Energetics of Quasiequivalence: Computational Analysis of Protein-Protein Interactions in Icosahedral Viruses

Vijay S. Reddy,* Heidi A. Giesing,* Ryan T. Morton,* Abhinav Kumar,* Carol Beth Post,#
Charles L. Brooks, III,* and John E. Johnson*
*Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, and #Department of Medicinal Chemistry, Purdue University, West Lafayette, Indiana 47907 USA

ABSTRACT Quaternary structure polymorphism found in quasiequivalent virus capsids provides a static framework for studying the dynamics of protein interactions. The same protein subunits are found in different structural environments within these particles, and in some cases, the molecular switching required for the polymorphic quaternary interactions is obvious from high-resolution crystallographic studies. Employing atomic resolution structures, molecular mechanics, and continuum electrostatic methods, we have computed association energies for unique subunit interfaces of three icosahedral viruses, black beetle virus, southern bean virus, and human rhinovirus 14. To quantify the chemical determinants of quasiequivalence, the energetic contributions of individual residues forming quasiequivalent interfaces were calculated and compared. The potential significance of the differences in stabilities at quasiequivalent interfaces was then explored with the combinatorial assembly approach. The analysis shows that the unique association energies computed for each virus serve as a sensitive basis set that may determine distinct intermediates and pathways of virus capsid assembly. The pathways for the quasiequivalent viruses displayed isoenergetic oligomers at specific points, suggesting that these may determine the quaternary structure polymorphism required for the assembly of a quasiequivalent particle.

INTRODUCTION

Virus crystallography is a mature area of biophysics, with nearly 20 years of productivity since the first high-resolution structure of tomato bushy stunt virus was reported (Harrison et al., 1978). High-resolution virus structures have had a large impact on virology, with detailed chemical mechanisms now defined for functions as diverse as the autocatalytic maturation cleavage in nodaviruses (Zlotnick et al., 1994) and the details of antibody-virus interactions (Smith et al., 1996). Particle-related functions in the viral life cycle frequently direct attention to particular regions of the structure, and in most cases either a rationale or a detailed mechanism has emerged for various experimental observations. A less studied aspect of virus structure/function is the importance of those parts of the particle that have not been associated with a particular biological function and that appear to form only the framework for spatially distributing the functional parts of the virion. Because the average small RNA virus structure contains more than 10^6 nonhydrogen atoms, there is a need to provide a reduced representation of the virus particle that is dependent on the three-dimensional structure and that addresses a characteristic common to all virus particles, namely the assembly. In this paper we report an initial effort to represent high-resolution virus structures as unique dimer association energies that occur between a single subunit type arranged on a $T = 3$ icosahedral lattice, or between three subunit types (picornavirus capsid) arranged on a pseudo $T = 3$ ($P = 3$) surface lattice. This representation addresses the energetics of quasisymmetry in the protein capsid and utilizes the high symmetry present in the shell to describe the features of particle assembly. Electrostatic and van der Waals contributions to the unique interface energies were computed with refined, energy-minimized structures of three different viruses determined by x-ray crystallography.

Two significant observations emerged from our analysis. First, it is possible to create an inventory of residues and their relative contributions to interface stability. Of particular interest in this category is the similar contribution toward stability that the same residue may make at quasiequivalent interfaces (e.g., an interface created by an icosahedral twofold axis compared to an interface created by a quasi-twofold axis) through remarkably different chemical interactions. In a second application, these association energies were used to compute and compare potential assembly pathways, employing a procedure introduced by Horton and Lewis (1992). Although the "combinatorial" approach to assembly is at best a first approximation to actual assembly, its dependence on ratios of the association energies among different interfaces suggests that these computed parameters represent a fundamental property of the particle. The combined results from these studies immediately suggest mutations that may affect the assembly and disassembly processes and provide important insight into events in the virus life cycle associated with these occurrences.

Received for publication 11 March 1997 and in final form 3 September 1997.

Address reprint requests to Dr. Charles L. Brooks or John E. Johnson, Department of Molecular Biology, MB31, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037. Tel.: 619-784-9705; Fax: 619-784-8660; jackj@scripps.edu.

0006-3495/98/01/546/13 $2.00

SUBUNIT INTERFACES IN QUASIEQUIVALENT LATTICES

Sixty copies of a single gene product distributed with the symmetry of an icosahedron generate the largest shell possible for a subunit of a given size. Such an arrangement places all 60 proteins in identical environments; hence the proper intersubunit complementarity leads naturally to assembly and a symmetrical particle. The large majority of plant and animal viruses display icosahedral symmetry; however, very few contain only 60 copies of the capsid protein. Most icosahedral viruses display quasisymmetry, leading to $60T$ subunits in the capsid, where T is an integer greater than 1, reflecting the selection rules for distributing hexamers and pentamers of proteins on a surface lattice (Caspar and Klug, 1962). Quasisymmetrical particles always display exact icosahedral symmetry, but they contain T copies of the capsid protein in the asymmetrical unit. Construction of a quasiequivalent capsid requires that subunits contain a molecular switch that facilitates polymorphic association into hexameric and pentameric oligomers. The 12 pentamers are always exact and are located at positions equivalent to the vertices of an icosahedron. The $10(T - 1)$ hexamers may form nearly perfect sixfold symmetrical associations, or they may form structures better described as a trimer of dimers (Fisher and Johnson, 1993) or a dimer of trimers (Munshi et al., 1996). The switching mechanism, which is usually multifaceted and different for each virus group, is a major determinant of the geometry of the hexamer structure. Depending on the nature of the hexameric (sixfold) association, the distribution of subunits on a $T = 3$ lattice can closely approximate well-defined geometric solids related to the icosahedron. Viruses with $T = 3$ symmetry and forming nearly exact sixfold symmetry interactions generate a structure closely similar to a truncated icosahedron (Speir et al., 1995; Canady et al., 1996), whereas subunits associating as a trimer of dimers on a $T = 3$ lattice (Harrison et al., 1978; Wery et al., 1994) have the geometric shape of a rhombic triacontahedron (Williams, 1979).

Fig. 1 illustrates a schematic drawing of the rhombic triacontahedron, a space-filling model, and ribbon drawings of the subunit organization of black beetle virus (BBV), a member of the nodaviridae group. The viral subunit, which contains 407 amino acids folded into an elaborate, trapezoid-shaped, β-sandwich structure (Fig. 1 *d*), is represented as a trapezoid in the diagram (Fig. 1 *a*). Subunits labeled A, B, and C have identical amino acid sequences, but are structurally unique (quasiequivalent) and occupy the central triangle, which is the icosahedral asymmetrical unit (Fig. 1 *d*). Subunits identified by the same letter (e.g., A, A5, A2) are related by icosahedral symmetry. Icosahedral symmetry elements (twofold, threefold, and fivefold axes) are labeled as a white oval, triangle, and pentagon, respectively, and the yellow oval and triangle refer to quasisymmetry elements. The A subunit and its fivefold related equivalents form the 12 pentamers, and the B and C subunits alternate to form a quasihexamer about the 20 threefold axes to generate a particle formed by 180 copies of a single gene product. Fig. 1 *e* shows clearly the difference in contacts between subunits forming pentamers and those present only at hexamers, illustrating how subunit interactions can switch between the two types of oligomers.

The distribution of hexamers and pentamers on a $T = 3$ surface lattice generates other subunit contacts that are similar, but not exactly the same. Examination of the C subunits related by the icosahedral twofold axis (C and C2) and the subunits labeled A and B5 shows that the same regions of the trapezoid make contact, but the latter twofold association is local or quasi in that a fivefold axis and a threefold axis would be superimposed by the same symmetry axis that relates A and B5 with reasonable fidelity. Likewise, the A, B, and C subunits in the central triangle lie in the same plane and are related by a nearly exact (quasi) threefold symmetry axis, identified by a yellow triangle (Fig. 1, *a* and *c*) that is perpendicular to the plane of the trapezoids and parallel to the icosahedral twofold axis defined by the white oval. The quasi-threefold axis is local because it does not pass through the center of the particle, and it only superimposes the trapezoids adjacent to it.

The difference between the C-C2 contact and the A-B5 contact is easily seen if the shape of the particle is considered (Fig. 1 *a*). The rhombic face containing A, B, C, A2, B2, and C2 is planar, and the C-C2 contact is referred to as flat. In contrast, the triangular face containing A, B, and C forms a dihedral angle of 144° with the triangular unit containing A5, B5, and C5, and the A-B5 contact is referred to as bent. Thus although the interactions at C-C2 and A-B5 are similar, as the same parts of the subunits are in contact (Fig. 1), they require that the subunits form different contacts along the line connecting threefold and fivefold axes (A-B5) and the two adjacent threefold axes (C-C2). It is the molecular switch, in the case of BBV a 15-amino acid region of the polypeptide and a 10-bp RNA duplex, that alters the subunit contacts (Fig. 1 *e*). Note that the $T = 3$ particle results from the perfect blend of the two contacts. If all were of the C-C2 type, a flat sheet would result. If all were bent like the A-B5 contact, a small particle with 60 subunits would result.

