Taylor & Francis
Taylor & Francis Group

A stochastic model of icosahedral capsid growth

R. KERNER*

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

A model of growth of icosahedral capsids is proposed. It is similar to the stochastic model of fullerene growth, but takes into account variations in the composition of pentamers and hexamers that are responsible for the final capsid size. We show that the observed high yield of capsid production implies a high level of self-organization of the elementary building blocks.

1. Introduction

Since Caspar and Klug [1] introduced simple rules predicting a sequence of viral capsids, several models of growth dynamics of these structures have been proposed, for example by Zlotnick [2]. It is amazing that these structures, known to mathematicians since Coxeter's [3] classification, are also observed in the so-called *fullerenes*, huge molecules composed exclusively of carbon atoms, predicted by Smalley and Kroto, and discovered in the eighties.

The common geometrical feature of many viral capsids and fullerenes is their icosahedral symmetry, with twelve pentagons found on the opposite sides of six five-fold symmetry axes, and an appropriate number of hexagons in between. The number of hexagons is given by the following simple formula: $N_6 = 10(T - 1)$, $T = (p^2 + q^2 + pq)$, where p and q are two non-negative integers. Although certain virus species grow medium-size capsids corresponding to $N_6 = 20$ (as in the C_{60} fullerene molecule), or $N_6 = 30$ and $N_6 = 60$, some of them form pure dodecahedral capsids (with exclusively pentamers as building blocks), e.g. certain *Comoviridae* or *Cowpea* virus [4], while some others, e.g. human *Adenovirus* [5], form very large capsids with $N_6 = 240$, corresponding to $p = 5$, $q = 0$.

In some cases, the similarity with the fullerene structure is striking: For example, the TRSV capsid is composed of 60 copies of a single capsid protein (56,000 Da, 513 amino acid residues) [6], which can be put in a one-to-one correspondence with 60 carbon atoms forming a fullerene molecule.

The process of fullerene formation differs essentially from the building of icosahedral viral capsids: Fullerenes are formed in a hot plasma around an electric arc between two graphite electrodes, from carbon atoms and small carbon agglomerates (C_2, C_3, up to C_9 or C_{10}), whereas capsids are built progressively in liquid medium, from giant protein aggregates displaying pentagonal or hexagonal symmetry, or from smaller units called *dimers* or *trimers*. It also 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 capsid structure, whereas another $T = 7$ phage (P22) appears to assemble its capsids directly from individual coat proteins (see [7,8]).

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 hexagon number is $N_6 = 10(T - 1)$.

In this article we shall apply the idea of *statistical self-similarity* [9] to icosahedral virus capsid growth. It states that statistical properties of randomly chosen parts of the construction must reflect the overall statistics obtained at final stages of agglomeration. Unlike the case of fullerene molecules, whose yield from the hot plasma is in the best case no higher than 10% of the total mass of carbon soot, viruses use almost 100% of pentamers and hexamers at their disposal to form perfect icosahedral capsid structures, into which the viral genetic material is densely packaged afterwards.

*Tel.: +33-1-44-27-72-98. Fax: +33-1-44-27-51-00. Email: rk@ccr.jussieu.fr

This means that the initial nucleation ratio of pentamers versus hexamers is very close to its final value in the capsids in order to minimize the waste. The final size of the capsid depends on particular assembly rules, many of which can be deduced from the statistical weights of various agglomeration steps, found by maximizing the final production rate.

2. Model of agglomeration and growth

A fullerene cage, composed of 60 carbon atoms, corresponds to a $T = 3$ geometry, i.e. to $(p, q) = (1, 1)$. The agglomeration and growth at high temperatures depend only on energetic and combinatorial factors. A high yield of perfect C_{60} molecules can be obtained only if during the agglomeration process the ratio of pentagons in the k-th intermediates, $P_5^{(k)}$, is very close to the value observed in the final structure, which is $P_5 = 12/32 = 3/8 = 0.385$.

The simplest stochastic model of growth successfully applied to fullerene formation [10,11], is based on the assumption that the dominant agglomeration processes consist of forming new polygons in the cavities between two polygons on the border of the existing cluster, by adjoining a C_2 or a C_3 molecule to a cavity found in a cluster already formed. One of such processes is shown in figure 1 below.

