Mathematical Models for Tubular Structures in the Papilloma-Polyoma family of viruses

R. Twarock
Centre for Mathematical Science, City University,
Northampton Square, London EC1V 0HB

Abstract

The surface lattices of all possible tubular structures in the papilloma-polyoma family of viruses are classified based on tiling theory and compared with experimental results.

1 Introduction

The papilloma-polyoma family of viruses contains tumour causing viruses, and mathematical models for the structure of these viruses and their tubular variants are hence of particular importance in view of anti-viral drug research. The distinctive feature of viruses in this family is the fact that the surface lattices of the viral capsids, that is the protein shells encapsidating the viral genome, as well as their tubular variants are formed from clusters of five proteins called pentamers. This is in contrast to most other families of viruses, in which capsids are usually formed from twelve pentamers and otherwise clusters of six proteins called hexamers. This structural peculiarity of viruses in the papilloma-polyoma family is the reason why Caspar-Klug theory [1], which explains the structure of most viruses with overall icosahedral symmetry, cannot be applied to viruses in this family [2, 3, 4]. Recently it has been shown that an approach based on tiling theory, that is a theory concerned with tessellations of surfaces by a set of basic building blocks called tiles, can be used to generalise Caspar-Klug theory and hence explain the surface lattices of viral capsids in the papilloma-polyoma family [5].

Already in [1] it has been suggested that the surface lattices of the tubular variants of a virus should have structural similarities with the surface lattices of the isometric virus particle. Therefore, it is natural to assume that if the tiling approach to viral capsids is the appropriate mathematical concept for the description of the isometric capsids, it should also lead in a natural way to models for their tubular variants. It is shown here that this is indeed the case, and that all possible tubular variants in the papilloma-polyoma family can be classified based on the tiling approach. The results are compared with the experimental data in [6].

2 The tiling approach for viral capsids applied to the papilloma-polyoma family

In the tiling approach for viral capsids [7], the surface lattices of viral capsids are represented as tilings, that is as tessellations given in terms of a set of basic building blocks called tiles. The tiles represent interactions between protein subunits in the capsid and have the interpretation of either dimer or trimer interactions, that is interactions between two, respectively three, protein subunits.

In [5] is has been shown that for the papilloma-polyoma family, the tiles are given by the shapes in Fig. 1. The dots indicate the location of the protein

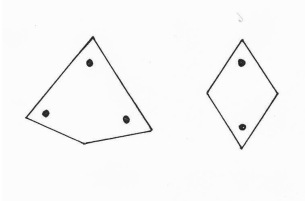

Figure 1: Tiles for spherical capsids in the papilloma-polyoma family.

subunits with respect to the shapes of the tiles, and are located precisely at the angles of size $\theta := \frac{2\pi}{5}$. The tile on the right is called a *rhomb* and represents a dimer interaction between the two subunits represented by the dots on the rhomb; the tile on the left is called a *kite* and represents a trimer interaction between the three subunits represented by the dots on the kite.

Moreover, it has been shown in this reference that the surface lattice of spherical capsids formed from 360 protein subunits in the papilloma-polyoma family can be modelled based on the tiles in Fig. 1 via the tiling in Fig. 2. The tiling accurately models the surface lattices of the (pseudo) $T = 7$ structures in this family, which cannot be explained in the framework of Caspar-Klug theory. In particular, vertices surrounded by decorations correspond to the location of the pentamers, and the tiling indicates both their location and orientation. Moreover, the tiling also predicts the location of the inter-subunit bonds that stabilise the capsid. Using the fact that tiles represent dimer and trimer interactions between the protein subunits that are represented as dots on the tiles, one immediately obtains the bonding structure in the capsid. It is superimposed schematically on the tiles in Fig. 2, where spiral arms represent the C-terminal arm extensions extended by the proteins. The tiling accurately predicts the experimentally found bonding structure observed in [8].

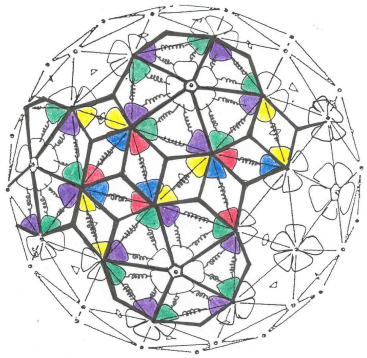

Figure 2: Tiling representing the surface lattice of spherical tilings in the papilloma-polyoma family.

3 Tubular structures implied by tiling theory

With tiles assumed to be building blocks of biological relevance in the tiling approach to viral capsids, it is natural to expect that they are the relevant building blocks for all structures that arise from the same types of protein subunits in a family of viruses. In particular, the surface lattices of all tubular structures in the papilloma-polyoma family should hence be expressible as tilings in terms of the tiles in Fig. 1. Therefore, a classification of tubular structures in this family can be obtained via a classification of all tubular structures that correspond to tilings in terms of the tiles in Fig. 1.

