Crystal structure of vitelline membrane outer layer protein I (VMO-I): a folding motif with homologous Greek key structures related by an internal three-fold symmetry

Toshiyuki Shimizu¹, Dmitry G.Vassilyev,
Shoko Kido², Yukio Doi² and
Kosuke Morikawa³

Protein Engineering Research Institute, 6-2-3 Furuedai, Suita,
Osaka 565 and ²Department of Food Science, Kyoto Women's
University, Higashiyamaku, Kyoto 605, Japan

¹Present address: Kirin Brewery Co. Ltd, Central Laboratories for Key
Technology, 1-13-5 Fukuura, Kanazawa-ku, Yokohama, Kanazawa
236, Japan

³Corresponding author

Communicated by C.H.Chothia

The crystal structure of vitelline membrane outer layer protein I (VMO-I), which is isolated from the vitelline membrane outer layer of hen's eggs, has been determined by the multiple isomorphous replacement method and refined to an R-factor of 18.8% at 2.2 Å resolution. The main chain folds into an unusual structure that consists of three β -sheets forming Greek key motifs, which are related by an internal pseudo three-fold symmetry. The internal portion surrounded by these three β -sheets is filled with hydrophobic side chains. This conformational feature coincides with three internal repeats in the sequence. Although a similar fold exists in the second domain of δ -endotoxin, there are significant structural differences between the two proteins, with the three-fold symmetry being most regular in VMO-I.

Key words: Greek key motif/hen's egg/vitelline membrane outer layer protein I/X-ray structure

Introduction

The vitelline membrane of a hen's egg is composed of two layers that are distinct in terms of their morphology and chemical composition (Bellairs *et al.*, 1963; Jensen, 1969). The two layers are synthesized in different organs; the inner layer is formed in the ovary before ovulation, whereas the outer layer is formed in the upper oviduct after ovulation (Bain and Hall, 1969). The fine fibrils of the outer layer are mainly constructed from huge macromolecules, ovomucin, to which soluble proteins are tightly bound, e.g. lysozyme, and vitelline membrane outer layer proteins I (VMO-I) (Back *et al.*, 1982; Kido and Doi, 1988) and II (VMO-II) (Kido *et al.*, 1992). In contrast, the inner layer is composed of several kinds of glycoproteins, GP-I, GP-II, GP-III and GP-IV (Kido *et al.*, 1975, 1976, 1977; Kido and Doi, 1988). Soluble proteins bound to the outer layer can be dissociated from ovomucin by a high salt buffer (Back *et al.*, 1982; Kido *et al.*, 1992) and are not involved in the maintenance of the physical strength of the vitelline membrane (Kido and Doi, 1988). The primary structure of VMO-I has been determined and revealed that VMO-I is a basic protein composed of 163 amino acids, with a molecular mass of 17 979 daltons (Y.Doï *et al.*, in

preparation). The cDNA encoding VMO-I has been cloned and sequenced (A.Uyeda *et al.*, in preparation). The VMO-I-encoding mRNA is expressed in a restricted region of hen oviducts, the area joining the infundibulum to the magnum. In contrast, no mRNA was observed in the ovary, the upper part of the infundibulum and the lower part of the magnum. Sequence analyses show that VMO-I is a β -protein composed of three internal repeats of ~ 53 residues. The circular dichroism (CD) spectrum of VMO-I is more characteristic of a random coil than a globular protein (Y.Doï *et al.*, in preparation). Recently it has been found that VMO-I has the ability to synthesize *N*-acetylchitooligosaccharides ($n = 14$ or 15) from hexasaccharides of *N*-acetylglucosamine, which is similar to the transferase activity of lysozyme without the hydrolysis activity (S.Kido, unpublished data), although the true physiological function of VMO-I remains unknown. In order to obtain more refined insight into the three-dimensional (3D) structure of this protein, we have determined the 3D structure at an atomic resolution by X-ray crystallography.

