PROTEIN INTERFACES AND INTERSUBUNIT BONDING

The Case of Tomato Bushy Stunt Virus

Stephen C. Harrison, *Gibbs Laboratory, Harvard University, Cambridge, Massachusetts 02138 U.S.A.*

ABSTRACT An atomic model of the subunit of tomato bushy stunt virus (TBSV) has been constructed to fit an electron density map at 2.9 Å resolution. Subunit interfaces show networks of polar residues forming H-bonds and salt bridges. The way in which alternative specific bonding geometries are built into a contact are described.

INTRODUCTION

Our ideas about the construction of protein assemblies are based on the notion that the same noncovalent contacts between subunits will be used over and over again. This reuse necessarily leads to symmetry, as Crick and Watson (1956) first observed. Moreover, in many cases, it also leads to what Caspar and Klug (1962) called "quasi-equivalence" – essentially identical bonding even at contacts not related by the overall symmetry of the structure. Such cases arise when the assembly is large enough that several chemically identical units comprise a single asymmetric unit of the complete structure. It can be seen that the plausibility of quasi-equivalence rests on the view that globular proteins must fold in a single, relatively precise way and that contacts between them must likewise be well determined in order to conserve specificity.

Icosahedral viruses are the classic examples of biological structures with several chemically identical subunits in the asymmetric unit. Caspar and Klug (1962) showed that certain designs permit conservation of noncovalent contacts, provided that these contacts, or the units themselves, have some degree of flexibility. Their argument may be summarized by reference to Fig. 1. The structure shown in the lower part of the figure contains sixty comma-shaped units, icosahederally related. Each comma makes three sorts of nearest-neighbor contacts, corresponding to the three kinds of rotation symmetry elements present: head-to-head contacts across a twofold axis, head-to-tail about a threefold, tail-to-tail about a fivefold. The structure in the upper part of the figure contains 180 units, with contacts that at first appear identical to those in the 60-unit shell. Closer inspection reveals that some of the tail-to-tail contacts are in rings of five, while others are in rings of six. The basic similarity of the contacts is preserved, and hence the fundamental biochemical specificity maintained, but distortions (e.g., arranging six tails instead of five around a particular axis) must be introduced to accommodate more than sixty subunits. The icosahederal symmetry of course implies that subunits must be strictly equivalent in sets of sixty; members of a given set must be flexed and bonded to their neighbors in exactly the same way, but members of different sets are only quasi-equivalent to each other. Caspar and Klug (1962) showed that certain (but not all) multiples of sixty subunits could be accommodated in such designs, the simplest being the 1 × 60 and 3 × 60-subunit arrangements just described. The permitted multiplicities can be

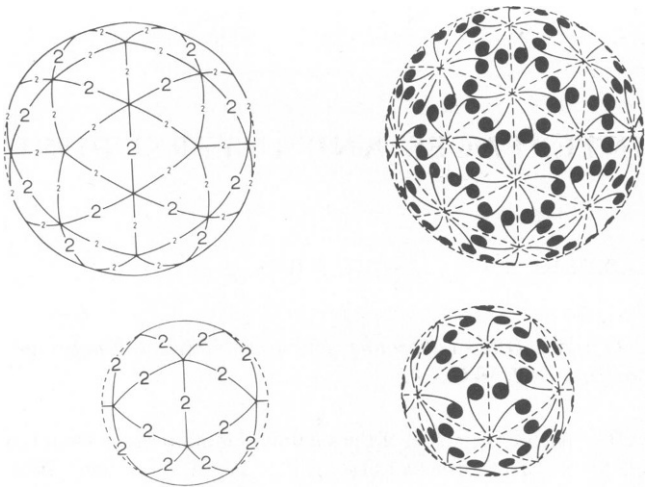

Figure 1 Icosahedral surface lattices, showing the packing of 60 strictly equivalent units in a T = 1 lattice (bottom) and 180 quasi-equivalent units in a T = 3 lattice (top). The left-hand figures show twofold symmetry axes; large numbers indicate strict dyads and small numbers indicate local ("quasi") dyads.

derived by considering a planar, hexagonal (p6) net as the limiting case of a very large shell. Introduction of curvature into such a shell is equivalent to transforming sixfold contacts into fivefolds, spaced at regular intervals so as to form a closed, icosahederally symmetric structure. Transforming adjacent sixfolds yields a simple, 60-subunit ("triangulation number", T, equal to one) design; transforming only second nearest-neighbor sixfolds yields a 180-subunit (T = 3) structure, and so on. Caspar and Klug (1962) argued that the deviations in bonding relations within a shell required to produce such structures would be within a range considered reasonable for interlocking amino-acid side chains. Their picture holds for shells of

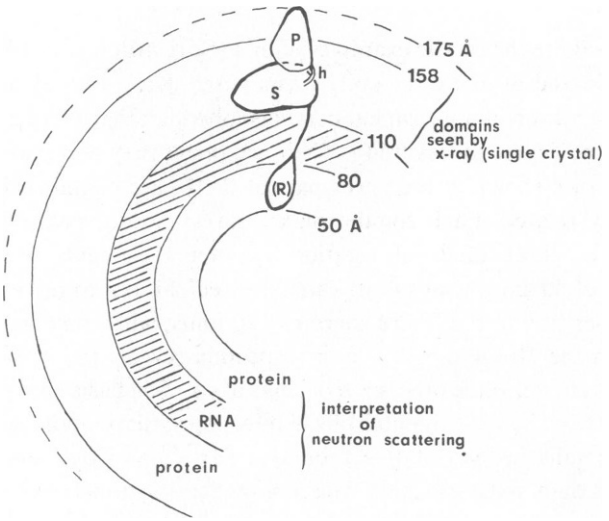

Figure 2 Radial organization of the TBSV particle, indicating distribution of protein and RNA. The "R domain" is inferred from the x-ray chain trace and the known chain length. Its location is implied by the results of neutron scattering (Chauvin et al., 1978) and small angle x-ray scattering (Harrison, 1969) although not all R-domains may lie in this shell. We do not know whether this portion of the chain is precisely folded, although several lines of evidence suggest that it is.

restricted radial extent, where variations in curvature changes local angular relations but not the distance between subunit surfaces. A similar argument led Pauling (1953) to predict that helically assembling subunits might form polymorphic structures.