In this paper we examine the energetics of the subunit associations at the unique interfaces of the $T = 3/P = 3$ surface lattices for three different viruses. Referring to Fig. 1, and grouped as quasiequivalent sets, these are [A-B, B-C, C-A], [A-B5, C-C2], and [A-A5, C-B5, B2-C]. The refined coordinates for BBV (Wery et al., 1994), southern bean mosaic virus (SBMV) (Silva and Rossmann, 1987), and human rhinovirus 14 (HRV14) (Arnold and Rossmann, 1990) were used to compute the van der Waals and electrostatic association energy contributions to each of the interfaces. BBV and SBMV are $T = 3$ RNA viruses with capsids that form a rhombic triacontahedron with an average diameter of ~300 Å. Switching between flat and bent contacts is regulated in a similar way for these two viruses, but the details are different. The HRV14 capsid also has a diameter of 300 Å; however, it differs from BBV and

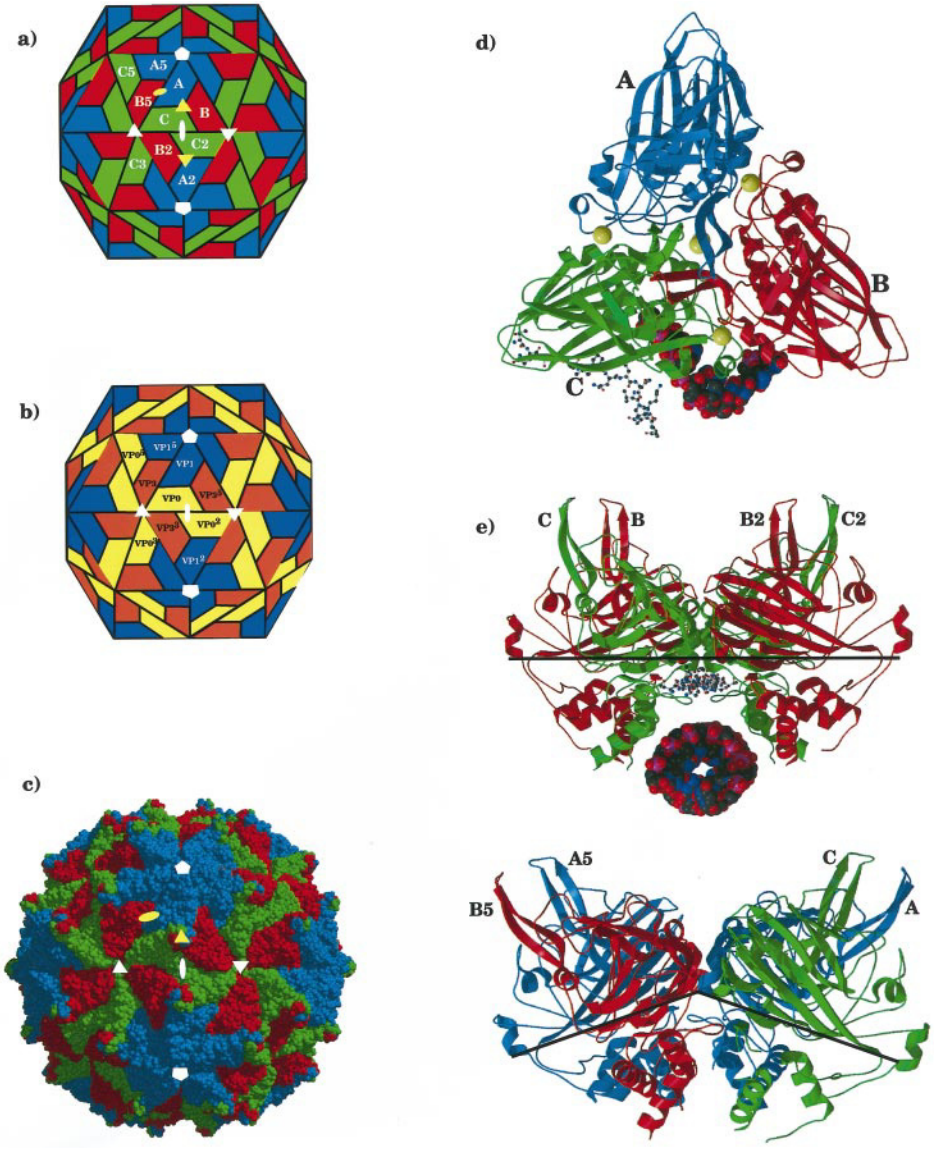


FIGURE 1 (*a*) Schematic representation of a $T = 3$ quasiequivalent lattice, corresponding to a rhombic triacontahedron (Williams, 1979), the geometrical architecture of black beetle virus (BBV). Each of the trapezoids represents a single subunit with the same amino acid sequence. The $T = 3$ particle is formed of 180 subunits that lie in three structurally unique positions (labeled A, B, and C). Subunits labeled with same letter are related by icosahedral symmetry axes corresponding to twofold, threefold, and fivefold rotations identified by white ovals, triangles, and pentagons, respectively (e.g., C and C2 related by icosahedral 2-fold axis). Subunits marked with different letters are related to one another by quasisymmetry axes corresponding to twofold and threefold local rotation axes identified, respectively, by yellow ovals and triangles (e.g., A and B5 are related by a quasi-twofold axis). The subunits labeled A, B, and C are related by quasi-threefold symmetry; they form an icosahedral asymmetrical unit (protomer) of the $T = 3$ particle. There are two different joints that a triangular asymmetrical unit makes with neighboring triangles, bent (*left* and *right*) and flat (*bottom*). (*b*) Schematic drawing of a pseudo $T = 3$ surface lattice, also known as a $P = 3$ lattice. In this lattice there are three types of trapezoids (labeled VP1, VP2, and VP3) representing subunits with different amino acid sequences; hence there is no quasisymmetry present. The $P = 3$ lattice is, in general, similar to the $T = 1$ lattice, where VP1, VP2, and VP3 are considered together as a single subunit. The subunits identified by the same label are related by icosahedral symmetry elements, twofold, threefold, and fivefold, identified by white ovals, triangles, and pentagons, respectively. (*c*) Space-filling model of black beetle virus (BBV). The A, B, and C subunits are identified by blue, red, and green, respectively. The average diameter of the particle is 312 Å, and the maximum diameter of the particle is 340 Å at the quasi-threefold protrusions. Icosahedral and quasisymmetry elements are identified by white and yellow labels, respectively. (*d*) Ribbon drawing of an icosahedral asymmetrical unit (protomer) of BBV made up of the A, B, and C subunits and a strand of partially ordered RNA of 10 bases, shown as a space-filling object. The molecular switch, required for quasiequivalent shell formation, is a segment of polypeptide corresponding to residues 20–34 of the C-subunit, drawn as a ball-and-stick model, the duplex RNA. There are five Ca^{2+} ions associated with each protomer, shown as yellow spheres. Three of the five are located at the quasi-threefold related interfaces (A-B, B-C, and C-A), and two of them are located on the quasi-threefold axes. (*e*) Two different quaternary contacts found between the triangular asymmetrical units of BBV and two other nodavirus structures. The top figure shows the flat contact along the line joining the two adjacent threefold axes, formed by the C and B and C2 and B2 subunits. This contact results from the selective presence of a segment of polypeptide of residues (20–34) of C (C2) subunits and 10 base pairs of duplex RNA. Together they act as a wedge (molecular switch) at the contact of protein subunits. The bottom figure shows the bent contact formed by C and A and B5 and A5 subunits along the line connecting adjacent threefold and fivefold axes. Both structural elements that act as the wedge at the flat contact are absent at this interface, allowing the subunits to pivot about the three- to fivefold line and form a bent contact. Both the flat and bent contacts are needed to assemble into virions with $T = 3$ quasisymmetry.

SBMV in that the amino acid sequences of the three capsid proteins are different. The surface lattice on which they are arranged is similar to the $T = 3$ lattice.