Results

Crystallography

The structure of VMO-I was initially determined at 3 Å resolution, using the multiple isomorphous replacement (MIR) method augmented with anomalous scattering information, a solvent flattening procedure and molecular averaging. Molecular averaging was particularly effective in improving the phases and enabled us to trace the main chain in the electron density map, although the redundancy of the molecule was only twofold. The functions for generating main chain atoms from $C\alpha$ positions and for producing side chain rotamers, available in the program O (Jones and Thirup, 1986), accelerated the model building. This initial model was subjected to cyclic crystallographic refinement using X-PLOR (Brünger *et al.*, 1987). The current model includes two polypeptide chains designated as A and B, and 144 water molecules. Experimental details and statistical information on the structure determinations and refinements are presented in Materials and methods. In molecule A, regions 41–48 and 92–101, which lie within two different loops, are poorly defined in the electron density map even at 2.2 Å resolution. On the other hand, in molecule B, the corresponding residues 92–101 are poorly defined, whereas residues 41–48 are well defined, probably due to stabilization by intermolecular interactions in the crystal. The average B-factors for all main chain atoms in each molecule are 29.4 Å² and 29.0 Å², respectively. These relatively high values are mainly due to poorly defined loop regions. The B-factors decrease to 23.1 Å² and 23.7 Å² when these regions are excluded. The r.m.s. deviation between the main chain atoms and the $C\beta$ positions in the two monomers is 0.41 Å, excluding the N-terminal segment (1–5) and regions 41–48 and 92–101 in loops of which structures are not well defined.

Fig. 1. (a) Ribbon representation of VMO-I drawn with the program MOLSCRIPT (Kraulis, 1991). One of the Greek key structures is colored red, one green and the third blue. (b) Topology diagram of VMO-I. The coloring of the β -strands is the same as that in panel (a). Thick lines represent the disulfide bonds. The numbers of three β -sheets are defined so as to correspond to those of δ -endotoxin domain II. (c) The sequence of VMO-I aligned by the three internal repeats. Identical amino acids among the three internal repeats are enclosed by solid lines. The coloring of the sequence corresponds to that in panel a. The ranges of the β -strands are underlined. Solid lines connecting cysteine residues represent the disulfide bonds. From the sequence determination, two combinations, (i) Cys133–Cys161 and Cys136–Cys162, and (ii) Cys133–Cys162 and Cys136–Cys161, were possible (Doi *et al.*, in preparation), but which combination is correct remained undetermined. This structural study revealed that the combination of Cys133–Cys161 and Cys136–Cys162 is the correct one.

Fig. 2. C α trace of VMO-I (stereo view). This illustration was drawn with the program MOLSCRIPT (Kraulis, 1991).

Fig. 3. $2(F_o - F_c)$ map of the region around Ser130 (stereo view). Contour levels are 1.0σ . The hydroxyl groups of the side chains of Ser 86, 88 and 130 form hydrogen bonds to the main chain atoms.

Molecular structure

The ribbon representation and the $C\alpha$ trace of VMO-I are shown in Figures 1a and 2. The molecule is shaped like a flower vase with overall dimensions of $30 \text{ \AA} \times 30 \text{ \AA} \times 45 \text{ \AA}$. The molecule consists of six long β -strands and eight short β -strands, and can thus be classified as an all β -protein. All the β -strands are anti-parallel and they are divided into three β -sheets. The numberings of β -sheets are defined in Figure 1b. The first sheet is composed of two long β -strands and three short β -strands, which are sequentially arranged to form a Greek key motif. The second sheet has essentially the same structure as the first sheet but it lacks one short β -strand (Figure 1b). However, hydrogen bonds are formed between the side chain atoms of Ser 86, 88, 130 and the main chain atoms (Figure 3). These interstrand interactions appear to have a similar role to that of the hydrogen bonds between main chain atoms in the other two sheets. The Greek key motif, originally defined by Richardson (1977), has recently been divided into several sub-classes by Hutchinson and Thornton (1993). Using this notation, the first and second sheets of VMO-I can be classified as $(4,0)_C$ flat Greek keys. The third sheet is composed of three short β -strands from the N-terminal region and two long β -strands from the C-terminal region. If we treat the two $(4,0)_C$ Greek keys as an insertion, the remaining strands from the N- and C-termini may be regarded as a $(4,0)_N$ flat Greek key. These three Greek key structures form the triangular prism walls of the 'flower vase'.

