Acta Cryst. (1996). D52, 1153–1160

Molecular Structure of a Proteolytic Fragment of TLP20

HAZEL M. HOLDEN,[a] GARY WESENBERG,[a] DEBORAH A. RAYNES,[a] DAVID J. HARTSHORNE,[b] VINCE GUERRIERO JR [b] AND IVAN RAYMENT[a]*

[a]*Institute for Enzyme Research, Graduate School and Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53705, USA, and* [b]*Department of Animal Sciences, University of Arizona, Tucson, Arizona 85721, USA. E-mail: ivan@enzyme.wise.edu*

(Received 29 March 1996; accepted 18 June 1996)

Abstract

Myosin light-chain kinase is responsible for the phosphorylation of myosin in smooth muscle cells. In some tissue types, the C-terminal portion of this large enzyme is expressed as an independent protein and has been given the name telokin. Recently, an antibody directed against telokin was found to interact with a protein derived from the baculovirus *Autographa californica* nuclear polyhedrosis virus. This protein was biochemically characterized and given the name TLP20 for telokin-like protein of 20000 molecular weight. The amino-acid sequence of TLP20 was determined on the basis of a cDNA clone and subsequent alignment searches failed to reveal any homology to telokin or to other known proteins. The three-dimensional structure of a proteolytic portion of TLP20 is reported here. Crystals employed in the investigation were grown from ammonium sulfate solutions at pH 6.0 and belonged to the space group $P2_13$ with unit-cell dimensions of $a = b = c = 76.3$ Å and one molecule per asymmetric unit. The structure was determined by multiple isomorphous replacement with three heavy-atom derivatives. Least-squares refinement of the model reduced the crystallographic R factor to 18.1% for all measured X-ray data from 30.0 to 2.2 Å. The overall fold of the molecule may be described as a seven-stranded antiparallel β-barrel flanked on the bottom by two additional β-strands and on the top by an α-helix. Quite surprisingly, the three-dimensional structure of this β-barrel is not similar to telokin or to any other known protein.

1. Introduction

Myosin light-chain kinase, hereafter referred to MLCK, catalyzes the phosphorylation of myosin. It is believed that this phosphorylation reaction serves as a major regulatory mechanism in smooth muscle (Hartshorne, 1987). MLCK is widely distributed in smooth muscle and has also been identified in striated muscle and several non-muscle cells (Hartshorne, 1987). As isolated from chicken gizzard, MLCK is composed of 972 amino-acid residues (Olson *et al.*, 1990). In some smooth muscle cells the C-terminal domain of MLCK is also expressed as an independent protein containing 157 amino acids and referred to as telokin (Ito, Dabrowska, Guerriero & Hartshorne, 1989).

At present, the function of telokin is unknown. Interestingly, telokin does show amino-acid sequence similarity to several quite different muscle proteins including titin, C-protein and twitchin. The three-dimensional structure of telokin was determined to 2.8 Å resolution by X-ray crystallographic analysis and shown to consist of seven strands of antiparallel β-pleated sheet wrapping around to form a barrel (Holden, Ito, Hartshorne & Rayment, 1992). This type of β-barrel, as seen in telokin, was first observed in the λ-type Bence-Jones protein and subsequently in other immunoglobulin constant domains. (Schiffer, Girling, Ely & Edmundson, 1973).

Recently, an antibody directed against smooth muscle telokin was found to bind to a protein in cell extracts from *Autographa californica* nuclear polyhedrosis virus (AcMNPV)-infected Sf9 cells and from *Manduca sexta* larvae that had been infected with AcMNPV (Raynes, Hartshorne & Guerriero, 1994). This baculovirus protein was named TLP20 for telokin-like protein of molecular weight 20000. Subsequently, a cDNA for TLP20 was isolated, sequenced and shown to code for a polypeptide chain of 180 amino-acid residues (Raynes *et al.*, 1994). The protein lies in a region of the viral genome that suggests expression late in the virus replication cycle. Quite surprisingly, both the nucleotide and the amino-acid sequences failed to reveal homologies with telokin or with other known proteins (Devereux, Haeberli & Smithies, 1984). In an effort to further characterize this protein from baculovirus-infected cells, a proteolytic fragment of TLP20 containing Gly7 to Arg108 was crystallized. Here we describe the three-dimensional structure of the N-terminal region of TLP20 determined and refined to 2.2 Å resolution. Contrary to expectations, TLP and telokin, while both composed primarily of β-sheet, are topologically very different. As such the antigenic similarities shared between these two proteins most likely reside in the C-terminal portion of TLP20 beyond Arg108. Structural homology searches have failed to reveal significant

three-dimensional relationships between TLP20 and any other protein.

