

Classification and evolutionary trends of icosahedral viral capsids

Richard Kerner*

LPTMC, Université Paris-VI-CNRS UMR 7600, Tour 24, 4-ème, Boite 121, 4 Place Jussieu, 75005 Paris, France

(Received 6 January 2008; final version received 31 March 2008)

A classification of icosahedral viral capsids is proposed. We show how the self-organization of capsids during their formation implies a definite composition of their elementary building blocks. The exact number of hexamers with three different admissible symmetries is related to capsids' sizes, labelled by their T -numbers. Simple rules determining these numbers for each value of T are deduced and certain consequences concerning the probabilities of mutations and evolution of viruses are discussed.

Keywords: viral capsid classification; self-organized agglomeration; symmetry

1. Introduction

Viruses provide a spectacular example of self-organization of proteins which can build up increasingly complicated structures. The first theoretical of quantitative analysis of such processes was set forth by Eigen in his book on biological self-organization [1].

The large group of spherical viruses [2], whose protective protein shells called capsids display perfect icosahedral symmetry, is of particular interest. These geometrical structures, known to mathematicians since Coxeter's classification [3]; are also observed in fullerenes, huge molecules composed exclusively of carbon atoms, predicted by Smalley and Kroto, and discovered in the 1980s.

Since Caspar and Klug [4] introduced simple rules predicting a sequence of observable viral capsids, several models of agglomeration dynamics of these structures have been proposed, *e.g.* Zlotnick's model [5], published in 1994.

The common geometrical feature of many viral capsids and fullerenes is their icosahedral shape, with 12 pentagons found on the opposite sides of six fivefold symmetry axes, and an appropriate number of hexagons in between. The number of hexagons is given by the following simple formula: $N_6 = 10(T - 1)$, with $T = (p^2 + pq + q^2)$, called the triangular number, and where p and q are two non-negative integers [3].

In capsids, the building blocks made of coat proteins are called monomers, dimers, trimers, pentamers and hexamers, according to the number of their constituent proteins, the bigger ones usually being assembled from smaller entities prior to further agglomeration into capsid shells [10]. Configurations with fivefold and sixfold symmetry are formed from 60 or 180 smaller subunits that by themselves do not have these symmetries, like in Cowpea Mosaic Virus and Cowpea Chlorotic Mottle Virus, respectively [6]. Although certain virus species grow medium-size capsids corresponding to $N_6 = 20$, or $N_6 = 30$ and $N_6 = 60$, *e.g.*, the TRSV capsid composed of 60 copies of a single capsid protein (56,000 Da, 513 amino acid residues) [12],

*Email: rk@ccr.jussieu.fr

which, from the geometrical point of view, can be put in a one-to-one correspondence with 60 carbon atoms forming a fullerene, some of them form pure dodecahedral capsids (with exclusively pentamers as building blocks), like certain Comoviridae or Cowpea virus [9]), while some others, like human Adenovirus [7], form huge capsids with $N_6 = 240$, corresponding to $p = 5, q = 0$.

It seems that there is no such thing as universal assembly kinetics: the way the capsids are assembled differs from one virus to another. The $T = 7$ phage *HK97* appears to build pentamers and hexamers first, then assemble these capsomers to form the final structure, whereas another $T = 7$ phage labelled P22 appears to assemble its capsids directly from individual coat proteins (see [11] and the references within).

The common point is the presence of pentagons and hexagons in the resulting structure, and the strict topological rules that result from Euler's theorem on convex polyhedra: $V - E + F = 2$, with V number of vertices, E number of edges and F number of faces. From this, one derives the fact that when only pentagonal and hexagonal faces are allowed, the number of pentagons is always $N_5 = 12$, while the number of hexagons is $N_6 = (10 = (T-1))$.

Minimizing the waste implies that the initial nucleation ratio of pentamers versus hexamers is very close to its final value in capsids. Second, the final size of the capsid must depend on particular assembly schemes. We investigate here the rules that define the type and the size of the capsids, simultaneously optimizing the production rate [13].

