

## GUEST COMMENTARY

# Virus Particle Explorer (VIPER), a Website for Virus Capsid Structures and Their Computational Analyses†

VIJAY S. REDDY,\* PADMAJA NATARAJAN, BRIAN OKERBERG, KEVIN LI, K. V. DAMODARAN, RYAN T. MORTON, CHARLES L. BROOKS III,\* AND JOHN E. JOHNSON\*

*Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037*

The number of icosahedral-capsid structures determined at a near-atomic level of resolution is growing rapidly as advances in synchrotron radiation sources, fast-readout detectors, and computer hardware and software are made. Hence, there is an increasing need to organize these mega-assemblies into a uniform and easy-to-use database. The coordinates of the icosahedral-capsid structures deposited in the Protein Data Bank (PDB) (2) follow a variety of conventions in which the icosahedral symmetry axes are oriented differently in the orthogonal coordinate system. While trying to analyze the various capsid structures en masse, we became aware of the need for a database in which all capsid structures (coordinates) are stored in a standard icosahedral orientation. Such a structural database of viral capsids would indeed facilitate the development of tools for high-throughput analyses of the virus structures. We report here the creation of a web-base (website and database) of virus structures, the Virus Particle Explorer (VIPER), which can be accessed through the World Wide Web (WWW) at the uniform resource locator (URL) <http://mmtsb.scripps.edu/viper/>. The organization of the VIPER database is shown in Fig. 1.

### TRANSFORMING THE PDB COORDINATES INTO VIPER CONVENTION

To describe icosahedral symmetry axes with respect to the orthogonal coordinate axes, two conventions,  $z(2)$ -3-5- $x(2)$  and  $z(2)$ -3-5- $y(2)$ , are commonly used, where 2, 3, and 5 correspond to the icosahedral 2-fold, 3-fold, and 5-fold symmetry axes, respectively, and  $x$ ,  $y$ , and  $z$  are the orthogonal coordinate axes. In the  $z(2)$ -3-5- $x(2)$  convention, two icosahedral 2-fold axes coincide with the  $z$  and  $x$  coordinate axes while a set of icosahedral 3-fold and 5-fold axes lie, in that order, between the  $z$  and  $x$  axes in the  $xz$  plane (Fig. 2). Similarly, in the  $z(2)$ -3-5- $y(2)$  orientation, a pair of 2-fold axes coincide with the  $z$  and  $y$  axes and a 3-fold and a 5-fold axis lie in between in the  $zy$  plane. These two orientations are related to each other by a 90° rotation around the  $z$  axis. All of the entries in the VIPER web-base are stored uniformly in the  $z(2)$ -3-5- $x(2)$  convention.

\* Corresponding author. Mailing address: Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, CA 92037. Phone for Vijay S. Reddy: (858) 784-8191. Fax: (858) 784-8660. E-mail: [reddiv@scripps.edu](mailto:reddiv@scripps.edu). Phone for Charles L. Brooks III: (858) 784-8035. Fax: (858) 784-8688. E-mail: [brooks@scripps.edu](mailto:brooks@scripps.edu). Phone for John E. Johnson: (858) 784-9705. Fax: (858) 784-8660. E-mail: [jackj@scripps.edu](mailto:jackj@scripps.edu).

† Manuscript number 14160-MB of The Scripps Research Institute.

Initially, the coordinates of each entry in the PDB were transformed to one of the two standard icosahedral orientations depending on the depositor's reference convention. They were then converted into the VIPER ( $z$ -3-5- $x$ ) convention. In addition, a reference icosahedral asymmetric unit was chosen for each type of triangulation (T) lattice (e.g., T=1, T=3, or T=4). Accordingly, the coordinates were transformed to the reference asymmetric unit and stored in the VIPER database. The product of these matrices is stored as a single transformation matrix, the PDB TO VIPER matrix, making the conversion between the VIPER and PDB coordinates straightforward.