The fact that all simple, isometric viruses conform to one of the designs predicted by Caspar and Klug (1962) shows the fundamental validity of their argument. At a more detailed level, it has remained for a high-resolution view of such a structure to determine how in practice such flexibility is achieved, to what extent the precise conservation of contacts is maintained, and how the ambiguity introduced by flexibility is avoided when assembling the particle. Such a view of tomato bushy stunt virus (TBSV) is now at hand (Harrison et al., 1978), and this paper presents a first analysis of the subunit interfaces in TBSV, with these questions in mind.

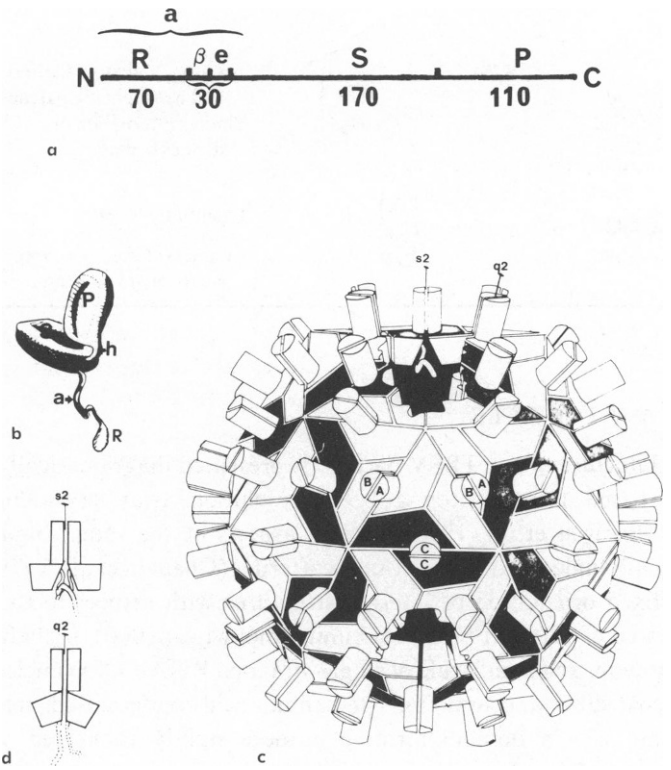

Figure 3 Architecture of the TBSV particle. (*a*) Order of domain in the polypeptide chain N terminus to C terminus. Approximate number of residues in each segment is indicated below the line. The letters indicate: arm (a), divided into R domain (possible RNA binding region), β-annulus, and extended arm; S domain; hinge; P domain. (*b*) Schematic view of the folded polypeptide chain. (*c*) Arrangement of subunits in the particle. A, B, and C denote distinct packing environments for the subunit; outer surfaces of C-subunit S domains are shaded. S domains of A subunits pack around fivefold axes; S domains of B and C alternate around threefolds. Examples of such axes are indicated by numerals 5 and 3. The local threefold axis (q3) relating S domains of an ABC trimer is nearly parallel to the adjacent strict twofold (s2), across which P domains of C subunits are paired. Trimers present a rather flat surface across the strict dyad and a distinctly sharper dihedral angle (~40°) across the quasi dyad (q2). A and B P-domains, paired across q2, therefore have a hinge angle with respect to their S domain that differs by ~20° from the angle on C. (*d*) The two principal states of the TBSV subunit viewed as dimers about s2 and q2. Subunits in C positions have the interdomain hinge "up" and a cleft between twofold related S domains, into which fold parts of the N-terminal arms. Subunits in quasi-twofold related A and B positions have hinge "down,"domains abutting, and a disordered arm.

TABLE I
INTERSUBUNIT CONTACTS IN TBSV

Class of contact	Domains participating	Code in Fig. 3	Packing details
Twofold			
Strict	P/P	D_{CC}	Extensive hydrophobic contact, conserved at strict and local twofolds
Quasi (local)		D_{AB}	
Strict	S/S	D_{CC}	Divided: Antiparallel helix/helix contact; alternating polar and hydrophobic; arm participates in divided D_{CC} contact
Quasi (local)		D_{AB}	Direct:
Quasi-threefold	S/S	T_{AC}, T_{CB}, T_{BA}	Conserved (all three identical); includes divalent cation sites
Fivefold/sixfold			
Fivefold	S/S	P_{AA}	Essentially identical, alternating polar and hydrophobic patches
Direct quasi-sixfold		H_{CB}	
Divided quasi-sixfold		H_{BC}	Similar alternation of residue types; arm divides contact
Fivefold		P'_{AA}	Essentially identical
Direct quasi-sixfold		H'_{CB}	
Divided quasi-sixfold		H'_{BC}	Contacts shifted with respect to other two; relatively loose contact in both states

THE TBSV PARTICLE

The general architecture of the TBSV particle is presented diagramatically in Figs. 2 and 3. The picture is based primarily on a 2.9-Å resolution x-ray crystallographic structure determination (Harrison et al., 1978); certain aspects of the radial organization must be inferred from small-angle neutron and x-ray scattering (Chauvin et al., 1978; Harrison, 1969) due to lack of fixed orientation of internal structures with respect to the outer shell. The particle is composed of 180 coat protein subunits (mol wt ~40,000), probably one chain of an 80,000 mol wt protein, and a molecule of single-stranded RNA (4,800 nucleotides) (Ziegler et al., 1972). The coat subunit, containing ~380 amino-acid residues, folds into three regions: a projecting domain (P), a domain forming a more tightly connected shell (S), and an N-terminal arm (a). The three symmetrically distinct environments for this subunit are denoted A, B, and C. The polypeptide accommodates to these three packing positions by flexion at the hinge between S and P and by an ordering or disordering of part of the arm. Units at positions A and B (60 of each) have one hinge configuration, and the entire N-terminal region appears to be spatially disordered. Subunits at positions C (60 in all) have another hinge position, and that part of the arm designated "β,e" in Fig. 3 *a* is folded in an ordered way along the bottom of the S domain. The portion marked "R" is spatially disordered in C subunits as well as in A and B. "Spatially disordered" is used here to mean that these regions adopt enough distinct positions (at least three) with respect to the fixed S and P domains, and hence to the crystal lattice, that they appear "smeared out" in the x-ray electron density map. The packing is tight enough that these parts of the molecule are probably not actually moving about. All the RNA is spatially disordered in this sense.