Notably, the three segments (41–48, 92–101 and 149–152) in the long loops connecting the long β -strands are flexible, as indicated by the relatively high temperature factors. The third flexible segment is slightly shorter than the others and its B-factors are also lower. The r.m.s. deviations of the $C\alpha$ positions between the respective three

Greek key structures were calculated, excluding the three flexible and N-terminal segments. Their values are 1.1 $\text{\AA}$, 1.2 $\text{\AA}$ and 1.1 $\text{\AA}$ between the first and second, second and third, and first and third, respectively. The three structures are also related by an internal pseudo three-fold symmetry, and thus it is of interest to determine the exactness of this pseudo three-fold axis. The polar angles, κ (rotation angles about the internal three-fold axis), among these structures are evaluated to be 122° , 125° and 115° , respectively, indicating almost exact internal three-fold symmetry. The superimposed backbone structures of these three Greek key structures are shown in Figure 4. The occurrence of three homologous Greek key structures in VMO-I coincides with the presence of three internal repeats of ~ 53 residues in the sequence (Figure 1c). The internal region of VMO-I surrounded by the wall of three β -sheets is filled with the side chains of hydrophobic residues, such as Phe, Ile, Leu and Val. Furthermore, hydrogen bond networks lie at the top and bottom regions of this triangular prism (Figure 5a and b). The three disulfide bonds (Cys26–Cys57, Cys79–Cys110 and Cys133–Cys161) tie the three sheets together (Figure 1b). No hydrogen bonds involving main chain peptide groups or other direct interactions are observed between the respective β -sheets.

Discussion

CD spectrum

The far-UV spectrum of VMO-I substantially differs from the spectra commonly observed for a variety of globular proteins. It exhibits a positive ellipticity from 215 nm to ~ 240 nm and a negative ellipticity below 215 nm. A similar spectrum is observed in the kringle of tissue plasminogen activator (Cleary *et al.*, 1989) and in human growth hormone receptor binding protein (Bass *et al.*, 1991). A spectrum

Fig. 4. Superimposition of three Greek key structures (stereo view) (Kraulis, 1991). Each chain is well superimposed, except for the loop regions.

showing a positive band around 230 nm is also observed in the gene V protein (Day, 1973). These protein structures have already been determined and found to be all β -proteins. However, the CD spectra for most all β -proteins, for example, neuraminidase (β -propeller) (A.Kortt, personal communication), γ II-crystallin (Greek key) and immunoglobulin (Perczel *et al.*, 1992), show spectral patterns typical of β -structures. All the β -sheets of VMO-I have the usual right-handed twists. Therefore, it is unlikely that the unusual spectral pattern of VMO-I arises from the spatial arrangement of β -strands. We presume that it may be derived from the aromatic, phenolic and disulfide chromophores and unordered secondary structures. In fact, many aromatic rings, many tyrosyl rings and all the disulfide bridges are concentrated in a small space within the VMO-I molecule.

Negatively charged cavity

A molecular graphics display of the electrostatic potentials on the solvent accessible surface (Nakamura and Nishida, 1987) of VMO-I reveals a cluster of negative charges near the top of the molecule. On the other hand, the side and bottom regions of the triangular prism-like molecule are covered with positive charges (Figure 6). This negatively charged cluster forms a groove-like cavity surrounded by three long loops (39–51, 92–104 and 143–155) containing the flexible segments in the top region. The approximate dimensions of the cavity are $25 \text{ \AA} \times 10 \text{ \AA} \times 5 \text{ \AA}$. In the crystal lattice, this cavity interacts with the positively charged surface of a symmetrically related neighboring molecule. The size of the cavity is close to that of the cleft in lysozyme, which is well-known as the hydrolase reaction center for the polysaccharide. The negative charges are derived from Asp 46, 47, 99, 100, 150 and 151, and Glu 39, 92 and 143, which are located at the roots of the long loops. Notably, these residues are conserved in the sequences of the three repeats. Considering that, in many proteins with Greek key structures, functional variability is often contributed by the flexible loops which connect the β -strands, and that the size of this negative cavity is suitable for binding ligands like oligosaccharides, it is likely that the VMO-I molecule is involved in an unknown enzymatic function which requires acidic residues.