The transformed coordinates were further processed so that the amino acid residues were placed at the top of the file, after which followed the nucleic acid residues, metal ions, and water molecules, if present. Atoms with missing chain identifiers (IDs) were given chain IDs, and these changes were documented in the REMARK lines at the beginning of the coordinate file. This rearrangement of the coordinates was required to facilitate the analysis of the capsid structures in a rapid and automatic fashion by using the molecular mechanics program CHARMM (3).

### ORGANIZATION AND DESCRIPTION OF THE ENTRIES

Presently there are 53 unique capsid structures available at the VIPER site. The coordinates of the capsids whose structures have been published but not yet deposited in the PDB were acquired through personal communications. The entries are listed in alphabetical order, as well as classified in terms of T number, which usually refers to the number of copies of the capsid protein subunits that occupy the icosahedral asymmetric unit. For each structure in the database, a web page was created with a pictorial description of the capsid and subunit structures as well as structural characteristics and related information available on the virus (e.g., diameter, T number, PDB ID, etc.).

Individual pages show the overall molecular surface rendering of the complete capsid, color coded according to the  $z$  coordinate, created to resemble the cryoelectron microscopy reconstruction. The tertiary fold of the capsid protein subunit(s), the capsomeric organization of the subunits, and a geometrical representation of the quasi-equivalent lattice are also provided. The rendered surface drawings were created with the program GRASP (9). The remaining figures were generated using the programs MOLSCRIPT (6) and RASTER3D (7). A typical web page at the VIPER site is shown in Fig. 3. While the individual VIPER pages show the figures in reduced sizes, each figure can be magnified to its original size by click-