2. Experimental procedures

2.1. *Crystallization and search for heavy-atom derivatives*

Intact TLP20, containing 180 amino-acid residues, was purified according to previously published procedures (Raynes *et al.*, 1994). For crystallization trials, the protein sample was concentrated to 10 mg ml^{-1} and contained 5 m*M* HEPES (pH 7.0) and 100 m*M* NaCl. Adventitious proteolysis upon storage cleaved TLP20 to a fragment containing the first N-terminal 108 amino-acid residues. Small crystals of this fragment were first observed growing in hanging drops equilibrated against 2 *M* ammonium sulfate, 50 m*M* K^+/Na^+ succinate (pH 6.0), and 0.5 m*M* NaN_3. Larger crystals were obtained by the technique of macroseeding (Thaller *et al.*, 1981, 1985). For such experiments, 8 µl of protein were mixed with 8 µl of 1.6–1.8 *M* ammonium sulfate, 50 K^+/Na^+ succinate (pH 6.0), 0.5 m*M* NaN_3 and equilibrated against 1.6–1.8 *M* ammonium sulfate for several days at room temperature. Small seed crystals were subsequently introduced to these hanging drops and grew to typical dimensions of 0.8 × 0.8 × 0.5 mm in approximately 1 week.

With a Siemens X1000D area-detector system and the software package *SADIE*, the crystals were shown to belong to the space group $P2_13$. The unit-cell dimensions were $a = b = c = 76.3$ Å, and there was one molecule per asymmetric unit. The solvent content, based on a V_m of 3.1 Å^3 Da^{-1}, was 60% (Matthews, 1968).

For heavy-atom derivative searches, crystals of TLP20 were transferred to a synthetic mother liquor containing 2.5 *M* ammonium sulfate, 50 m*M* K^+/Na^+ succinate (pH 6.0) and various heavy-metal reagents. Three isomorphous heavy-atom derivatives were prepared by soaking crystals in either 3 m*M* triethyllead acetate, 3 m*M* uranyl acetate, or 0.2 m*M* mercuric acetate.

2.2. *X-ray data collection and processing*

Three-dimensional X-ray data sets for the native and heavy-atom derivative crystals were collected at 277 K using a Siemens X1000D area-detector system. The X-ray source was Cu $K\alpha$ radiation from a Rigaku RU200 rotating-anode generator operated at 50 kV and 50 mA and equipped with a 200 µm focal cup. Only one crystal was required for each data set. The X-ray data were processed with the *XDS* data-reduction package (Kabsch, 1988*a*,*b*), and internally scaled according to the algorithm of Fox & Holmes (Fox & Holmes, 1966) as implemented by Dr Phil Evans. All heavy-atom derivative data sets were collected to 2.2 Å resolution, although only data to 2.3 Å was used in the subsequent phase calculations. The native data set was 99.8% complete to 2.2 Å resolution. Relevant X-ray data collection statistics may be found in Table 1.

2.3. *Structure determination and least-squares refinement*

The positions of the heavy-atom derivatives were determined by inspection of appropriate difference Patterson maps calculated with X-ray data from 30 to 5.0 Å and placed on a common origin by difference Fourier maps. All positions and relative occupancies for the heavy atoms were refined according to the origin-removed Patterson function correlation method with the program *HEAVY* (Rossmann, 1960; Terwilliger & Eisenberg, 1983). The positions of the heavy-atom

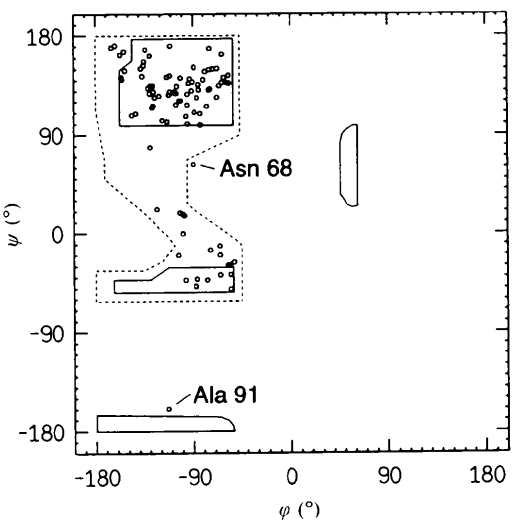

Fig. 1. A Ramachandran plot of all non-glycinyl main-chain dihedral angles. Fully and partially allowed dihedral angles are enclosed by the solid and dashed lines, respectively.

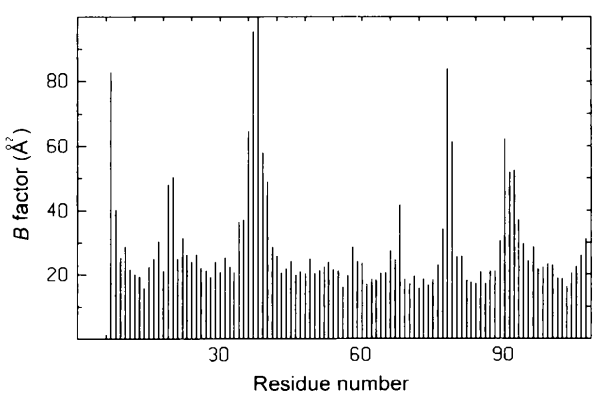

Fig. 2. Plot of the mean *B* value *versus* amino-acid residue for all main-chain atoms.