This description of subunit packing in TBSV emphasizes that there are really only two large-scale conformational states of the subunit: C (hinge "up" and ordered arm) and A/B (hinge "down" and disordered arm). There are significant detailed differences between A and B, however, which are best described by considering "conformational states" of the interfaces and the extent to which side-chain conformation and interaction is conserved or not when comparing symmetrically distinct versions of chemically homologous contacts. Each local symmetry element gives a class of subunit contact, and it is useful to subdivide these classes further, according to the domains involved (Table I; Fig. 4). In general, contacts are either fully conserved (identical side-chain packing at quasi-equivalent positions) or distinctly altered. The fivefold/sixfold contacts (P and H in Fig. 4) are a good illustration. Pairwise contacts of the six S domains packed about a particle threefold axis alternate in character: C→B contacts (proceeding anticlockwise when viewed from outside the particle) are essentially identical to A→A contacts about a fivefold axis, while B→C contacts involve quite different interactions, due to intervention of the C-subunit arm. In the notation of Fig. 3, P_{AA} and H_{CB} are identical in local side-chain configuration and bonding and different from H_{BC}. As we show in more detail below, however, even H_{BC} retains a number of side-chain interactions present at H_{CB}. Note that the similarity of H_{CB} and P_{AA} implies that the H_{CB} face of C has A-like side-chain conformations, while the corresponding (H_{BC}) face of B has a

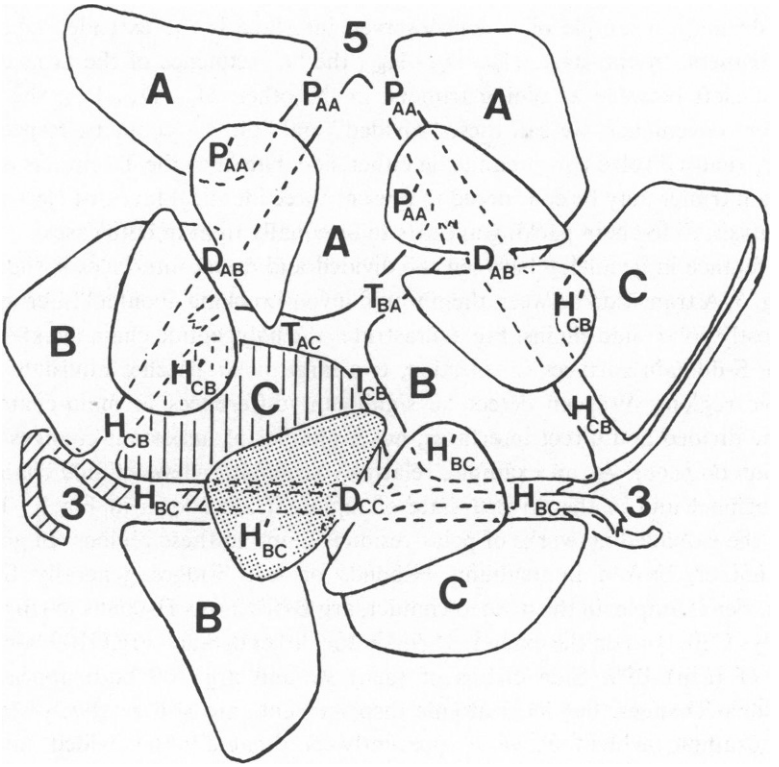

Figure 4 Diagram showing subunit contacts. Domains are indicated in outline. Barrel-shaped P domains would protrude outward from the page, making only twofold contacts, while the more extensive S domains make contacts across all adjacent symmetry axes. The contacts are lettered, as described in Table I. The large letters indicate the class of contact (D, T, P, H for dimer, trimer, pentamer, hexamer, respectively), and the subscripts indicate participating subunits.

somewhat different detailed arrangement. It is at this level that A and B represent different conformers of the protein subunit.

PROTEIN INTERFACES IN TBSV

Our analysis of the 2.9 Å resolution electron density map of TBSV has yielded an atomic model of the subunit, but our as yet incomplete knowledge of the amino-acid sequence[1] means that most of the side chains in this model are guesses based on the shape of density features and on chemical good sense. The map is clear enough that these guesses are likely to be reasonable: where chemical confirmation exists, we are correct ~60% of the time, most errors being "benign" in terms of our initial interpretation (ser for asn, etc.). The description that follows is therefore preliminary, but we have great confidence that the character of contacts derived from this analysis is correct, independent of specific assignments. (For example, since main-chain carbonyl groups are quite clearly defined, a side chain making an obvious H-bond to such a group must be polar. Clear networks of H-bonds involving main-chain as well as side-chain atoms establish the character of a large polar patch, even if individual residue assignments confuse asn for ser, asp for glu, etc.). Residue numbers below correspond to a count beginning at the first ordered residue on C. The entire "R" sequence is therefore excluded from the numbering.

S-Domain Contacts along the Trimer Boundary

The most dramatic example of a nonconserved interface is the extended contact between S-domain trimers. In one state, H_{BC}–D_{CC}–H_{BC}, the "e" sequence of the arms of C subunits folds into a cleft between adjoining trimers; in the other, H_{CB}–D_{AB}–P_{AA}, the trimers abut directly. For convenience, we call these "divided" and "direct" contacts, respectively. These interfaces actually involve two subunits on either side, but since the T contacts are conserved, the S-domain trimer may be considered to present three identical faces, at least at the present level of analysis. Side-chain packing appears to be equally tight in both cases.

The difference in geometry between the divided and direct interfaces is shown schematically in Fig. 5. A transition between them would involve rocking about a "fulcrum" formed by various, mostly polar, side chains. Fig. 6 illustrates the polypeptide chain backbone and shows that at the S-domain surfaces in question, two large polar patches alternate with smaller hydrophobic regions. We can detect no significant differences in main-chain coordinates between the divided and direct interfaces, but a number of important changes in side-chain conformation do occur. As an example, relatively schematic views of side chains in the two states of region I and in the divided state of region II are shown in Fig. 7. The drawings emphasize the extended networks of polar residues found in these regions. In going from one state to another, broken intersubunit H-bonds or salt bridges generally find an exact alternative. For example, in the divided contact, arg B109 forms H-bonds to (thr) C19[2] and to the CO of lys C20, both on the extended arm; in the direct contact, arg C109 bonds to CO and side-chain of (asn) B99. Side chains of (asn) 99 and arg 109 both appear to undergo conformational changes, but local atomic displacements are still relatively slight. Another interesting example involves the water apparently coordinated in the divided contact at region II by CO B131, CO C25, (thr) C27 and (thr) C42. In the direct contact, residues 25 and 27

[1]Sauer, R. Work in progress.

