Structural rigidity in the capsid assembly of cowpea chlorotic mottle virus

2004 J. Phys.: Condens. Matter 16 S5055

(http://iopscience.iop.org/0953-8984/16/44/003)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **16** (2004) S5055–S5064 PII: S0953-8984(04)85835-1

Structural rigidity in the capsid assembly of cowpea chlorotic mottle virus

B M Hespenheide[1], D J Jacobs[2] and M F Thorpe[1]

[1] Department of Physics and Astronomy, Arizona State University, PO Box 871504, Tempe, AZ 85287-1504, USA
[2] Department of Physics and Astronomy, California State University, 18111 Nordhoff Street, Northridge, CA 91330-8268, USA

Received 31 August 2004
Published 22 October 2004
Online at stacks.iop.org/JPhysCM/16/S5055
doi:10.1088/0953-8984/16/44/003

Abstract

The cowpea chlorotic mottle virus (CCMV) has a protein cage, or capsid, which encloses its genetic material. The structure of the capsid consists of 180 copies of a single protein that self-assemble inside a cell to form a complete capsid with icosahedral symmetry. The icosahedral surface can be naturally divided into pentagonal and hexagonal faces, and the formation of either of these faces has been proposed to be the first step in the capsid assembly process. We have used the software *FIRST* to analyse the rigidity of pentameric and hexameric substructures of the complete capsid to explore the viability of certain capsid assembly pathways. *FIRST* uses the 3D *pebble game* to determine structural rigidity, and a brief description of this algorithm, as applied to body–bar networks, is given here. We find that the pentameric substructure, which corresponds to a pentagonal face on the icosahedral surface, provides the best structural properties for nucleating the capsid assembly process, consistent with experimental observations.

1. Introduction

The life cycle of a virus that results in productive infection generally consists of four steps:

(1) entry into a host cell and release of the viral genetic material from the viral packaging,
(2) reproduction of the viral genome and production of new packaging proteins,
(3) assembly of new virus particles in which copies of the viral genome are repackaged, and
(4) release of the new virions from the cell.

The details for each step vary widely depending on the specific virus and are in many cases unknown, a fact that limits development of broad-spectrum anti-viral therapies. However, some systems, such as *influenza* and the *human immunodeficiency virus* (HIV), have been studied in great detail.

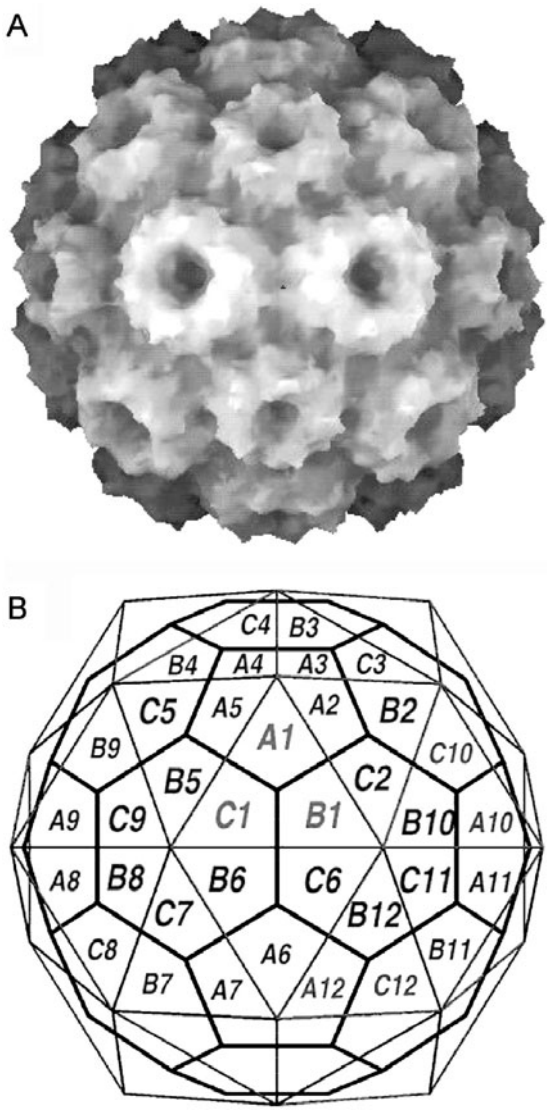

Figure 1. (A) The surface topology of the native form of CCMV [21]. (B) A schematic diagram indicating the threefold, fivefold and sixfold symmetry axes. The pentagonal and hexagonal faces are outlined in dark lines, indicating the isomorphism to a buckyball and/or a soccer ball.

