from intersubunit nuclear Overhauser effects (NOEs), the structure determination was carried out on samples of uniformly labeled



G. M. Clore, J. G. Omichinski, A. M. Gronenborn, Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

K. Sakaguchi, N. Zambrano, H. Sakamoto, E. Appella, Laboratory of Cell Biology, Building 37, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA.

386