2. The assembly rules for icosahedral capsids

If viruses were using undifferentiated hexamers with equivalent sides, then nothing would forbid the creation of all kinds of structures as shown in Figure 1, and the final yield would be very low (like in fullerenes, less than 10%).

Therefore, we should assume that basic building blocks, organized in pentamers and hexamers, have differentiated and recognizable sides in order to encode the successive build-up information. From symmetry considerations (and confirmed by chemical analysis) it results that pentamers are composed of five identical dimers, their five edges being equivalent. All the five sides of a pentamer must be equivalent (identical), because five is a prime number, and any division into parts will break the symmetry. Concerning the hexamers, as 6 is divisible by 2 and 3, one can have the following four situations:

- All six sides equivalent, denoted as $(aaaaaa)$
- Two types of sides, denoted as $(ababab)$

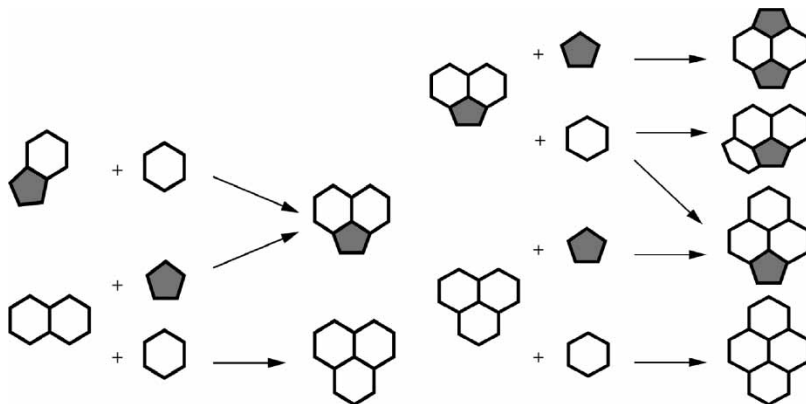


Figure 1. Random agglomeration of five- and six-sided polygons.

- Three types of sides, denoted as $(abcabc)$
- Six different sides, denoted as $(abcdef)$

The hexamers are also oriented, with one face becoming external, and the other one turned to the interior of the capsid. The three differentiations of hexamers are represented in Figure 2 below.

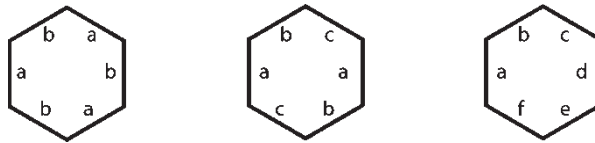


Figure 2. Three differentiations of hexamers.

Let us consider the simplest case first, the hexamers of the type $(ababab)$. Let us denote pentamers' sides by the symbol p , whereas two different kinds of sides on hexamers' edges will be called a and b (Figure 2). Suppose that a hexamer can stick to a pentamer with only $(p + a)$ -combination; then two hexamers must stick to each other only through a $(b + b)$ combination, with both $(p + b)$ and $(a + b)$ combinations being forbidden by a chemical potential barrier. If a second type of differentiation is being used, $(abcabc)$ say, the allowed sticking rules are $(a + p)$ and $(b + c)$. This determines the issue completely, as shown in Figure 3.

Note that in both cases we show only one of the 'basic triangles' forming the capsid, which is always made from 20 identical triangles sticking together to form a perfect icosahedral shape. These sticking rules can be summarized in tables that we shall call the 'affinity matrices', displayed on the right in Figure 3. Here, a '0' is put at the crossing of two symbols whose agglomeration is forbidden, and a '1' when the agglomeration is allowed. By construction, a '1' can occur only once in each line or column. This implies that each kind of protein appearing at capsomer's edges occurs exactly 60 times in the final construction. There are $12 \times 5 = 60$ p -type edges in 12 pentamers; they stick to only one kind of hexamers' edges, labelled 'a', which therefore must appear 60 times, too; and so forth, as imposed by the structure of the affinity matrix. In the case when hexamers are replaced by combinations of dimers or trimers, as shown in Figure 3, it is easy to count that each dimer will occur 30 times, and each trimer will appear 20 times.