BBV is a member of the nodavirus group that infects insects, mammals, and fish. The subunit contains 407 amino acids, and the particle packages two RNA molecules (1.4 kb and 3.1 kb) that contain the genes for three proteins; the capsid subunit, an RNA-directed RNA polymerase and a 10-kDa protein of unknown function, believed to be associated with RNA replication. SBMV is a plant virus with a 4.2-kb RNA molecule encoding four proteins. The coat protein subunit contains 260 amino acids. HRV14 is the causative agent of the common cold, with an RNA genome of ~7.5 kb, which among a variety of other proteins, encodes three major capsid proteins with molecular masses of 30–40 kDa. The subunits of BBV, SBMV, and three capsid proteins of HRV14 have tertiary folds similar to canonical viral β-sandwiches (Rossmann and Johnson, 1989).

HRV14 has a capsid that is qualitatively similar to BBV and SBMV, but the A, B, and C positions of the $T = 3$ capsid are occupied by different gene products with very similar folds. The surface lattice for HRV14 is referred to as $P = 3$ because it is a pseudo $T = 3$ lattice formed of three different major proteins (Fig. 1 *b*). The $P = 3$ lattice is assumed to have evolved from a $T = 3$ lattice by a triplication of the coat protein gene, followed by independent evolution of the three major subunits, in which the fold of the structure has been retained but the sequences have changed dramatically (no significant sequence similarity is detectable among the three subunits). In this process the need for a switch has been lost, because the subunit interfaces at the quasiequivalent contacts are now defined by altogether unique amino acids, and assembly may be similar to that of a simple icosahedron in which the "subunit" size (~100 kDa) is three times that of the $T = 3$ subunits in BBV and SBMV. It is useful to include HRV14 in this analysis, because assembly intermediates have been well characterized and it therefore serves as a control, and because of the qualitative differences observed in the assembly pathway of $P = 3$ and $T = 3$ particles. The latter require points of quaternary structure polymorphism, which are apparent in the computationally determined assembly pathways, whereas the former do not because there is no conformational switching.

METHODS

Computational procedures to analyze quasiequivalence and assembly were initially performed with the coordinates of BBV refined at 2.8-Å resolution. The model of an icosahedral asymmetrical unit including the A, B, C subunits, a segment of ordered RNA, five Ca^{2+} ions, and 202 ordered water molecules, was subjected to energy minimization to relieve any unrealistic interactions in the starting x-ray model. The program CHARMM (Brooks et al., 1983) was used to compute four rounds of 1000 cycles of energy minimization with a tolerance in the energy (change in the energy between two consecutive steps of minimization) of 0.0001kcal/mol. Initially harmonic restraints (force constant = 25 kcal/mol Å^2) were imposed on all of the nonhydrogen atoms, and the force constant was reduced by 5 kcal/mol Å^2 at the end of each of the four rounds of minimization. The positions of the Ca^{2+} ions were held fixed during the minimization procedure. To represent the environment of the protein surrounding the asymmetric unit, periodic images with icosahedral symmetry were employed, using the IMAGE facility in CHARMM. The all-hydrogen parameters and topologies from version 22 of CHARMM were used in all energy calculations. The energy minimization was carried out using atom-based cutoffs: CUTNB = 13 Å, CUTOFNB = 12 Å, CTONNB = 10 Å, and the nonbonded pairs were updated every 100 steps of minimization. A distance-dependent dielectric constant was assumed to represent some aspects of solvation during the minimization. At the end of minimization the model had moved only 0.04 Å (Cα rms deviation) from the original structure.

The energy-minimized model was used to calculate association energies for the unique subunit interfaces in the orientation found in the assembled particle (Fig. 1 *c*). The procedure described by Gilson and Honig (1988) was used to calculate association energies. In addition, the contribution of van der Waals interactions calculated using a Lennard Jones 6–12 potential, without any distance cutoffs imposed, was included in the evaluation of the association energy. The assocation energies were calculated as a sum of energies due to desolvation, electrostatic, and van der Waals interactions between the interacting subunits. The finite difference solutions to the Poisson-Boltzmann (PB) equation were computed with the program DelPhi (Sharp and Honig, 1990). The energy of desolvation and the electrostatic interactions between protein-protein, protein-Ca^{2+}, and protein-RNA were calculated in the presence of bulk solvent. For the purpose of reliably comparing the protein-protein interactions at quasiequivalent interfaces, all of the ordered water molecules were omitted from the calculation. In initiating the Poisson-Boltzmann calculations, a grid size of 0.7 Å and dielectric constants of 4.0 (within the protein) and 78.5 (solvent) were used. Furthermore, the solvent ionic strength was set at the physiological value (0.145 M), and a 2-Å stern layer was employed. We note that the qualitative predictions of interface interactions and association energies appeared to converge at the grid size used (0.7 Å). For systems of this size, this also represents a practical limit. Interaction energies upon binding to the adjacent subunit were estimated for each residue as a summation of electrostatic and van der Waals contributions between all of the atoms in the residue and all of the atoms of the neighboring subunit. The electrostatic interaction energy at an atom was calculated as a product of partial charge at the atom times the electrostatic potential at the atom position due to the interacting subunit. The plot of these energies as a function

of residue numbers facilitates the comparison of interfaces related by quasisymmetry; these are shown in Figs. 2–4.

The subunit-subunit association energies were then used as a basis set to examine assembly intermediates and the resultant pathways in the virus assembly. We have used the combinatorial approach suggested by Horton and Lewis (1992) to assemble the viral capsid by adding one subunit at a time, starting with a single subunit. At each step, all possible unique ways of adding a subunit within the context of $T = 3$ icosahedral symmetry were explored, with the objective of maximizing the contacts between the precursor structure and the newly joining unit. The cutoff limit of 9000 structures/configurations was used. The association energy of the newly formed structure was calculated as a sum of binding energies of all possible dimeric interfaces present in the structure. The number of configurations increases exponentially as a function of number of associating subunits. All of the configurations for an oligomer of N subunits were sorted in increasing order of energy, and the top 9000 configurations were chosen as precursors for generating the structures containing $N + 1$ subunits. However, at each stage on average, only 2500 structures were required as precursors to generate 9000 structures of the next higher oligomer. The cutoff limit of 9000 was selected to make the calculation of the pathways computationally economical, and a further increase in this limit did not affect the pathway significantly. Hence each new assembly is not computed completely de novo, but is assumed to be derived from one of the energetically likely configurations of the previous partial assembly. Considering only the lowest energy con-

