J. Mol. Biol. (1987) 193, 693–707

Correlation of Co-ordinated Amino Acid Substitutions with Function in Viruses Related to Tobacco Mosaic Virus

D. Altschuh†, A. M. Lesk‡, A. C. Bloomer§ and A. Klug

MRC Laboratory of Molecular Biology
Hills Road
Cambridge CB2 2QH, England

(Received 14 February 1986, and in revised form 10 September 1986)

Sequence data are available for the coat proteins of seven tobamoviruses, with homologies ranging from at least 26% to 82%, and atomic co-ordinates are known for tobacco mosaic virus (TMV) *vulgare*. A significant spatial relationship has been found between groups of residues with identical amino acid substitution patterns. This strongly suggests that their location is linked to a particular function, at least in viruses identical with the wild-type for these residues. The most conserved feature of TMV is the RNA binding region. Core residues are conserved in all viruses or show mutations complementary in volume. The specificity of inter-subunit contacts is achieved in different ways in the three more distantly related viruses.

"Give us more details", he said to his son,
"Give us more details, there is no novelty and truth in anything but the details"
translated from: Stendahl
Lucien Leuwen (Chapter 56)

1. Introduction

Tobacco mosaic virus (TMV)‖ is the best studied example of a self-aggregating system (reviewed by Butler, 1984). The coat protein subunit of TMV aggregates reversibly to give a helix formed of 2130 identical protein subunits packaging the RNA. Several polymorphic aggregates of the protein subunit, and the relationships between them, are well-characterized. Detailed structural studies have focused upon two of these aggregates. The intact virus forms highly ordered gels from which a fibre diffraction analysis has allowed the structure to be determined using data to 4 Å resolution (Stubbs *et al.*, 1977), and more recently to calculate an image of the structure at 3·6 Å resolution (Namba & Stubbs, 1985). The only form of the protein that has been crystallized is the two-layered disk of 34 subunits, the structure of which is now known from crystallographic refinement at 2·8 Å (Bloomer *et al.*, 1978; Mondragon, 1984). The atomic co-ordinate set used in this work was that of the refined subunit of the upper ring of the TMV protein disk (Mondragon, 1984).

The constraints on the overall size, shape and character of those regions of the subunit surface which are involved in quaternary structure(s) are expected to be somewhat more severe than the constraints on typical monomeric proteins, and this should be reflected in the nature and pattern of acceptable amino acid substitutions. These constraints are likely to be even more severe in the case of the coat protein of TMV, which forms two distinct quaternary aggregates, the helical form as in the intact virus and the cylindrical form as in the disk, which have some of the inter-subunit interactions in common whilst others are quite different.

The characteristics of this group of viruses are described by Gibbs (1977), who used the sequence data then available to study both evolutionary relationships within the tobamovirus group and also the correlation between the serological differentiation indices of these viruses (Van Regenmortel, 1975) and their degree of primary sequence homology (Gibbs, 1980). The present study uses additional sequence data now available for two members of the group of seven viruses, together

† Permanent address: Institute de Biologie Moleculaire et Cellulaire, 15, rue Descartes, 6700 Strasbourg, France.

‡ Permanent address: Fairleigh Dickinson University, Teaneck-Hackensack Campus, Teaneck, NJ 07666, U.S.A.

§ Author to whom correspondence should be addressed.

‖ Abbreviations used: TMV, tobacco mosaic virus; DAHL, Dahlemense virus; U2, virus U2; HR, Holmes ribgrass virus; SHMV, Sunn-hemp mosaic virus; ORSV, orchid ringspot virus; CGMMV-W, cucumber green mottle mosaic virus (water melon strain).

0022-2836/87/040693-15 $03.00/0

with the detailed molecular conformation of the coat protein molecule of TMV *vulgare* (Bloomer *et al.*, 1978). This paper describes the results of a detailed analysis of the patterns of amino acid substitution among these seven sequences and the relationship of these to the structure of TMV *vulgare* coat protein.

2. Methods

(a) *Sequence and alignment*

The viruses considered in this study are listed in Table 1 together with references to their alternative names, closely related viruses and individual descriptions. Sequence data are from the P.I.R. Protein Databank (Barker *et al.*, 1986) except where specifically referenced in Table 1. Seven distinct members of the tobamovirus group were identified by Gibbs (1980). There are primary sequence data now available for 6 complete sequences and one partial sequence so that 7 viruses have been included in this study: tobacco mosaic virus *vulgare* (TMV), Dahlemense virus (DAHL), virus U2 (U2), Holmes ribgrass virus (HR), Sunn-hemp mosaic virus (SHMV), orchid ringspot virus (ORSV), and cucumber green mottle mosaic virus, water melon strain (CGMMV-W).

The 7 sequences are shown in Table 2. The incomplete sequence for ORSV is shown with the punctuation given by the P.I.R. Protein Databank (Barker *et al.*, 1986): a full stop (period) separating 2 residues indicates that the residue to the left has been placed with 90% confidence; separation by a comma indicates that the residue to the left could not be positioned with confidence by homology; an equality sign is equivalent to ")(," whilst preserving proper spacing within the printed sequence. However, since this sequence was entered into the databank over 10 years ago, additional data suggest that some of the unknown regions can now be confidently assigned on the basis of homology with other members of the group. Residues in ORSV that still cannot be assigned with confidence are printed in italics in Table 2.

The 6 related virus sequences were aligned individually with the TMV *vulgare* sequence. Two viruses (SHMV and CGMMV-W) have short additions to the sequence at the carboxyl terminus. The only internal additional residues are found in the external loop of residues 64 to 66 where there are 1 (SHMV) or 2 (ORSV) extra residues. These are shown in Table 2 at opposite ends of this short loop but the precise equivalence of residues between the shorter TMV loop and these 2 longer loops remains uncertain. Similarly the equivalence of the pair of adjacent threonine residues (103 and 104 in TMV *vulgare*) to the pair found in 5 other sequences suggests that this region of the polypeptide may be folded differently in U2 from the other viruses, so enabling any specific interactions made by this Thr-Thr dipeptide to be maintained in U2 as in the other 5 viruses. The only deletions occur towards the carboxyl terminus at residue 151 in SHMV and in similar positions in ORSV and HR though the exact alignment here again remains uncertain. These deletions are expected to involve slight

Table 1

The tobamovirus group

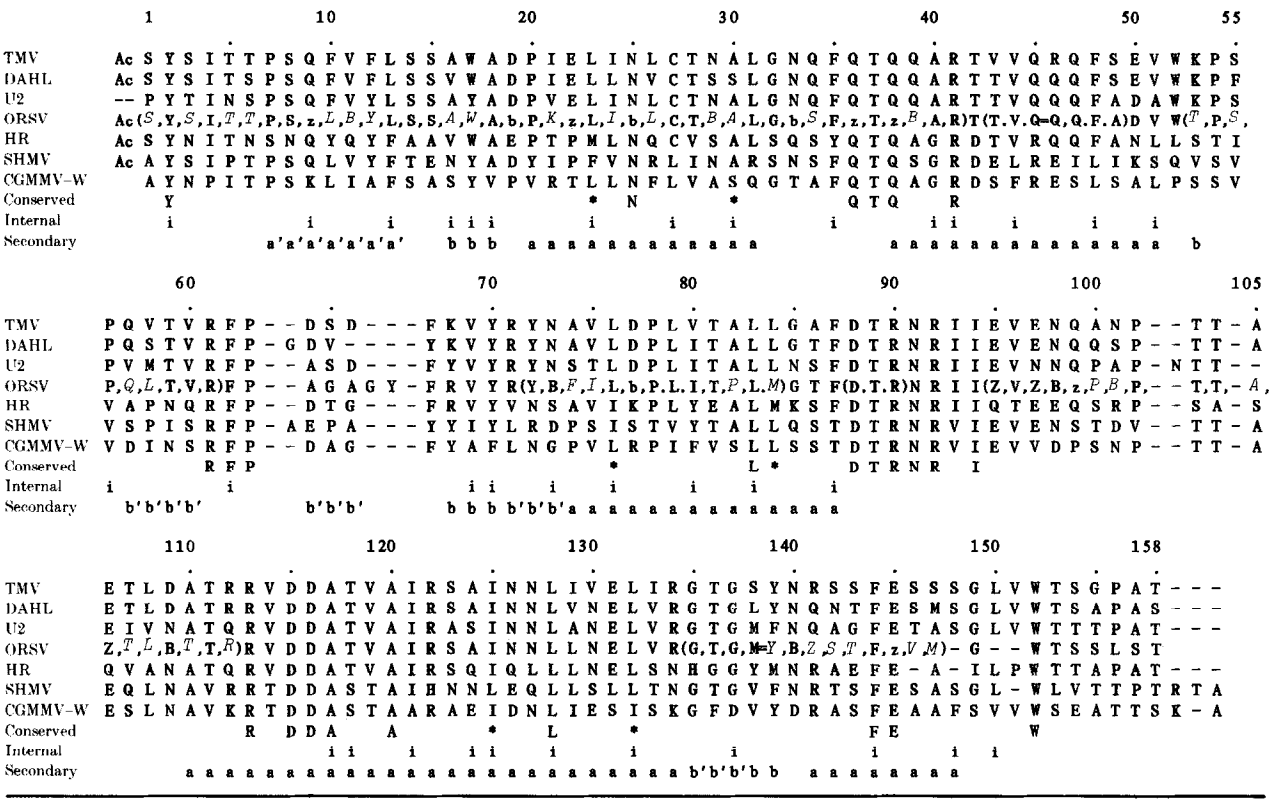
Tobacco mosaic virus *vulgare*—the type member of the tobamovirus group	
Abbreviation:	TMV
Description:	Zaitlin & Israel (1975)
Other viruses:	OM strain (129-Val 153-Asn); a third change in OM (15-Ala) is shown by the RNA sequence of Takamatsu *et al.* (1983).
Tomato mosaic virus—Dahlemense strain	
Abbreviation:	DAHL
Description:	Hollings & Huttinga (1976)
Other viruses:	TOMATO-L strain (20-Asn 86-Ala) Y-TAMV (yellow tomato atypical mosaic virus)
U2 tobacco mosaic virus—isolate of T2MV group	
Abbreviation:	U2
Sequence:	corrections to Databank sequence are from Altschuh *et al.* (1981)
Other viruses:	G-TAMV (green tomato atypical mosaic virus)
Holmes ribgrass virus—wide host range	
Abbreviation:	HR
Description:	Oshima & Harrison (1975)
Other viruses:	Japanese isolate (95-Glu 99-Glu 109-Asp 140-Asp 143-Gln)
Sunn-hemp mosaic virus—host range is legume family; formerly known as Cowpea strain of TMV	
Abbreviation:	SHMV
Description:	Kassanis & Varma (1975)
Odontoglossum ringspot virus—orchid family	
Abbreviation:	ORSV
Description:	Paul (1975)
Cucumber green mottle mosaic virus—water melon strain	
Abbreviation:	CGMMV-W
Description:	Hollings *et al.* (1975)
Sequence:	from the RNA sequence of Meshi *et al.* (1983)
Other viruses:	CV4—cucumber virus