Since tubular structures can be obtained by rolling appropriate planar structures into cylinders with different cylindrical curvature, the first step is to derive all planar tilings that can be formed from the tiles in Fig. 1. As a second step, all tubular structures derivable from these planar tessellations are classified.

3.1 Planar tessellations

The main result of this subsection is the following Theorem.

Theorem 3.1 *The only planar tiling based on the tiles in Fig. 1 is the tessellation in Fig. 3.*

The reminder of this subsection is devoted to a proof of this statement. Any reader interested primarily in the results of this paper may skip this part and continue reading in the following subsection. The proof of the Theorem relies on a series of Lemmas and Propositions, which are based on concepts from tiling theory.

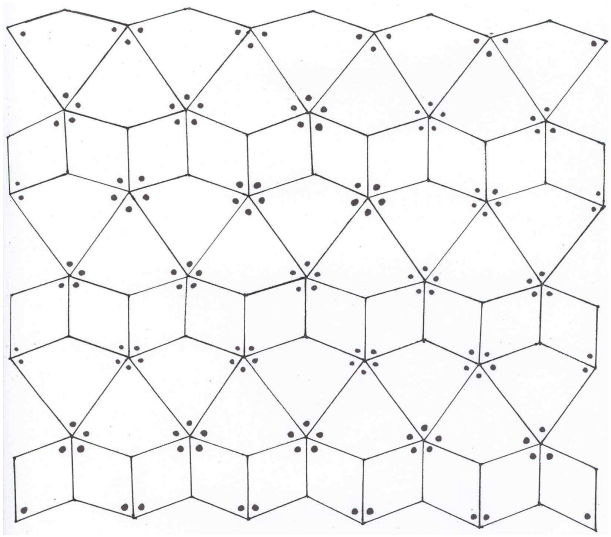

Figure 3: Planar tiling based on the tiles in Fig. 1.

Definition 3.2 *A tiling of a surface S in Euclidean space E^3 is defined as a finite family $\mathcal{T} = \{T_1, \ldots, T_n\}$ of sets T_i (called tiles) which cover S without gaps and overlaps, that is*

- $\cup_{i=1}^{n} T_i = S$
- *$T_i \cap T_j$ has zero measure whenever $i \neq j$.*

In order to interpret the tiles from a biological point of view, rules are required which relate the mathematical objects to the structures they describe. This is done via so-called *decorations*, that is dots on the tiles indicating the location of protein subunits with respect to the shapes of the tiles. Even if two tiles T_i in Definition 3.2 have the same geometrical shape, they are considered as different tiles if their decorations do not coincide.

A tiling may also be viewed as a graph composed of vertices and edges. A vertex is called *decorated* if the angles of all tiles meeting at this vertex are decorated by dots. It is called *undecorated*, respectively *partially decorated*, if none, respectively all such angles are decorated by a dot.

In the case of the papilloma-polyoma family of viruses, the set of tiles consists of the two tiles in Fig. 1. Conditions on how tiles may be assembled are called *matching rules* in tiling theory. The tiling approach to viral capsids relies on a set of assumptions which are motivated by biological considerations and imply the matching rules. In particular, due to the fact that vertices in the tilings correspond either to capsomers (clusters of proteins) or are empty, we have the following matching rule:

Definition 3.3 *The matching rule for the papilloma-polyoma family of viruses is given by the following statement:*
A vertex must be either decorated or undecorated, and partly decorated vertices cannot occur.

In particular, any capsid structure in the papilloma-polyoma family is hence given as a tiling in terms of the tiles in Fig. 1 and the matching rule in Def. 3.3.

Definition 3.4 *Any subset of a tiling is called a patch. A patch is called non-tilable with respect to a given set of tiles and matching rules if it cannot be completed to a tiling of the whole plane using these tiles and rules. It is called tilable otherwise.*

Proposition 3.5 *In any tiling that is tilable with respect to the tiles in Fig. 1 and the matching rule in Def. 3.3, long edges of kite tiles always meet long edges of kite tiles.*

PROOF: Suppose a long edge of a kite tile meets the short edge of a rhomb tile as shown in Fig. 4. Then the insertion of either a rhomb or a kite tile at

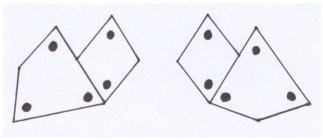

Figure 4: Examples of two non-tilable patches.

the undecorated vertex leads to the creation of a partially decorated vertex. As this is excluded by the matching rule for the papilloma-polyoma family, such configurations are non-tilable, which proves the claim. □