Table 1. *Intensity statistics for native and heavy-atom derivative data*

		Resolution range (Å)						
	Overall	50.00–4.40	3.49	3.05	2.77	2.57	2.42	2.30	2.20
Native									
No. of measurements	36733	6678	6123	5242	4847	4420	3897	3123	2394
No. of independent reflections	7732	1038	970	960	979	960	961	937	927
Average intensity	4356	10000	8064	4093	2061	1116	949	720	637
Average σ	406	400	457	416	393	380	385	390	397
R factor* (%)	4.8	2.4	3.2	5.4	9.6	16.3	18.3	23.9	24.2
Triethyllead acetate									
No. of measurements	22581	4124	3747	3223	2988	2706	2383	1938	1472
No. of independent reflections	7599	1034	969	958	978	955	954	914	837
Average intensity	3839	10000	6909	2975	1484	856	664	515	436
Average σ	407	422	449	405	395	383	384	386	394
R factor (%)	6.1	3.4	4.5	7.6	13.7	21.3	26.1	31.5	33.9
$UO_2(OCOCH_3)_2$									
No. of measurements	23760	4326	3978	3416	3183	2869	2509	2024	1455
No. of independent reflections	7624	1027	970	958	978	960	959	921	851
Average intensity	3938	10000	6888	3388	1586	871	720	601	534
Average σ	449	464	511	466	433	415	411	410	410
R factor (%)	7.2	4.9	5.6	8.1	13.8	21.3	23.9	28.0	28.9
Mercuric acetate									
No. of measurements	21241	3925	3535	3057	2812	2527	2227	1816	1342
No. of independent reflections	7459	1031	970	957	978	956	931	870	766
Average intensity	4247	10000	7334	3867	2059	1201	957	758	654
Average σ	436	430	485	437	421	415	418	425	436
R factor (%)	6.1	3.9	4.8	6.7	10.5	15.6	18.9	24.3	25.7

* R factor $= (\sum |I - \bar{I}| / \sum I) \times 100$.

sites are listed in Table 2. Anomalous difference Fourier maps calculated from 30 to 5 Å were employed for determining the correct hand of the heavy-atom constellation. Protein phases were calculated with *HEAVY* and relevant phase statistics are given in Table 3. The phasing included the anomalous scattering information from all three heavy-atom derivatives.

An electron-density map, using centroid protein phases based on the three heavy-atom derivatives and a structure-factor weighting scheme based on the figure of merit (Blow & Crick, 1959), was calculated with X-ray data from 30.0 to 2.3 Å. This map was plotted onto transparencies and then stacked on thin Plexiglas sheets. The course of the polypeptide chain was

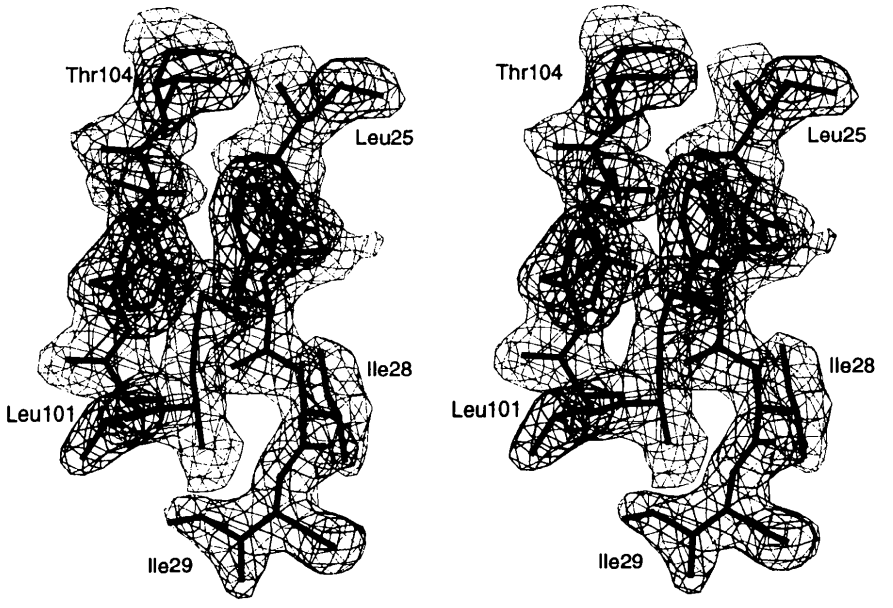

Fig. 3. Representative electron density. The electron density shown was calculated to 2.2 Å resolution with coefficients of the form $(2F_o - F_c)$ where F_o and F_c were the native and calculated structure-factor amplitudes, respectively. The map was contoured at 1σ.

immediately obvious from visual inspection of this map. A protein model was built into the electron density with the program *FRODO* as implemented on an Evans and Sutherland computer graphics system (Jones, 1985). Small peaks of electron density were considered to be ordered solvent molecules if they were positioned within approximately 3.2 Å of potential hydrogen-bonding groups. A total of 40 water molecules were included in the final coordinate file. The average B value for the solvent was 48.6 Å^2. The solvent continuum feature of *TNT* was used to further improve the fit of the model to the low-resolution X-ray data. Several cycles of least-squares refinement with the program package *TNT* (Tronrud, Ten Eyck & Matthews, 1987) and manual model building reduced the crystallographic R factor to 18.1% at 2.2 Å resolution. Relevant refinement statistics may be found in Table 4.