The recommended terminology used here differs from that in the earlier literature on TMV as virus taxonomy is still a developing subject, (cf. review by Matthews, 1985). The properties of the group as a whole, and the distinctive members within the group, have been described by Gibbs (1977). The 7 viruses included in this study are listed here together with references to their descriptions; strains or closely related viruses which are referred to in recent literature are also shown with the positions of any amino acid differences from the sequence shown in Table 2.

All sequence data are from the P.I.R. Protein Databank (Barker *et al.*, 1986) except where specifically referenced.

Table 2

Alignment of the seven tobamovirus coat protein sequences



For references to virus descriptions and sequences, see Table 1. Lower case symbols (z, b) in the ORSV sequence indicate those positions where a sequence ambiguity in ORSV between acid and amide residues has been assumed to be the same residue as that in all (25, 36, 38, 145) or most (9, 19, 22, 33, 77, 99) other viruses. Italic symbols represent positions where homology does not allow a confident assignment within an incompletely sequenced region of ORSV; those in normal type are assigned with high confidence, notwithstanding the commas given in the Databank sequence. The CGMMV-W sequence comes from the RNA; its amino-terminal is expected to be acetylated but this remains unknown. Residues conserved in all viruses are emphasized in a separate line; those regarded here as semi-conserved are marked with an asterisk; those with secondary structure (Mondragon, 1984) are shown a, a′, b, b′ for α, 3_{10} helices, β-sheet and β-bends. Internal or buried residues are marked (i) for those in the isolated subunit with accessible surface area (Lee & Richards, 1971) for each residue of less than 20 Å^2, as calculated by the program of Richmond (1984) using a probe sphere 1·4 Å in diameter.

rearrangement of the loop following the high radius helix (141–148).

Alignment of the sequences (Table 2) reveals that only 25 of the positions (16%) contain the same amino acid in all 7 of the viruses. The degree of homology within this family of 7 viruses is shown in Table 3. Those residues in ORSV that can confidently be assigned by homology with all 3 of TMV, DAHL and U2 are shown in normal type in

Table 3

Sequence homology between the tobamovirus coat proteins

	TMV	DAHL	U2	ORSV†	HR	SHMV	CGMMV-W
		% sequence homology					
TMV	:	82	72	59(75)	45	41	36
DAHL	28	:	70	61(73)	47	38	35
U2	44	48	:	61(70)	46	40	33
ORSV†	65(40)	62(43)	63(48)	:	40(46)	29(40)	26(32)
HR	87	85	86	96(85)	:	35	35
SHMV	95	100	98	115(98)	105	:	43
CGMMV-W	103	106	108	120(110)	105	92	:
		Number of differences					

The number of differences and percentage homology are calculated from the alignment shown in Table 2, where the number of residue comparisons for each pair of sequences lies between 158 and 163. The 2 diagonal halves of the Table show the percentage homology and the number of differences, respectively.

† Figures in this row and column represent the minimum homology and maximum number of differences between ORSV and other viruses, assuming that the residues shown in Table 2 in italics or as being unassigned (upper case) Glx or Asx residues are all differences from each of the other 6 viruses. The figures in parentheses are the maximum homology and minimum number of differences using the ORSV sequence as in Table 2, but without punctuation and with resolution of Glx and Asx uncertainties.

Table 4

Groups of residues with identical patterns

Pattern (text ref. and Fig.)	Residue number	Residue type	Pattern (text ref. and Fig.)	Residue number	Residue type
1 1 1 1 1 1 1	2	Y Y Y Y Y Y Y	1 1 1 1 2 2 2	13	L L L L F F F
Section 4(a)	25	N N N B N N N	Section 4(d)(i)	40	A A A A G G G
Fig. 3	36	Q Q Q Z Q Q Q		42	T T T T D D D
	37	T T T T T T T		45	Q Q Q Q R R R
	38	Q Q Q Z Q Q Q		56	P P P P V V V
	41	R R R R R R R			
	61	R R R R R R R	1 1 1 1 2 2 3	22	E E E Z P P T
	62	F F F F F F F	Section 4(d)(ii)	134	R R R R N N K
	63	P P P P P P P			
	83	L L L L L L L	1 1 1 1 2 3 2	15	S S S S A E A
	88	D D D D D D D	Section 4(d)(iii)	28	T T T T V I V
	89	T T T T T T T	Fig. 8	59	T T T T N I N
	90	R R R R R R R		72	Y Y Y Y N R N
	91	N N N N N N N			
	92	R R R R R R R	Consider also		
	94	I I I I I I I	. 1 1 1 2 3 2	133	I V V V S T S
	113	R R R R R R R			
	115	D D D D D D D	1 1 1 1 2 3 3	54	P P P P T S S
	116	D D D D D D D	Section 4(d)(iv)	60	V V V V Q S S
	117	A A A A A A A	Fig. 9	71	R R R R V L L
	120	A A A A A A A			
	128	L L L L L L L	1 1 1 1 2 3 4	52	W W W W L Q P
	144	F F F F F F F	Section 4(d)(iv)	77	D D D B K S R
	145	E E E Z E E E	Fig. 9	126	N N N N Q E D
	152	W W W W W W W			
			1 1 2 2 3 4 5	50	E E D D N K A
1 1 1 1 2 1 1	7	P P P P S P P			
Section 4(b)	8	S S S S N S S	1 1 1 1 1 1 2	4	I I I I I I P
Fig. 4	35	F F F F Y F F		9	Q Q Q Z Q Q K
	96	V V V V T V V		18	A A A A A A V
	103	T T T T S T T		70	Y Y Y Y Y Y F
	104	T T T T A T T		121	I I I I I I A
	135	G G G G H G G		132	L L L L L L I
				137	G G G G G G D
1 1 1 1 1 2 2	27	C C C C C L L			
Section 4(c)(i)	48	F F F F F L L	1 1 1 1 1 2 1	78	P P P P P T P
Fig. 5	87	F F F F F T T		102	P P P P P V P
	93	I I I I I V V		122	R R R R R H R
	111	T T T T T V V		125	I I I I I L I
	114	V V V V V T T			
	118	T T T T T S S	1 1 1 1 2 3 1	14	S S S S A T S
	119	V V V V V T T		23	L L L L M F L
				127	N N N N L Q N
1 1 1 1 1 2 3	20	P P P P P Y V			
Section 4(c)(ii)	31	L L L L L R Q	1 1 1 1 2 1 3	19	D D D B E D P
Fig. 6	44	V V V V V L F		33	N N N B Q N T
	47	Q Q Q Q Q I S		81	T T T T E T V
	69	V V V V V I A		136	T T T T G T F
	79	L L L L L V I		149	G G G G I G S
	99	Q Q Q Z Q S P			
	131	E E E E E L S	1 1 1 2 1 1 3	82	A A A *P* A A S
	153	T T T T T L S		140	N N N B N N D
				150	L L L – L L V
1 1 1 1 . . .					
Fig. 7			1 1 2 1 3 4 5	57	Q Q V *Q* A S D
				85	G G N G K Q S
				107	T T I *T* V Q S
				124	A A S A Q N E

See the text for an explanation and the corresponding paragraph and Figure numbers where these groups are specifically discussed. All groups represented by more than 2 residues are listed. The 7 columns to the right indicate the residue type in TMV *vulgare*, DAHL, U2, ORSV, HR, SHMV and CGMMV-W, respectively. Italic symbols are as for Table 2.

Table 2, with appropriate punctuation, and Tables 4 and 5, where the punctuation has to be omitted. These residues are included in the homology calculations shown in Table 3. Other residues, printed in italics in Tables 2, 4 and 5, were excluded from the calculations. The homologies within this group of 7 viruses range from at least 26% to 82%.