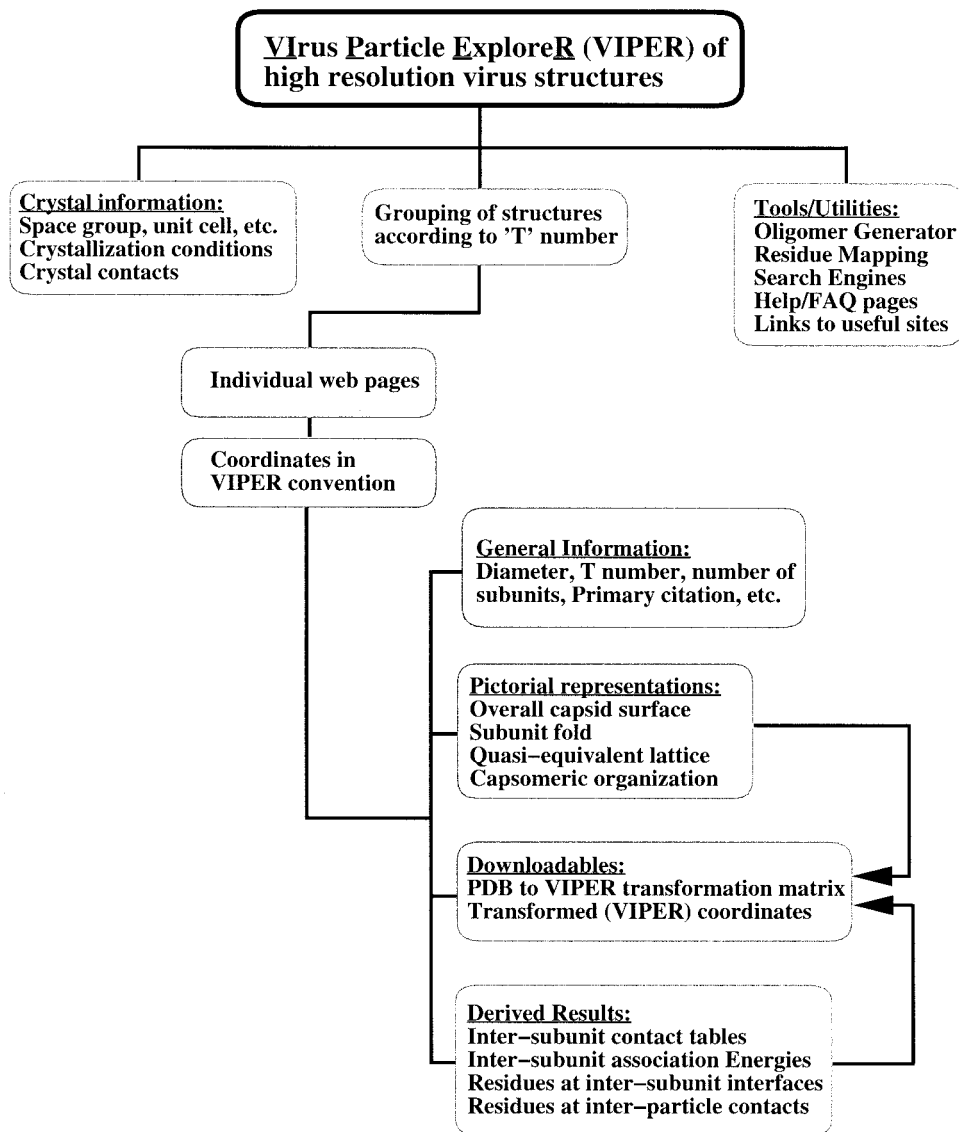


FIG. 1. Flow chart showing the organization of the contents of the VIPER site. The VIPER database contains the structures of viral capsids determined at a nearly atomic-level resolution. Coordinates of the capsid structures are stored in the  $z(2)$ -3-5- $x(2)$  convention. For each structure in the database, a web page showing the molecular structure, general information, and derived results of various structural and computational analyses was created. FAQ, frequently asked question.

ing on the figure. The overall surface renderings were generated on a fixed scale, enabling the comparison of the relative sizes of different capsids as well as their surface features. General information on each capsid structure (such as diameter, T number, resolution at which the structure was determined, and crystallization conditions) that is available in the literature was also catalogued. In addition, hyperlinks to the corresponding entries in different databases, such as PDB (2) and SWISS-PROT (1), and links to the abstract of the primary reference in the PubMed database (<http://www.ncbi.nlm.nih.gov/>) are listed. All of the figures, data, and results provided in these pages can be readily downloaded.

Even though different serotypes of viruses were not considered to be part of the unique set of entries, separate pages were created for these structures, and they can be accessed from the main lists of entries. However, VIPER pages were generated only

for a few mutants and capsid structures complexed with small ligands, and these pages can be accessed through the corresponding main (unique) virus entries. Including the different serotype and mutant structures, there are presently 72 entries in the VIPER web-base. The database is updated when a new virus or capsid entry gets deposited in the PDB.

#### ANALYSIS AND WEB TOOLS

A number of tools were developed for the analysis of the entries in the VIPER web-base, primarily by employing the molecular mechanics program CHARMM (3) and web-based scripting and programming languages such as practical extraction and report language and Java. These were used to derive contact tables, examine the fidelity of quasi-symmetry, compute association energies, and estimate the residuewise contri-