A Ramachandran plot of all main-chain dihedral angles is given in Fig. 1. Only two amino-acid residues, Asn68 and Ala91, have dihedral angles that fall slightly outside of the allowed regions. A temperature-factor plot for the polypeptide chain backbone is shown in Fig. 2. The first six amino-acid residues at the N terminus were disordered and not included in the protein model. Weak electron density was also evident for Gly7 and Arg108. Both the surface loop defined by Lys36 to Val40 and the region delineated by Asn78 to Gly79 were weak in electron density. Excluding these residues from the calculation, the average B value for the polypeptide chain backbone was 24.9 Å^2. A representative portion of the electron-density map is displayed in Fig. 3. The only residues that had disordered side chains were Lys36, Lys37, Arg38, Lys46, Arg54, Asn78, Glu92, Lys94 and Arg108. Electron density for the polypeptide chain ends at Arg108. To prove that the remaining C-terminal residues of TLP20 had been proteolytically cleaved rather than merely disordered in the electron-density map, crystals were dissolved and the protein sample subjected to sodium dodecyl sulfate gel electrophoresis. The results were conclusive that the protein sample crystallized was not that of intact TLP20 but rather a proteolytic portion. Subsequent crystallization trials of intact TLP20 were unsuccessful.

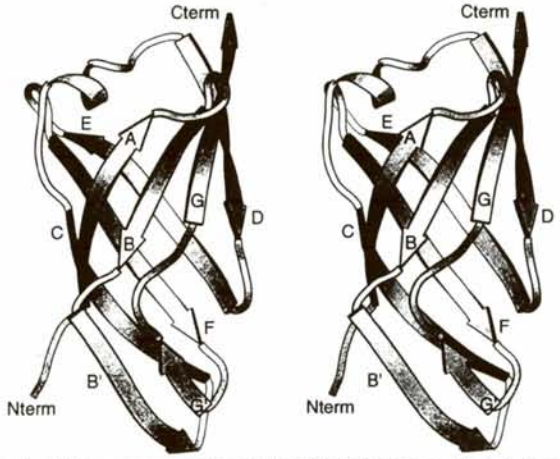

Fig. 4. Ribbon representation of the TLP20 fragment. Figs. 4–6 and 8 were prepared with the program *MOLSCRIPT* (Kraulis, 1991).

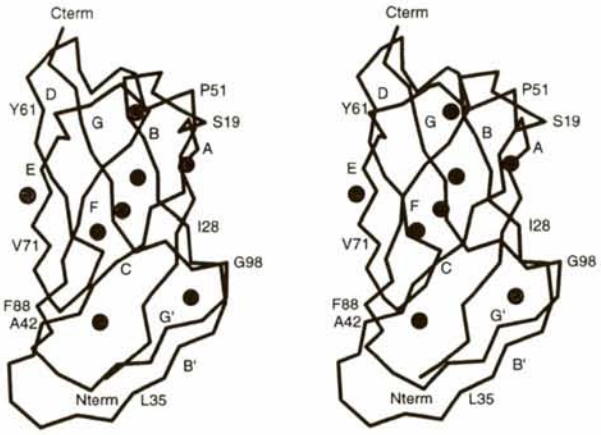

Fig. 5. Cα trace of the TLP20 fragment. The large spheres indicate the positions of the ordered water molecules that play important structural roles.

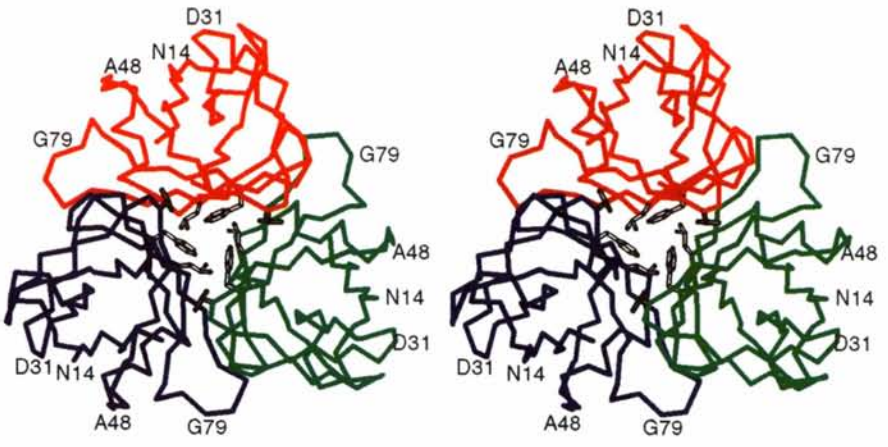

Fig. 6. Cα trace of the TLP20 fragment trimer. Residues Val40, Leu72, and Phe88 are depicted in a ball-and-stick representation.