[2]Side chains known from chemical sequence data or confidently assigned from electron density are indicated by the usual threeletter code; side chains assigned from the map but with some uncertainty are given in parentheses. The numbering is based on model-building, but there are very few places where uncertainty about Cα and CO exists.

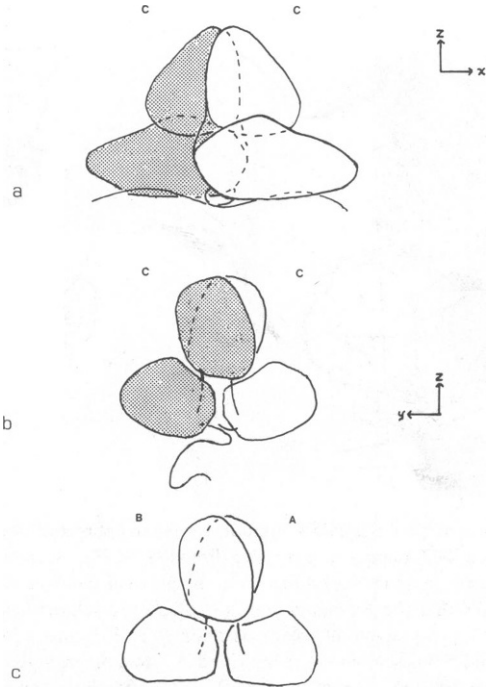

Figure 5 A C/C dimer is viewed along the direction from twofold axis to fivefold axes in *a* and from threefold to twofold in *b*. The latter corresponds to Fig. 3 *d*. The shape of the subunit along the cleft is reasonably accurately represented. In *c*, a corresponding view of an A/B dimer shows that the cleft closes down by rotating the S-domain contact about a fulcum.

"drop out," and CO C131 forms a direct H-bond with (thr) B42. In general we note the importance of arginine, tryosine, and bound water as "spacers," the prevalence of polar interactions, and the precise positioning of alternative contacts for the two states. A few side chains are ordered in one state and disordered in the other. A particularly striking example is arg 43 (residue "*x*" in Fig. 3 *a* of Harrison et al., 1978), ordered by H-bond to CO 23 of the arm in C subunits (divided interface) and barely discernable as weak density in A and B subunits (direct interface).

P-Domain Dimer Contacts

The face of the P-domain is such that one of the β-sheets presents a relatively flat surface for dimer bonding with a corresponding sheet (Fig. 6). The P-domain contact between A and B subunits appears to be identical to the corresponding C-subunit interface, and small-angle x-ray scattering suggests conservation of this contact in the expanded form as well.[3] Most of the buried side chains appear to be hydrophobic, with an external "rim" of polar interactions. The general character of bonding is thus quite different from S-domain interfaces, where striking buried polar patches appear to contribute critically to stability.

S-Domain Trimer Contacts: the Divalent Cation Sites

Contacts T_{AB}, T_{BC}, and T_{CA} are all well conserved in the TBSV structure, with no differences detectable at the present stage of analysis. An important heavy-atom site (UO_2^{++}) lies at this

[3]Robinson et al. Work in progress.

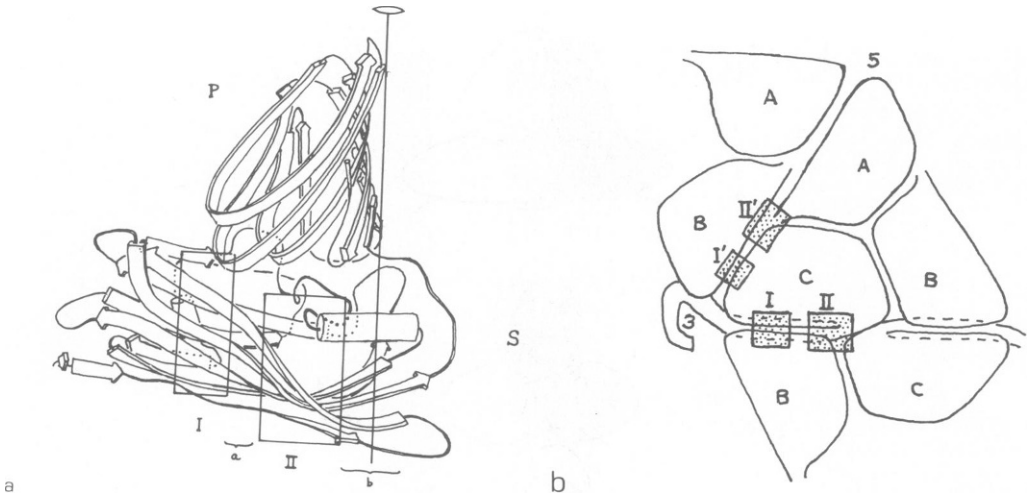

Figure 6 (*a*) "Ribbon diagram" of the TBSV subunit, viewed in a direction corresponding to the view of shaded subunit in Figs. 4 *a*, i.e., along *y* in a positive direction. In Fig. 4, the view would correspond to sighting the shaded C subunit in an upward direction in the plane of the page. The twofold axis is shown; note the extensive, flat face that the P domain presents to its dyad-related neighbor. The H_{BC} and D_{CC} surface is seen face on. It has two regions of polar side chains, I and II, separated by narrower columns of hydrophobic contacts, a and b. These regions are indicated in *b* on a diagram that corresponds to Fig. 4. I', II', a', b' are the same contracts in the "direct" state of this interface. Both S and P domains are essentially two-layer β structures, with "classical" hydrophobic interiors.

contact and nearby is a peak previously identified as a probable divalent cation ("Ca^{++} site"; Harrison et al., 1978). We have recently computed a difference map, using data collected from crystals soaked in 0.01 M EDTA at pH 5.5. The only significant features are two essentially equal troughs corresponding to the "UO_2^{++} site" and the "Ca^{++} site" respectively; these troughs are present at homologous positions at all three interfaces in the trimer. Thus two ions capable of chelation by EDTA appear to lie at each such contact. The occupancy is such that both could be Ca^{++}, and we believe these to be the sites involved in divalent-cation regulated expansion of TBSV[3]. Partial sequence information presently available shows two aspartate residues liganding one ion and at least one aspartate in contact with the other (Fig. 8). Ligands for the ions come from both subunit surfaces participating in the contact.