The ORSV sequence also includes 12 residues identified only as Asx and 10 residues as Glx. For this analysis 10 of these have been assigned on the basis of the residues

Table 5

External residues involved in lateral contact in the disk

A subunit		*C* subunit	
Residue number	Residue type in the 7 sequences	Residue number	Residue type in the 7 sequences
7†	P P P P S P P	20†	P P P P P Y V
8†	S S S S N S S	21	I I V *K* T I R
10	F F F *L* Y L L	24	I L I *I* L V L
11	V V V *B* Q V I	25†	N N N B N N N
14†	S S S S A T S	28†	T T T T V I V
15†	S S S S A E A	31†	L L L L L R Q
45†	Q Q Q Q R R R	32†	G G G G S S G
49	S S A A A I S	34	Q Q Q *S* S S A
71†	R R R R V L L	36‡	Q Q Q Z Q Q Q
72†	Y Y Y Y N R N	67	F Y F F F Y F
81†	T T T T E T V	106	E E E Z Q E E
84	L L L M M L L	107	T T I *T* V Q S
85	G G N G K Q S	111†	T T T T T V V
88‡	D D D D D D D	112	R R Q *R* Q R K
90‡	R R R R R R R	122†	R R R R R H R
92‡	R R R R R R R	126†	N N N N Q E D
93†	I I I I I V V	129	I V A L L L I
94‡	I I I I I I I	133	I V V V S T S
113‡	R R R R R R R	135†	G G G G H G G

The residues are listed separately for the neighbouring subunits denoted *A* and *C* in the nomenclature of Fig. 1.

Italic symbols are as for Table 2.

The difference in accessible surface area (probe sphere 1·4 Å diameter) between the isolated subunit and the dimer is larger than 20 Å^2. The residue type for each position is given for the 7 viruses TMV *vulgare*, DAHL, U2, ORSV, HR, SHMV and CGMMV-W, respectively.

† Residues conserved in at least the 4 viruses TMV, DAHL, U2 and ORSV.

‡ Residues conserved in all 7 sequences.

being conserved either in all of the other 6 viruses (residues 25, 36, 38 and 145) or in the 3 most closely related viruses (residues 9, 19, 22, 33, 77 and 99). The remaining Asx and Glx residues are not further considered here.

(b) *Analysis strategy*

Some regions of the sequence are highly variable. At other positions, only restricted substitutions are observed. Conservation of individual residues, or complementary mutations occurring at residues adjacent within the 3-dimensional structure of the folded protein, may be indicative of these residues having specific roles in the function of the protein, e.g. maintaining a particular tertiary or quaternary interaction between parts of the chain(s), or binding to RNA.

There is no objective way of deciding whether or not any change is conservative because the constraint on a residue depends primarily upon its function in the protein: the substitutions allowed for a given residue will be different if that residue is at the protein surface, at subunit interfaces or partly buried. Therefore only identical residues are here considered as being conserved.

The strategy used to detect possible complementary mutations was as follows.

(1) If a mutation at a given residue is complemented by a mutation at another residue in a neighbouring location, these 2 residues will consistently be replaced simultaneously in 2 or more viruses, with an identical pattern of changes. Therefore, the family of 7 sequences was searched for residues with identical conservation patterns, i.e. residues having 1 amino acid in a given set of sequences, but different amino acids in the remaining ones. The pattern of each residue in the 7 sequences is symbolized by 7 numbers. The residue type found in the first member (TMV *vulgare*) of the family of 7 sequences aligned as in Table 2 is symbolized by number '1'. This number is incremented by unity for each new type of amino acid found at that position in one of the following sequences. Thus, the pattern is '1111122' for residue 27 (Cys Cys Cys Cys Cys Leu Leu) and '1111232' for residue 15 (Ser Ser Ser Ser Ala Glu Ala).

(2) The environment in space of residues with an identical conservation pattern was examined in order to find whether these residues were related in pairs or groups close to one another, either within one subunit or at subunit interfaces within the disk and/or virus.

By first analysing the sequences to find repeated patterns of changes and only then examining the spatial location of residues with similar sequence characteristics, no assumptions were made *a priori* about the factors likely to influence protein stability and interaction with other subunits and RNA.

(c) *Relationship of disk to virus*

The lateral interface between subunits is similar in the disk and in the virus, but the axial interface is quite different in these 2 types of aggregate (Fig. 1). Axial contacts occur only at high radius in the disk (Fig. 2) but are far more extensive in the virus (Fig. 1(b)). In the absence of any co-ordinates for the virus structure, axial contacts between subunits in the virus were examined by transforming the atomic model of the disk as a rigid body into the helical lattice of the virus, using a transformation obtained (Mondragon & Bloomer, unpublished results) from refining the atomic co-ordinates from the protein disk against the intensities from fibre diffraction photographs (Stubbs *et al.*, 1977) of the virus.

The subunit structure in the virus is not known to such high resolution as that of the disk but, since it is the proximity at the residue level and not at the atomic level that is being investigated here, the conclusions are unlikely to change when the virus structure is known at higher resolution. Nor do the results depend on possible small conformational changes occurring in distantly related viruses. The gross conformation of most of these related proteins is known to be very similar from a comparison of their diffraction patterns (cf. Holmes & Franklin (1958) for the radial density distributions of viruses similar to TMV, U2, SHMV, and CGMMV-W; cf. Caspar (1963) for small differences (1 Å or less) seen in the inter-subunit interactions of DAHL virus). Because no high resolution structural investigations of tobamoviruses other than TMV *vulgare* are in progress, it is appropriate to proceed with this study using information currently available.

3. Sequence Location of Residues with Identical Conservation Patterns

All the patterns of co-ordinated sequence changes observed in three or more residues in the tobamovirus sequence alignment, together with the corresponding residue numbers and types are listed in Table 4. Residues of most of the groups listed in Table 4 have been found to have significant spatial relationship and these are described below.

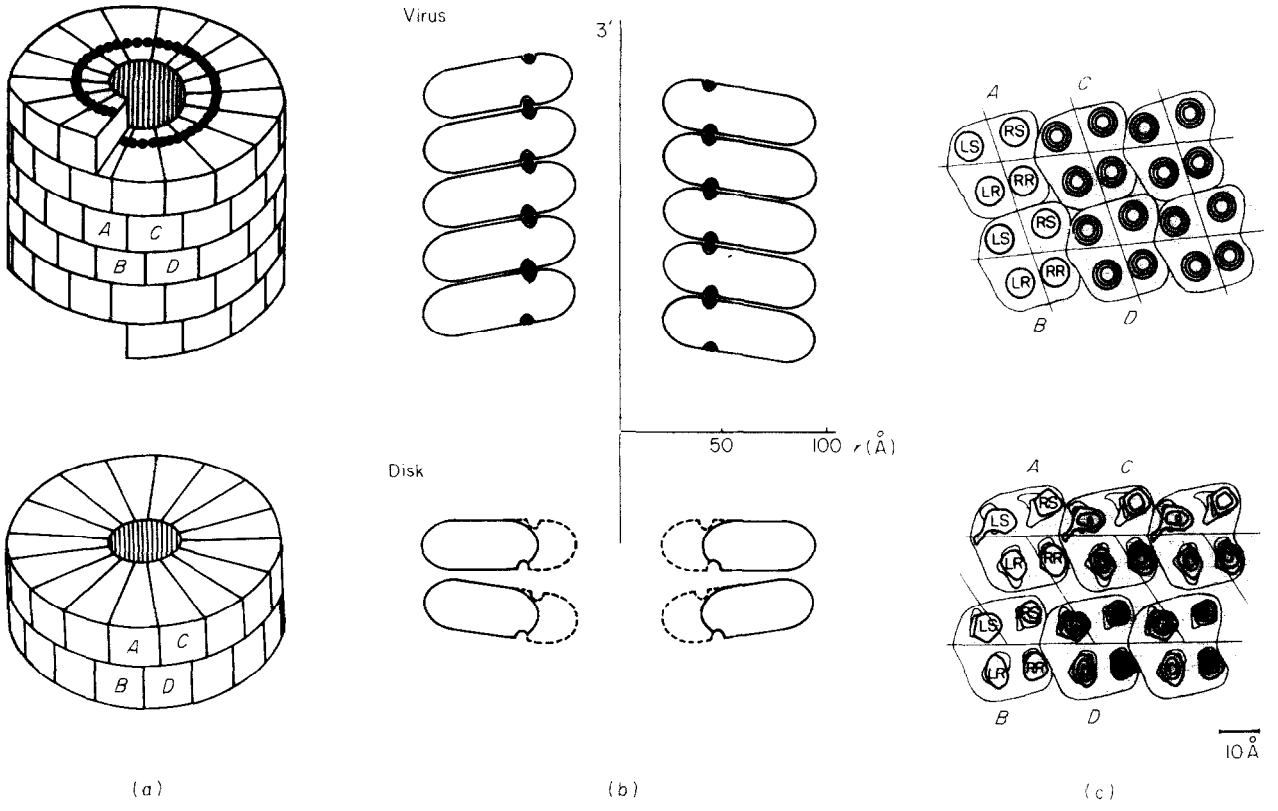


Figure 1. Packing of subunits in the TMV protein disk (below) and in the virus (above). (a) The disk contains 34 subunits in 2 rings of 17; the virus $16\frac{1}{3}$ subunits/turn of the helix. RNA is present in the virus particle, sandwiched between the helically arranged subunits. Individual subunits are labelled *A B C D* to denote their relative positions, as referred to in the text and later in the Figure legends. (b) Subunits in the upper and lower layers of the disk have different tilts with respect to the axis, and subunits in the virus have a still different tilt. The innermost part of the subunit appears to be disordered in the disk (broken outline), perhaps to facilitate incorporation of RNA in the transition from disk to helix that occurs during assembly, when this part of the protein becomes ordered around the RNA. (c) Cylindrical sections through subunits in the disk and in the virus, viewed end-on. The disk sections are taken from the 6 Å electron density map of Champness *et al.* (1976); the virus section is schematic. The major part of each subunit consists of 4 α-helical rods, running in a roughly radial direction, and denoted LS, RS, RR, LR (cf. Fig. 2); the sections have been taken through these rods. The Figure shows that in the transition from a 17-fold disk to a $16\frac{1}{3}$-fold helix, subunits slide over each other by about 10 Å. Because of the change in subunit tilt, lateral contacts will be bent somewhat on transition from disk to helix, but not completely altered like the axial bonds. Adapted from Bloomer & Butler (1986).