Table 2. *Refined heavy-atom parameters*

Derivative	Site No.	Relative occupancy	x	y	z	B (Å^2)	R_{iso}* (%)
Triethyllead acetate	1	6.44	0.506	0.714	0.177	29.1	20.9
$UO_2(OCOCH_3)_2$	1	6.55	0.966	0.080	0.613	29.7	22.7
Mercuric acetate	2	8.69	0.304	0.869	0.038	26.1	29.3
		2.96	0.505	0.708	0.181	30.3	

* $R_{iso} = (\sum ||F_N| - |F_H||/\sum |F_N|) \times 100$, where F_N is the native structure-factor amplitude and F_H is the heavy-atom derivative structure-factor amplitude. x, y and z are the fractional atomic coordinates.

Table 3. *Phase calculation statistics*

	Resolution range (Å)							
	∞–8.33	5.24	4.09	3.47	3.06	2.77	2.55	2.37
No. of reflections	345	571	721	849	929	1026	1070	1156
Figure of merit	0.84	0.86	0.79	0.74	0.70	0.63	0.57	0.22
Phasing power*								
Triethyllead acetate								
Centric reflections	2.17	1.40	1.20	1.16	1.11	1.20	0.91	0.91
Acentric reflections	2.08	1.71	1.48	1.38	1.39	1.20	1.08	0.83
$UO_2(OCOCH_3)_2$								
Centric reflections	1.54	1.65	1.16	1.22	1.04	1.11	0.77	0.67
Acentric reflections	2.06	2.03	1.42	1.52	1.40	1.23	1.11	0.92
Mercuric acetate								
Centric reflections	1.31	1.40	1.37	1.03	1.22	1.15	1.26	1.03
Acentric reflections	1.86	1.84	1.46	1.31	1.48	1.44	1.44	1.18

* Phasing power is the ratio of the root-mean-square heavy-atom scattering-factor amplitude to the root-mean-square lack-of-closure error.

3. Results and discussion

A ribbon representation of TLP is shown in Fig. 4, and a list of the secondary-structural elements is given in Table 5. Nearly 79% of the polypeptide chain folds into classical secondary-structural elements. As can be seen, the TLP20 fragment is elongated with overall dimensions of approximately 20 × 20 × 43 Å. The structural architecture of the TLP20 fragment is dominated by an antiparallel β-barrel of seven strands. There are two additional β-strands at the base of the barrel and one short α-helix at the top. Four classical reverse turns are observed in the protein: three type I and one type II. With the exception of Thr104, the barrel is composed completely of hydrophobic amino-acid residues. The β-strands range in length from four to eight amino-acid residues. β-strands *A* and *B* are connected by two type I turns. The three β-regions, *B*, *B'* and *C*, form nearly one long β-strand except for the interruptions caused by Glu30 and Gly39 with dihedral angles of $\varphi = -77.0$, $\Psi = -42.3$ and $\varphi = 95.2$ and $\Psi = 176.8$, respectively. β-strands *C* and *D* are connected *via* the one small α-helix which terminates in a type I turn. The only type II turn in the TLP20 fragment serves to connect β-strands *G'* and *G*.

The β-barrel can be envisioned almost as two layers of sheets. One sheet contains β-strands *A*, *B*, *D* and *G* while the other is composed of β-strands *C*, *E* and *F*. Of the 40 water molecules located in the electron-density map, eight play important structural roles in the formation of these β-sheets. The positions of these solvents are shown in Fig. 5. They are all located on one side of the protein. Three of these water molecules serve to bridge β-strands *C* and *F* together by forming hydrogen bonds to backbone carbonyl O atoms and amide N atoms. As can be seen, two of the water molecules lie at the beginning and the end of the β-sheet region formed by β-strands *A* and *C*. Another water links β-strands *A* and *B'* together by hydrogen bonding to N of Ile12, N of Asp31 and O of Asp31. One solvent lies within hydrogen-bonding distance to O_γ of Thr63, O of Val71, and N of Cys73 thereby functioning as a link between β-strands *D* and *E*. The polypeptide chain between β-strands *E* and *F* folds into a non-classical reverse turn. This region is stabilized by a water molecule that hydrogen bonds to O of Asn78, N of Tyr81, and $O_{\delta 1}$ of Asp82.