It is clear that the resulting (655) cluster is wasted for further fullerene formation, whereas a (656) cluster can be used in the next agglomeration step. Because the (666) clusters are also absent in the final fullerene cage, it is easy to see that at each consecutive agglomeration step the yield of "proper" clusters, useful for further fullerene construction, is exactly 1/2. After about 23 to 24 steps leading from the initial three-polygon structure to an almost finished fullerene cage with 27 to 28 (out of a total of 32) polygons already in place, the total yield would approach $2^{-24} \simeq 10^{-8}$ instead of the observed 10^{-1}, i.e. 10%. This means that there is a mechanism that favours the creation of "correct" structures versus the "incorrect" ones, so that the average yield of "proper" clusters becomes close to $q = 0.957$, ensuring q^{24} to be of order 10^{-1}.
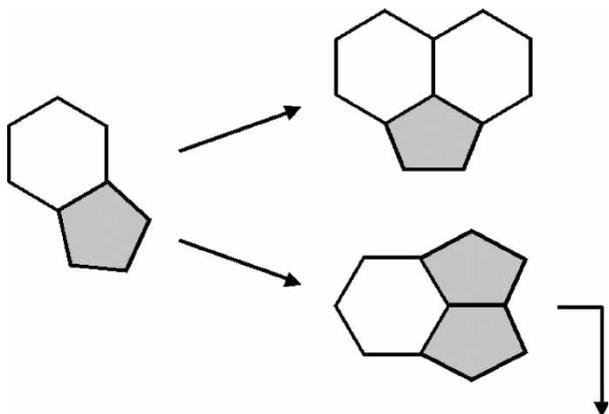
Figure 1. Adding a pentagon to a ((5,6)) doublet.

In the case of fullerenes, the correction is due to the Boltzmann factors that reflect the energy differences between the following basic processes: Creation of a new pentagon in a (6,6) cavity, or the creation of a new hexagon in a (5,6) or in a (6,6) cavity, under the assumption that the energy barrier against creation of two or three pentagons sticking together is so large that the corresponding Boltzmann factor is close to 0. These factors could be evaluated by requiring that the successive probabilities of finding pentagons among all polygons in clusters of a given size (after the n-th agglomeration step) and the corresponding yields form a geometric progression [10,11].

3. Combinatorics of icosahedral capsid growth

We investigate here a model in which purely statistical factors play the decisive role in ensuring that the "correct" configurations are produced at each consecutive step almost without exception, i.e. practically with a 100% yield. Let us show now how these statistical factors can be evaluated, and what constraints they imply on the structures of the capsomers.

From symmetry considerations (and confirmed by chemical analysis) it follows that the pentamers are composed from five identical building blocks, so that their five edges are perfectly equivalent.

Let us denote the concentration (or the nucleation rate) of pentamers by c, that of hexamers by $(1 - c)$. Then the probabilities of doublets are readily calculated as follows:

$$P_{56} = 2 \cdot W_{56}c(1 - c)/Q; \quad P_{66} = W_{66}(1 - c)^2/Q, \quad (1)$$

where $W_{jk}, j, k = 5, 6$ are the statistical weights depending on the virus type and chemical barriers between various sides, and $Q = 2 \cdot W_{56}c(1 - c) + W_{66}(1 - c)^2$ is the normalizing factor. Note that we exclude two pentagons coming together, i.e. $W_{55} = 0$. Similarly, the probabilities of admissible "triplets" in the next agglomeration step displayed in figure 2 are given by:

$$P_{566} = P_{56} + P_{66}W_{66,5}c/Q_2,$$

$$P_{666} = P_{66}W_{66,6}(1 - c)/Q_2, \quad (2)$$

where $W_{66,5}$ and $W_{66,6}$ denote the statistical weights of the corresponding agglomeration processes, and $Q_2 = W_{66,5}c + W_{66,6}(1 - c)$.

Now, we can evaluate the average pentamer ratio $c^{(k)}$ in clusters of a given size after the kth agglomeration step. The first three values are given by:

$$c^{(1)} = \frac{1}{2}P_{56}, \quad c^{(2)} = \frac{1}{3}P_{566},$$

$$c^{(3)} = \frac{1}{4}(2P_{5665} + P_{5666} + P_{6665}), \quad \text{etc.}$$

(In the expression for $c^{(3)}$ we use the probabilities for three different allowed clusters, which are not presented here due to lack of space).
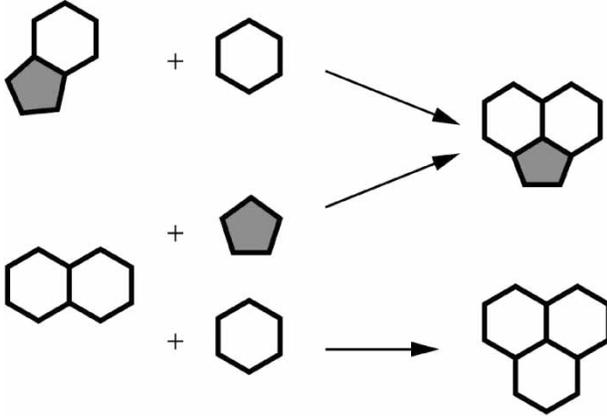
Figure 2.   Second agglomeration step.

Figure 3.   Three distinct hexamer structures.