4. Spatial Location of Residues with Identical Conservation Patterns

(a) *Residues conserved in all sequences (pattern '1111111')*

Twenty-five residues are conserved in all seven sequences. Fifteen of these are charged or polar whilst ten are neutral. Figure 3 shows the distribution of these residues within the protein monomer. Twenty of these residues (80%) are concentrated into just two areas of the molecule, involving five regions of the polypeptide chain: near the RNA binding site (36–41, 88–94 and 113–120) and at high radius (61–63 and 144–145).

Of the charged residues, the only two at high radius form an intra-subunit salt-bridge (Arg61, Glu145) and are adjacent to the two invariant phenylalanine residues (62 and 144) and proline 63 considered below, thus forming a tight conserved cluster. At intermediate radii, Asn25 is involved in the lateral interface where it interacts with two water molecules and a main-chain nitrogen (Ser15) from a neighbouring molecule (Mondragon, 1984). The remaining 12 of the 15 invariant polar or charged residues are all located at about 45 Å radius where they may be involved in RNA binding as well as in stabilizing the conformation of the molecule or the quaternary aggregates. Details of the interactions between protein and RNA remain unclear but the three phosphate groups are probably neutralized by the invariant arginine residues 41, 90 and 92 (Mondragon *et al.*, 1985). It has been suggested (Holmes, 1979) that Asp115 and Asp116 interact with two of the ribose rings of the RNA. The invariant side-chains in the hairpin joining the slewed helices (Fig. 2(a)) interact (Mondragon, 1984) with the main chain to stabilize that hairpin (37, 41 and 36 in one of its alternative positions); with both the main-chain, just beyond the end of the RR helix, and the side-chain of an invariant arginine in one of its observed conformations (38); or with the hairpin region of neigh-

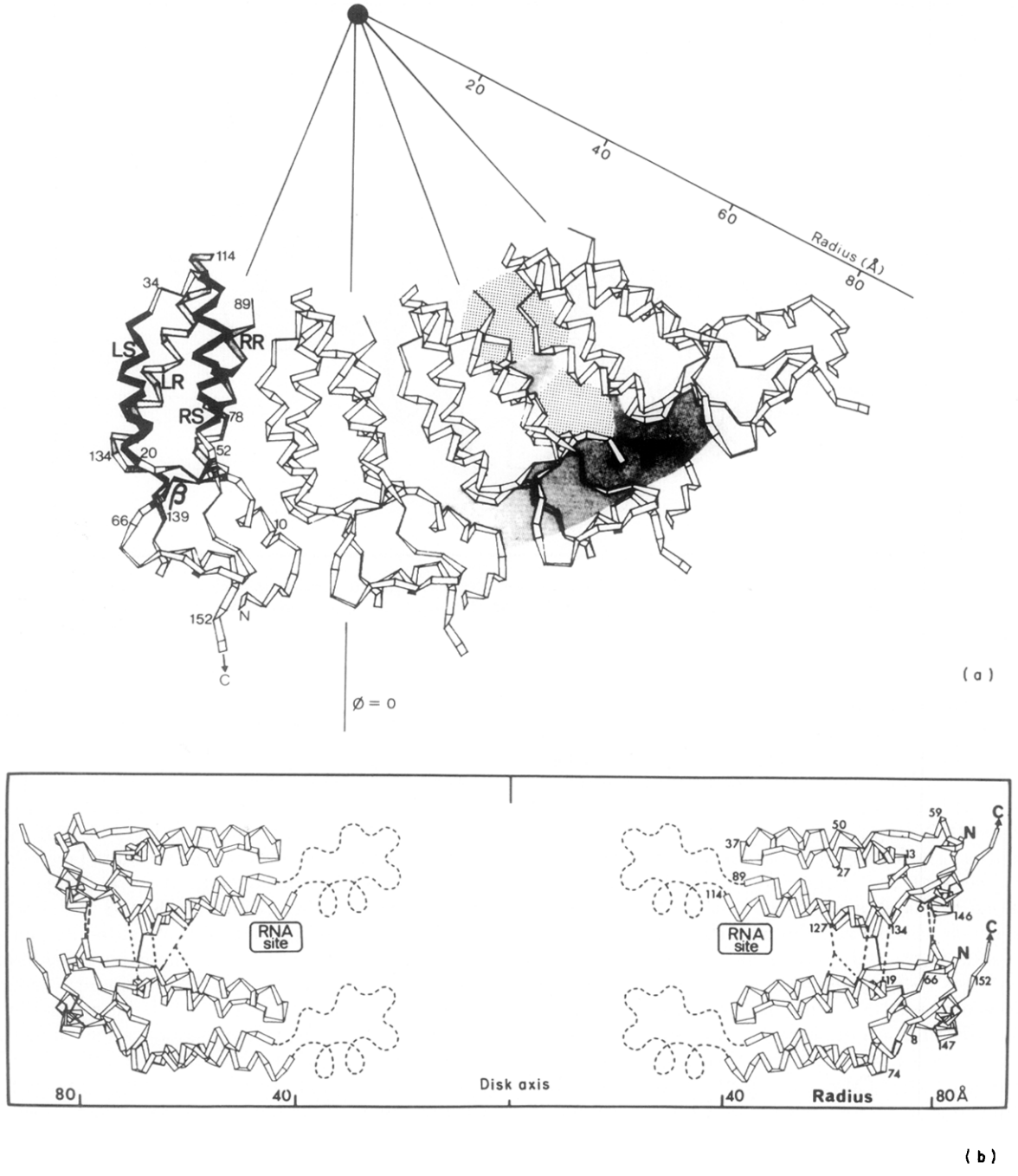


Figure 2. Ribbon drawing of the α-carbon positions in the protein subunits in the disk of TMV *vulgare*. Numbers refer to amino acid position in primary sequence of TMV *vulgare*. (a) Four adjacent subunits in 1 ring viewed from above. Regions of the longer α-helices (labelled LS and RS for left and right slewed and RR and LR for right and left radial, respectively) and β-sheet are indicated on the left-hand subunit, and the alternating patches of polar (light shading) and hydrophobic (heavy shading) contact are indicated in the interface. The "hydrophobic girdle" of interactions between and within subunits, which continues around the whole circumference of the disk, is indicated by shading. (b) Side view of subunits in an axial section through the disk showing vertical contacts between subunits in the 2 rings. The flexible loop at inner radius is shown as a broken line and the RNA-binding site is indicated. Adapted from Bloomer *et al.* (1978).

bouring molecules (36 and 38). The other conformation of Gln36, seen in the A-ring of the disk, interacts with the hydroxyl of Thr118 (Mondragon, 1984), a hydrogen bond which could be maintained when this is replaced by Ser118 in both SHMV and CGMMV-W. In the lower part of the molecule, the role of the side-chains of 89, 91 and 113 is ambiguous in the absence of a good model for the RNA binding. However, Asp88 makes an inter-subunit salt-bridge with Arg122 (Bloomer *et al.*, 1978), which is conserved in all of these seven viruses except SHMV, where the salt bridge is expected to involve Asp88 and, from the neighbouring molecule, Arg31 and His122 (Bloomer *et al.*, 1981).

The ten neutral conserved residues are part of the

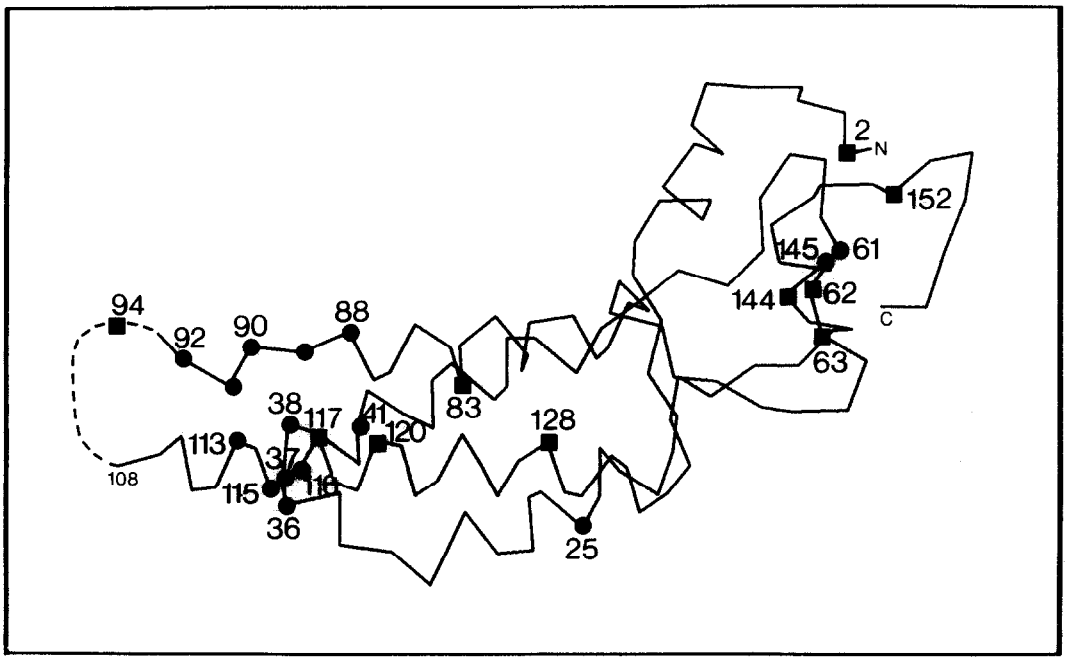


Figure 3. Residues invariant in all 7 viruses. The α-carbon chain tracing of 1 subunit is shown viewed down the disk axis as in Fig. 2(a), and the positions are marked for (■) hydrophobic and (●) hydrophilic residues, which are invariant in all 7 viruses.