Finally, let us use the maximally differentiated hexamers of the $(abcdef)$ -type. Starting with pentamers surrounded by the hexamers sticking via the $(p - a)$ -pairing, we discover that now two choices are possible, leading to left- and right-hand sided versions, as shown in Figure 4.

Figure (5) shows the construction of $T = 9$ and $T = 12$ capsids.

In order to grow capsids with T -numbers greater than 7, one has to introduce new types of hexamers that would never stick to pentamers, but are able to associate themselves with certain sides of the previous maximally differentiated hexamers. The result is shown in Figure 5; the corresponding affinity matrices are not displayed because of their large size, but they can be easily reconstructed from the scheme shown in the Figure (see Figure 5) (see [14]). Bigger capsids, with T -numbers 13, 16, etc. are constructed in a similar manner. In order to describe the structure of the $T = 25$ adenovirus capsid, one must introduce no less than four hexamer types, out of which only one type can agglomerate with pentamers.

These examples suggest that strict association rules may exist, providing precise agglomeration pathways for each kind of icosahedral capsid. Let us analyse these rules in more detail.

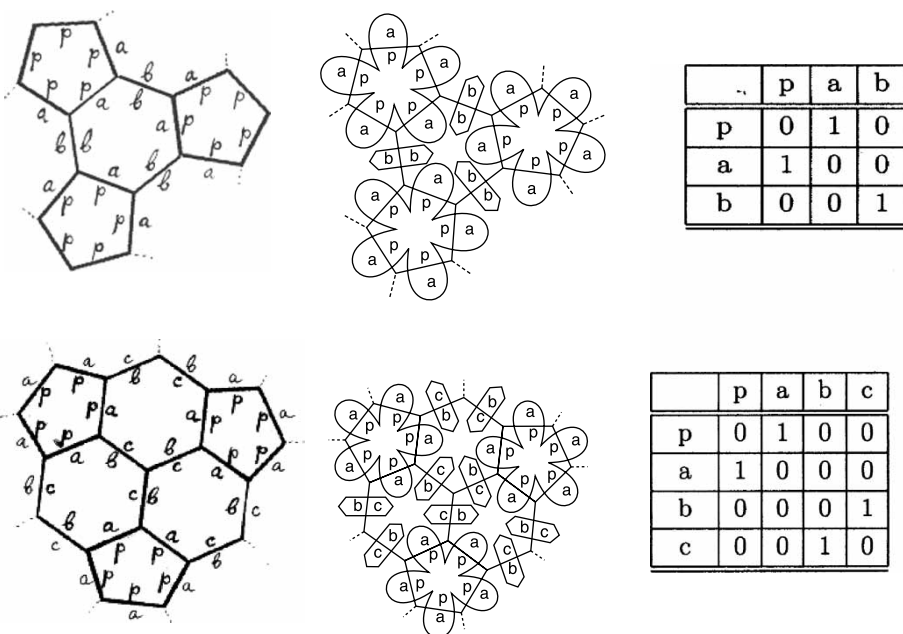


Figure 3. The building schemes and affinity matrices for the $T = 3$ and $T = 4$ capsids.

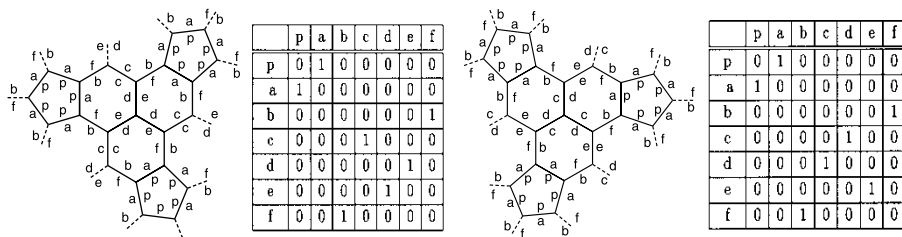


Figure 4. The building schemes of the $T = 7$ capsids.