Structural similarity to δ -endotoxin domain II

Although sequence homology searches (Nishikawa *et al.*, 1987) and a 3-D–1-D compatibility assessment (Nishikawa and Matsuo, 1993) did not detect any related proteins, the topology of VMO-I is similar to that of the second domain in δ -endotoxin, whose crystal structure was determined by Li *et al.* (1991). This domain is composed of two four-stranded β -sheets and one three-stranded β -sheet including one α -helix. Although these sheets are also packed around a triangular hydrophobic core, more detailed comparisons with VMO-I reveal important differences. Firstly, the internal pseudo three-fold symmetry in VMO-I is much more regular than that of δ -endotoxin domain II. Secondly, the third sheet of δ -endotoxin domain II contains one α -helix, and the hydrophobic core comprises the four-stranded β -sheets and the three-stranded β -sheet (Figure 7). In contrast to VMO-I, δ -endotoxin domain II has a low level of internal sequence repeat only in the first and second sheets. In addition, no sequence similarities have been found between VMO-I and δ -endotoxin domain II. δ -Endotoxin domain II eluded detection during 3-D–1-D compatibility assessments and 3D profile searches (Bowie *et al.*, 1991), which used VMO-I as the test structure. Nevertheless, an overall structural similarity does exist between VMO-I and δ -endotoxin domain II. Roughly 60% of the residues are matched with an r.m.s. deviation of $2.6 \text{ \AA}$, when the residues in loop regions and the helical insertion are excluded. It is very unlikely that this remarkable structural similarity is accidental. δ -Endotoxin domain II is thought to bind with high affinity to specific membrane receptors (Widner and Whiteley, 1990; Schnepf *et al.*, 1990). VMO-I also has been reported to adhere to endothelial cells (T.Wakamatsu, personal communication), in addition to the transferase-like activity. Therefore, these two proteins may have somewhat similar functions.

Recent progress in protein crystallography and NMR analyses has revealed that proteins can be classified into superfamilies with common structural motifs. Chothia (1991) predicted that the number of protein families would be limited at most to 1000 in the biological world. From such a viewpoint of structural classification, the identification of new fundamental protein topologies has a central significance for the understanding of protein folding and evolution. To

(a)

(b)

Fig. 5. (a) Ribbon representation of a cross-section viewed from the bottom (left) and top (right), showing the pseudo three-fold symmetry. White lines indicate hydrogen bonds which are formed between main chain atoms in neighboring two β -sheets. (b) Schematic diagram of the main-chain hydrogen bonding scheme. Striped rods indicate only the hydrogen bonds between neighboring two β -sheets. These hydrogen bonds correspond to the white lines in panel (a). Broken lines show hydrogen bonds between main chain atoms. Amino acids are indicated using the single-letter code.

date, numerous topologies, such as the four helix bundle, the α/β barrel, the 'jelly roll', β -sandwiches, and most recently, the parallel β -helix, have been found. The topology

observed in VMO-I and δ -endotoxin domain II seems to constitute a new superfamily which has never been found before. In fact, the database analyses including 3D profile

Fig. 6. Distribution of electrostatic potential on the solvent-accessible surface of VMO-I. Red indicates electrostatic potential values below -0.1 V, while blue indicates values above $+0.1$ V. Other color codes correspond to values between 0.1 and -0.1 V, sequentially following the spectrum from blue to red. The calculation was performed under neutral conditions.

searches and 3-D–1-D compatibility assessments have not been successful in detecting any related proteins in terms of folding motifs, except for δ -endotoxin domain II whose similarity was found by visual inspection. VMO-I appears to be a typical case of this topology, because it forms a triangular prism-like structure consisting of three typical Greek key motifs, related by an almost exact internal three-fold symmetry. We assume that VMO-I and δ -endotoxin domain II are descended from a common ancestor. It is also possible that this topology is prevalent in many proteins, when it is considered that it has already been found in hen's eggs and *Bacillus thuringiensis*, which belong to evolutionarily distant species.

Although the function of VMO-I remains obscure, the structural features and the strikingly biased charge distribution imply that it may have an enzymatic activity related to saccharides. Extensive screening experiments will be required to make a definite conclusion about the function of VMO-I.