The quaternary structure of TLP20 in solution is unknown. Examination of the crystalline lattice shows that the TLP20 fragment packs as a trimer along a crystallographic threefold rotation axis as shown in Fig. 6. The subunit:subunit interface is formed primarily by β-strands *C*, *D*, *E* and *F* from each subunit and is decidedly hydrophobic with no charged side chains. There is a hydrophobic patch formed by Val40, Leu72 and Phe88 as displayed in Fig. 6. The surface area lost upon trimer formation is approximately 1269 Å^2 as calculated according to the method of Lee and Richards with a probe sphere of 1.4 Å (Lee & Richards, 1971). This is in the same range as the buried

Table 4. *Refinement statistics*

Resolution limits (Å)	30.0–2.2
R factor (%)*	18.1
No. of reflections used	7727
No. of protein atoms	811
No. of solvent atoms	40
Weighted root-mean-square deviations from ideality	
Bond length (Å)	0.014
Bond angle (°)	1.93
Planarity (trigonal) (Å)	0.004
Planarity (other planes) (Å)	0.013
Torsional angle (°)†	18.7

* R factor $= \sum |F_o - F_c| / \sum |F_o|$ where F_o is the observed structure-factor amplitude and F_c is the calculated structure-factor amplitude.
† The torsional angles were not restrained during the refinement.

Table 5. *List of secondary-structural elements*

Amino-acid residues	Type of structure
Ile11–Ile17	β-sheet (*A*)
Asn18–Asp21	Type I turn
Asp21–Val24	Type I turn
Leu25–Ile29	β-sheet (*B*)
Asp31–Arg38	β-sheet (*B'*)
Val40–Val47	β-sheet (*C*)
Pro51–Arg54	α-helix
Leu55–Lys58	Type I turn
Asn59–Ser65	β-sheet (*D*)
Tyr69–Val76	β-sheet (*E*)
Asp82–Asn89	β-sheet (*F*)
Glu92–Leu95	β-sheet (*G'*)
Asn96–Gln99	Type II turn
Phe102–Arg108	β-sheet (*G*)

surface area of other oligomeric proteins of comparable size (Miller, Lesk, Janin & Chothia, 1987). Thus, on the basis of the observed packing it can be speculated that TLP20 exists as a trimer under some physiological conditions.

One of the purposes for pursuing the three-dimensional structure of TLP20 was to shed light on its biological function in the baculovirus. In light of the fact that it was first discovered with an antibody directed against telokin, it was expected to have a β-barrel of similar overall structure. A comparison of the topological diagrams and ribbon representations for the TLP20 fragment and telokin are shown in Figs. 7 and 8, respectively. Although the structures of both proteins are best described as being β-barrels formed from two layers of sheet it is immediately clear that topology of these proteins is quite different. Whereas telokin exhibits a classic Greek key topology (Richardson, 1981) and belongs to the immunoglobulin family (Bork, Holm & Sander, 1994), TLP20 exhibits the topology of a jelly roll (Richardson, 1981). The latter topology is common among viral coat proteins and given that TLP20 is a viral protein it is appropriate to examine if it belongs to the same family. Figs. 7 and 8 also compare the topology and ribbon representations of southern bean mosaic virus (SBMV) (Abad-Zapatero *et al.*, 1980) with those of TPL20. Although both proteins exhibit the jelly roll topology the location of the strands in the two layers of sheet are quite different. SBMV coat protein contains eight strands of antiparallel sheet whereas TLP20 contains seven. Even after consideration of the additional strand of sheet present in SBMV coat protein the order of the strands in the sheet are different. In particular the threading of the strands through the layers of sheets is inverted. This can be seen by examining the orientation of the strands adjacent to the ends of each layer as shown by the dashed lines in the center of the topology diagrams