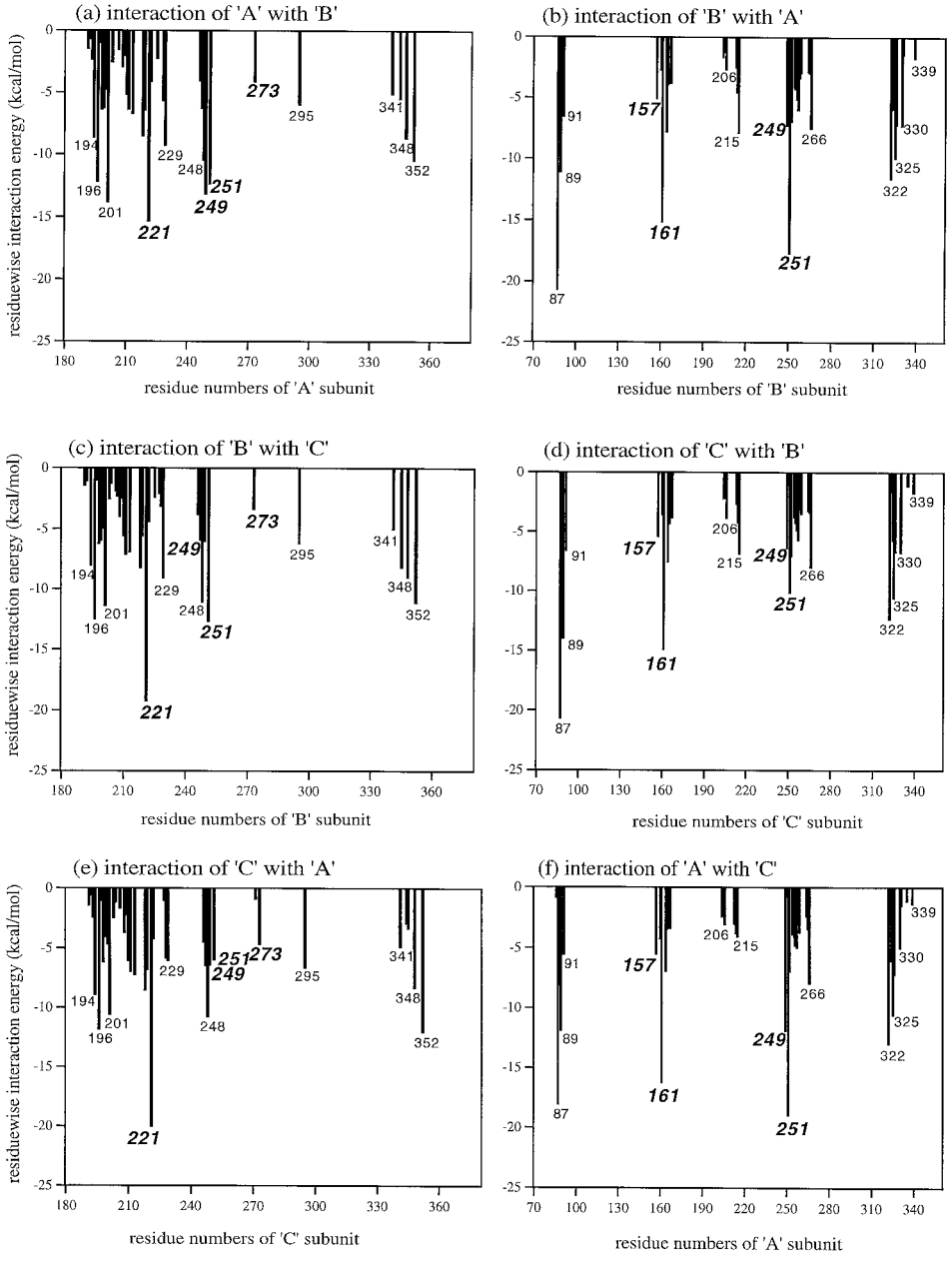


FIGURE 2 (*a*) Plots showing residues of the A subunit that contribute significantly to the interaction energy while associating with the subunit B at the quasi-threefold axes in BBV. Some of the residues are identified by their residue numbers. (*b*) A similar plot showing the contribution of residues in the B-subunit when interacting with the A-subunit at the quasi-threefold axes. Residues plotted in the left panels interact with the residues plotted on the right. *c* and *d* refer to the residue contributions at the B-C interface, and *e* and *f* show the energies at the C-A interface. A comparison of all of the plots in the figure suggests that, in general, the same residues of the subunits interact at each of the interfaces, indicating the similarity of contacts along the quasi-threefold related A-B, B-C, and C-A interfaces. Residues shown in bold italics coordinate Ca^{2+} ions. Residues 249 and 251 of the A, B, and C subunits together coordinate two Ca^{2+} ions that lie on the quasi-threefold axes. Residues 221 and 273 of the A subunit and residues 157 and 161 of the B-subunit coordinate calcium ions at the A-B interface. These same residues coordinate Ca^{2+} ions at the B-C and C-A interfaces. Major contributions to the interaction energy at these interfaces comes from the residues that coordinate Ca^{2+} ions and those involved in salt links (e.g., 341–87; 352–89; 229–91 and 295–322).

figuration at every stage of assembly to generate higher oligomeric structures (i.e., the method employed by Horton and Lewis) did not always yield the lowest energy configuration at each stage of assembly. This affects the sampling of configurations as well as the course of the assembly. Typically, the highest energy structure ignored was ~2000 kcal/mol above the most stable structure of N subunits.

The absolute difference in the association energies of the first and second most stable configurations for a given oligomer of N subunits is considered as a measure of the "preference" for the most stable assembly of the particular association. The greater the difference between the most stable and the next most stable configurations, the higher the chance that the most stable configuration is the sole intermediate. Assemblies where two or more configurations have similar energies are potential points of conformational polymorphism, where the path of assembly may take different directions. Plotting the difference in stability between the most stable association and the next most stable association allows one to identify the best configuration for each oligomeric association, thus leading to a potential assembly pathway. The same procedure of calculating association energies and the assembly pathway was employed with structures of SBMV and HRV14.

RESULTS AND DISCUSSION

Association energies at subunit interfaces

The association energies for the unique dimer subunit interfaces of BBV are shown in Table 1. The binding energies for BBV at the quasi-threefold related interfaces, A-B, B-C, and C-A, are very similar, -134 ± 3 kcal/mol, showing that the quasi-threefold symmetry is maintained at the atomic level. Various contributions to the binding energies are comparable at these interfaces, except for the electrostatic energy contribution due to protein-protein and protein-Ca^{2+} interactions at the C-A interface (Table 1). This difference is due to small differences (<1.0 Å) in a few interacting charged atoms. These differences in the structure are real, supported by the encompassing electron density from the x-ray structure analysis. The loss of protein-protein interaction energy at the C-A interface is somewhat compensated for by the strong interactions of the protein with the Ca^{2+} ions. The similar binding energies for the C-B5 and A-A5 interfaces demonstrate that the quasi-twofold symmetry is well maintained. However, quasi-sixfold symmetry that relates the bent contact to the flat contact is not well conserved, as suggested by the geometric shape discussed above. This is due to the molecular switches, segments of partially ordered peptide and RNA, that are present selectively at the C-B2, C-C2 contact (i.e., the joint along the line between adjacent icosahedral threefold axes) and absent at B5-C, B5-A, and A-A5 interfaces (i.e., the joint along the line connecting adjacent threefold and fivefold icosahedral axes). Likewise, the association energies for SBMV reflect not only quasi-twofold and threefold symmetries, but also quasi-sixfold symmetry (Table 2). The calculated association energies for HRV14 reflect no quasisymmetry, in agreement with the $P = 3$ architecture (Table 2 and Fig. 1 *b*).

Figs. 2–4 show plots displaying the contribution of every relevant residue to the interaction energy at each interface of BBV. Similarity in the plots corresponding to the quasiequivalent interfaces accounts for fidelity of quasiequivalence, and the differences determine deviations from quasiequivalence. The plots in Fig. 2 show that the same residues at each side of the contact at the A-B, B-C, and C-A interfaces have similar residue-wise contributions to the interaction energy (compare *a*, *c*, and *e*; *b*, *d*, and *f*). This suggests that the quasi-threefold symmetry is well conserved. Residues plotted in the left panels interact with the residues in the right panels, and the similarity in the plots results from closely similar contributions for the same individual residues in each subunit. The differences in the residuewise contributions are due to small differences in the geometry of the interacting residues, especially involving charge interactions with Ca^{2+} ions. The interactions be-

TABLE 1 Association energies* for subunit-subunit contacts for BBV

	interaction energy*					
Interface	Energy of desolvation*	E^1_{elec} #	E^2_{elec} §	E_{vdw} ¶	Binding energy*	Buried surface area (Å^2)
A-B; Ca^{2+}	180.0	−32.0	−142.0	−142.0	−136.0	3150.0
B-C; Ca^{2+}	178.0	−35.0	−130.0	−148.0	−135.0	3140.0
C-A; Ca^{2+}	184.0	−20.0	−147.0	−148.0	−131.0	3115.0
A-A5	42.0	−19.0	—	−95.0	−72.0	1500.0
C-B5	39.0	−20.0	—	−93.0	−74.0	1570.0
C-B2	42.0	−18.0	—	−122.0	−98.0	2115.0
A-B5	51.0	−25.0	—	−82.0	−56.0	1585.0
C-C2	20.0	−7.0	—	−77.0	−64.0	1220.0
C-C2-dRNA	73.0	−6.0	−46.00	−132.0	−111.0	2375.0

*All energies are in kcal/mol. Electrostatic and van der Waals energies were calculated using DelPhi and CHARMM, respectively.
#Electrostatic energy due to protein-protein interactions alone.
§Electrostatic energy due to protein-Ca^{2+} or protein-RNA interactions.
¶Van der Waals energy between all of the components interacting at the interface.

TABLE 2 Comparison of dimer association energies* of BBV, SBMV, and HRV14

$T = 3$ interface	BBV	SBMV	$P = 3$ interface#	HRV14
A-B	−136.0§	−64.0§	VP1-VP3^5	−111.0
B-C	−135.0§	−64.0§	VP3^5-VP0	−70.0
C-A	−131.0§	−53.0§	VP0-VP1	−246.0
A-A5	−72.0	−78.0	VP1-VP1^5	−81.0
C-B5	−74.0	−74.0	VP0-VP3	−189.0
C-B2	−98.0	−71.0	VP0-VP3^3	−109.0
A-B5	−56.0	−59.0	VP1-VP3	−348.0
C-C2	−111.0¶	−66.0	VP0-VP0^2	−43.0
A-C5	—	—	VP1-VP0^5	−44.0
B-B5	—	—	VP3-VP3^5	−56.0
A-C3	—	—	VP1-VP0^3	−55.0

*In kcal/mol.
#VP1, VP0, and VP3 are the capsid protein subunits of HRV14, equivalent to A, C, and B subunits in the $T = 3$ lattice. VP0 includes both VP2 and VP4. Superscripts refer to icosahedral symmetry elements, and the subunits are symmetry-related partners.
§Including the interaction energy due to Ca^{2+} ions.
¶Including the interaction energy due to duplex RNA.

tween quasi-threefold related subunits is substantially stronger than the other subunit interfaces (Table 1). This is due to the large number of residues that contribute to the interaction energy, the presence of Ca^{2+} ions, and salt links that are only at these interfaces.