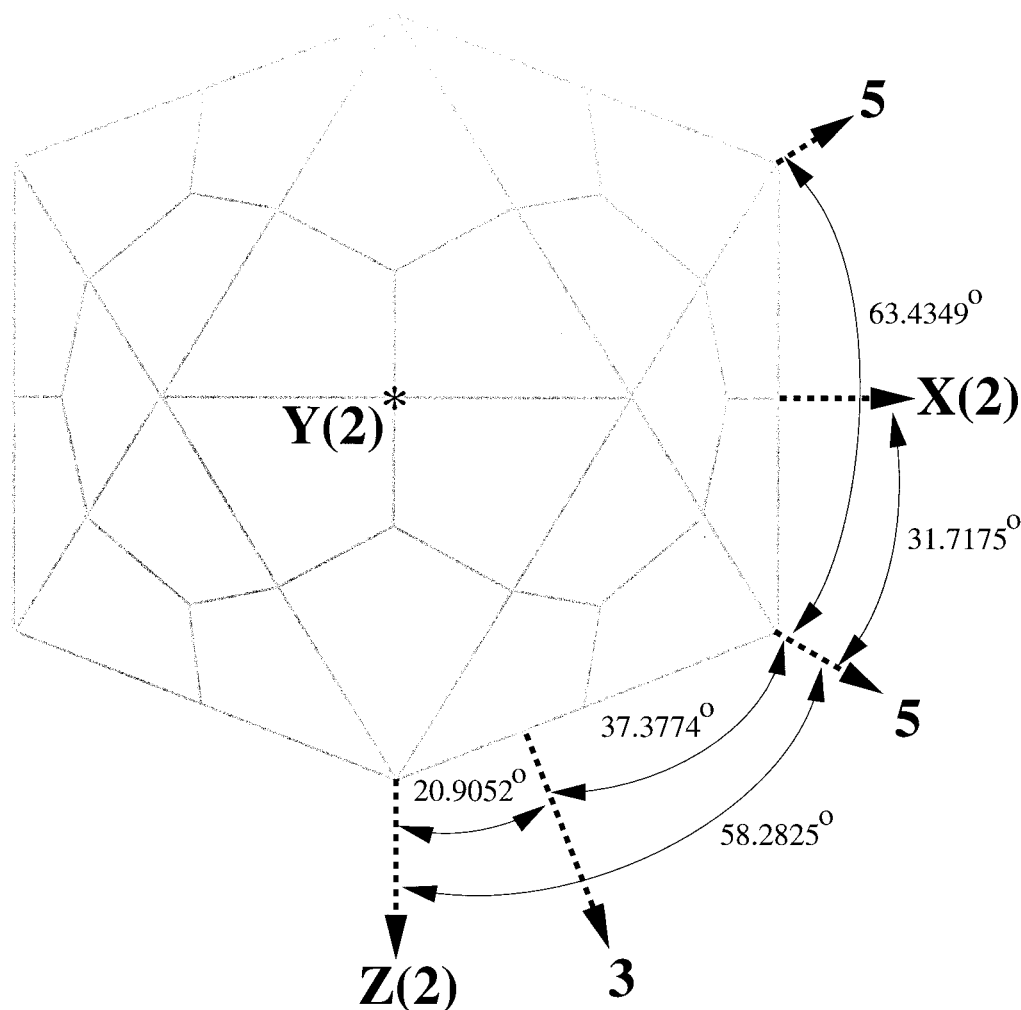


FIG. 2. Schematic diagram of the VIPER convention,  $z(2)$ - $3$ - $5$ - $x(2)$ .  $x$ ,  $y$ , and  $z$  are the orthogonal coordinate axes, and the numbers 2, 3, and 5 correspond to 2-fold, 3-fold, and 5-fold icosahedral symmetry axes, respectively. In this convention, the icosahedron is oriented such that a pair of icosahedral 2-fold axes are coincident with the  $z$  and  $x$  axes and a 3-fold/5-fold axis pair lies between the  $z$  and  $x$  axes in the  $xz$  plane. The coordinates of the all entries in the database are stored in this convention.

butions at the unique subunit-subunit interfaces. The contact tables were generated with the residue-residue contacts determined by the simple distance criterion at the unique intersubunit interfaces. In addition, these contacts were used to quantify the extent of quasi-equivalence (K. V. Damodaran, V. S. Reddy, J. E. Johnson, and C. L. Brooks, submitted for publication). The intersubunit association energies were estimated by multiplying the loss of surface area for each atom by the atomic solvation parameter (4, 5, 10). The buried surface areas were calculated by using the program CHARMM (3) and a probe radius of 1.4 Å. Residuewise contributions were extracted from the above-noted association energies. Interparticle packing contacts in the crystal lattice were generated as described by Natarajan and Johnson (8). All of these tools are being adapted to the WWW and will be available, through the VIPER site, for the analysis of user-supplied virus coordinates.