core of the molecule. Five of them are located at high radius: Tyr2 and Trp152 each make links between all three regions of the polypeptide chain; the ring of Pro63 is next to Phe62 and the aromatic cluster which includes Phe144. Two residues (Leu83 and Leu128) are in the centre of the bundle of four α-helices. The conserved alanine residues (117 and 120) occur at the inner end of the LR helix and Ile94 at the beginning of the flexible loop. Alanine 110, which occurs even closer to the inner end of the LR helix, is the only additional residue which may prove also to be conserved in all viruses, once all regions of the ORSV sequence are fully determined.

The distribution of these 25 absolutely conserved residues shows that the most strongly conserved features of the TMV coat protein are the RNA binding region at low radius, and the hydrophobic core at high radius.

(b) *Residues conserved in all viruses except HR (pattern '1111211')*

Seven additional residues are conserved in six viruses only, but differ in HR: Pro7, Ser8, Phe35, Val96, Thr103, Thr104 and Gly135. Of these residues, Pro7 and Ser8 of one subunit face Gly135 of the laterally adjacent subunit in the *vulgare* disk (Fig. 4). Three other residues (96, 103 and 104) are in a flexible loop at low radius. Residue 35 changes in HR to Tyr35, which may here be considered a "conservative" change as the aromatic ring is buried between the LS and RS helices (cf. Fig. 2). Uncertainties in the ORSV sequence may add three more residues to this group: 95 and 106, which are adjacent to other residues having this pattern within the flexible loop, and 84, which does not interact with others of this group.

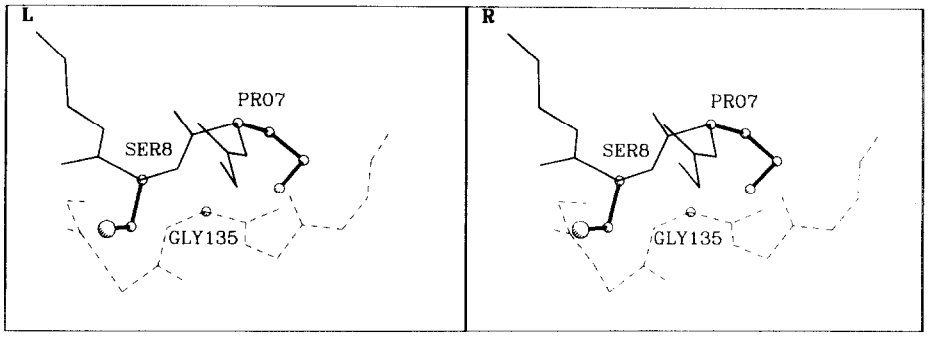


Figure 4. Stereo view of a lateral contact in the TMV *vulgare* disk. The side-chains of Pro6 and Ser8 of 1 molecule (*A* subunit: backbone continuous line) are shown with Gly135 of the adjacent molecule (*C* subunit: backbone broken line). The view is perpendicular to the disk axis and approximately along the long axis of the molecule towards the centre.

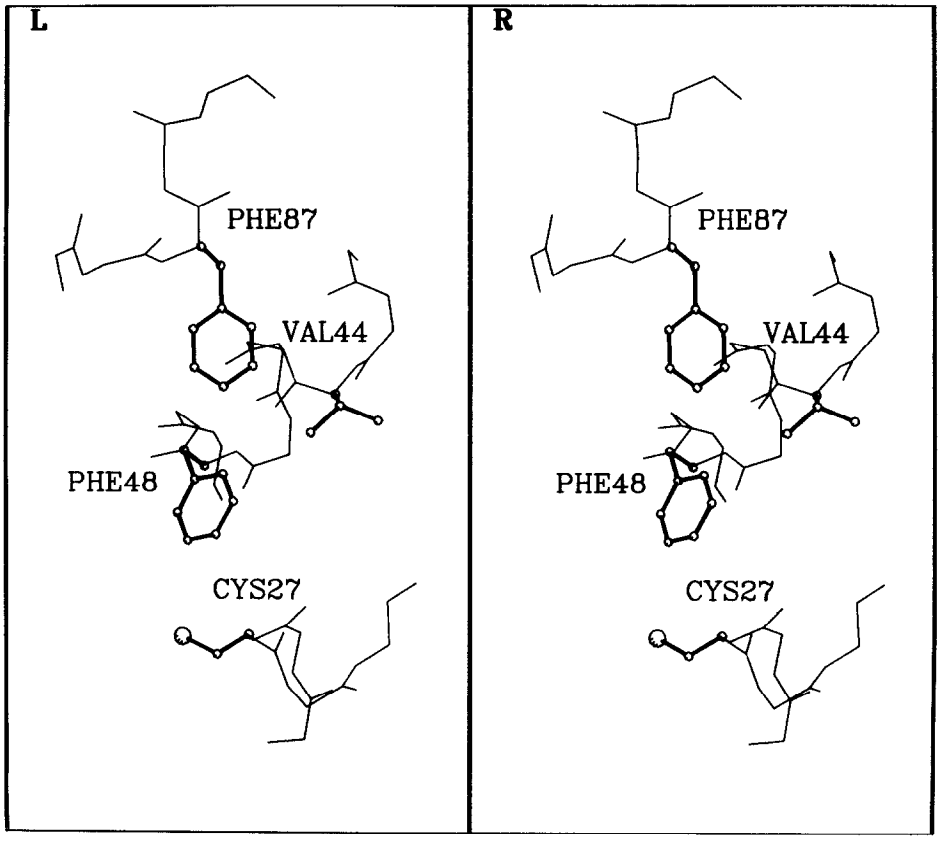


Figure 5. Stereo view of the group of residues preserving a similar total side-chain volume in all 7 tobamoviruses. The side-chains drawn are those of TMV *vulgare* (Cys27, Val44, Phe48 and Phe87), viewed from beneath the subunit.

(c) *Residues invariant in TMV, DAHL, U2, ORSV and HR (pattern '11111..')*

(i) *Those common to SHMV and CGMMV-W (pattern '1111122')*

Of the eight residues of this group, four maintain a small size and a non-polar character in all viruses (111, 114, 118 and 119). They are located at the inner end of the LR helix, close to the conserved alanine residues (117 and 120).

Residues 27, 48 and 87 are part of the core. Phe48 in TMV is in van der Waals' contact with Cys27 and Phe87, as is Phe87 also with Val44 (Fig. 5). Val44 has the similar pattern '1111123' and so may be considered with this group of spatially related residues. In the other two viruses, SHMV and CGMMV-W, these four positions have different side-chain volumes, but the total volume of the four side-chains (Chothia, 1984) is similar in all viruses: 654 Å^2 for the sequences identical with TMV *vulgare*, 626 Å^2 for SHMV and 661 Å^2 for CGMMV-W. These mutations thus appear to compensate each other. The remaining member of this group (Ile93, or Val93) is at the beginning of the flexible loop and so its contacts remain uncertain.

(ii) *Those different in SHMV and CGMMV-W (pattern '1111123')*

In the virus model residues 131 and 79 of one subunit are in a good position to make axial contacts with residue 47 of the subunit beneath (Fig. 6). Residues 20 and 69 are neighbours inside one subunit although not in van der Waals' contact but, in this case, the sum of the volumes of the residues at these two amino acid positions is not conserved. Residue 69 is buried, but residue 20 is not. Residues 31, 99 and 153 are not in spatial proximity to other residues of this group. Residue 44 was considered in the previous paragraph, as if the roles of leucine and phenylalanine at this position are equivalent. The main-chain of residue 31 has a hydrogen bond from Arg122 (see section (a), above), except in SHMV where the side-chain of residue Arg31 is expected to be involved in this salt-bridge (Bloomer *et al.*, 1981).

(d) *Residues invariant in TMV, DAHL, U2 and ORSV (pattern '1111...')*

Many of the 18 residues shared by only these four viruses are located in the middle region of the subunit at the distal end of the α-helices or near the β-sheet (Fig. 7). These make many contacts with lateral or axial neighbours. The detailed pattern of interactions for TMV *vulgare* (Mondragon, 1984; Mondragon & Bloomer, unpublished results) suggests that many features of the interfaces are likely to be conserved amongst these four viruses.