One such system that has been well studied is the *cowpea chlorotic mottle virus* (CCMV). A member of the *Bromoviridae* family, CCMV is quite simple compared to other viruses, and has provided a model system for exploring various stages of the virus life-cycle. In particular, there is a wealth of experimental information pertaining to the assembly of the new virions inside the infected cell. This assembly process can be quite complicated, as it involves concurrently building a small package out of proteins, known as the *capsid*, and placing the genetic material of the virus inside this capsid. In the case of CCMV, the capsid is composed of 180 copies of a single protein that are symmetrically arranged to form an icosahedron (figure 1(A)).

There are several experiments that have led to our current understanding of how the CCMV capsid assembles. Perhaps the most important has been the determination of the structure of the completely formed capsid by using a combination of electron microscopy and x-ray diffraction techniques [1]. Figure 1 shows the complete capsid (pdb code: 1cwp), which has a diameter

of ∼280 Å, along with a schematic diagram illustrating the icosahedral symmetry. Additional experiments have provided insight into the assembly mechanism. Adolph and Butler [2] showed that the capsid protein can be isolated as a stable homodimer, and it is this dimer that represents the smallest 'building block' during CCMV capsid assembly. Subsequent studies of assembly *in vitro* by using light scattering suggest that the first step in the assembly process is the formation of a pentamer of dimers [3]. The absence of any other stable substructures early in the assembly process led to the conclusion that the pentamer of dimers is the nucleation structure for capsid assembly.

In this paper we present a theoretical analysis of CCMV capsid assembly. We show how the changes in the structural flexibility that occur when two substructures assemble qualitatively favour a pathway that begins with the pentamer of dimers binding a free dimer, consistent with experiment. Structural flexibility is measured by using the program *FIRST* ('floppy inclusions and rigid substructure topography'), which maps the chemical bonding information of the protein onto a generic 3D graph, in which the edges represent distance constraints between atoms. This graph is then decomposed into rigid and flexible regions. Since the number of atoms in viruses can become quite large, an improved version of the *pebble game* algorithm that requires less memory and runs more efficiently is implemented. These improvements are made possible by completely representing the molecular structure as a body–bar network. The new implementation of *FIRST* eliminates the use of ghost atoms to model hydrophobic constraints, and thus helps facilitate the analysis of networks containing millions of atoms, important in the study of viruses and other supramolecular assemblies.

2. Methods

2.1. FIRST flexibility analysis and the 3D pebble game

The program *FIRST* performs two general tasks. The initial step is to read in structural and chemical information for a protein, such as can be found in any x-ray crystal structure (available via the Protein Data Bank [4]). Processing this information yields a mechanical representation of the protein as a set of constraints on the distance between atom pairs. The second step performs an analysis on this distance constraint network using the 3D *pebble game* algorithm to identify regions that are overconstrained (hyperstatic or stressed), isostatically rigid, or underconstrained (hypostatic or flexible). This information is then mapped back onto the protein structure (now protein assembly). Earlier versions of the 3D pebble game used the bar–joint representation in which atoms are points with three degrees of freedom, whereas more recent implementations has used the body–bar representation in which the atoms are bodies with six degrees of freedom. These are believed to be equivalent as discussed in more detail below.

The details of how *FIRST* generates a network of distance constraints have previously been described [5, 6]. The resulting 3D network of distance constraints represents pairs of atoms that are at fixed distances from each other. The mathematical analysis of structural rigidity in such networks dates back to Maxwell [7]. In 1970 an important theorem by Laman [8] was established that provides combinatorial criteria for identifying independent constraints in 2D generic bar–joint networks. On the basis of Laman's theorem, efficient and exact algorithms for 2D graphs that test for network rigidity have been developed, the most popular being the 2D *pebble game* [9] where the atoms are represented as points that have two degrees of freedom. An analogous 3D bar–joint *pebble game* was constructed for a limited class of generic bond-bending networks [10], where now the atoms represented as points have three degrees of freedom. The original impetus for the 3D *pebble game* was to study rigidity percolation in million-atom 3D covalent glass networks [11].