The trimer interface (Fig. 8) has certain features in common with the hex/pent surface described earlier. The two cations form part of an extensive network of negatively charged and polar residues, as well as what appear to be several structural water molecules. These networks are separated from each other by a set of hydrophobic interactions just around the local threefold axis. Although well conserved in the native TBSV structure, several lines of evidence indicate that these contacts move apart in the expanded form.

P-Domain/S-Domain Contacts

Interdomain contacts occur between domains of the same chain as well as between domains of adjacent chains (e.g., H'_{BC} in Fig. 4). These contacts appear as if designed to vary, being much less tightly packed with side chains than the interfaces described so far. At least one key arginine residue is involved, reaching out from the P domain to form H-bonds with S-domain atoms, in one case via a solvent molecule.

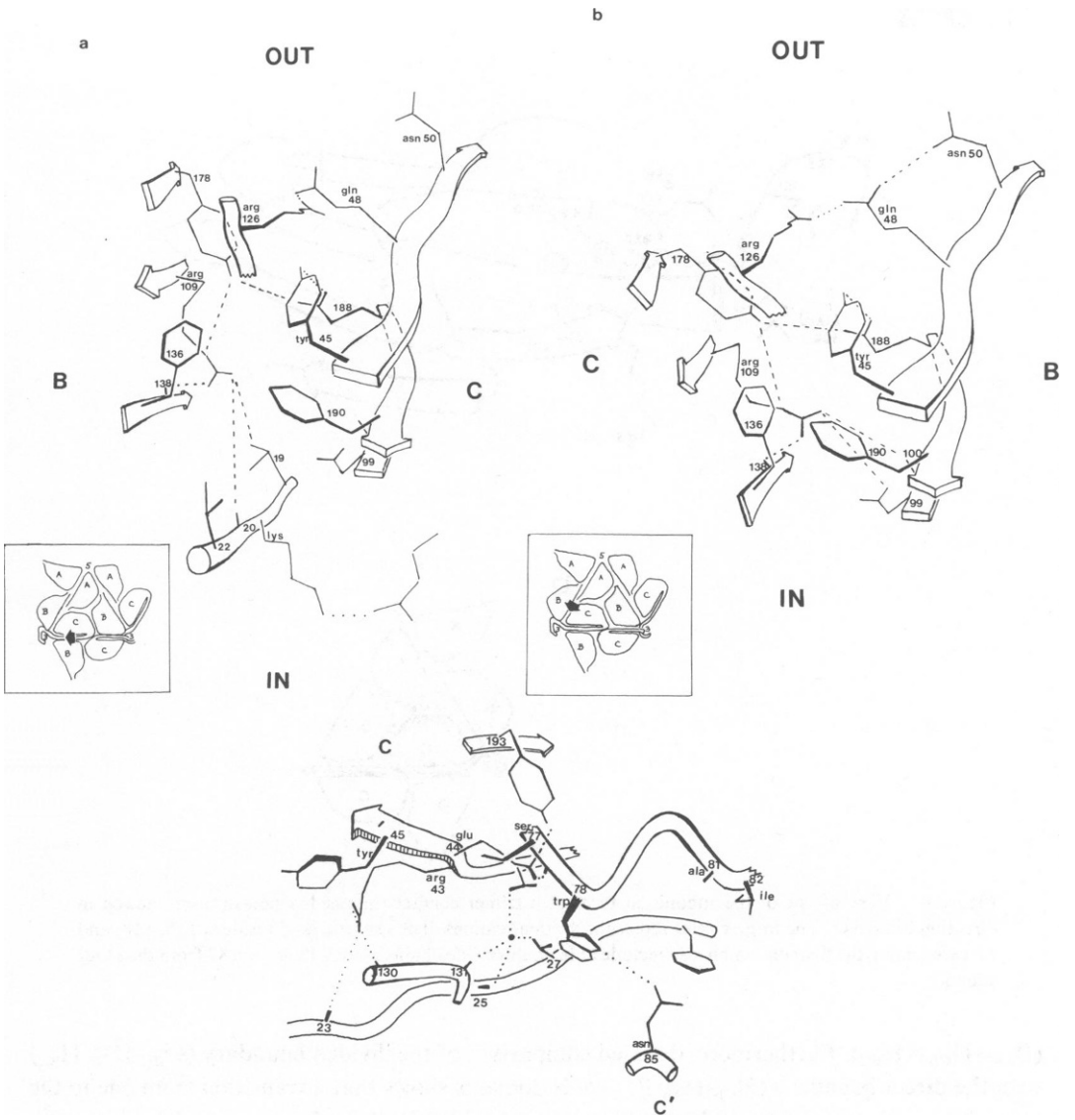

Figure 7 Side chains in regions I and II of the intertrimer interface (Fig. 6). Region I is seen in the divided contact in *a* and in the direct contact (I') in *b*. The view in *a* and *b* is along the interface, from the twofold (or quasi-twofold) toward the threefold (see arrows on insets). The divided contact at region II is shown in *c*, viewed from outside the particle looking inwards. Note the tetrahedrally coordinated bound solvent molecule. Numbers indicate sequence position, beginning at the β-annulus. The ~70 residues of R are not included in this count. Residues know from chemical sequence data or from confident assignment from the electron density map are labeled.

QUASI-EQUIVALENCE

How does the intersubunit bonding in TBSV achieve a compromise between conservation of detailed noncovalent contacts and inevitable nonequivalence, and to what extent does the compromise fit Caspar and Klug's original description (1962) of "quasi-equivalent" bonding? We observe that there is indeed extensive conservation, and, as just described, the geometry is such that even where a contact must change, it appears to have two rather than three states