(i) *Pattern '1111222'*

Five of these residues are identical in the other three viruses SHMV, HR and CGMMV-W. The

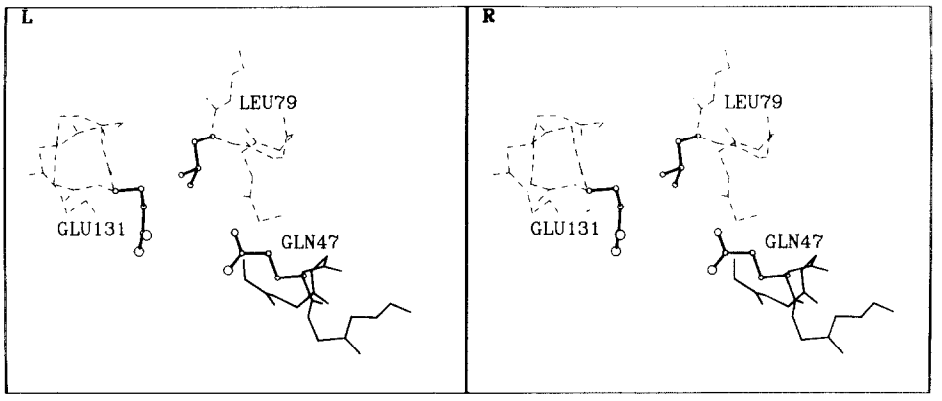


Figure 6. Stereo view of an axial proximity of 3 residues in the virus structure of TMV *vulgare*. The view is perpendicular to the axis of the viral helix and approximately along the long axis of the molecule towards the centre, showing Gln47 of one subunit (*B*: backbone continuous line) with Glu131 and Leu79 of the subunit above to the left (*A*: backbone broken line).

side-chains of residues 42 and 45 are close together at the surface of the α-helix and can form an intramolecular salt-bridge in SHMV, HR and CGMMV-W, whereas residues 13 and 56 are in van der Waals' contact at the distal end of the α-helices.

(ii) *Pattern '1111223'*

The only two residues with this pattern (Glu22 and Arg134) form an axial salt-bridge in the TMV *vulgare* disk. In CGMMV-W, the change to Thr22 and Lys134 will give an altered hydrogen bonding pattern. However, in SHMV and HR, the presence of Pro22 may disrupt the structure of the first long α-helix.

(iii) *Pattern '1111232'*

Four residues are shared by HR and CGMMV-W. Three of these (15, 28 and 72) are located at intermediate radius within the molecule and form lateral contacts between subunits in the TMV *vulgare* disk (Fig. 8). Residue 133 (pattern '.111232') has also been included with this group in Figure 8 because the substitution of Ile (in TMV) by Val (in DAHL virus, U2 and ORSV) is "conservative" on the basis of both shape and chemical character.

(iv) *Patterns '1111233' and '1111234'*

The remaining residues identical in only TMV, DAHL, U2 and ORSV include positions 71, 77 and 126. These may form contacts at the interface between three subunits in the virus (Fig. 9), together with position 50 (Glu or Asp in the first 4 sequences). A charge interaction such as this one is thought to play a crucial role in controlling the assembly of TMV (see section 5(e)(ii), below), although the precise identity of all of the residues involved has changed since the model was based on unrefined disk co-ordinates (Bloomer *et al.*, 1981), to

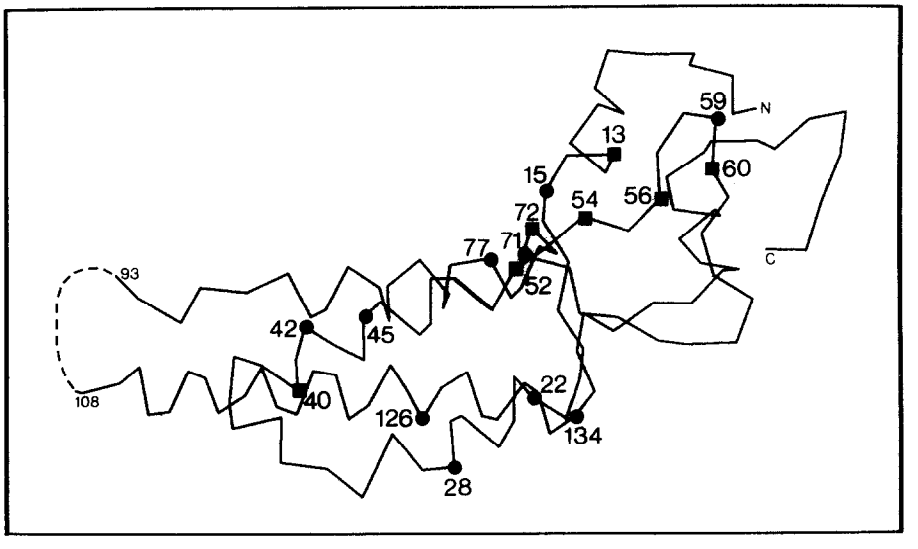


Figure 7. Residues common to only the 4 most similar viruses (in addition to the invariant residues shown in Fig. 3). The α-carbon chain tracing of 1 subunit is viewed down the disk axis, showing the locations of (■) hydrophobic and (●) hydrophilic residues that are identical in only the 4 viruses TMV *vulgare*, DAHL, U2 and ORSV.

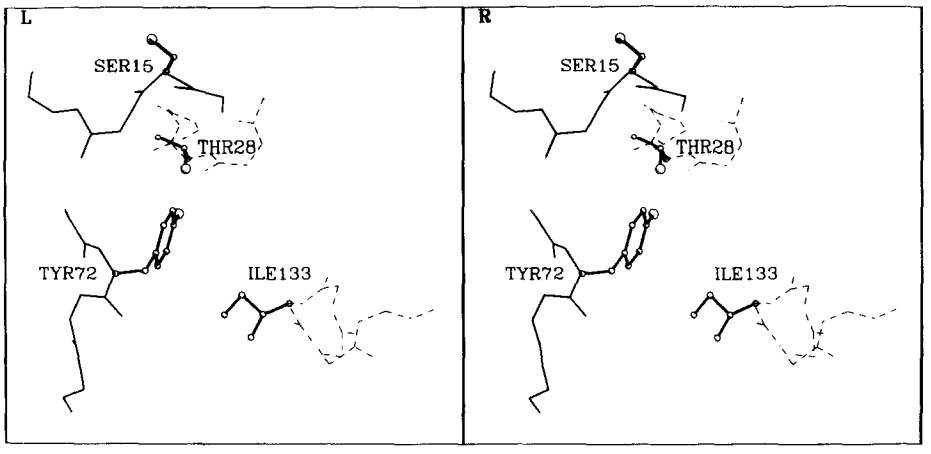


Figure 8. Stereo view of a lateral interaction in the disk of TMV *vulgare*. The view is as in Fig. 4 and shows residues 15 and 72 of one subunit (*A*: backbone continuous line) and residues 28 and 133 of the laterally adjacent subunit (*C*: backbone broken line).

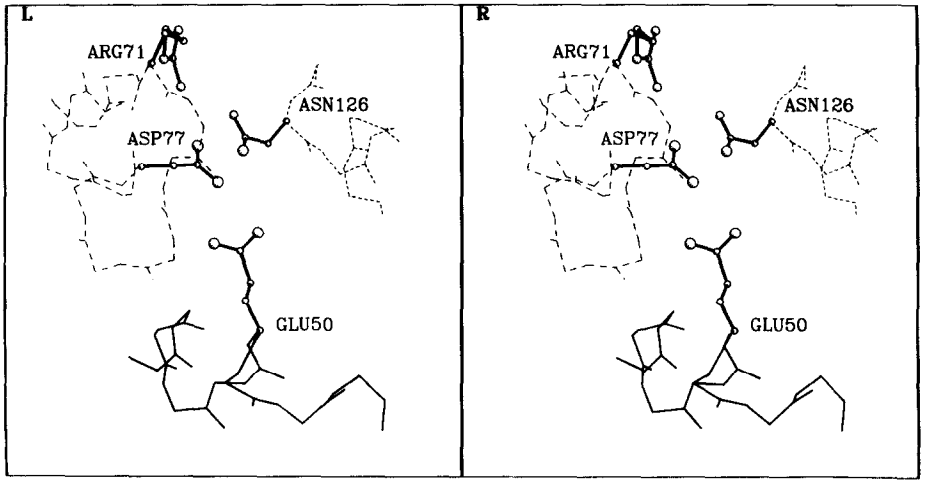


Figure 9. Stereo view of an axial proximity of 3 subunits in the virus structure of TMV *vulgare*. The view is as in Fig. 6 and shows: Glu50 of one subunit (*B*: backbone continuous line), Asn126 of the subunit above right (*C*: backbone short broken lines), and Arg71 and Asp77 of the subunit above left (*A*: backbone long broken lines).

the present model based on fitting to the fibre diffraction data (Mondragon, 1984) and on further refinement to improved fibre data (Namba & Stubbs, 1986).†

The role of such an interaction between charged and polar residues from three subunits is discussed below (section 5(e)(ii)).

(e) *Other groups*

The other groups shown in Table 4 yield few significant spatial relationships. For the seven residues of pattern '1111112', which are identical in six viruses and differ only in CGMMV-W, and the further three (pattern '1112113'), which may belong to this group when the ORSV sequence is fully known, only residues 4 and 9 are close together. The change of Ile4 to Pro4 may reduce the flexibility of this region of the polypeptide sufficiently that the hydrogen bond between the main-chain of residue 5 and the amide carbonyl group of Gln9 is no longer necessary to maintain the molecular conformation. When this residue becomes Lys9 it retains the possibility of making the second hydrogen bond observed in TMV *vulgare* from the amide nitrogen of Gln9 to the main-chain carbonyl of residue 148. This is serine in four sequences but becomes Phe148 in CGMMV-W when residue 9 is different. However, the deletion(s) observed in this region for both ORSV and HR indicate rearrangement of the backbone conformation from that seen in TMV *vulgare* where the hydrogen bonds from Gln9 were identified (Mondragon, 1984).