In formulating the 3D *pebble game*, a connection was made [10] to earlier work from the 1980s by Tay [12] and Whiteley [13] on the '*molecular framework conjecture'*. The molecular framework conjecture provides a mathematical scaffold for representing molecular structure as a body–bar framework for identifying independent constraints, similar to Laman's theorem. A 3D *pebble game* was constructed for the body–bar representation, and (unpublished) extensive direct comparisons showed that the two kinds of 3D pebble games produce identical results when local rigid clusters of atoms are represented as bodies with six degrees of freedom. Although it is a pity that proofs are still unavailable, there is overwhelming supporting evidence that the algorithms for the two versions of the 3D *pebble game* are equivalent and exact. The body–bar representation was discussed in prior publications [10, 11, 14], and the 3D *pebble game* used in the program *FIRST* incorporates the body–bar implementation. Over the years, many extensions to *FIRST* have been made. In particular, hydrophobic constraints were modelled using extra 'ghost' atoms [15]. An undesirable consequence of adding ghost atoms is an increase in the effective size of the network. With recent application of *FIRST* to viruses, it has become prudent to eliminate ghost atoms. We present an improved modelling scheme using a (body–bar) 3D *pebble game* as it is currently implemented in *FIRST* (http://flexweb.asu.edu).

In the body–bar representation, rigid bodies, each having six degrees of freedom, define a set of vertices, and the set of generic bars that connect those bodies defines a network. The most essential element of any *pebble game* algorithm is the test for an independent constraint. Moreover, the identification of the set of independent constraints across the entire network is determined in a recursive fashion by building the network up by placing one constraint (bar) at a time. Part of this procedure requires basic operations such as pebble covering and pebble rearrangements. These basic pebble operations are common to all kinds of pebble game algorithms for which details can be found elsewhere [10, 11, 14, 16]. Now, each vertex is assigned six pebbles representing the three translations and three rotations associated with rigid body motions. Next, in arbitrary order, a stack of constraints is defined in the order in which they are to be placed in the network, as is done in the 2D pebble game [16]. A constraint can consist of 1–6 bars. Working down the stack, one constraint at a time, the following series of tests is applied [14]. Note that six constraints between two bodies lock the two bodies with respect to each other, so having more than six constraints would be redundant (and hence unnecessary).

The 3D body–bar pebble game algorithm:

(1) Place a constraint consisting of g bars between two vertices v_o and v_f.
(2) Check whether v_o and v_f are marked to belong to a Laman subgraph. If they belong to the same Laman subgraph go to step (7); otherwise continue.
(3) Rearrange the pebble covering to collect six pebbles on vertex v_o.
(4) Rearrange the pebble covering to collect g pebbles (or the maximum possible) on vertex v_f while holding the six pebbles on vertex v_o in place.
(5) If g pebbles are collected on vertex v_f, then all bars are independent. Extend the pebble covering by placing one pebble on each of the g-bars. Go to step (7).
(6) When q pebbles are collected on vertex v_f, for $q < g$, then q bars are independent. The other $(g-q)$ bars are redundant. The failed pebble search for the $(q+1)$th pebble defines a Laman subgraph (overconstrained region), which is recorded after merging the identified region with all overlapping Laman subgraphs previously recorded. After the merging, all vertices in the union of Laman subgraphs are condensed to a single vertex, selected to be the minimal label.