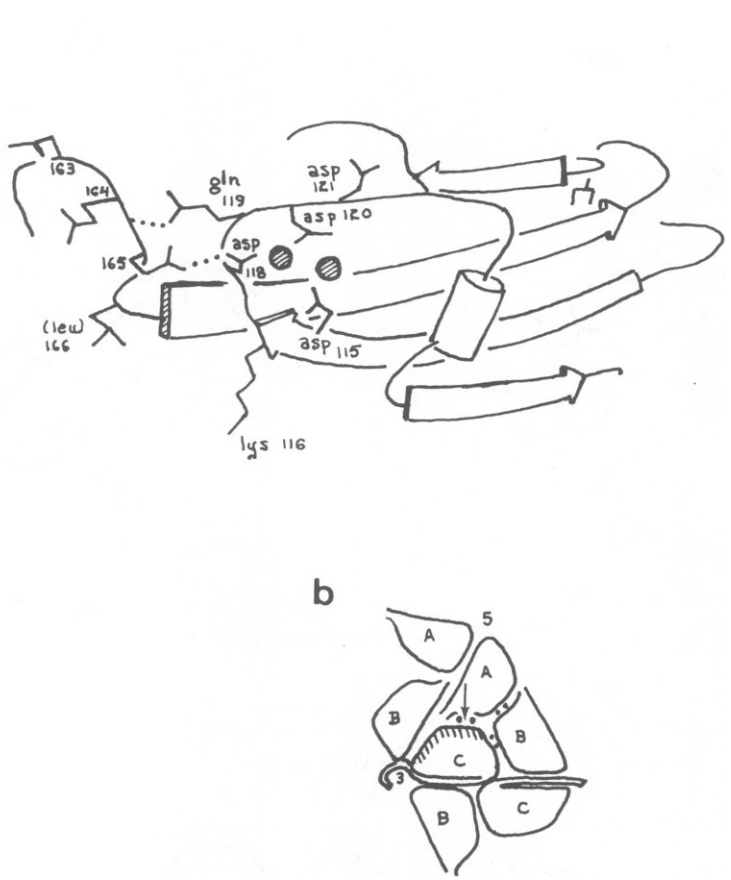

Figure 8 View of one of the subunit surfaces at a trimer contact (hatched region in inset, viewed in direction of arrow). The large circles represent divalent cations. The aspartic acid residues 115, 118, and 120 are among the ligands, which also include at least three side chains among them, asp 87 from the other subunit.

(P_{AA}–H_{CB}≠H_{BC}). Furthermore, detailed comparison of the divided boundary (H_{BC}-D_{CC}-H_{BC}) with the direct boundary (H_{CB}-D_{AB}-P_{AA}) of S domains shows that a transition from one to the other has something of the fulcrum-like character imagined by Caspar and Klug for such interfaces. The considerable internal flexibility conferred on a subunit by the interdomain hinge is not part of the original description. This does not, however, alter the fundamental idea of conserved, specific bonding; rather, it allows a much larger radial extent for these bonds by localizing some of the "give" at a site within the subunit itself. The consequence of an internal hinge is that contacts are not conserved between different domains of the same subunit.

A more fundamental departure from earlier descriptions of icosahedral structures is the interdigitation of C-subunit arms around each particle threefold axis. These arms extend along the H_{BC}-D_{CC}-H_{BC} cleft, describing as they reach the threefold an inward-spiraling course (Fig. 9). Since three such arms reach each threefold, the spirals run parallel to each other, displaced by 120° (Fig. 9). The resulting structure, which we have called a "β-annulus" because of the permuted three-stranded β-sheets that stabilize it, is a remarkable bit of protein folding. Its function is to determine the triangulation number precisely. All sixty C subunits form a coherent network, laced together at threefold positions by β-annuli and at

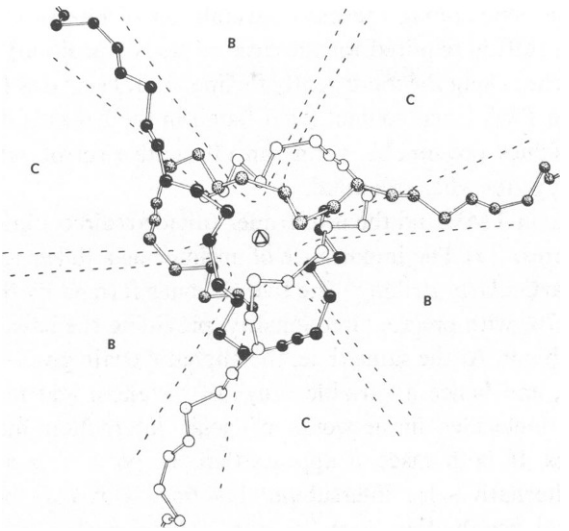

Figure 9 Configuration of interdigitated arms, the "β annulus," viewed down a threefold axis. The balls represent α-carbon positions. The residues here are the sequence "β" (Fig. 3 *a*) and a part of "e" extending into the intersubunit cleft.

twofold positions by P-domain contacts. The shell thus generated has gaps to accommodate sixty A and sixty B subunits, with unambiguous geometry. The arms of the C subunits form, in a sense, an internal T = 1 framework. It is likely that all self-assembling icosahedral structures will have such a T = 1 scaffold, for error-free size determination. A number of the T = 3 viruses have, like TBSV, a rhombic triacontrahedral packing; the quasi-threefold is parallel to a nearby twofold, generating particles that bulge near the fivefold axis. A cleft along the line joining particle threefold axes (H_{BC}-D_{CC}-H_{BC} in Fig. 3) must therefore occur, and it is reasonable to suggest that extended arms, interdigitating to some extent at the threefold axes, will also be present. Rossmann and colleagues[4] have already found such a structure in SBMV. Shells of larger triangulation number may have a separate protein core as an internal scaffold rather than a flexibly tethered arm. The bacteriophages with a T = 7 surface lattice, (λ, P22, T7, etc.) assemble around a core that is subsequently eliminated from the particle to make way for DNA (Casjens and King, 1975). The tumor viruses (polyoma, SV 40) have, in addition to 420 copies of VP1 arranged in a T = 7 outer shell, ~60 copies each of proteins VP2 and VP3, probably internal (Murakami and Schaffhausen 1974). One or both of these might serve an analogous function. The situation with T = 4 particles is unclear, since relatively few such structure are known. Generation of such a lattice from a p6 net involves reducing the symmetry of a sixfold position to a twofold, and one might therefore imagine that a twofold interlinking of arms from tight, threefold-related clusters could provide a framework analogous to the C-subunit network in TBSV.

VARIABLE AND CONSERVED CONTACTS: SOME COMPARISONS

The networks of polar interactions at variable interfaces in TBSV are not peculiar to this structure. In tobacco mosaic virus (TMV) disk, the protein shows an extensive chain of salt

[4]Personal communication.

bridges and hydrogen bonds along the most variable set of intersubunit contacts (between layers, where a 10-Å shift is required for the disk → helix transition) and alternating polar and hydrophobic patches along the more gently flexing lateral surfaces (Bloomer et al., 1978). The other state of the TMV axial contact (that found in the helix) is not yet known, but the array of charged residues presumably forms an alternative set of salt links and H-bonds, including or expelling water where required.