The following tools are presently available for use at the VIPER site. Oligomer Generator is a tool that provides the coordinates of a chosen oligomer structure or the entire capsid for any given entry in the VIPER database. The Residue Mapping tool

can be employed to identify a particular residue and to view its location in the structure, using the Virtual Reality Markup Language viewer. In addition, a number of web tools, such as search engines, help pages, and frequently asked questions, were included for ease of navigation through the VIPER web-base.

## DERIVED RESULTS

The results generated by the various analyses mentioned above are available as web pages for individual entries. Table 1 shows a representative list of residue-residue contacts at the quasi-3-fold related interfaces of black beetle virus (PDB-ID: 2BBV). A number of contacts between the same residues at the latter interfaces are aligned quite well, indicating that the fidelity of the quasi-3-fold symmetry is highly conserved. Figure 4 shows the association energies and buried surface areas at the unique interfaces for the entry 2BBV. The derived results of various other analyses (e.g., interacting residues at the interfaces and residues involved in interparticle contacts in the crystal lattice) can be readily accessed from the website.



## Black Beetle Virus

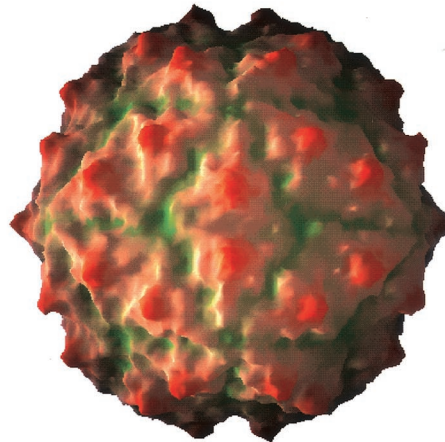
### Resume:

PDB-ID :2BBV  
 Resolution : 2.8Å  
 A.A. Seq.Acc.# :P04329  
 T number : 3  
 # of Subunits :180  
 Diameter (Max) :346.5Å

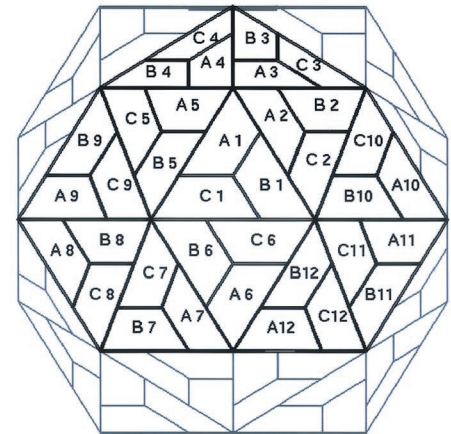
### Primary Citation

### Download Files:

[Header](#)  
[PDB to VIPER matrix](#)  
[Transformed Coordinates](#)



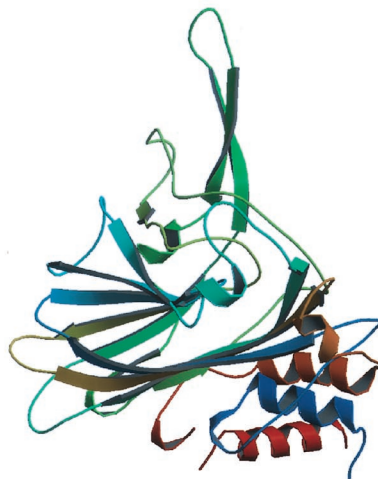
Rendered Surface



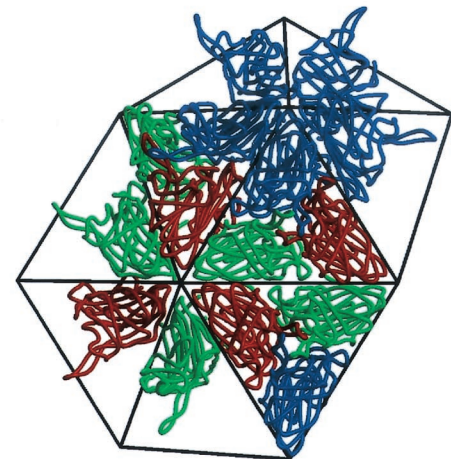
T=3 lattice

### Computations:

[Contact Tables](#)  
[Inter Subunit Association energies](#)  
[Residues at the interfaces](#)  
[Crystal Contacts](#)