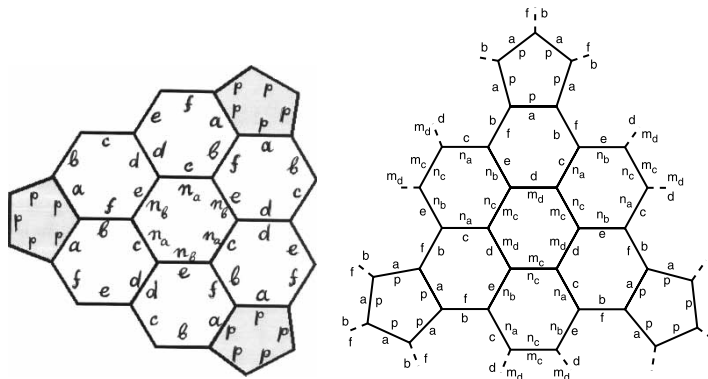


Figure 5. The building schemes for the $T = 9$ and $T = 12$ capsids.

3. Classification of icosahedral capsids

Now we can organize all these results in a single table that follows in Figure 6. To each value of triangular number T corresponds a unique partition into $1 + (T - 1)$, where the '1' represents the unique pentamer type and $(T - 1)$ is partitioned into a sum of certain numbers of different hexamer types according to the formula

$$(T - 1) = 6\alpha + 3\beta + 2\gamma$$

with non-negative integers α, β and γ . More precisely, the numbers β and γ can take on the values 0 or 1 only; this results from the fact that the corresponding hexamers are centred on

$N = p + q$	(p, q)	T	N_6	Partition	Chirality
1	(0, 1)	1	0	1	0
2	(1, 1)	3	20	1 + 2	0
2	(0, 2)	4	30	1 + 3	0
3	(1, 2)	7	60	1 + 6	+
3	(0, 3)	9	80	1 + 6 + 2	0
4	(2, 2)	12	110	1 + 6 + 2 + 3	0
4	(3, 1)	13	120	1 + 6 + 6	+
4	(4, 0)	16	150	1 + 6 + 6 + 3	0
5	(3, 2)	19	180	1 + 6 + 6 + 6	+
5	(4, 1)	21	200	1 + 6 + 6 + 6 + 2	+
5	(5, 0)	25	240	1 + 6 + 6 + 6 + 6	0
6	(3, 3)	27	260	1 + (4 × 6) + 2	0
6	(4, 2)	28	270	1 + (4 × 6) + 3	+
6	(5, 1)	31	300	1 + (5 × 6)	+
6	(6, 0)	36	350	1 + (5 × 6) + 2 + 3	0
7	(4, 3)	37	360	1 + (6 × 6)	+
7	(5, 2)	39	380	1 + (6 × 6) + 2	+
7	(6, 1)	43	420	1 + (7 × 6)	+
7	(7, 0)	49*	480	1 + (8 × 6)	0
8	(4, 4)	48	470	1 + (7 × 6) + 2 + 3	0
8	(5, 3)	49*	480	1 + (8 × 6)	+
8	(6, 2)	52	510	1 + (8 × 6) + 3	+
8	(7, 1)	57	560	1 + (9 × 6) + 2	+
8	(8, 0)	64	630	1 + (10 × 6) + 3	0