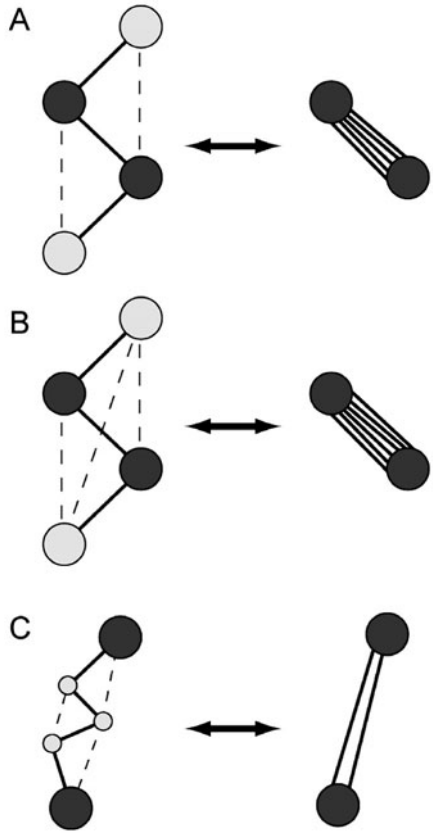

Figure 2. A series of mappings from the bar–joint (left-hand diagrams, with three degrees of freedom per site) to the body–bar representation (right-hand diagrams, with six degrees of freedom per site) for key interactions modelled in protein structures. The bond being modelled is between the two dark shaded spheres in each panel. The dashed lines represent bond-bending or angular constraints required in the bar–joint model. (A) A covalent bond is mapped to a constraint with five bars. (B) A peptide bond is mapped to a constraint with six bars, which prohibits dihedral rotation. (C) A hydrophobic tether, which previously required ghost atoms to account for their limited effect on the rigidity of a protein structure, is now mapped to a constraint with two bars. In all cases, the length of the bars does not affect the results, as all constraints are generic.

(7) If more constraints remain to be placed, return to step (1); otherwise the procedure is finished.

As in the bar–joint (2D or 3D) *pebble game* [11, 14, 16], here the pebble data structure only accounts for independent constraints. Overconstrained regions (or Laman subgraphs) are recorded (and merged with previously recorded regions) with an additional data structure as soon as they are identified, as described previously [11]. The condensation process is implemented by using the minimum vertex label within a Laman subgraph to replace all other vertex labels that belong within the same Laman subgraph. In practice, applying condensation of overconstrained regions allows the algorithm to perform nearly linearly with the number of vertices. Details pertaining to step (2) and the process of condensation in step (6) can be found elsewhere [10, 11, 14]. Correlated motions are determined in the same way as described previously [11].

This algorithm gives identical results to the bar–joint 3D *pebble game* for the number of independent degrees of freedom (floppy modes), rigid cluster decomposition, overconstrained regions, and correlated motions. Its advantage is that it is easier to implement and runs approximately 30% faster on identical input molecular structures. Moreover, the body–bar 3D *pebble game* couples to the molecular conjecture of Tay [12] and Whiteley [13] which is therefore invoked as the physical modelling scheme for representing molecular interactions. Molecular interactions are represented as b-bars, where $1 \leqslant b \leqslant 6$, between objects having six degrees of freedom, or even objects with less than six degrees of freedom as discussed in the physics literature [10, 11, 17]. We restrict ourselves to having exactly six pebbles associated