Materials and methods

Structure determination

The protein material for crystallization was purified from hen's eggs, according to the procedure described previously (Kido *et al.*, 1992). Crystals were grown by a hanging drop mode of the vapor diffusion technique, as already reported by Shimizu *et al.* (1994). The best crystals were obtained when a crystallization drop containing 25–30 mg/ml protein in 50 mM KCl, 3% (w/v) polyethylene glycol (PEG) 4000 and 46.7 mM acetate buffer (pH 4.7) was equilibrated with 5% (w/v) PEG 4000 and 100 mM KCl at 20°C. The crystals were harvested into a stabilization solution [150 mM KCl, 15% PEG 4000, 50 mM HEPES (pH 7.5)] containing 0.025% sodium azide. They belong to the orthorhombic space group $P2_12_12_1$, with unit cell parameters $a = 62.42$ Å, $b = 110.51$ Å, $c = 44.15$ Å, and contain two molecules in the asymmetric unit. The medium resolution intensity data to 2.8 Å were collected for the native and all derivative crystals at 4°C, using an automated oscillation camera system, DIP100 (Mac Science), on a rotating anode generator operated at 50 kV and 90 mA with CuK α radiation. The data were processed using the program ELMS (Tanaka *et al.*,

Fig. 7. Superimposed ribbon representation of VMO-I (red, green and blue) and δ -endotoxin domain II (yellow). The red sheet of VMO-I has four strands, while δ -endotoxin domain II has three strands and one α -helix. As a result, the three-fold symmetry in VMO-I is much clearer in δ -endotoxin domain II.

1990) and merged using the program PROTEIN (Steigemann, 1974). For crystallographic refinement, a higher resolution data set was collected to 2.2 Å, using a Weissenberg camera for macromolecules (Sakabe, 1991) installed on the beamline 6A2 at the Photon Factory (Tsukuba, Japan). The wavelength was set to 1.00 Å, and the diffraction path was filled with helium gas to avoid air scattering. Two crystals, with different orientations relative to the beam, were used for data collection to cover the blind region. The data were processed by the program WEIS (Higashi, 1989). Heavy-atom

Table I. Crystallographic data

(a) Structure determination statistics										
Data set ^a	Instrument (resolution)	No. of unique reflections (completeness)		$R_{\text{merge}}^{\text{b}}$ (all data)		$R_{\text{iso}}^{\text{c}}$	Phasing power ^e	$R_{\text{Cullis}}^{\text{d}}$	No. of sites	Anomalous data
Native	PF (2.2 Å)	13 489 (83.4%)		4.59%						
Native	DIP100 (2.8 Å)	7279 (90.8%)		3.47%						
OSO1	DIP100 (3.0 Å)	5882 (89.8%)		5.82%		13.0%	1.18	0.782	3	Yes
OSO4	DIP100 (3.0 Å)	5721 (87.3%)		7.07%		17.2%	1.04	0.825	3	Yes
GPNN	DIP100 (3.0 Å)	5991 (91.5%)		3.92%		14.1%	0.91	0.871	4	Yes
PTNN	DIP100 (3.0 Å)	5886 (89.8%)		5.27%		14.9%	1.06	0.796	3	Yes
MERS	DIP100 (3.0 Å)	6000 (91.6%)		4.80%		19.6%	0.83	0.884	2	Yes
MER2	DIP100 (3.0 Å)	5876 (89.7%)		4.67%		9.5%	0.76	0.878	2	Yes
Resolution	8.73	6.86	5.65	4.80	4.17	3.69	3.31	3.00	Total	
Figure of merit	0.86	0.85	0.81	0.79	0.73	0.67	0.61	0.57	0.68	
		Averaging								
Averaging R -factor ^f		0.135								
Real space R -factor ^g		0.227								
Correlation coefficient ^h		0.947								
(b) Refinement statistics										
Model status				Refinement statistics						
Residues	Non-hydrogen atoms	Water molecules		R -factor			r.m.s. deviations			
							bond lengths	bond angles	improper torsion	
326	2530	144		18.8% (6–2.2 Å)			0.012 Å	1.834°	1.443°	

^aOSO1: 1 mM K_2OsO_4 ; OSO4: 5 mM K_2OsO_4 ; GPNN: 1.5 mM $(\text{NH}_4)_2\text{Pt}(\text{NO}_2)_4$ + 1.5 mM glutathione; PTNN: 1 mM $(\text{NH}_4)_2\text{Pt}(\text{NO}_2)_4$; MERS: 2 mM Mersalyl; MER2: 1 mM Mersalyl.