Proposition 3.6 *The patches in Fig. 5 are non-tilable with respect to the tiles in Fig. 1 and the matching rule in Def. 3.3.*

PROOF: In both cases, a completion of the tiling is not possible, because an insertion of either of the two tiles into the gap would lead to a configuration which would enforce the creation of a partially decorated vertex in the next step. As this is excluded by the matching rule in Def. 3.3, these patches are non-tilable. □

Lemma 3.7 *A patch is non-tilable with respect to the tiles in Fig. 1 and the matching rules in Def. 3.3 if it contains kite tiles that meet at a short edge or at their undecorated vertices.*

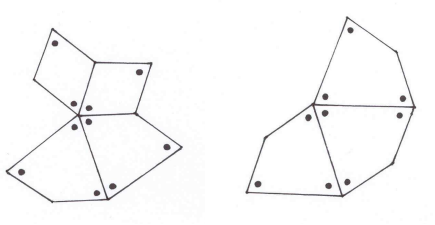

Figure 5: Examples of two non-tilable patches.

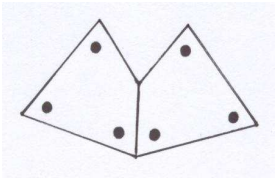

Figure 6: A non-tilable configuration.

PROOF: *Case 1:* Suppose that two kite tiles meet at a short edge as shown in Fig. 6. It is not possible to complete this patch to a tiling, because the insertion of either a rhomb or a kite tile would create a partially decorated vertex which is excluded by the matching rule in Def. 3.3.

Case 2: Suppose that two kite tiles meet at their undecorated vertices but do not share an edge. Then the gap created between the tiles is smaller in angle that the smallest angle of any of the tiles in Fig. 1, for example as shown in Fig. 7. The configuration is hence non-tilable. □

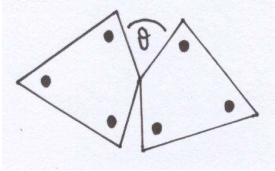

Figure 7: A non-tilable configuration.

It is now possible to prove the main theorem of this subsection.

PROOF: Starting the tiling with a kite tile, there are only two possibilities to continue the tiling according to Proposition 3.5. These correspond to the configurations in Fig. 8. They will be called *parallel kite configuration* (shown on the left) and *anti-parallel kite configuration* (shown on the right) and will be discussed separately in the following.

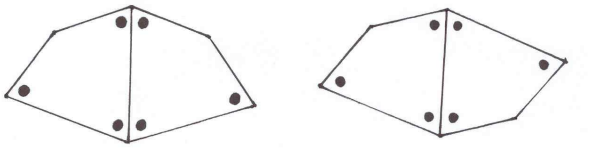

Figure 8: The two possible start configurations based on kite tiles.

Case 1: the parallel kite configuration.

Given the parallel kite configuration, the next steps in completing the tiling must be to surround the short-edges of the kites with rhomb tiles due to Lemma 3.7. Like this, one necessarily arrives at the configuration in Fig. 9. In particu-

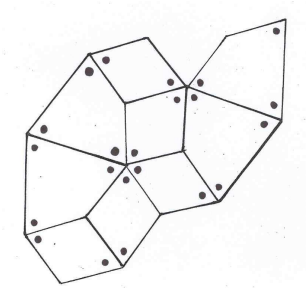

Figure 9: Configuration arising from the parallel kite configuration.

lar, the addition of a further kite that shares its small edges with two rhombs is mandatory. Furthermore, the long edge of this kite must meet the long edge of another kite due to Proposition 3.5, and can only be inserted in a parallel way as it otherwise leads to a non-tilable patch according to Proposition 3.6. From here, one obtains the configuration in Fig. 10. It corresponds to a non-tilable patch according to Proposition 3.6.

Hence, patches involving the parallel kite configuration are non-tilable and a tessellation of the entire plane that contains the parallel kite configuration does not exist.

Case 2: the anti-parallel kite configuration.

Given the anti-parallel kite configuration, it is again necessary to match the long edges of the kites with the long edges of other kites due to Proposition 3.5. Since the creation of any parallel kite configuration is excluded according to case 1, one obtains infinite strips of anti-parallel kite configurations. Due to Lemma 3.7, these different strips are not allowed to have any common edges or vertices and hence the gaps in between the strips must be filled by rhombs. As there is only one possible way of doing this, which corresponds to the tiling

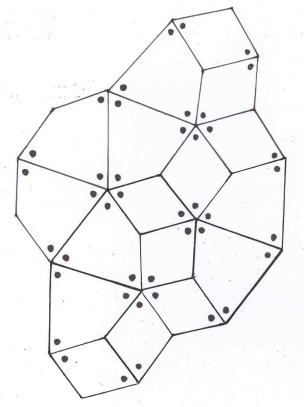

Figure 10: Configuration arising from Fig. 9.

in Fig. 3, the anti-parallel kite configuration always leads to the tiling in this figure.