Fig. 3, *c–f*, shows the plots corresponding to the B5-C and A-A5 interfaces, which are related by quasi-twofold symmetry. These plots show that similar residues interact at these interfaces, indicating that the quasi-twofold symmetry is well conserved energetically. However, at the C-B2 interface (Fig. 3, *a* and *b*), there are a number of residues (20–34 and 146 of the C-subunit and residues 108, 130, and 300–305 of the B2 subunit) that interact only at this interface, whereas residues 76–78 of the B5 and A subunits make significant contributions only at the B5-C and A-A5 interfaces. This difference in the interactions at the C-B2 contact compared to C-B5 and A-A5 breaks the quasi-sixfold symmetry.

Fig. 4 shows the interactions at the C-C2 interface in the absence and the presence of duplex RNA (Fig. 4, *a–d*). The C and C2 subunits are related by icosahedral twofold symmetry, which is reflected in these plots, where the same residues of each side of the interface of the C and C2 subunits interact with each other (Fig. 4, *a* and *b*). The residues 56–68, 371, and 375, present only in the middle panels (*c* and *d*) and not in the top panels (*a* and *b*), interact with RNA. Residues that interact at the B5-A interface (*bottom panels, e* and *f*) are similar to the residues that interact at the C-C2 interface in the absence of RNA (*a* and *b*). It is interesting to note that the same residues of the A and B5 subunits interact at the A-B5 interface, indicating that the quasi-twofold symmetry at this interface is nearly perfect (Fig. 4, *e–f*). In this case local symmetry is not necessarily inexact symmetry. The analysis demonstrates that the quasi-twofold and threefold axes are individually and locally highly symmetrical, but the symmetry breaks down when the icosahedral twofold and quasi-twofold axes are compared with each other (they are related by a quasi-threefold axis, but are at a significant distance from it). It is the lack of equivalence of these axes that makes the hexamer structure a trimer of dimers rather than a true hexamer around the icosahedral threefold axes, as observed in some other $T = 3$ lattices (Speir et al., 1995).

The inventory of individual residues and their contribution to interface stability provide an important guide for the construction of potential assembly mutants and the prediction of their phenotypes. Of particular interest are those residues that make unique contributions at quasiequivalent interfaces. The availability of both baculovirus expression vectors (Schneemann et al., 1993) and reagents for transfection with clonal transcripts (Dasmahapatra et al., 1986) makes the nodavirus system particularly attractive for the analysis of these phenotypes.

Combinatorial viral assembly based on subunit association energies

The relative ratios of association energies among the categories of quasisymmetry related interfaces are probably the determinants of assembly pathways. In the case of BBV, the association energies favor the formation of the quasi-threefold related interfaces over others (Table 2). In contrast, for SBMV, also a $T = 3$ virus, the formation of quasi- and icosahedral twofold related interfaces is favored. Examination of the ratios of the dimer interface binding energies for a particular virus in Table 2 and the variation of these ratios for different viruses suggested that they are an intrinsic property of the particle. These ratios will probably be similar for different viruses in the same family, thus representing a reduced characteristic of the individual virions, which is potentially valuable for categorizing and comparing viruses. It is likely that the stability of the individual interfaces is also related to the assembly properties of the subunits, and this was explored for a number of different protein assemblies, including HRV14, by Horton and Lewis (1992). Assembly is a complex process for any large oligomer, and for quasiequivalent capsids it certainly involves dynamic switching between flat and bent joints that is influenced by the extent of the growing shell to which subunits are being added. It is also certain that RNA serves as a scaffold for assembly in some viruses and must be important in BBV assembly, where it serves as part of the molecular switch generating the quasiequivalent capsid (Wery et al., 1994). Furthermore, it is clear that icosahedral symmetry can only exist in the final assembled particle and that there must be sizable deviations from this symmetry in the course of assembly. Despite the obvious dangers in ignoring the role of disordered protein and RNA and significant asymmetry that must occur in virus assembly, we have used the approach of Horton and Lewis to examine assembly pathways for BBV, SBMV, and HRV14, based on the subunit asso-

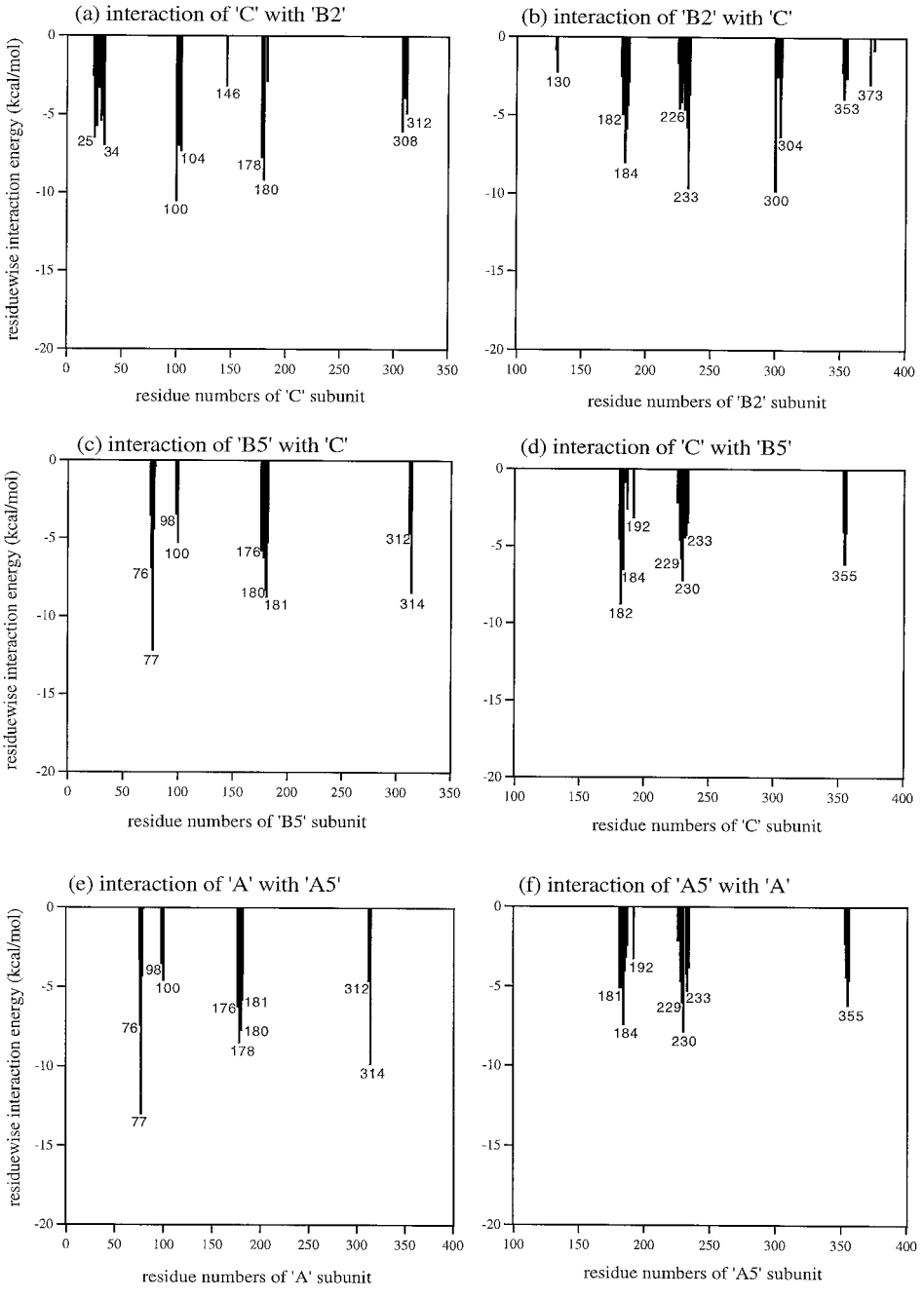


FIGURE 3 (*a–f*) Plots showing the residues that contribute to the interaction energy at the C-B2, B5-C, and A-A5 interfaces in BBV. The similar patterns at the B5-C and A-A5 interfaces (*c–f*) indicate that the same residues are making contact, inferring the quasi-twofold symmetry is well conserved. The differences in the plots for the C-B2 contact compared to the B5-C and A-A5 interfaces are consistent with the flat and bent nature of the contacts at these interfaces. Residues that interact at the C-B2 interface but not at the B5-C and A-A5 interfaces form contacts with the ordered peptide, residues 20–34 of the C-subunit, present only at the flat contact. There are certain interactions absent at the C-B2 interface compared to B5-C and A-A5 interfaces (e.g., with residues 76 and 77). These interactions are unique to the bent contact and partially compensate for the interactions with molecular switch at the C-B2 contact.

ciation energies in Table 2. This approach assumes that subunits can only assemble into an icosahedral surface lattice and that all interactions between subunits (i.e., the association energies in Tables 1 and 2) are determined from residues visible in the high-resolution x-ray structure. With these criteria, the shell is allowed to grow as described in the Methods section, and the relative stabilities of structures formed by the same number of subunits are assessed and compared. The free energy difference between the two most stable structures, as each subunit is added, is computed for analysis of assembly pathways of different viruses. The objectives of carrying out this analysis are the identification of potential assembly intermediates and the points of potential quaternary structure polymorphism required for quasiequivalence.