The group of residues for which SHMV differs

† Since this paper was submitted for publication, further results from the fibre diffraction analysis of the intact virus of TMV *vulgare* have been published (Namba & Stubbs (1986). *Science*, **231**, 1401–1406). Only a preliminary consideration of this paper is given here, however, until the further refinement of their model has been completed.

from the other six viruses has two members (122 and 125) on adjacent turns of the long α-helix. The change from Arg122 to His is discussed elsewhere in connection with the inter-subunit salt-bridge. The change from Ile125 to Leu122 is likely to give better packing of the hydrophobic parts of these side-chains.

None of the other groups listed in Table 4 shows any special features, even those where the residues are common to five of the seven viruses. Amongst the patterns seen for only one or two residues, which are not shown in Table 4, some conclusions can be drawn from the existing structural data.

(1) Arg141 and Asp64 form a salt-bridge in TMV *vulgare*, which is conserved in SHMV and CGMMV-W, but not in DAHL, U2 and ORSV. In HR the salt-bridge is between Arg141 and Glu64′ (where the prime indicates that an additional residue has been inserted at the preceding position). The loop of residues 64 to 66 is one of only two places where insertions occur in these sequences and thus the exact positioning is unclear (64′ or 65) for the aspartate residue in DAHL. There is no corresponding basic residue in DAHL for a homologous salt-bridge.

(2) When Trp17 in the protein interior is replaced by Tyr, this is accompanied by a change from a basic residue at position 68 to a tyrosine. Provided that the side-chain orientation of residue 68 is changed, this aromatic group would compensate for the volume change caused by loss of the Trp17. Residues 17 and 68 are the only two, apart from the 25 totally conserved residues, which are identical in the three viruses U2, SHMV and CGMMV-W.

5. Implications for the Function of TMV Protein

(a) *Hydrophobic core*

(i) *Buried residues*

The 35 buried residues of TMV coat protein (accessible surface area less than 20 Å^2) are indicated in Table 2. Seven are absolutely conserved (2, 41, 62, 83, 117, 128 and 144) and six conserve an identical side-chain volume (30, 35, 70, 76, 125 and 132) (Chothia, 1984).

Among the other residues, four are conserved (or semi-conserved) except in CGMMV-W. These are 18 (Ala and Val), 69 (Val or Ile and Ala), 121 (Ile and Ala) and 137 (Gly and Asp). This virus is the only one which seems to differ significantly from TMV *vulgare* in the protein core. The end of the side-chain of some variable residues is near the surface: Gln9, Ala16, Val51, Asn73, Ala124. Two residues are in a loop near the viral surface which undergoes rearrangement (deletions) in different viruses (148 and 150). Mutations in other residues seem to be complementary (e.g. 17 and 68) and several have been described above, e.g. 44, 48, 27 and 87; 13 and 56.

(ii) *α-Helix interfaces*

The residues which constitute the α-helix interfaces are shown in Figure 10, which distinguishes between those residues that interact with one or both of the neighbouring α-helices within the four-helix bundle. They are either conserved, or maintain a similar side-chain volume, or show complementary volume changes. The residues that are most conserved are those which are in contact with residues of two other helices. This strong conservation of the side-chains in van der Waals' contact at the α-helix interfaces suggests that the detailed geometry of these tertiary structural features of the protein is maintained.

Residues buried in the TMV *vulgare* structure and which change volume by large amounts in other viruses may cause structural changes that would disrupt either the folding of the monomer, or its assembly. Some buried residues in the helix bundle mutate in SHMV and CGMMV-W, but these mutations are not random and seem to maintain a constant volume of the core. The hydrophobicity of protein cores is usually well-conserved (Perutz *et al.*, 1965; Bajaj & Blundell, 1984). However, it is known that monomeric proteins accommodate mutations that change the size of buried residues by shifts of elements of secondary structures (Lesk & Chothia, 1980; Chothia & Lesk, 1984).

(b) *RNA binding site*

Because the refined structure is that of the disk without RNA, and because the region 94 to 105 is disordered in the disk, no detailed information is yet available on the RNA binding site. However, most of the absolutely conserved charged residues are in this region (see section 4(a), above) and some are believed to interact directly with RNA *via* the phosphate groups for arginines 41, 90 and 92 (Mondragon *et al.*, 1985) and *via* the ribose rings for aspartates 115 and 116 (Holmes, 1979). The precise role of the other invariant residues in this region has still to be determined.

(c) *Lateral contacts*

The lateral contacts are expected to be identical in the disk and in the virus except at low radius. The residues which lose more than 20 Å^2 of their accessible surface area, on side-by-side interaction in the disk, are listed in Table 5. It is clear from this Table that only the four viruses TMV, DAHL, U2 and ORSV have a conserved lateral interface. Some features of the high and low radius contact regions are conserved among all viruses.

(1) The residues that are part of the hydrophobic girdle at high radius (Fig. 2(a)) conserve a hydrophobic character, but some residues change in size: 10 (Phe, Tyr, Leu), 20 (Pro, Tyr, Val), 135 (Gly, His).

(2) At low radius, a cluster of conserved residues, mostly charged or polar, may play a role in both

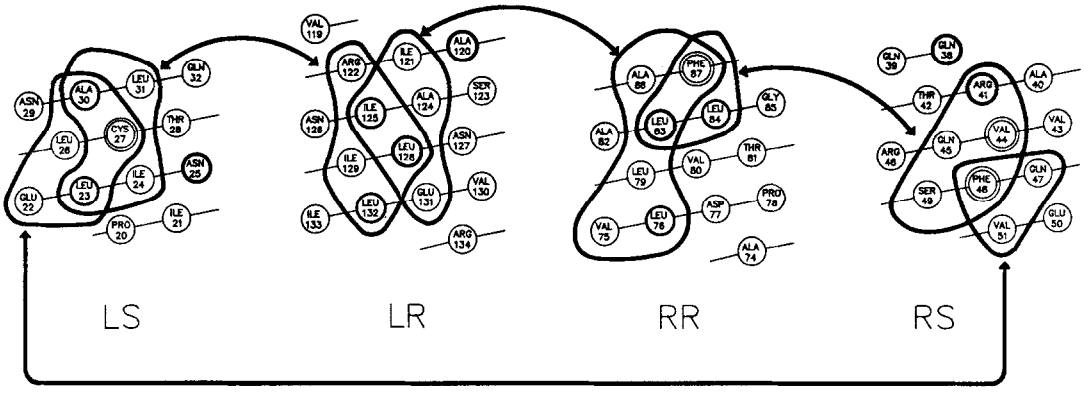


Figure 10. Schematic drawing of the intra-subunit contacts between α-helices in TMV protein. Residues within each patch are in van der Waals' contact with those residues within the patch on a neighbouring α-helix. The contacts between α-helices are indicated by arrows. Residues totally invariant in the 7 viruses are marked with the thickest circles; those which are semi-conserved are marked with intermediate circles. Double circles denote the group of residues (27, 44, 48, 87) with complementary volume changes (section 4(c)(i)). The *vulgare* sequence is illustrated. The low radius end of each α-helix in this alternating parallel and anti-parallel α-helical bundle is at the top of the diagram.

protein–RNA and protein–protein interactions. (Gln36, Thr89, Arg90 and 92, and Asp115).

(3) At intermediate radii, the only invariant residues are Asp88 and Asn25. Asp88 forms an inter-subunit salt-bridge and Asn25 forms a hydrogen bond with the main-chain of a neighbouring subunit.

(4) The other contact residues are generally conserved in the four most closely related sequences, but highly variable in SHMV, HR and CGMMV-W.

This variability could mean either that different viruses use different contacts or that the high and low radius interactions are sufficient for the specificity of assembly. Since mixed aggregation experiments have shown that TMV and DAHL can co-assemble (Otsuki & Takebe, 1978; Sarkar, 1960) but not TMV and U2 (Taliansky *et al.*, 1977) or TMV and HR (Sarkar, 1960), it is likely that the different viruses use contacts that are different enough to be mutually incompatible. This view is supported by the fact that clusters of residues in the same neighbourhood at subunit interfaces mutate together: residues 7 and 8 with 135 (Fig. 4): residues 15 and 72 with 28 and 133 (Fig. 8).

(d) *Axial contacts*

The axial contact residues in the disk are extremely variable from one virus to the other. The more extensive axial contacts in the virus can only be inferred from the presently available crystallographic data. Contact residues which might be important can be deduced from sequence comparisons, e.g. Glu22 and Arg134 form a salt-bridge between subunits in the TMV *vulgare* disk, which is expected to be maintained in the four most closely related viruses, and residues 131 and 47 might form an axial contact in the helical form of these same four viruses and also in HR.

(e) *Charged groups*

(i) *Salt-bridges*

The intra-subunit salt-bridge Glu145–Arg61 is the only one conserved in all tobamoviruses. The lateral salt-bridge Asp88–Arg122 is conserved in all viruses except SHMV, where Arg is replaced by His, but this change is accompanied by the change from Leu31 to Arg31, which will be involved in this salt-bridge *via* its side-chain, rather than *via* main-chain only as for Leu31. A certain number of salt-bridges are conserved in TMV, DAHL, U2 and ORSV. These are the intra-subunit salt-bridges Glu131–Arg134, Asp77–Arg71, Glu22–Lys53 (except in ORSV) and, in the disk, the inter-subunit axial salt-bridge Glu22–Arg134 (Mondragon, 1984).