We can use these formulae in two different ways. Either we impose the statistical weights, and determine the consecutive pentamer concentrations in growing clusters, starting from a given initial concentration c, or we treat the statistical weights as unknowns and determine them from self-similarity equations for successive pentamer concentrations:

$$\frac{c^{(n+2)} - c^{(n+1)}}{c^{(n+1)} - c^{(n)}} = \frac{c^{(n+1)} - c^{(n)}}{c^{(n)} - c^{(n-1)}}, \quad n = 2, 3, 4, \ldots \quad (3)$$

where the general expression for pentamer concentration in all clusters resulting from the nth agglomeration step is given by the recurrent formula:

$$c_n = \frac{1}{n} \sum_k k P_{(k \times 5)((n-k) \times 6)}, \quad (4)$$

the summation going up to the maximal admissible number of pentamers in clusters with n capsomers; this maximal number depends on n and we can get it only by direct inspection of all configurations with edge-sharing pentamers excluded. The resulting solutions for the limit values of c and for the auxiliary variables $\xi = (W_{56})/(W_{66})$, $\eta = (W_{56,6})/(W_{66,5})$, $\zeta = (W_{66,6})/(W_{66,5})$, etc. although usually not in the form of simple fractions, give very good hints concerning the assembly rules leading to particular capsid structures.

For example, if the sides of all hexamers were equivalent and could stick with equal probability to the sides of the pentamers as well as between themselves, the rate of production of proper ($T = 3$) capsids would be close to 2^{-23}, which is not the case. Therefore, hexamers and pentamers must display sticking rules discriminating against the undesired configurations.

Now, as the capsid production rate from initial protein material is close to 100%, for fullerene-like ($T = 3$) capsids we should have an initial ratio of pentamers versus all capsomers of 3:8. Based on symmetry considerations, we suggest the three possible scenarios for hexamers shown in figure 3.

Let us denote the edges of the pentamers by the symbol κ, and the two different edges of the hexamers as α and β (first case in figure 3). Supposing that a hexamer can stick to a pentamer only with a combination of κ and α edges, whereas two hexamers can stick to each other only with a
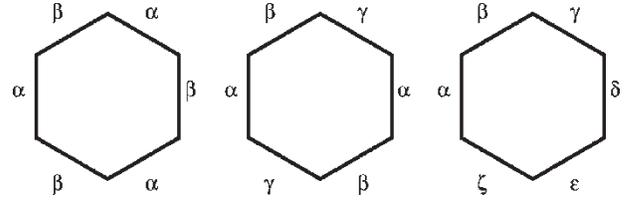
combination of two β edges, with a combination of both κ and β edges, and α and β edges, being forbidden by chemical potential barriers, we get the following statistical factors: $W_{56} = 15$, $W_{66} = 9$ $W_{56,6} = 3$, $W_{66,5} = 5$ and $W_{66,6} = 0$. With these rules the statistics in clusters naturally converges to the final value of $c = 3/8$.

Similarly, with the second scheme ($\alpha\beta\gamma\alpha\beta\gamma$) and with the assembling rules allowing only the sticking of κ and α edges and bond β and γ edges, respectively, we get with probability 1 the $T = 4$ capsid ($c = 2/7$), while for the third possibility, ($\alpha\beta\gamma\delta\epsilon\zeta$), displayed in figure 3, and the sticking rules allowing combinations of κ and α edges, β and ζ edges, δ and ϵ edges, and two γ edges, we obtain the $T = 7$ capsid ($c = 1/6$).

For bigger capsids, in which the ratio of pentamers is lower, one cannot obtain proper probabilities unless more than one type of hexamers is present, out of which only one is allowed to agglomerate with pentamers. In the case of two different types of hexamer, one obtains either the $T = 9$ capsid, or, with more exclusive sticking rules, the $T = 12$ capsid. Finally, in order to get the $T = 25$ adenovirus capsid, one must introduce no less than *four* hexamer types, out of which only two can agglomerate with pentamers.

Acknowledgements

References

[1] Caspar, D.L.D. and Klug, A., 1962, *Symp. Quant. Biol.*, **27**, 1.
[2] Zlotnick, A., 1994, *J. Mol. Biol.*, **241**, 59–67.
[3] Coxeter, M.C.M., 1948, *Regular Polytopes* (London: Methuen and C).
[4] Lin, T. and Johnson, J.E., 2003, *Adv. Virus Res.*, **62**, 167–236.
[5] Stewart, P.L., Burnett, R.M., Cyrklaff, M. and Fuller, S.D., 1991, *Cell*, **67**, 145–154 October 4.
[6] Buckley, B., Silva, S. and Singh, S., 1993, *Virus Res.*, **30**, 335–349.
[7] Prevelige, P.E., Thomas, D. and King, J., 1993, *Biophys. J.*, **64**, 824–835.
[8] Schwartz, R., Prevelige, P.E. and Berger, B., 1998, *Biophys. J.*, **765**, 2626–2636.
[9] Kerner, R., 1998, The principle of self-similarity. In: J. Moran-Lopez (Ed.) *Current Problems in Condensed Matter*, pp. 323–341.
[10] Kerner, R., Bennemann, K.H. and Penson, K.A., 1992, *Europhys. Lett.*, **19**, 363.
[11] Kerner, R., 1994, *Comput. Mater. Sci.*, **2**, 500–508.