Examination of the association energies of the best structure at each oligomeric association of subunits ranging from 2 to 180 demonstrated a continuous monotonic decrease in the association energy, going from the Nth association to $(N + 1)$th association. This suggests that assembly of subunits into a growing shell is a thermodynamically favorable process. A large difference in association energy between the two most stable structures possible for an association indicates the preference of a particular configuration over others, whereas small differences in energies reflect points of polymorphism, where a number of configurations might

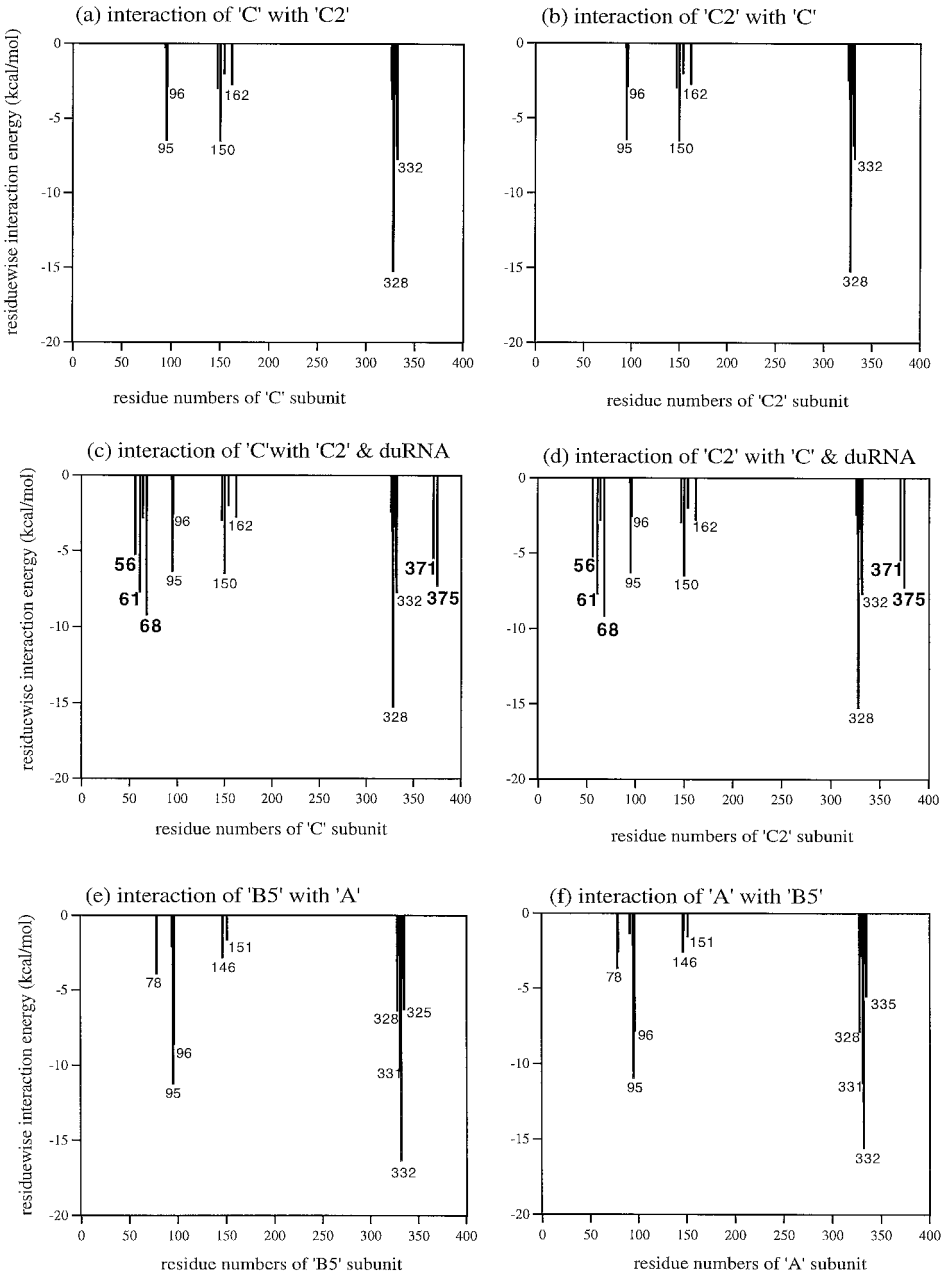


FIGURE 4 (*a–f*) Plots identifying the residues and their contributions to interaction energy at the C-C2 and B5-A interfaces in BBV. *a* and *b* show the residues that are responsible for protein-protein interactions alone at the C-C2 interface. *c* and *d* show the residues that interact with the duplex RNA (identified by *bold letters*), along with the residues that contact with the protein at the C-C2 interface. These plots show perfect twofold symmetry, as the C and C2 subunits are related by an icosahedral twofold axis. *e* and *f* show the high-fidelity quasi-twofold symmetry at the A-B5 interface, which relates A with B5. There are certain residues that are common to both the C-C2 and A-B5 interfaces. The quasi-sixfold symmetry that relates the A-B5 contact with the C-C2 is poorly maintained because of the absence/presence of the molecular switch at the respective interfaces.

be in equilibrium. It is important to note that each new assembly is computed nearly de novo. It is not assumed that the subunit is added to the previous most stable oligomer, but the most stable 2500 structures from the *N*th assembly product are tested for the $N + 1$ assembly product. Such an exploration is necessary to avoid losing an important configuration that might give rise to a stable structure down the assembly pathway.

Fig. 5 *a* is a plot of the difference in the free energies of the two most stable structures in BBV, as each subunit (A, B, or C) is added to the configurations from the previous associations. From an examination of the results in Figs. 5 *a* and 6, it is clear that there is no preference for the initial dimer formation among the three subunits that form a quasi-threefold interaction; however, each of these dimers is much more stable than any other dimer, as can be seen in Table 1. The first oligomer showing significant relative stability among its population of configurations for a given association of *N* subunits is the trimer. The preferred configuration of the trimer is the triangular asymmetrical unit, made up of subunits A, B, and C ($\Delta\Delta G = -60$ kcal/mol; Figs. 5 and 6). Even though the 4mer and 5mer associations have structures significantly preferred over their next most stable structure, the association of multiples of three subunits ("ABC" protomer) consistently gives rise to preferred structures. The regular peaks in the pattern correspond to the structures made up of multiples of trimeric protomers, and this is an indication that the assembly of BBV, and probably nodavi-