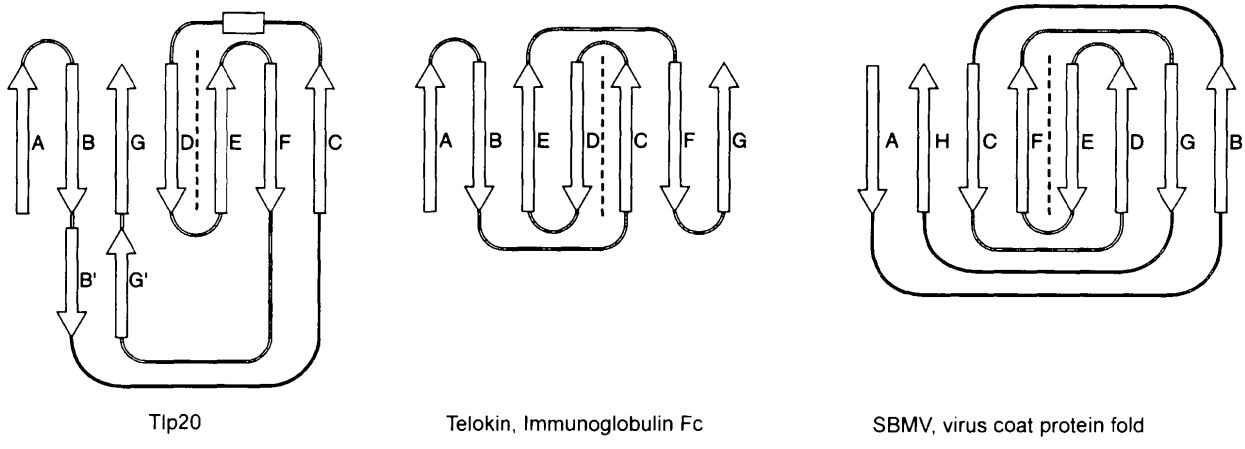

Fig. 7. Topological diagrams. Shown are the topological schemes for (*a*) the TLP20 fragment, (*b*) telokin, and (*c*) a portion of the coat protein from southern bean mosaic virus. The designation of the strands for SBMV coat protein shown here is different from that normally used to describe this fold. In one subunit of the three quasi-equivalent copies of this protein that form the icosahedral asymmetric unit there is an additional β-strand at the N terminus that is usually designated as *A*. This additional a strand is not shown here. Rather the first strand in the β-barrel is given the designation *A* to provide easier comparison of the folding patterns and emphasize the differing topologies of these proteins. The vertical dashed lines indicate the boundaries between the two layers of sheet observed in these proteins.

shown Fig. 7. The difference in topology cannot be resolved by simply moving one strand from one layer to the next.

In addition to the topological differences between TLP20 and telokin and SBMV coat protein there is considerable difference in the disposition of the strands relative to the barrel axis. The strands of TLP20 are inclined at a significantly greater angle than those in either telokin or SBMV. The inclination observed in TLP20 is more closely related to the arrangement of strands seen in the chaperonin protein GroES and ribosomal protein L14 (Davies, White & Ramakrishnan, 1996; Hunt, Weaver, Landry, Gierasch & Deisenhofer, 1996) that both contain a five-stranded β-barrel. However again the topology of the sheets in GroES and ribosomal protein L14 differ from that observed in TLP20.

Thus, it is clear that the globular domains of telokin and TLP20 differ considerably in their tertiary and primary structures which would suggest that their antigenic similarly does not reside in these domains. Preliminary investigation of the antibody-binding properties of the proteolytic fragments of TLP20 shows that the antigenic similarly between these proteins resides in a part of the molecule that is missing from the present structure (Raynes *et al.*, 1996).

In recent years it has become increasingly apparent that certain structural motifs are frequently associated with specific biological functions. Often these motifs provide clues as to the biochemical function of the molecule. For example, the presence of a homeodomain would indicate that an unknown protein is a transcription factor since this type of architecture is responsible for some types of DNA binding in eukaryotic systems. Likewise, a Rossmann fold in an unknown protein would suggest that the molecule binds a dinucleotide. It appears that the β-barrel of the TLP20 fragment is unique among the barrels observed thus far in protein structures. A search with the program *DEJAVU*, written by G. J. Kleywegt (University of Uppsala, Uppsala, Sweden), failed to reveal any significant homology of the TLP20 fragment with other known protein structures. Consequently, the three-dimensional structure of the TLP20 fragment, at present, provides virtually no clues as to its biological function. It is anticipated that as additional three-dimensional structures of proteins with known function are solved, however, this type of β-barrel will be observed again. Further studies to