The axial interface in TMV and the intertrimer (divided/direct) interface in TBSV share several detailed features. (*a*) The importance of arginine as a linker residue (and to a lesser extent, tyrosine) is particularly striking. The five H-bonds formed by the guanidinium group fan out over nearly 270° with precise directionality, providing the possibility of rigid docking to sites on another subunit. At the same time, the aliphatic chain gives several possibilities for rotational isomerism, and hence a variable range of distances and target angles. (*b*) The inclusion of solvent molecules in networks of polar interaction may assist in creating alternative geometries. In both cases, it appears that the polar networks are constructed to provide two exact alternatives for intersubunit bonding. This may be difficult to achieve without the additional possibilities from "spacer" solvent molecules. The presence of 13 waters at the $\alpha_1\beta_1$ (invariant) interface in Hb suggests, however, that inclusion of water may be a standard feature of protein interactions, whether or not multiple states are required (Ladner et al., 1977).

Many descriptions of bonding at subunit contacts emphasize the role of hydrophobic bonding (e.g., Chothia and Janin, 1975). It is true that even interfaces with numerous polar residues contain large numbers of nonpolar contacts. For example, aliphatic chains of arginine residues in TBSV clearly participate in van der Waals interactions, while the guanidinium group forms salt bridges (cf. arg 126 in Fig. 7). Nonetheless, it is hard to avoid the impression that some net thermodynamic stability might be generated by the H-bond, salt-bridge network. Conventional estimates of net stability of protein H-bonds with respect to bonds to water vary from 0 to 1.5 kcal/mol (Schellman, 1955; Klotz and Franzen, 1962). Thus the ~6 intersubunit H-bonds at the H_{BC} contact could alone contribute as much as 9 kcal to this interaction.

The rather uniformly non-polar character of the P-domain contact stands in striking contrast to other interfaces in TBSV. This dimer contact is probably also the only one conserved at all positions in native and expanded virus. Are "conventional" hydrophobic contacts less capable of useful alternative geometries than networks of polar bonds? Relatively few known structures can provide suitable comparison. The $\alpha_1\beta_1$ (variable) interface in Hb contains a significant number of polar side chains, including two arginines (αFG4; βC6), but it is unclear from existing descriptions to what extent the two defined states of this interface are determined by alternative hydrogen bonds and to what extent simply by alternative "knobs-into-holes" packing. One interesting similarity between the deoxy → oxy transition at $\alpha_1\beta_2$ and the divided/direct difference at the trimer surface in TBSV is a "flexible joint" or fulcrum that provides a relatively constant region of contact (Baldwin and Chothia, 1979). Such a relationship may be important if a transition involving a steric barrier is to occur without dissociation. In Hb, side chains of arg α92, pro α95 interlock with trp β37, arg β40, constituting a sort of pivot (Baldwin and Chothia, 1979); in TBSV, a direct → divided transition would involve a hinge along a series of contacts including tyr 178 → (glu)188;(phe)137, tyr 45;asn 85 → arg 78, as well as insertion of the arm. We do not know whether this motion actually occurs (in a dynamic sense) in TBSV assembly, but the interface does appear designed to do so without requiring complete loss of contact. Another parallel

between these regions of Hb and TBSV is the construction of discrete alternatives. In Hb, there appears to be a barrier to the 6-Å slide of βFG past αC, concentrated where side chains of his β97 must pass thr α41 (Baldwin and Chothia, 1979). In TBSV, the alternatives are determined by intercalation of the extended arm.

I am grateful to Arthur Olson, Robert Ladner, Ian Robinson, James Hogle, and Robert Sauer, who helped generate the information contained in this account, and to Don Wiley for many helpful discussions.
This work was supported by National Institutes of Health Grant CA-13202 and by an Alfred Sloan Fellowship.

Received for publication 7 December 1979.

REFERENCES

Baldwin, J., and C. Chothia, 1979. Haemoglobin: the structural changes related to ligand binding and its allosteric mechanism. *J. Mo. Biol.* **129**:175–220.

Bloomer, A., J. N. Champness, G. Bricogne, R. Staden, and A. Klug. 1978. Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunit.

Casjens, S., and J. King. 1975. Virus assembly. *Annu. Rev. Biochem.* **44**:555–611.

Caspar, D. L. D., and A. Klug. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**:1–24.

Chauvin, C., J. Witz, and B. Jacrot. 1978. Structure of the tomato bushy stunt virus: a model for protein RNA interaction. *J. Mol. Biol.* **124**:641–651.

Chothia, C., and J. Janin. 1975. Principles of protein-protein recognition. *Nature (Lond.).* **256**:705–708.

Crick, F. H. C., and J. D. Watson. 1956. Structure of small viruses. *Nature (Lond.).* **177**:473–476.

Harrison, S. C. 1969. Structure of tomato bushy stunt virus, I: The spherically averaged electron density. *J. Mol. Biol.* **42**: 457–483.

Harrison, S. C., A. J. Olson, C. E. Schutt, F. K. Winkler, and G. Bricogne. 1978. Tomato bushy stunt virus at 2.9 Å resolution. *Nature (Lond.).* **276**:368–373.

Klotz, I. M., and J. S. Franzen. 1962. Hydrogen bonds between model peptide groups in solution. *J. Am. Chem. Soc.* **84**: 3461–3466.

Ladner, R. C., E. Heidner, and M. F. Perutz. 1977. The structure of horse methaemoglobin at 2.0 Å resolution. *J. Mol. Biol.* **114**:385–414.

Murakami, W. T., and B. Schaffhausen. 1974. Protein composition of SV 40 and polyoma viruses. *In* Molecular Studies in Viral Neoplasia. Twenty-Fifth Annual Symposium, M. D. Anderson Hospital and Tumor Institute. Williams & Wilkins Company, Baltimore, 43–62.

Pauling, L. 1953. Aggregation of Globular Proteins. *Disc. Faraday Soc.* **13**:170–176.

Schellman, J. A. 1955. *C. R. Trav. Lab. Carlsberg Ser. Chim.* **29**:223–229.

Ziegler, A., S. C. Harrison, and R. Leberman. 1974. The minor proteins in tomato bushy stunt and turnip crinkle viruses. *Virology* **59**:509–515.

DISCUSSION

Session Chairman: David Eisenberg *Scribe*: Michael Lennick

BUTLER: Would you please comment on a related virus, turnip crinkle, which will form small, T = 1, particles in the absence of RNA. There, the form of the particle seems to be RNA dependent.