^b $R_{\text{merge}} = \sum \sum |I_{h,j} - \langle I_h \rangle| / \sum \sum \langle I_h \rangle$

^c $R_{\text{iso}} = \sum |F_{\text{PH}}| - |F_{\text{P}}| / \sum |F_{\text{P}}|$

^d $R_{\text{Cullis}} = \sum |F_{\text{PH}} \pm F_{\text{P}} - F_{\text{H(CALC)}}| / \sum |F_{\text{PH}} - F_{\text{P}}|$ (for centric reflections)

Where $I_{h,j}$ = measured diffraction intensity; $\langle I_h \rangle$ = mean value of all intensity measurements of (h,k,l) reflection; F_{PH} = structure amplitude of a derivative; F_{P} = structure amplitude of the native crystal; $F_{\text{H(CALC)}}$ = the calculated contribution of the heavy atoms.

^ePhasing power: the ratio of the root mean square (r.m.s.) heavy-atom scattering factor amplitude to the r.m.s. lack of closure error.

^fAveraging R -factor: the crystallographic R -factor between observed structure factor amplitudes and structure factor amplitudes calculated from the averaged map.

^gReal space R -factor: the R -factor between non-crystallographically related electron density.

^hCorrelation coefficient: the correlation coefficient of each electron density from the mean electron density.

derivatives for MIR analysis were prepared by soaking native crystals in the harvesting buffer containing various heavy-atom compounds. Numerous heavy-atom compounds were screened to prepare suitable heavy-atom derivatives. However, most of them produced too serious non-isomorphism of crystals, and thus no interpretable difference Patterson maps were obtained. We tried to soak the heavy-atom compounds with β -mercaptoethanol and glutathione, since we anticipated that these additives might contribute to a decrease in non-isomorphism. $(\text{NH}_4)_2\text{Pt}(\text{NO}_2)_4$ containing glutathione and K_2OsO_4 crystals were found to produce interpretable difference Patterson maps. The positions of the heavy-atom sites were also confirmed on the basis of calculation using the direct method program MULTAN (Main *et al.*, 1980), according to essentially the same procedure as reported by Navia and Sigler (1974), and Wilson (1978). The difference Patterson map of the mersalyl derivative was very noisy, but the difference Fourier maps from the refined phases using K_2OsO_4 and $(\text{NH}_4)_2\text{Pt}(\text{NO}_2)_4$ revealed the different major sites, and hence this derivative was included for phasing. Heavy-atom parameters were refined with the program PROTEIN at 3 Å, including anomalous data from all derivatives. The heavy-atom sites were essentially identical between OSO1 and OSO4, between GPNN and PTNN, and between MERS and MER2, whereas these sites were different from each other among the three heavy-atom compounds (Table Ia). The average figure-of-merit was 0.68. However, this MIR map was too poor to trace the chain, although some β -strands were recognizable. Solvent flattening (Wang, 1985) was used to improve the phases. All calculations were performed using programs from the CCP4 package, as described by Leslie (1988). Sufficient cycles of calculation were performed to achieve convergence, as judged by the mean phase change for successive cycles. Then, the heavy-atom parameters were refined again against the solvent flattened phases with an

option of the program MLPHARE, and new MIR phases were calculated from those refined heavy-atom parameters (Cura *et al.*, 1992; Rould *et al.*, 1992). The whole procedure described above was repeated three times. However, tracing of the polypeptide was still difficult, because of a lot of breaks in the electron density. We then applied the molecular averaging technique, since the crystal contains two molecules in the asymmetric unit. All calculations for local symmetry averaging were performed in real space (Bricogne, 1976) using the program GAP (D.I. Stuart and J. Grimes, unpublished data). An initial estimate of non-crystallographic symmetry was obtained from the refined heavy-atom positions. To define the molecular envelope for one non-crystallographic asymmetric unit, a tentative model was constructed on the basis of the best map before averaging. Then, a 5 Å sphere around each atom of the model was generated and the pixels within the union of these spheres were considered as an envelope. Although this model was essentially incorrect, it was useful enough to define the envelopes for averaging. Using some routines in the GAP program which allowed editing of the envelope, we partially corrected the border of the protein molecule. Averaging was iterated 20 times until convergence was achieved. This procedure led to a substantial improvement of the map, which allowed us to trace continuous chains. The structure determination statistics are summarized in Table Ia.