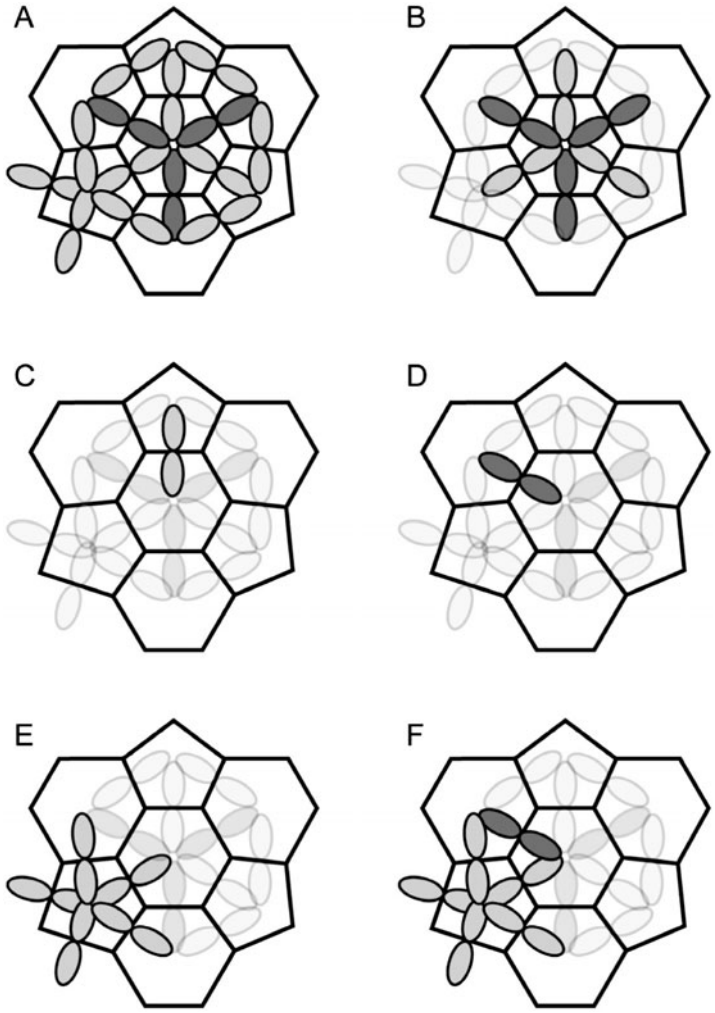

Figure 3. Projection of a portion of the icosahedral surface of the CCMV capsid showing the arrangement of the protein dimer building blocks. (A) The icosahedron surface can be divided into 20 hexagons and 12 pentagons. The protein dimer building blocks span the interface between the two kinds of polygons. Those across a hexamer–hexamer interface are shown as dark grey; those across the pentamer–hexamer interface are shown as light grey. (B) A hexamer of dimers, (C) a dimer from the pentamer–hexamer interface, (D) a dimer from the hexamer–hexamer interface, (E) a pentamer of dimers, (F) a pentamer of dimers plus one dimer.

with each atom, as this has an easy physical interpretation. Figure 2 shows the correspondence between the bar–joint and body–bar pictures for the interactions currently modelled in *FIRST*. For this set of interactions, the difference between these two pictures is only in perception. However, once the body–bar picture is adopted, a faster algorithm and more freedom in the way molecular interactions can be modelled result. There is the additional advantage that bars are only required between nearest neighbour bodies. The hydrophobic interaction is modelled by three pseudo-atoms in the bar–joint representation, and by two bars in the body–bar representation in figure 2(C). Only two constraints are used in the body–bar representation of the hydrophobic interaction on the right, as this allows hydrophobic atoms to be tethered