T and (p,q)	1 + (k×6)	1 + (k×6)+2	1 + (k×6)+3	1 + (k×6)+2+3
1 (1,0)	1 + (0×6)			
3 (1,1)		1 + (0×6)+2		
4 (2,0)			1 + (0×6)+3	
7 (2,1)	1 + (1×6)			
9 (3,0)		1 + (1×6)+2		
12 (2,2)				1 + (1×6)+2+3
13 (3,1)	1 + (2×6)			
16 (4,0)			1 + (2×6)+3	
19 (3,2)	1 + (3×6)			
21 (4,1)		1 + (3×6)+2		
25 (5,0)	1 + (4×6)			
27 (3,3)		1 + (4×6)+2		
28 (4,2)			1 + (4×6)+3	
31 (5,1)	1 + (5×6)			
36 (6,0)				1 + (5×6)+2+3
37 (4,3)	1 + (6×6)			
39 (5,2)		1 + (6×6)+2		
43 (6,1)	1 + (7×6)			
48 (4,4)				1 + (7×6)+2+3
49 (7,0)	1 + (8×6)			
49 (5,3)	1 + (8×6)			
52 (6,2)			1 + (8×6)+3	
57 (7,1)		1 + (9×6)+2		
61 (5,4)	1 + (10×6)			
63 (6,3)		1 + (10×6)+2		
64 (8,0)			1 + (10×6)+3	
67 (7,2)	1 + (11×6)			
73 (8,1)	1 + (12×6)			
75 (5,5)		1 + (12×6)+2		
76 (6,4)			1 + (12×6)+3	
79 (7,3)	1 + (13×6)			
81 (9,0)		1 + (13×6)+2		

Figure 6. (Left) Classification of icosahedral capsids (isomers marked with *). (Right) The four types of icosahedral capsids, up to $T = 81$.

T and (p,q)	1 + (k×6)	1 + (k×6)+2	1 + (k×6)+3	1 + (k×6)+2+3
84 (8,2)				1 + (13×6)+2+3
91 (6,5)	1 + (15×6)			
91 (9,1)	1 + (15×6)			
93 (7,4)		1 + (15×6)+2		
96 (8,3)				1 + (15×6)+2+3
100 (10,0)			1 + (16×6)+3	
103 (9,2)	1 + (17×6)			
108 (6,6)				1 + (17×6)+2+3
109 (7,5)	1 + (18×6)			
111 (10,1)		1 + (18×6)+2		
112 (8,4)			1 + (18×6)+3	
117 (9,3)		1 + (18×6)+2		
121 (11,0)	1 + (20×6)			
124 (10,2)			1 + (20×6)+3	
127 (7,6)	1 + (21×6)			
129 (8,5)		1 + (21×6)+2		
133 (11,1)	1 + (22×6)			
133 (9,4)	1 + (22×6)			
139 (10,3)	1 + (23×6)			
144 (12,0)				1 + (23×6)+2+3
147 (7,7)		1 + (24×6)+2		
147 (11,2)		1 + (24×6)+2		
148 (8,6)			1 + (24×6)+3	
151 (9,5)	1 + (25×6)			
156 (10,4)				1 + (25×6)+2+3
157 (12,1)	1 + (26×6)			
163 (11,3)	1 + (27×6)			
169 (13,0)	1 + (28×6)			
169 (8,7)	1 + (28×6)			
171 (9,6)		1 + (28×6)+2		
172 (12,2)			1 + (12×6)+3	
181 (11,4)	1 + (30×6)			
183 (13,1)		1 + (30×6)+2		
189 (12,3)		1 + (31×6)+2		
192 (8,8)				1 + (31×6)+2+3