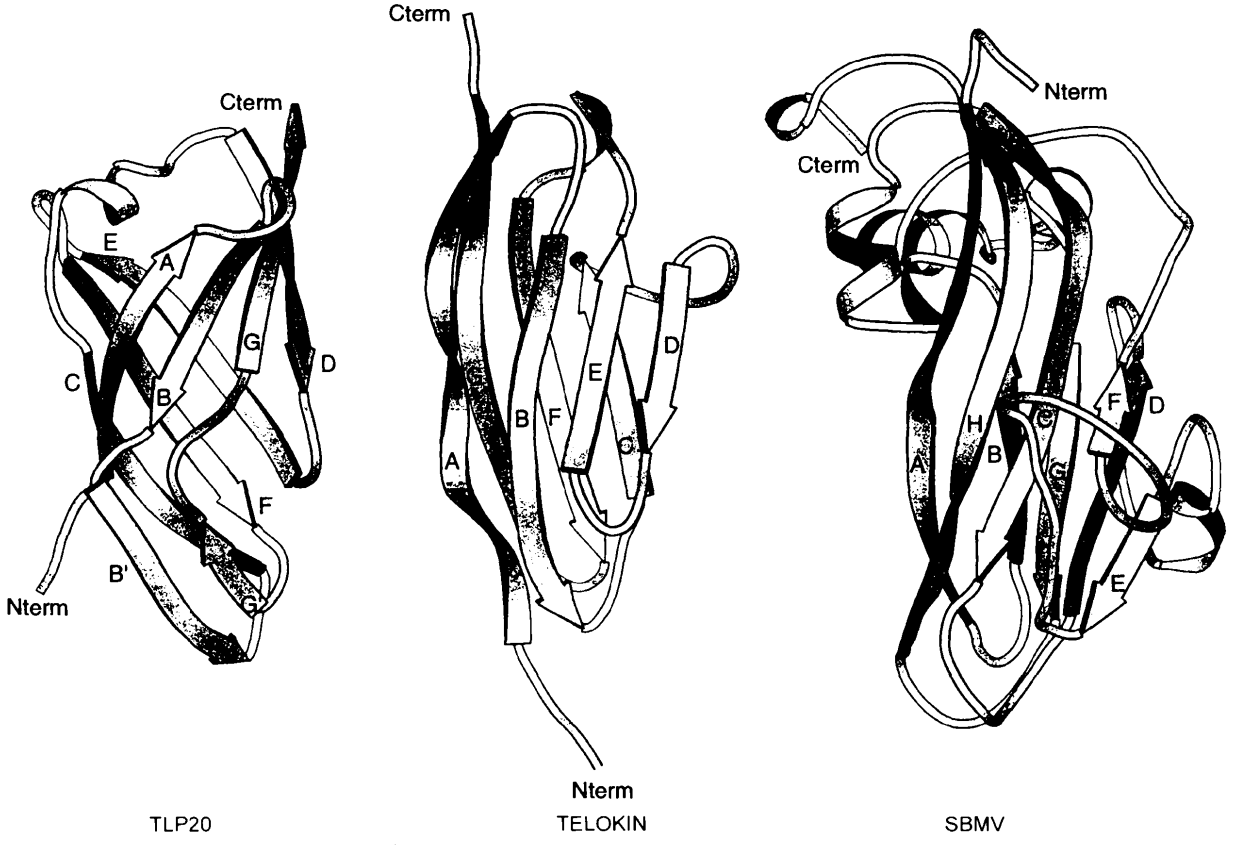

Fig. 8. Ribbon drawings of representative β-barrels. Those β-barrel motifs observed in (*a*) the TLP20 fragment, (*b*) telokin, and (*c*) southern bean mosaic virus are shown in the same orientation. As noted in Fig. 7 the first strand in the β-barrel is designated as strand *A* to facilitate comparison of these folding motifs.

address the biochemical function of TLP20 are in progress.*

We would like to thank Kelsey M. Rayment for sleeping through this investigation. This research was supported in part by a grant from the NIH (AR35186 to IR, HL43651 to VG and HL23615 to DJH).

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory. Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AM0046).

References

Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G. W., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). *Nature (London)*, **286**, 33–39.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

Bork, P., Holm, L. & Sander, C. (1994). *J. Mol. Biol.* **242**, 309–320.

Davies, C., White, S. W. & Ramakrishnan, V. (1996). *Structure*, **4**, 55–66.

Devereux, J., Haeberli, P. & Smithies, O. (1984). *Nucleic Acids Res.* **12**, 387–395.

Fox, G. C. & Holmes, K. C. (1966). *Acta Cryst.* **20**, 886–891.

Hartshorne, D. J. (1987). *Physiology of the Gastrointestinal Tract*, edited by L. R. Johnson, pp. 423–482. New York: Raven Press.

Holden, H. M., Ito, M., Hartshorne, D. J. & Rayment, I. (1992). *J. Mol. Biol.* **227**, 840–851.

Hunt, J. F., Weaver, A. J., Landry, S., Gierasch, L. & Deisenhofer, J. (1996). *Nature (London)*, **379**, 37–45.

Ito, M., Dabrowska, R., Guerriero, V. J. & Hartshorne, D. J. (1989). *J. Biol. Chem.* **264**, 13971–13974.

Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

Kabsch, W. (1988*a*). *J. Appl. Cryst.* **21**, 67–71.

Kabsch, W. (1988*b*). *J. Appl. Cryst.* **21**, 916–924.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Miller, S., Lesk, A. M., Janin, J. & Chothia, C. (1987). *Nature (London)*, **328**, 834–836.

Olson, N. J., Pearson, R. B., Needleman, D. S., Hurwitz, M. Y., Kemp, B. E. & Means, A. R. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 2284–2288.

Raynes, D. A., Hartshorne, D. J. & Guerriero, V. J. (1994). *J. Gen. Virol.* **75**, 1807–1809.

Raynes, D. A., Hartshorne, D. J. & Guerriero, V. J. (1996). Personal communication.

Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–329.

Rossmann, M. G. (1960). *Acta Cryst.* **13**, 221–226.

Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973). *Biochemistry*, **12**, 4620–4631.

Terwilliger, T. C. & Eisenberg, D. (1983). *Acta Cryst.* A**39**, 813–817.

Thaller, C., Eichele, G., Weaver, L. H., Wilson, E., Karlsson, R. & Jansonius, J. N. (1985). *Methods Enzymol.* **114**, 132–135.

Thaller, C., Weaver, L. H., Eichele, G., Wilson, E., Karlsson, R. & Jansonius, J. N. (1981). *J. Mol. Biol.* **147**, 465–469.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.