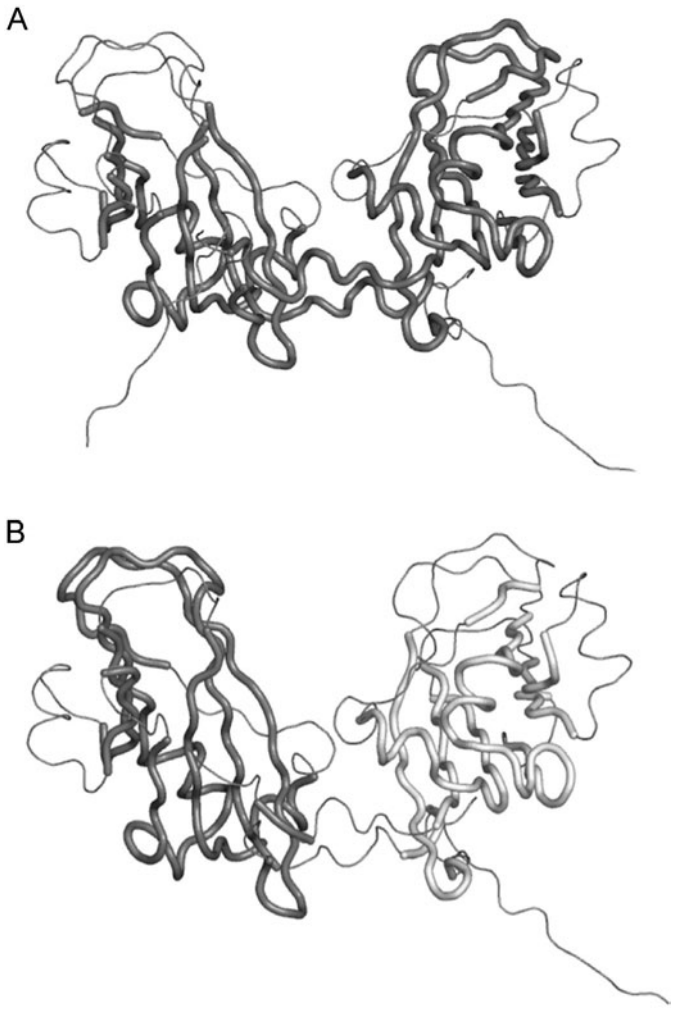

Figure 4. Flexibility results for two kinds of dimers. (A) A hexamer–hexamer dimer with a single combined rigid region and (B) a pentamer–hexamer dimer with two separate rigid regions. Rigid clusters are depicted by thick tubes; flexible bonds are shown with thin tubes.

locally while retaining considerable freedom of motion. The equivalence can be seen in that locking the four bonds in the bridge in the bar–joint representation of the left is equivalent to adding four additional bars to the body–bar representation on the right. In both cases a rigid link is established with no dihedral rotation allowed around a line connecting the two original atoms.

Analysis of CCMV assembly products. Hydrogens were added to the polar atoms of the complete capsid structure by using the software WhatIf [18, 19]. The following substructures (shown schematically in figures 3(B)–(F)) were isolated from the structure of the complete capsid: (1) a hexamer of dimers, (2) a dimer from the pentamer–hexamer interface, (3) a dimer from the hexamer–pentamer interface, (4) a pentamer of dimers, and (5) a pentamer of dimers with one additional dimer bound. Buried water molecules were predicted by using the software ProAct [20]. An energy cut-off of -0.35 kcal mol^{-1} was used for including hydrogen bonds in the *FIRST* analysis. This cut-off was chosen to be consistent with the hypothesis that the inner ring of proteins in the pentamer substructure forms a large rigid region providing stability to the pentamer, while each dimer partner is independently rigid. This energy cut-off was used in all *FIRST* analyses presented here.

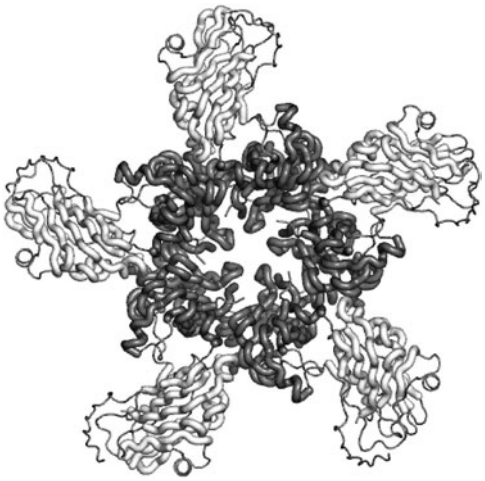

Figure 5. Flexibility results for the pentamer of dimers substructure from the CCMV capsid. The side-chain bonds are not displayed. Rigid clusters are depicted by thick tubes; flexible bonds are shown with thin tubes. The energy cut-off of -0.35 kcal mol^{-1} was specifically tuned to produce the result shown, in which the inner ring of proteins form a single rigid cluster (dark shade), and each of the outer proteins is rigid (light shade), but independent from the inner ring. Each of the five outer proteins is connected to the rigid core via flexible bonds.