T and (p,q)	1 + (k×6)	1 + (k×6)+2	1 + (k×6)+3	1 + (k×6)+2+3
193 (9,7)	1 + (32×6)			
196 (14,0)				1 + (32×6)+3
196 (10,6)				1 + (32×6)+3
199 (13,2)	1 + (33×6)			
201 (11,5)		1 + (33×6)+2		
208 (12,4)			1 + (34×6)+3	
211 (14,1)	1 + (35×6)			
217 (9,8)	1 + (36×6)			
217 (13,3)	1 + (36×6)			
219 (10,7)		1 + (36×6)+2		
223 (11,6)	1 + (37×6)			
225 (15,0)			1 + (37×6)+2	
228 (14,2)				1 + (37×6)+2+3
229 (12,5)	1 + (38×6)			
237 (13,4)		1 + (39×6)+2		
241 (15,1)	1 + (40×6)			
243 (9,9)		1 + (40×6)+2		
244 (10,8)			1 + (40×6)+3	
247 (11,7)	1 + (41×6)			
247 (14,3)	1 + (41×6)			
252 (12,6)				1 + (41×6)+2+3
256 (16,0)				1 + (42×6)+3
259 (13,5)	1 + (43×6)			
268 (14,4)				1 + (44×6)+3
271 (10,9)	1 + (45×6)			
273 (11,8)		1 + (45×6)+2		
273 (16,1)		1 + (45×6)+2		
277 (12,7)	1 + (46×6)			
283 (13,6)	1 + (47×6)			
289 (17,0)	1 + (48×6)			
291 (14,5)		1 + (48×6)+2		
292 (16,2)			1 + (48×6)+3	
301 (11,9)	1 + (50×6)			
301 (15,4)	1 + (50×6)			
303 (16,3)		1 + (50×6)+2		
304 (12,8)				1 + (50×6)+3

Figure 7. The next two classification tables, up to $T = 304$.

a threefold or a twofold symmetry axis, so that the first type must be found at the centre of icosahedron's triangular face, whereas the second type must be found in the centre of an edge between elementary triangles. The number α of maximally differentiated hexamers follows then from the corresponding partition of a given triangular number, as shown in the following table (Figure 6, left).

It follows that all icosahedral capsids can be divided into four separate groups following their internal symmetry, dictated by the presence of three and twofold centres of symmetry inside the elementary triangles or on the edges in between. The result looks like a periodic table of capsids, arranged in four columns, according to the composition of constitutive hexamers: pure $(abcdef)$ hexamers exclusively in the first column; the $(abcdef)$ -type hexamers with one $(ababab)$ -type hexamer in the second column; with one $(abcabc)$ -type hexamer in the third column, and finally with both $(ababab)$ and $(abcabc)$ -type in the fourth column. Symmetric icosahedral capsids (marked in gold) correspond to specific values of T with $(p, 0)$ or (p, p) ; the chiral icosahedral capsids, existing in left- and right-handed versions, are marked in magenta. The statistics appearing to the naked eye in Figure 6 show that the most differentiated capsids are the rarest. This tendency continues in the following tables (Figure 7). There are 46 capsids (including isomers) in the first column, 27 capsid species in the second column, 19 species in the third column and only 11 species in the fourth column.

Trying to define evolutionary trends from these capsid classification tables seems to be a risky endeavour. One can recall, however, that non-negligible information was drawn from a careful study of animal skeletons, fish scales, and similar secondary features of living organisms.

It seems plausible that the major evolution trend is from smaller towards bigger forms, as it supposes progressive differentiation among the constitutive hexamers. From a purely mathematical point of view, the evolution would mean then an addition of a new hexamer type, or a transformation of one of the constitutive hexamers into another one with higher differentiation. An addition of a new $(abcdef)$ hexamer to one of the capsids of any column results in one step down the same column; an addition of several new maximally differentiated hexamers results in the same number of steps down the same column.

In order to better understand the tables in Figures 6 and 7, one should imagine them as cylinders, with the right edge of the right column glued to the left edge of the first column. Then one can imagine all possible transitions from any column to another one, not necessarily its immediate neighbour, but always towards the right side and down. For example, adding a new $(ababab)$ -type hexamer to a capsid from the first column creates a species belonging to the second column, etc.