Subunit fold



Capsomeres

FIG. 3. A typical web page of an entry at the VIPER site. The web page for black beetle virus (PDB ID:2BBV) shows the molecular surface of the capsid, the tertiary fold of the constituent subunits, a geometrical representation of the quasi-equivalent lattice, and the representative oligomeric organization of the subunits in the capsid. The molecular surface of the capsid is color coded as a function of the z coordinate. The tertiary structure of the subunit fold is color coded from blue to red going from the N to the C terminus. In the capsomeric representation, the A, B, and C subunits are shown in blue, red, and green, respectively. Listed on the left are related information on the virus; links to corresponding entries in the PDB, SWISS-PROT, and MEDLINE databases; transformed coordinates; and derived results. A.A. Seq. Acc. #, amino acid sequence accession number in the SWISS-PROT database.

Subunit Interface	Quasi/icosahedral Symmetry	Association Energies (kcal/mol)	Buried Surface Area (Å <sup>2</sup> )
A1_B1	Quasi 3-fold	-84.0	4262.0
B1_C1	Quasi 3-fold	-70.0	3621.0
C1_A1	Quasi 3-fold	-72.0	3662.0
B5_A1	Quasi 2-fold	-36.0	1877.0
C1_C6	Icos. 2-fold	-21.0	1402.0
A1_A5	Icos. 5-fold	-36.0	1982.0
B5_C1	Quasi 5-fold	-37.0	1992.0
C1_B6	Quasi 6-fold	-26.0	1624.0
A1_C6	Quasi 6-fold	-18.0	1207.0
C1_R1	--	-6.0	562.0
C1_R6	--	4.0	119.0
R1_R6	--	-4.0	816.0

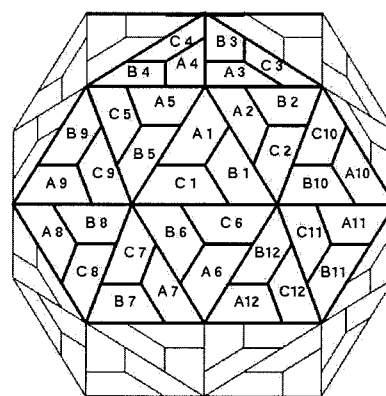


FIG. 4. Intersubunit association energies for the unique subunit interfaces present in the capsid of black beetle virus (PDB-ID: 2BBV). On the left is a table listing the different interfaces, quasi-symmetry associated with the interface, association energies, and buried surface areas. Shown on the right is the geometric representation of the T=3 quasi-equivalent lattice. Each trapezoid represents a subunit. The letters A, B, and C identify the trapezoids representing three different (unique) environments that subunits occupy in the T=3 capsid, and the numbers (1, 2, 3, ...) refer to different icosahedral asymmetric units. The A1, B1, and C1 subunits form the reference asymmetric unit.