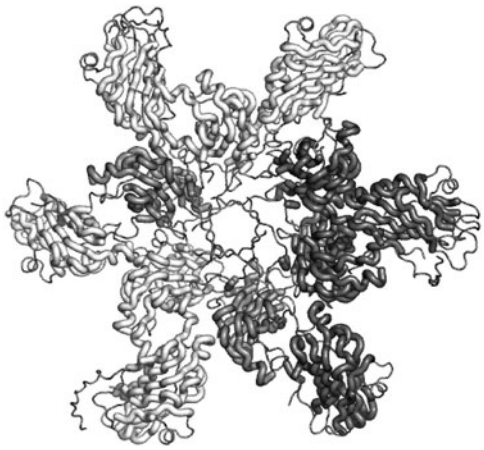

Figure 6. Flexibility results for the hexamer of dimers from the CCMV capsid. The side-chain bonds are not displayed. Rigid clusters are depicted by thick tubes; flexible bonds are shown with thin tubes. In contrast to the pentamer of dimers case shown in figure 5, there is no single rigid core that encompasses all of the dimers. Instead, the hexamer is non-symmetrically decomposed into five rigid clusters. Beginning with the darkest shaded rigid cluster on the right and continuing clockwise around the hexamer there are: a cluster of four proteins (darkest shade), a cluster of a single protein (medium shade), a cluster of three proteins (light shade), a cluster of one protein (medium shade) and a cluster of three proteins (light shade). The symmetry is broken due to water-protein interactions.

3. Results and discussion

The results of *FIRST* flexibility analysis indicate which of the bonds in the protein are rigid, and which bonds are flexible. Rigid bonds that share a common vertex (an atom in this case) are grouped together to form a rigid cluster. The presence of noncovalent interactions in our model allows for rigid regions to span across the interface between to proteins. Two or more rigid regions, connected via flexible bonds, may be present in the protein, and these are referred to as independently rigid clusters. In figures 4–7 the rigid regions are shown as thick tubes, and the flexible bonds are shown as thin lines. Independent rigid clusters are given different shades of grey.

Figures 4(A) and (B) show the rigid region decomposition of the dimer isolated from the hexamer–hexamer and pentamer–hexamer interface, respectively. In the case of the hexamer–hexamer dimer, both domains belong to a single rigid cluster, although the rigidity is not symmetric. For example, the loops at the top of the left-hand monomer are flexible, while they are rigid in the right-hand monomer. The result is different for the dimer from the pentamer–hexamer interface, in which each domain of the dimer is independently rigid; the intervening bonds are flexible. In the two cases the amino-acid sequences are identical; however, they must have different structural environments within the context of the complete capsid.

The rigid region decomposition for the pentamer of dimers is shown in figure 5. This result represents the baseline for the simulation, as the energy cut-off for all the *FIRST* analyses was

Figure 7. Flexibility results for the pentamer of dimers plus one dimer. The side-chain bonds are not displayed. Rigid clusters are depicted by thick tubes; flexible bonds are shown with thin tubes. The additional dimer is in the lower left side of the structure (see figure 5 for reference). The results shown that inter-protein bonds that form when the dimer binds to the pentamer cause the dimer to become part of the rigid core of the protein (dark shaded rigid region). Additionally, one of the outer ring proteins of the pentamer has become part of the rigid core. The other four outer ring proteins remain independently rigid.

chosen to produce this result. As expected, the pentamer of dimers has six large rigid regions. The five proteins in the centre of the pentamer (one from each of the five dimers) form one large rigid cluster. The remaining five domains located along the outer edge of the pentamer are all rigid, but independently of each other. Each of these regions is linked to the rigid cluster in the centre via flexible bonds.

An interesting structural feature of the hexamer of dimers (shown in figure 6) is the 12-stranded β-barrel that forms in the centre of the hexamer. Two strands from each of the six dimers contribute to this β-barrel, which is a stable and common motif within the set of known proteins. It was previously proposed that formation of the hexamer of dimers nucleated the capsid assembly process, in part because of the potential role that the β-barrel could play in structural stability and capsid function. It was subsequently shown that it is not the hexamer that nucleates capsid assembly in CCMV, and the flexibility results qualitatively support this conclusion. Figure 6 shows that there are four rigid regions that span across dimer–dimer interfaces; however, there is no single rigid cluster that encompasses all six of the dimers present in the hexamer.

Figure 7 shows the flexibility results when the pentamer is analysed in a complex with an additional dimer building block. In this conformation, the dimer becomes part of the rigid core of the pentamer, along with one of the outer ring domains that it is in contact with.