Model building and crystallographic refinement

Model building was accomplished by use of the interactive computer program FRODO (Jones, 1978) on an Evans & Sutherland PS390 graphics system. Only $\text{C}\alpha$ atoms were placed into the 3 Å averaged map in one molecule. The main chain was constructed from a database of refined structures (Jones and Thirup, 1986), and the side chains were added in their preferred rotamer

conformation (Jones *et al.*, 1991) using the program O. The full model was then generated by application of the non-crystallographic symmetry operator. Crystallographic refinement was carried out using the X-PLOR package, version 3.1. Conventional energy minimization and simulated annealing with molecular dynamics (Brünger *et al.*, 1987) were performed. In the simulated annealing step, the slow-cooling protocol was applied (Brünger *et al.*, 1990) starting from 4000 to 300 K (0.5 fs time-step; 25 K temperature decrement; 50 steps at each temperature; 0.2 Å tolerance). In all steps two molecules were handled and treated independently. The current R-factor is 18.8% for all data with $|F| \geq 2\sigma_F$ between 6 and 2.2 Å resolution, and the overall geometry of the model is satisfactory, as shown in Table Ib. The average B-factors for all main chain atoms are 29.4 Å² and 29.0 Å², respectively. The relative orientations of the two molecules are related by Euler angle (-10.2° , 5.6° , 16.3°). Examination of the Ramachandran plot shows that ϕ and ψ torsion angles for non-glycine residues lie mostly within the allowed regions. However, each of the two protein molecules in the asymmetric unit contains a few residues (Thr48, Ser65 and Ser94 in molecule A; Thr2, Ser65 and Ser94 in molecule B) whose torsion angles are observed outside of the favorable regions. The electron densities corresponding to these residues are poorly defined, presumably because of their mobility. In fact, all of them are located in loop regions, except for Thr2B. The coordinates of VMO-I have been deposited in the Brookhaven Protein Data Bank as IVMO.

Acknowledgements

We thank Drs D.I. Stuart, E.Y. Jones and J. Grimes for useful advice about the molecular averaging technique and for providing the averaging program, GAP. Dr A. Murzin is acknowledged for first pointing out the similarity between VMO-I and δ -endotoxin domain II. The generous support of Drs N. Sakabe and A. Nakagawa in data collection is appreciated. We also thank Drs P. Colman and A. Kortt for measuring the CD spectrum of neuraminidase, Dr J. Li for providing the coordinate of δ -endotoxin domain II, and Dr D. Eisenberg for providing the 3D profile search program. Dr H. Nakamura is acknowledged for calculating the electrostatic potential and Dr K. Nishikawa and Y. Matsuo are also acknowledged for calculating the 3-D-1-D compatibility assessment. We thank Drs K. Katayanagi, O. Nureki, T. Ohkubo, M. Swindells, N. Alexandrov and A. Pähler for helpful discussions.