TABLE 1. Intersubunit contacts for the entry 2BBV<sup>a</sup>

A1-B1	B1-C1	C1-A1
W:191_P:325 (H-H)	W:191_P:325 (H-H)	W:191_P:325 (H-H)
W:191_P:325 (H-H)	W:191_P:325 (H-H)	W:191_P:325 (H-H)
P:194_S:164 (H-P)	P:194_S:164 (H-P)	P:194_S:164 (H-P)
P:194_R:322 (H-B)	P:194_R:322 (H-B)	P:194_R:322 (H-B)
P:194_N:324 (H-P)	P:194_N:324 (H-P)	P:194_N:324 (H-P)
K:196_S:164 (B-P)	K:196_S:164 (B-P)	K:196_S:164 (B-P)
K:196_S:165 (B-P)	K:196_S:165 (B-P)	K:196_S:165 (B-P)
K:196_D:254 (B-A)	K:196_D:254 (B-A)	K:196_D:254 (B-A)
K:196_R:322 (B-B)	K:196_R:322 (B-B)	K:196_R:322 (B-B)
S:198_D:254 (P-A)	S:198_D:254 (P-A)	S:198_D:254 (P-A)
S:198_L:256 (P-H)	S:198_L:256 (P-H)	S:198_L:256 (P-H)
N:199_H:215 (P-B)	N:199_H:215 (P-B)	N:199_H:215 (P-B)
N:199_L:256 (P-H)	N:199_L:256 (P-H)	N:199_L:256 (P-H)
V:200_E:257 (H-A)	V:200_E:257 (H-A)	V:200_E:257 (H-A)
Q:201_H:215 (P-B)	Q:201_H:215 (P-B)	Q:201_V:214 (P-H)
Q:201_G:258 (P-H)	Q:201_I:259 (P-H)	Q:201_H:215 (P-B)
Q:201_I:259 (P-H)	Q:201_I:259 (P-H)	Q:201_I:259 (P-H)
Q:201_P:264 (P-H)	Q:201_P:264 (P-H)	Q:201_P:264 (P-H)
P:203_A:265 (H-H)	P:203_A:265 (H-H)	P:203_A:265 (H-H)
P:203_N:266 (H-P)	P:203_N:266 (H-P)	

<sup>a</sup> Shown are representative residue-residue contacts that stabilize different quasi-3-fold interfaces (A1-B1, B1-C1, and C1-A1) in the capsid of black beetle virus (PDB-ID:2BBV). A1, B1, and C1 correspond to the subunits occupying the structurally unique environments in the T=3 icosahedral lattice (Fig. 4). A1-B1, B1-C1, and C1-A1 refer to the intersubunit interfaces between the A1 and B1, B1 and C1, C1 and A1 subunits, respectively. The interresidue contacts are also distinguished by the types of residues involved (e.g., hydrophobic [H], polar [P], acidic [A], and basic [B]). The same residue-residue contacts are aligned across different interfaces such that absence of contacts corresponds to breakdown of the quasi-3-fold symmetry. The fidelity of contacts across these interfaces suggests that the quasi-3-fold symmetry in 2BBV is quite well maintained. The complete lists of contacts are available at the VIPER website.

**Future directions.** With the above-described framework in place, we plan to study various aspects of structure, stability, and assembly of spherical viruses by using computational methods. These analyses will facilitate the local and global structural comparisons of virus capsid structures within and between the families of viruses. The derived results will be made available at the VIPER website as they are obtained.

VIPER is being developed as a training/service and dissemination component of the Research Resource: Multi-Scale Modeling Tools for Structure Biology (MMTSB), which is fully supported by the National Center for Research Resources of the National Institutes of Health (RR12255).

#### REFERENCES

- Bairoch, A., and R. Apweiler. 1997. The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucleic Acids Res.* **25**:31–36.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**:235–242.
- Brooks, B., B. Brucoleri, 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.
- Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature* **319**:199–203.
- Horton, N., and M. Lewis. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**:169–181.
- Kraulis, P. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**:946–950.
- Merritt, E. A., and J. D. Bacon. 1997. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**:505–524.
- Natarajan, P., and J. E. Johnson. 1998. Molecular packaging in virus crystals: geometry, chemistry, and biology. *J. Struct. Biol.* **121**:295–305.
- Nicholls, A., K. A. Sharp, and B. Honig. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**:281–296.
- Reddy, V. S., H. A. Giesing, R. T. Morton, A. Kumar, C. B. Post, C. L. Brooks III, and J. E. Johnson. 1998. Energetics of quasiequivalence: computational analysis of protein-protein interactions in icosahedral viruses. *Biophys. J.* **74**:546–558.