The important question then is how many mutations are necessary to accomplish one of these transformations? This would give a hint as to how can one conceive the notion of 'distance' between different capsids. It seems reasonable to assume that the closest species are those separated by a single addition of a maximally differentiated hexamer, which can be symbolized by the transition $(abcdef) \rightarrow (d'b'c'd'e'f')$, because such a mutation does not alter hexamer's character and can be obtained by a common modification (adding a particular radical, for example). This is why we should expect the viruses with triangulation numbers 7 and 13 to be close parents, both belonging to the first column and both skew-symmetric; we would also expect the Adenovirus ($T = 25$, $(p, q) = (5, 0)$) to be related to one of the isomers of the $T = 49$, $(p, q) = (7, 0)$ capsid, whose respective schemes are $25 = 1 + 4 \times 6$ and $49 = 1 + 8 \times 6$, which can be obtained by a common modification of all the four differentiated hexamers, thus doubling their number. The isomer $T = 49$, $(p, q) = (5, 3)$ should be quite distant from these two.

It is also plausible that the evolution by mutations keeps the capsid types inside the same column, with the same symmetry type, as the symmetry change $(abcdef) \rightarrow (ababab)$ requires several more specific mutations at once. This is probably why the capsid type

$T = 16 = 1 + 2 \times 6 + 3$ in the third column (Figure 6), which appears to be isolated, corresponds to a single and unique family of Herpesvirus, which admits many mutations and variations but always remains inside the same capsid species and size [8,15].

Acknowledgements

The author is greatly indebted to Dr. Nicola Stonehouse for many enlightening discussions and for her extremely useful suggestions and remarks. Thanks are also due to Drs. Reidun Twarock and Jack Johnson for interesting discussions, and to Dr. Maja Nowakowski for the help in getting acquainted with current literature.

References

- [1] M. Eigen, *Classification and evolutionary trends of icosahedral viral capsids in Selforganization of Matter and the Evolution of Biological Molecules*, 10, Die Naturwissenschaften 58 heft 10, Springer-Verlag, 1971.
- [2] D.D. Richman, R.J. Whitley, and F.G. Hayden, *Clinical Virology*, 2nd ed, ASM Press, Washington, DC, 2002.
- [3] M.C.M. Coxeter, *Regular Polytopes*, Methuen and C., London, 1948.
- [4] D.L.D. Caspar and A. Klug, *Symp. Quant. Biol.* 27 (1962), p. 1.
- [5] A. Zlotnick, *J. Mol. Biol.* 241 (1994), pp. 59–67.
- [6] S.B. Larson, J. Day, A. Greenwood and A. McPherson, *J. Mol. Biol.* 277 (1998), pp. 37–59.
- [7] P.L. Stewart, R.M. Burnett, M. Cyrklaff, and S.D. Fuller, *Cell* 67 (1991), pp. 145–154.
- [8] B.L. Trus, J.B. Heymann, K. Nealon, N. Cheng, W.W. Newcomb, J.C. Brown, D.H. Kedes and A.C. Steven, *J. Virol.* 75(6) (2001), pp. 2879–2890.
- [9] T. Lin and J.E. Johnson, *Adv. Virus Res.* 62 (2003), pp. 167–236.
- [10] H.R. Hill, N.J. Stonehouse, S.A. Fonseca, and P.G. Stockley, *J. Mol. Biol.* 266 (1997), pp. 1–7.
- [11] P.E. Prevelige, D. Thomas, and J. King, *Biophys. J* 64 (1993), pp. 824–835.
- [12] B. Buckley, S. Silva, and S. Singh, *Virus Res.* 30 (1993), pp. 335–349.
- [13] R. Kerner, *J. Theor. Med.* (2) (2005), pp. 95–97.
- [14] ———, *Models of Agglomeration and Glass Transition*, Imperial College Press, London, 2007.
- [15] D.J. McGeogh and A.J. Davison, *The Molecular Evolutionary History of the Herpesviruses: Origins and Evolution of Viruses*, Academic Press, London, 1999.