References

- Back, J.F., Bain, J.M., Vadehra, D.V. and Burley, R.W. (1982) *Biochim. Biophys. Acta*, **705**, 12–19.
- Bain, J.M. and Hall, J.M. (1969) *Austral. J. Biol. Sci.*, **22**, 653–665.
- Bass, S.H., Mulkerrin, M.G. and Wells, J.A. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 4948–4502.
- Bellairs, R., Harkness, M. and Harkness, R.D. (1963) *J. Ultrastruct. Res.*, **8**, 339–359.
- Bowie, J.U., Lüthy, R. and Eisenberg, D. (1991) *Science*, **253**, 164–170.
- Bricogne, G. (1976) *Acta Crystallogr.*, **A32**, 832–847.
- Brünger, A.T., Kuriyan, J. and Karplus, M. (1987) *Science*, **235**, 458–460.
- Brünger, A.T., Krukowski, J. and Erickson, J. (1990) *Acta Crystallogr.*, **A46**, 585–593.
- Chothia, C. (1992) *Nature*, **357**, 543–544.
- Cleary, S., Mulkerrin, M.G. and Kelly, R.F. (1989) *Biochemistry*, **28**, 1884–1891.
- Cura, V., Krishnaswamy, S. and Podjarny, A.D. (1992) *Acta Crystallogr.*, **A48**, 756–764.
- Day, L.A. (1973) *Biochemistry*, **12**, 5329–5339.
- Higashi, T. (1989) *J. Appl. Crystallogr.*, **22**, 9–18.
- Hutchinson, E.G. and Thornton, J.M. (1993) *Protein Engng*, **6**, 233–245.
- Jensen, C. (1969) *J. Embryol. Exp. Morphol.*, **21**, 467–484.
- Jones, T.A. (1978) *J. Appl. Crystallogr.*, **11**, 268–272.
- Jones, T.A. and Thirup, S. (1986) *EMBO J.*, **5**, 819–822.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
- Kido, S. and Doi, Y. (1988) *Poultry Sci.*, **67**, 476–486.
- Kido, S., Janado, M. and Nunoura, H. (1975) *J. Biochem. (Tokyo)*, **78**, 261–268.
- Kido, S., Janado, M. and Nunoura, H. (1976) *J. Biochem. (Tokyo)*, **79**, 1351–1356.
- Kido, S., Janado, M. and Nunoura, H. (1977) *J. Biochem. (Tokyo)*, **81**, 1543–1548.
- Kido, S., Morimoto, A., Kim, F. and Doi, Y. (1992) *Biochem. J.*, **286**, 17–22.

- Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
- Leslie, A.G.W. (1988) In Bailey, S., Dodson, E. and Phillips, S. (eds), *Proceedings of the Daresbury Study Weekend, 5–6 February 1988*. SERC Daresbury Laboratory, pp. 25–31.
- Li, J., Carroll, J. and Ellar, D.J. (1991) *Nature*, **353**, 815–821.
- Main, P., Fiske, S.J., Germain, G., Hull, S.E., Declercq, J.-P., Lessinger, L. and Woolfson, M.M. (1980) MULTAN 80. A system of computer programs for the automatic solution of crystal structures from X-ray diffraction data Universities of York, UK and Louvain, Belgium.
- Nakamura, H. and Nishida, S. (1987) *J. Phys. Soc. Japan*, **56**, 1609–1622.
- Navia, M.A. and Sigler, P.B. (1974) *Acta Crystallogr.*, **A30**, 706–712.
- Nishikawa, K. and Matsuo, Y. (1993) *Protein Engng*, **6**, 811–820.
- Nishikawa, K., Nakashima, H., Kanehisa, M. and Ooi, T. (1987) *Protein Seq. Data Anal.*, **1**, 107–116.
- Perczel, A., Park, K. and Fasman, G.D. (1992) *Proteins*, **13**, 57–69.
- Richardson, J.S. (1977) *Nature*, **268**, 495–500.
- Rould, M.A., Perona, J.J. and Steitz, T.A. (1992) *Acta Crystallogr.*, **A48**, 751–756.
- Sakabe, N. (1991) *Nucl. Instrum. Methods Phys. Res.*, **A303**, 448–463.
- Schnepf, H.E., Tomczak, K., Ortega, J.P. and Whiteley, H.R. (1990) *J. Biol. Chem.*, **265**, 20923–20939.
- Shimizu, T., Morikawa, K., Kido, S. and Doi, Y. (1994) *J. Mol. Biol.*, in press.
- Steigemann, W. (1974). Thesis, Technische Universität, München.
- Tanaka, I., Yao, M., Suzuki, M., Hikichi, K., Matsumoto, T., Kozasa, M. and Katayama, C. (1990) *J. Appl. Crystallogr.*, **23**, 334–339.
- Wang, B.C. (1985) *Methods Enzymol.*, **115**, 90–112.
- Widner, W.R. and Whiteley, H.R. (1990) *J. Bacteriol.*, **172**, 2826–2832.
- Wilson, K.S. (1978) *Acta Crystallogr.*, **B34**, 1599–1608.

Received on November 8, 1993; revised on December 7, 1993