4. Conclusions

The key result of these simulations is the pentamer of dimers in a complex with one additional dimer (figure 7). In this case, the extra dimer becomes locked into the large rigid core of the pentamer. This large rigid cluster serves two purposes: it maintains the proper curvature of the substructure consistent with the icosahedral shape of the complete capsid, and it provides a stable scaffold upon which additional free dimers, or even larger substructures that have the

proper shape complementarity, bind. In contrast, the lack of a central structurally rigid core seen in the hexamer of dimers implies that if an additional free dimer, or other substructure, were to bind the hexamer, the complex would not form a larger rigid cluster that spanned all six dimers in the hexamer. In the absence of a rigid core, the flexibility between the dimer subunits would inhibit fast formation of the complete capsid.

The body–bar representation provides a more general way of modelling molecular interactions as distance constraints, even though at first sight this may seem strange. Within the scheme of how hydrogen bonds, hydrophobic interactions, and torsion forces have been modelled, both the bar–joint and body–bar 3D *pebble games* provide complete and equivalent analyses of the network rigidity. However, the current version of *FIRST* (http://flexweb.asu.edu) is less restricted than before. Defining alternative distance constraint representations of molecular interactions (going beyond previous bond-bending networks) is now possible, although these alternative modelling schemes, and their consequences, will take some time to explore.

Acknowledgments

We would like to thank Trevor Douglas for introducing us to viral capsids and Walter Whiteley for continuing discussions on the mathematical underpinning of rigidity. We would also like to thank Adam Zlotnick for his many insights into capsid assembly. This work was supported by NSF grant DMR-0078361 and NIH grant GM067249.

References

[1] Speir J A, Munshi S, Wang G, Baker T S and Johnson J E 1995 *Structure* **3** 63
[2] Adolph K W and Butler P J 1974 *J. Mol. Biol.* **88** 327
[3] Zlotnick A, Aldrich R, Johnson J M, Ceres P and Young M J 2000 *Virology* **277** 450
[4] Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 *Nucleic Acids Res.* **28** 235
[5] Jacobs D J, Rader A J, Kuhn L A and Thorpe M F 2001 *Proteins* **44** 150
[6] Jacobs D, Kuhn L A and Thorpe M F 1999 *Rigidity Theory and Applications* ed M F Thorpe and P M Duxbury (Dordrecht: Kluwer–Academic) (New York: Plenum) p 357
[7] Maxwell J C 1864 *Phil. Mag.* **27** 294
[8] Laman G 1970 *J. Eng. Math.* **4** 331
[9] Jacobs D J and Thorpe M F 1995 *Phys. Rev. Lett.* **75** 4051
[10] Jacobs D 1998 *J. Phys. A: Math. Gen.* **31** 6653
[11] Thorpe M F, Jacobs D, Chubynsky M V and Rader A J 1999 *Rigidity Theory and Applications* (Dordrecht: Kluwer–Academic) (New York: Plenum)
[12] Tay T-S and Whiteley W 1984 *Structural Topology* **9** 31
Tay T-S 1984 *J. Comb. Theory* B **26** 95
[13] Whiteley W 1996 *Contemp. Math.* **197** 171
Whiteley W 1999 *Rigidity Theory and Applications* ed M F Thorpe and P M Duxbury (Dordrecht: Kluwer–Academic) (New York: Plenum) p 21
[14] Jacobs D and Thorpe M F 1998 *Computer-Implemented System for Analyzing Rigidity of Substructures Within a Macromolecule* (USA: Board of Trustees Operating Michigan State University)
[15] Rader A J, Hespenheide B M, Kuhn L A and Thorpe M F 2002 *Proc. Natl Acad. Sci. USA* **99** 3540
[16] Jacobs D and Hendrickson B 1997 *J. Comput. Phys.* **137** 346
[17] Moukarzel C 1996 *J. Phys. A: Math. Gen.* **29** 8079
[18] Vriend G 1990 *J. Mol. Graph.* **8** 52
[19] Hooft R W, Sander C and Vriend G 1996 *Proteins* **26** 363
[20] Williams M A, Goodfellow J M and Thornton J M 1994 *Protein Sci.* **3** 1224
[21] Reddy V S, Natarajan P, Okerberg B, Li K, Damodaran K V, Morton R T, Brooks C L III and Johnson J E 2001 *J. Virol.* **75** 11943