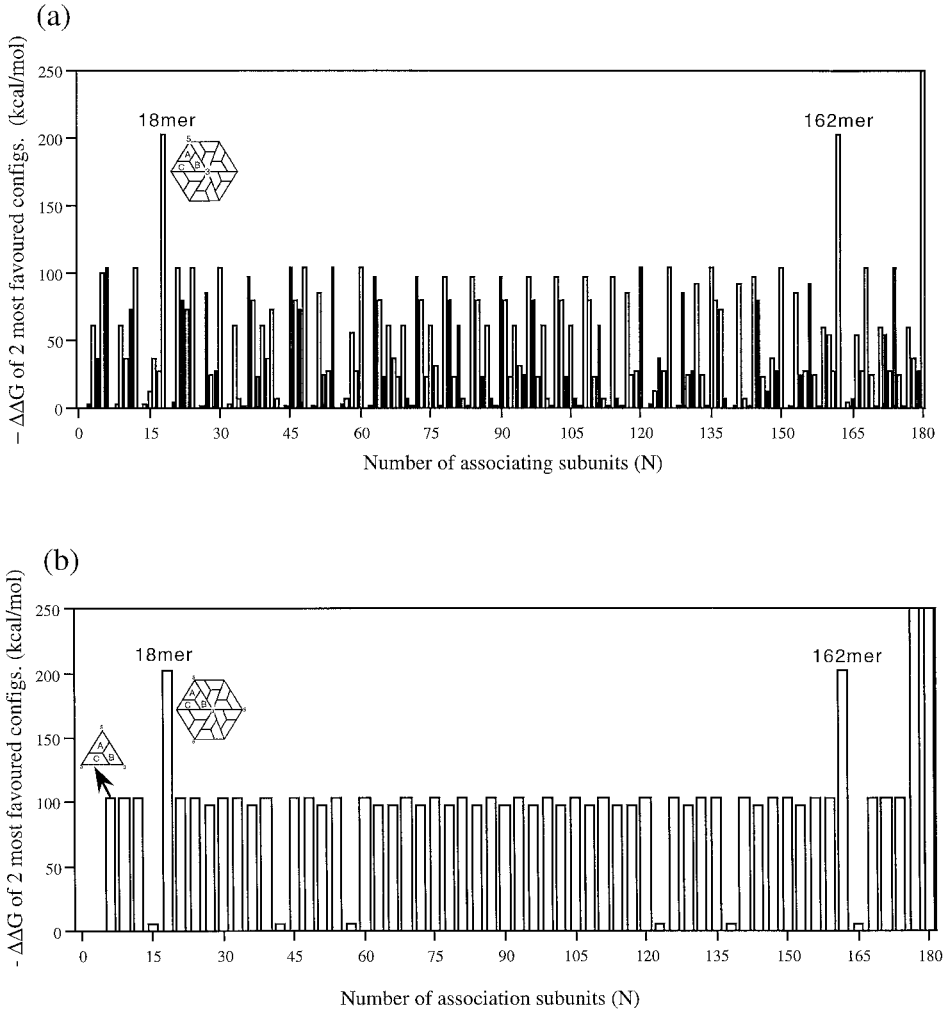


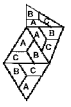
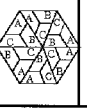
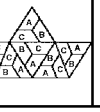
FIGURE 5 (*a*) A pathway of BBV assembly derived from the combinatorial procedure, assuming a subunit monomer as the assembling unit. The (negative) difference in association energies of the best and second best configurations is plotted as a function of number of associating subunits. The larger the peak, the greater the likelihood of that particular configuration becoming a stable intermediate; smaller differences refer to conformational polymorphism. A schematic representation of the promiment intermediates is shown. (*b*) A pathway of BBV assembly, assuming the trimeric (ABC) protomer as the assembling unit. A peak at each multiple of trimer of subunits in *a* suggests that it is a preferred assembling unit. Near-zero values of $\Delta\Delta G$ for certain structures, which are also made up of multiples of trimers, suggest that these structures are polymorphic in nature and are essential for the assembly of quasiequivalent capsids.

ruses in general, proceed with the addition of trimeric asymmetrical units. Fig. 5 *b* shows the assembly of BBV, with the ABC trimer as an assembling unit. The repetitive nature of the trimer assembly and the similar preferred structures found in both monomer and trimer assembly pathways suggest that the trimeric protomer is formed readily and often, creating a reservoir of these units, so that BBV assembly proceeds as the continuous addition of trimers (Fig. 6).

Fig. 6 shows that there are interesting patterns as the trimers assemble. The 15mer is of interest because there are two structures of nearly identical energy, a closed pentamer of trimers and a "hexamer" (of trimers) with one trimer missing. The similarity in stability of these two structures is not intuitively obvious, because it appears that the closed pentamer has much more buried surface area than the open "hexamer" (Table 1). The closed pentamer, however, is composed of all bent joint contacts that are less stable than the flat joint contacts by nearly 100 kcal/mol, and the combination of bent and flat joint contacts in the open "hexamer" is equivalent to the contacts determining the true pentamer. The occurrence of isoenergetic structures that fluctuate between a hexamer and pentamer is consistent with the assembly properties of a quasiequivalent shell in which such switching must occur. Thus the pathway defined by this approach has revealed a point of conformational polymorphism that is expected in the assembly pathway of a quasiequivalent shell. The energetics also reveal that the subunits that form the pentamer as an open "hexamer" will move into the most stable intermediate in the assembly, the closed hexamer of trimers (18mer), which is a notable organization found in the particle that is stabilized by RNA. Experimental results demonstrate that deleting residues 1–50 or residues 364–407, both regions that interact with RNA and appear essential for 18mer formation, disrupts the assembly of the virus particle (Schnemann et al., unpublished results).

Similar calculations were performed with the association energies of SBMV and HRV14, and the assembly pathways based on the combinatorial approach are shown in Fig. 7. Each of these viruses has a novel pathway when the plots are compared, and it is clear that these are determined by the ratios of the intersubunit association energies listed in Table 2. Experimental data for assembly are only available for HRV14, where trimer protomers (but not the same trimer as observed in BBV) and pentamers of trimers (15mer) are

(a)

# units	1st	$-\Delta\Delta G_{12}$	2nd	$-\Delta\Delta G_{23}$	3rd
2		1.0		4.0	
3		58.0		80.0	
4		34.0		45.0	
5		102.0		1.0	
6		105.0		32.0	
9		58.0		47.0	
12		105.0		32.0	
15		16.0		8.0	
18		202.0		8.0	

(b)

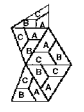
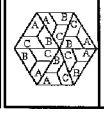
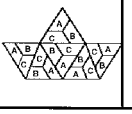
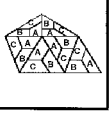
1st	$-\Delta\Delta G_{12}$	2nd	$-\Delta\Delta G_{23}$	3rd
	105.0			
	105.0			
	105.0		105.0	
	8.0		97.0	
	202.0		8.0	

FIGURE 6 (*a*) A table showing the top three preferred configurations for each association of subunits in the computed assembly pathway for BBV, with the monomer as the assembling unit. The first column shows the number of associating monomers. Columns 2, 4, and 6 show a schematic of the three best structures for each association. $-\Delta\Delta G_{12}$ and $-\Delta\Delta G_{23}$ refer to the negative differences of the association energies of the first and second and second and third configurations. (*b*) The preferred structures, with the trimer as the assembling unit. It is important to note that the best configurations for both assembly pathways are nearly always the same; in some cases even the second best is the same, emphasizing that the trimer is the likely assembling unit. An exception is the best structure of the 15mer association. In this case the most stable monomer assembly is not made up of a multiple of protomers, but its preference, compared to the second and third most stable structures, which are made of protomers, is marginal.

stable intermediates that can be visualized on sucrose gradients, and the 15mer is known to be the module that adds to the growing shell. These assembly intermediates are clearly identified in the combinatorial approach. It is also obvious that HRV14 assembly is much more directed than the quasiequivalent viruses, with larger free energy differences between the most stable and the next most stable structures throughout the particle assembly process. This is consistent with no requirement for conformational polymorphism in a $P = 3$ shell. The differences in the energies between the two best structures for any given oligomer in the SBMV pathway are at least sevenfold less than those calculated for HRV14, suggesting a less directed pathway for SBMV.

CONCLUSIONS

The analysis reported here demonstrates a strategy for the computation of intersubunit association energies and the role of individual residues in stabilizing these interfaces. A comparison of the relative stabilities of the unique dimer interfaces within each virus suggested that the relative ratios of association energies provide a novel and reduced description of the structure that is unique for each of the three viruses examined. As subunit interfaces must relate in some way to assembly, an attempt to understand the significance of the relative stabilization of dimer contacts was made by examining potential pathways of assembly with these energies and the combinatorial approach of subunit association (Horton and Lewis, 1992).

Although there are substantial approximations in this approach to virus assembly, properties of the pathways that emerged are consistent with expectations for the formation of $T = 3$, quasiequivalent shells for BBV and SBMV and the $P = 3$, pseudoequivalent shell of HRV14. The presence of conformational polymorphism in the former is consistent with switching that must occur in quasiequivalent structures, whereas a highly directed pathway is consistent with the latter, where each of the dimer interfaces is formed by a unique set of amino acids. The isoenergetic states observed in the quasiequivalent capsid assembly pathways are analogous to observations in the computational studies of protein folding. In the latter studies, a large gap between the ground state (native structure) and numerous compact nonnative states of comparable energy provides information about the overall stability of the native state (Shakhnovich and Gutin, 1993).