HARRISON: You're right that if you try to make shells out of TCV subunits without RNA, the best you can get is a T = 1 small particle (although it is possible that small particle formation requires proteolytic cleavage of the arms). The N-terminal part of the arm may have to interact with a nucleating sequence on RNA or with another minor protein bound to RNA in order to fold up correctly. This would be analogous to other systems where control by heterogeneous nucleation has evolved, such as flagella and phage tail. It would make sense in this case to ensure coating of RNA. The structural feature of the protein that would enable this control mechanism to dictate size is the correct interdigitation of the arms. We now know from a 5 Å map of TCV computed by Jim Hogle, that its structure is sensibly identical to that of TBSV.

J. KING Toward the end of the paper you describe the divalent cation location on the trimer positions. From looking at the model, can you see how the subunits would have to change for the shell to expand? Is it only one/three that has to change or does the whole thing have to go?

HARRISON: The expansion appears to be cooperative, jumping from the compact state described here to a reasonably well-defined expanded state with an increase of ~15% in radius. We do not yet have any estimate of how many of the 180 cation-liganding interfaces must be stripped for the expansion to occur. I can represent schematically a possible structure for the expanded particle, based on three assumptions: (*a*) the interfaces where divalent cations are bound move away from each other (in effect, are "blown apart" by clustered negative charge); (*b*) the divided and direct contacts about fivefold and quasi-sixfold positions are conserved, particularly including the β annulus and its relation to adjacent B and C subunits; (*c*) twofold clustering of P domains is conserved. A diagram of this expansion is shown in the figure where only S-domains are shown. An additional degree of freedom for the S-P hinge is required in order to position P-dimers over strict and local twofold axes relating S domains. This model conserves as many as possible of the bonding contacts found in native TBSV. We hope to know the real answer quite soon, from crystals of expanded TBSV being studied by Ian Robinson.

BLOOMFIELD: Quasi-equivalence costs you some energy to flip into the second configuration, but you gain that back in more complete binding to neighbors. Realizing from this meeting's "Forces" discussion how difficult it is to recognize what these energies are for subunit interactions, is it clear from the structures of the two forms of the protein that there are large sources of unfavorable energy in one relative to the other. Do they seem in general to be of equal energy?

HARRISON: There's no obvious difference to the crystallographic eye between the C/C dimer and the A/B dimer. Folding up the arm, which is really what makes a C/C dimer, doesn't seem favorable except in the completed assembly where it can make the β-annulus and contacts with adjacent B subunits.

BLOOMFIELD: I know you can't see water in your structure, but from the total intensity of the scattering is it possible to say what water content there might be in the protein shell or in the RNA region of the virus?

HARRISON: The interior of TBSV needs to accommodate the parts of the protein subunit that we don't see, and it needs to accommodate the RNA. If the RNA is hydrated at 1.5 grams of H_2O per gram of RNA and the protein is hydrated at 0.3 grams per gram of protein, then components will just fill the interior defined by observed inner surfaces of S domains. That's a reasonable figure for hydration of RNA, from various other measurements. Things are very compact inside. I should add that Chris Dobson has shown with ^{31}P NMR that relaxation of phosphorous nuclei is determined by tumbling of the virus.

MAKOWSKI: I'm fascinated by how these things put themselves together. You describe in your paper that the sixty C-subunits form a T = 1 "scaffolding." Do you know of any evidence for intermediates in the assembly of the whole virus? Do you mean to imply that the C subunits interact first?

HARRISON: No, I don't mean to imply that, because I just don't know. We do know that free TCV protein exists as a dimer in solution. Presumbably this is a P-clamped dimer which doesn't associate further; protein crosslinking doesn't reveal any further steps in the ladder beyond monomer and dimer. At the other extreme, we have the expanded and compact particle structures. What exists between is still unexplored.

MAKOWSKI: During the expansion, do the projecting domains remain in contact?

HARRISON: There is evidence from low single x-ray scattering that the pattern of icosohedral projections is maintained, so I believe the clustering is maintained. That retention requires another degree of freedom of the hinge. In order to remain positioned over the twofold axis the P domains have to change their positions over their corresponding S-domains. The P/S contacts look tenuous, like what Huber calls longitudinal contacts between domains in immunoglobulin, so more states besides the A/B and C may be possible.

POTSCHKA: Is the structure static or is there interconversion of the quasi-equivalent positions?

HARRISON: In the compact structure there is no interconversion. The β-annuli certainly don't unwrap and refold in alternative ways.

HENDRICKSON: You mentioned something about the Southern Bean Mosaic Virus lacking P domains. Do the other aspects of quasi-equivalence that you've seen in Bushy Stunt carry over?

HARRISON: Yes. SBMV, a structure determined recently at 2.8 Å by Rossman and colleagues at Purdue, does have an arm that looks strikingly like the one in TBSV, although some of the details of the interdigitating structure are different. In other words, the quasi-equivalance problem is solved in an identical fashion. The fold of the S domains is so similar that one wonders to what extent they are biologically independent structures.

GREER: Is there any evidence of flexibility at the hinge in the dimer in solution? Also, is the R piece disordered in the dimer?

HARRISON: We don't know the answers to either of those questions yet. We do know from work of John Golden and Bob Sauer, that there are two points on the arm which are sensitive to proteolysis in the dissociated TCV dimer. One of them is at the R-arm junction, ~7,000 mol wt units in from the N-terminus. The other is where we imagine the S domain to begin, ~11,000 mol wt units from the N-terminus. I imagine that there is considerable flexibility in this region. Whether the R piece is well-folded in solution is still a question we're trying to answer.

GREER: Do you think it is folded in the virus particle in different orientations?

HARRISON: I'm not sure.

FLETTERICK: Do you know if the STNV, a T = 1 system, has an arm?

HARRISON: Well, all T = 3 viruses are not likely to have the same fold and all T = 1 viruses are not likely to be T = 1 versions of the same subunit. STNV, for which a 4 Å structure is known (Strandberg and colleagues, Uppsala), appears to have an N-terminus which extends into the center and becomes disordered. It's as if the N-terminal portion is responsible for interaction with the RNA. There is no interdigitating structure of the kind we see in TBSV nor a need for one, since it's a T = 1 system. We can make T = 1 particles, as Jo Butler reminded us, from TCV. We don't expect any interaction between the arms. What the arms will do, if they are indeed present in the units that make T = 1 particles, I can not easily predict. It is quite a squeeze to get them all into the center.