The charged residues are most variable in the three distantly related viruses. An attempt to pair the new charged residues of the different sequences suggests that they may be part of a complex hydrophilic interaction between residues at the distal end of the α-helices. This interaction in the virus (Fig. 9) could include residues of the region 43 to 50 of one subunit (B), 71 to 77 of the subunit above left (A) and 126 to 134 of the subunit above right (C) as viewed from the outside.

Since there are charged residues on the surface of the protein which are buried in the virus, salt-bridges have to be made in virus assembly. The variability of these residues between the seven sequences suggests that these viruses achieve the specificity of assembly in different ways. No detailed model can yet be proposed because the side-chains of these charged residues can be oriented in many different conformations.

(ii) *The question of carboxyl–carboxylate pairs*

Figure 9, which has been derived by applying a rigid body transformation to the co-ordinates of a subunit in the disk, shows a contact between three subunits involving Asn126 of the upper subunit

together with Asp77 and Glu50, one from each of the two lower subunits.

There is the possibility here of forming a carboxyl–carboxylate pair with an abnormal pK value such as those postulated by Caspar (1963). Caspar found two abnormally titrating groups near to pH 7. This was confirmed by measurement of the difference between the titrations of the locked disk and the helix by Butler *et al.* (1972), who showed a total difference of two protons. One candidate long suspected as being the second possible carboxyl–carboxylate pair is the invariant pair Asp115 and Asp116, which are naturally close together, but these may be involved in interacting with the ribose groups of the RNA (Holmes, 1979). Another possibility for the second pair, if there are discrete rather than distributed effects, has recently been identified by Namba & Stubbs (1986) as the pair Glu95 and Glu106.

6. Discussion

The positions in the sequence that show an identical pattern of variation in seven related viruses are mainly close together in the three-dimensional structure. This strongly suggests that the residues in these positions have not become stabilized independently of each other, and that their location is linked to a particular function (e.g. maintaining RNA binding sites or tertiary and/or quaternary interactions of the protein), at least in viruses identical to TMV *vulgare* at these residues. The most obvious examples are the residues forming salt-bridges in TMV *vulgare*.

In viruses that are different at these residues, it is likely that the same function is assumed by other types of amino acids because these alternative amino acids are generally found in more than one sequence. For example the same alternatives are found in the core of SHMV and CGMMV-W (27, 48 and 87 with almost the same at 44), in an axial contact for SHMV and HR (22 and 134), and in a lateral contact for HR and CGMMV-W (15, 28, 72 and 133), etc.

For these new types of amino acids, the relationship between residue structure and function is clear in the case of buried residues that show changes complementary in size. A link can also be found in some mutations involved in inter-subunit interactions. For example residues 15 and 28, which are part of a lateral contact, are either both hydrophilic or both hydrophobic in all sequences except that of SHMV. This virus has Asp15 and Ile28, but at position 72 is an arginine which could form an intra-subunit salt-bridge neutralizing both of these charges. Residues 131 and 47, which are part of an axial contact in the virus, are also both hydrophobic or both hydrophilic.

In other cases the new pattern of bonding of the side chains is not obvious. The spatial proximity of some residues with identical conservation pattern may be fortuitous. But also, the known rules for protein interactions and folding are too general even to predict how molecules of known structure will interact. In particular, protein chemistry cannot be described in terms of pairs of residues. A more subtle arrangement of many atoms gives to a surface region the properties which allow specific assembly with another surface.

The method used here relies on the validity of the sequence alignment and will not detect covariant residues which are at different positions of the sequence in different viruses. The covariant residues undetected can, of course, be looked for in the structure. In the case of these tobamoviruses, for example, the interface between three subunits in the virus at the distal end of the helices is conserved in the four viruses TMV, DAHL, U2 and ORSV, indicating that this contact is necessary for helix stability. These residues are variable in the three other sequences, but (except HR) maintain a hydrophilic character and the possibility of forming inter-subunit salt-bridges.

From this study, the evolutionary order for tobamoviruses is not obvious. For example, HR shares the protein core with TMV *vulgare*, one axial contact with SHMV, some contacts at the trimer interface with CGMMV-W, and has some unique lateral contacts.

We have developed a simple method of detecting possible complementary substitutions in proteins. This strategy may be particularly successful with TMV protein because a large proportion of the residues is involved in some function: protein core, RNA binding, axial and lateral interactions with other subunits in various types of aggregates. However, several potential applications might result from a systematic study of homologous sequences in other protein families: understanding of protein folding and interactions, choice of residue for site-directed mutagenesis and study of evolutionary relations.

We thank Drs C. Chothia and K. Nagai for helpful discussion. D.A. thanks the European Molecular Biology Organisation, and A.M.L. thanks the U.S. National Science Foundation (research grant PCM83-20171) for support.

References

Altschuh, D., Reinbolt, J. & Van Regenmortel, M. H. V. (1981). *J. Gen. Virol.* **52**, 363–366.

Bajaj, M. & Blundell, T. (1984). *Annu. Rev. Biophys. Bioeng.* **13**, 453–492.

Barker, W. C., Hunt, L. T., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Seibel-Ross, E. I., Elzanowski, A., Hong, M. K., Ferrick, D. A., Bair, J. K., Chen, S. L. & Ledley, R. S. (1986). *Protein Sequence Database of the Protein Information Resource, release 8.0*, National Biomedical Research Foundation, U.S.A.

Bloomer, A. C. & Butler, P. J. G. (1986). In *The Plant Viruses* (Van Regenmortel, M. H. V. & Fraenkel-Conrat, M., eds), vol. 2, pp. 19–57, Plenum Publishing Corp., New York and London.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Nature (London)*, **276**, 362–368.

Bloomer, A. C., Graham, J., Hovmoller, S., Butler, P. J. G. & Klug, A., (1981). In *Structural Aspects of Recognition and Assembly of Biological Macromolecules* (Balaban, M., Sussman, J. L., Traub, W. & Yonath, A., eds), Balaban ISS, Rehovot and Philadelphia.

Butler, P. J. G. (1984). *J. Gen. Virol.* **65**, 253–279.

Butler, P. J. G., Durham, A. C. H. & Klug, A. (1972). *J. Mol. Biol.* **72**, 1–18.

Caspar, D. L. D. (1963). *Advan. Protein Chem.* **18**, 37–123.

Champness, J. N., Bloomer, A. C., Bricogne, G., Butler, P. J. G. & Klug, A. (1976). *Nature (London)*, **259**, 20–24.

Chothia, C. (1984). *Annu. Rev. Biochem.* **53**, 537–572.

Chothia, C. & Lesk, A. M. (1984). *J. Mol. Biol.* **182**, 151–158.

Gibbs, A. J. (1977). *C.M.I./A.A.B. Description of Plant Viruses*, no. 184.

Gibbs, A. (1980). *Intervirology*, **14**, 101–108.

Hollings, M. & Huttinga, H. (1976). *C.M.I./A.A.B. Description of Plant Viruses*, no. 156.

Hollings, M., Komuro, Y. & Tochihara, H. (1975). *C.M.I./A.A.B. Description of Plant Viruses*, no. 154.

Holmes, K. C. (1979). *J. Supramol. Struct.* **12**, 305–320.

Holmes, K. C. & Franklin, R. E. (1958). *Virology*, **6**, 328–336.

Kassanis, B. & Varma, A. (1975). *C.M.I./A.A.B. Description of Plant Viruses*, no. 153.

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

Lesk, A. M. & Chothia, C. (1980). *J. Mol. Biol.* **136**, 225–270.

Matthews, R. E. F. (1985). *Annu. Rev. Microbiol.* **39**, 451–474.

Meshi, T., Kiyama, R., Ohno, T. & Okada, Y. (1983). *Virology*, **127**, 54–64.

Mondragon, A. (1984). Ph.D. thesis, University of Cambridge.

Mondragon, A., Bloomer, A. C., Gebhard, W. & Holmes, K. C. (1985). In *Proc. FEBS Congress*, pp. 363–373, VNU Science Press, Utrecht.

Namba, K. & Stubbs, G. (1985). *Acta Crystallogr. sect. A*, **41**, 252–262.

Namba, K. & Stubbs, G. (1986). *Science*, **231**, 1401–1406.

Oshima, N. & Harrison, B. D. (1975). *C.M.I./A.A.B. Description of Plant Viruses*, no. 152.

Otsuki, Y. & Takebe, I. (1978). *Virology*, **84**, 162–171.

Paul, H. L. (1975). *C.M.I./A.A.B. Description of Plant Viruses*, no. 155.

Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965). *J. Mol. Biol.* **13**, 669–678.

Richmond, T. J. (1984). *J. Mol. Biol.* **178**, 63–89.

Sarkar, S. (1960). *Z. Naturforsch.* **15b**, 778–786.

Stubbs, G., Warren, S. & Holmes, K. (1977). *Nature (London)*, **267**, 216–221.

Takamatsu, N., Ohno, T., Meshi, T. & Okada, Y. (1983). *Nucl. Acids Res.* **11**, 3767–3778.

Taliansky, M. E., Atabekova, T. I. & Atabekov, J. G. (1977). *Virology*, **76**, 701–708.

Van Regenmortel, M. H. V. (1975). *Virology*, **64**, 415–420.

Zaitlin, M. & Israel, H. W. (1975). *C.M.I./A.A.B. Description of Plant Viruses*, no. 151.

Edited by R. Huber