It is also clear that each set of the dimer association energies (Table 1) and their ratios gives rise to a unique assembly pathway. Studies of combinatorial assembly using

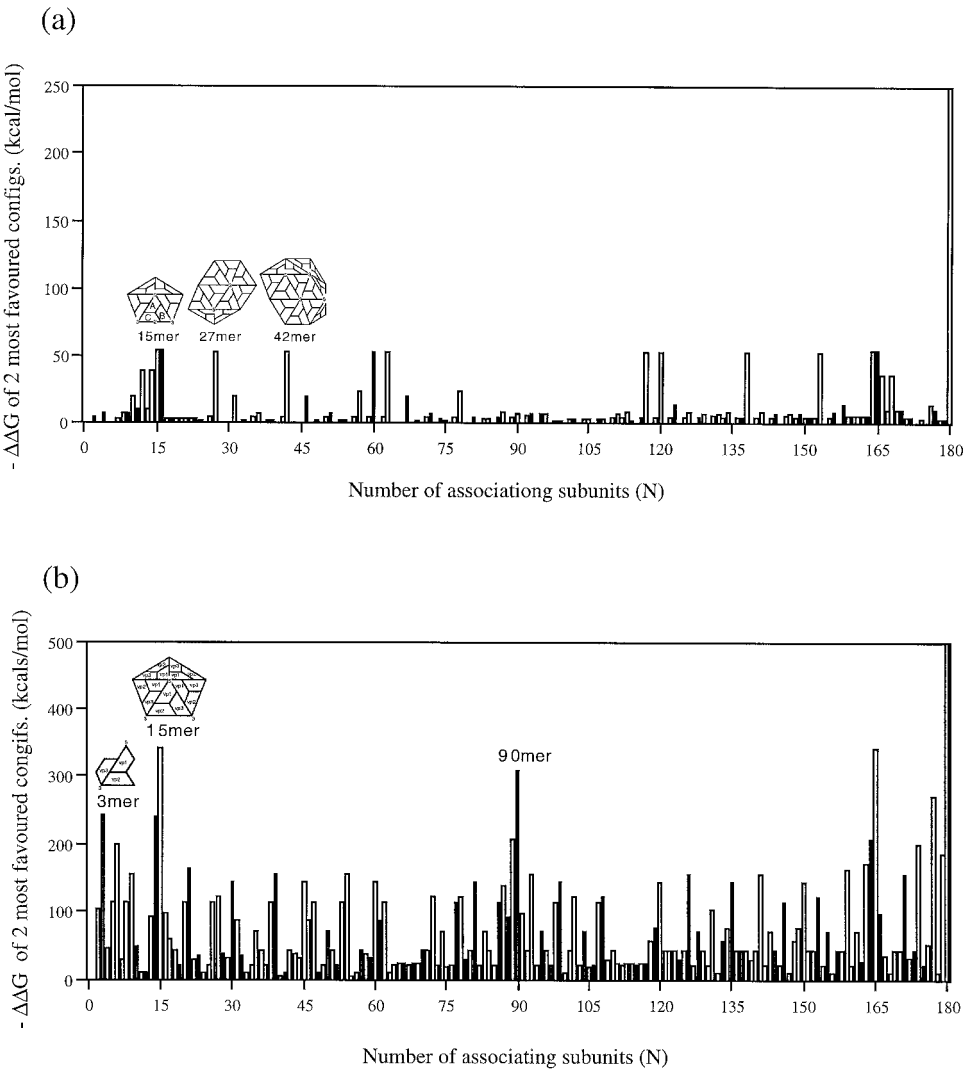


FIGURE 7 (*a*) Assembly pathway for SBMV with monomers as the assembling unit. Small $\Delta\Delta G$ values suggest that the SBMV pathway might not have any stable intermediates and would be less directed and more polymorphic in nature. Schematic representations of some of the likely intermediates are shown. (*b*) Pathway of assembly for HRV14. The protomer made up of VP1, VP0, and VP3 is the first preferred structure. The pentamer of trimers (15mer) is a strong intermediate ($-\Delta\Delta G$ = 350 kcal/mol), in agreement with experimental results, where 5s and 14s particles correspond to the protomer (3mer) and 15mer structures, respectively (Rueckert, 1985).

totally hypothetical ratios of dimer stabilities consistent with features of quasiequivalence found in different virus crystallographic structures indicated that pathways of assembly change with particular sets of ratios. This suggests an experimental approach to validate the computational results.

An inventory of interfacial residues and their contribution to the dimer association energies is provided by our calculations. The effect of site-directed mutations of these residues can be estimated if it is assumed that side-chain orientations are approximately similar in the native and mutant structures. It should be possible to alter the association energies in a predictable way with this approach and to employ a heterologous expression system to examine the assembly of site-directed mutants. A baculovirus expression system has been used for mutational studies of the maturation cleavage (Schneemann et al., 1993) and protein-RNA interactions (Schneemann et al., unpublished results) in nodaviruses. Straightforward use of this system will permit the study of alterations in the assembly pathways of nodaviruses. It is anticipated that when the ratios of dimer interfaces are altered beyond a particular point, the assembly pathway will change and aberrant associations will result. These are easily detectable by sucrose gradient analysis of the baculovirus expression products, and the experimental studies will provide an important assessment of the computational work described.

We thank Profs. Jeffrey Skolnick and Anette Schneemann for helpful discussions and Dawn Dunsmore for help in preparation of the manuscript.

This work was supported by grants GM34220 to JEJ, GM48807 and GM37554 to CLB, and AI39639 to CBP. This is manuscript number 10689MB at The Scripps Research Institute.

REFERENCES

Arnold, E., and M. G. Rossmann. 1990. Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 Å. *J. Mol. Biol.* 211:763–801.

Brooks, B., B. Bruccoleri, D. Olafson, D. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculation. *J. Comp. Chem.* 4:187–217.

Canady, M., S. Larson, J. Day, and A. McPherson. 1996. Crystal structure of turnip yellow mosaic virus. *Nature Struct. Biol.* 3:771–781.

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

Dasmahapatra, B., R. Dasgupta, K. Saunders, B. Selling, T. Gallagher, and P. Kaesberg. 1986. Infectious RNA derived by transcription from cloned cDNA copies of the genome RNA of an insect virus. *Proc. Natl. Acad. Sci. USA.* 83:63–66.

Fisher, A., and J. Johnson. 1993. Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature.* 361:176–179.

Gilson, M., and B. Honig. 1988. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies and conformational analysis. *Proteins.* 4:7–18.

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

Horton, N., and M. Lewis. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* 1:169–181.

Munshi, S., L. Liljas, W. Cavarelli, W. Bomu, B. McKinney, V. Reddy, and J. Johnson. 1996. The 2.8 Å structure of $T = 4$ animal virus and its implications for membrane translocation of RNA. *J. Mol. Biol.* 261: 1–10.

Rossmann, M., and J. Johnson. 1989. Icosahedral RNA virus structure. *Annu. Rev. Biochem.* 58:533–573.

Rueckert, R. 1985. Picornaviruses and their replication. *In* Virology. Raven Press, New York. 705–738.

Schneemann, A., R. Dasgupta, J. Johnson, and R. Rueckert. 1993. Use of recombinant baculoviruses in synthesis of morphologically distinct virus-like particles of Flock House virus, a nodavirus. *J. Virol.* 67: 2756–2763.

Shakhnovich, E. I., and A. M. Gutin. 1993. Engineering of stable and fast folding sequences of model proteins. *Proc. Natl. Acad. Sci. USA.* 90: 7195–7199.

Sharp, K., and B. Honig. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–332.

Silva, A., and M. Rossmann. 1987. Refined structure of southern bean mosaic virus at 2.9 Å resolution. *J. Mol. Biol.* 197:69–87.

Smith, T., E. Chase, T. Schmidt, N. Olson, and T. Baker. 1996. Neutralizing antibody to human rhinovirus 14 penetrates the receptor-binding canyon. *Nature.* 383:350–354.

Speir, J., S. Munshi, G. Wang, T. Baker, and J. Johnson. 1995. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by x-ray crystallography and cryo electron microscopy. *Structure.* 3:63–78.

Wery, J., V. Reddy, M. Hosur, and J. Johnson. 1994. The refined three-dimensional structure of an insect virus at 2.8 Å resolution. *J. Mol. Biol.* 235:565–586.

Williams, R. 1979. The Geometric Foundation of Natural Structure. Dover Publications, New York.

Zlotnick, A., V. Reddy, R. Dasgupta, A. Schneemann, W. Ray, R. Rueckert, and J. Johnson. 1994. Capsid assembly in a family of animal viruses primes an autoproteolytic maturation that depends on a single asparatic acid residue. *J. Biol. Chem.* 269:13680–13684.