Moreover, a planar tiling based only on the rhomb tiles in Fig. 1 is excluded by the matching rules in Def. 3.3, because it would lead to partially decorated vertices. Since this together with case 1 and case 2 above exhausts all possibilities of constructing tilings based on the tiles in Fig. 1 and the matching rules in Def. 3.3, the proof of the theorem is complete. □

3.2 Classification of tubular structures

Theorem 3.1 implies that the only planar tessellation compatible with the tiles in Fig. 1 and the matching rules in Def. 3.3 is the tiling in Fig. 3. In this section we classify all distinct possibilities of obtaining cylindrical structures from this planar tiling and hence classify all possible surface lattices for tubular structures in the papilloma-polyoma family. In this way, the tiling approach provides in a natural way an indexing system for experimental data.

In the following, the tiling in Fig. 3 will be considered as a tiling with every other rhomb, or equivalently every rhomb of the same orientation, being shaded in grey as shown in Fig. 11. Grey rhombs will be labeled as follows: Let one of the grey rhombs in Fig. 11 be labelled as $(0, 0)$. Then any other grey rhomb is labelled as (a, b), where $a \in \mathbb{Z}$ denotes the number of steps that it takes in vertical direction, and $b \in \mathbb{Z}$ the number of steps that it takes in horizontal direction, in the sublattice formed by the grey rhombs to reach the rhomb labelled (a, b) from the rhomb $(0, 0)$. Based on this terminology, one obtains the following result.

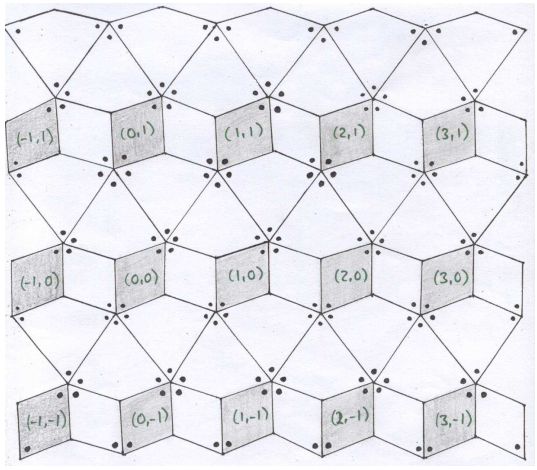

Figure 11: Labelling system for grey shaded rhombs.

Theorem 3.8 *The different tubular structures obtainable from the tiling in Fig. 3 are in a one-to-one correspondence with the cylindrical tilings obtained by identifying the grey rhomb* $(0,0)$ *in Fig. 11 with any grey rhomb* (a,b)*, with* $a \in \mathbb{Z}$ *and* $b \in \mathbb{Z}^{\geq 0}$*.*

PROOF: All different cylindrical tilings corresponding to the planar tiling in Fig. 3 are obtained by identifying tiles of the same type, but at different locations, in the tiling. Since all distinct tiles share a common vertex in the tiling, it is enough to consider only one type of tiles, because cylindrical tilings obtained via an identification of any other type of tiles would lead to the same cylindrical surface lattice. Without loss of generality, the rhombs shaded in grey in Fig. 11 are chosen.

Let any of the grey rhombs be denoted as $(0,0)$ according to the labelling system introduced above. Then an identification of this rhomb with any other grey rhomb labeled (a,b) with $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$ leads to a cylindrical tiling. Due to symmetry properties of the planar lattice, the cylindrical tiling obtained via identifying the grey rhomb at $(0,0)$ with the grey rhomb at (a,b) coincides with the cylindrical tiling obtained by identifying the grey rhomb at $(0,0)$ with the grey rhomb at $(-a,-b)$, but otherwise, all such cylindrical tilings are distinct.

This implies that all cylindrical tilings that are obtained by identifying the grey rhomb $(0,0)$ in Fig. 11 with any grey rhombs (a,b), where $a \in \mathbb{Z}$ and $b \in \mathbb{Z}^{\geq 0}$, are distinct and exhaust all possibilities of forming cylindrical tilings from the planar tiling in Fig. 3. This proves the claim of the Theorem. □

Let $T(a,b)$ denote the cylindrical tiling obtained via an identification of the grey rhomb $(0,0)$ with the grey rhomb (a,b). Then one has the following result.

Corollary 3.9 *The tubular variants of viruses in the papilloma-polyoma family are in a one-to-one correspondence with the cylindrical tilings $T(a, b)$, where $a \in \mathbb{Z}$ and $b \in \mathbb{Z}^{\geq 0}$.*

3.3 The bonding structure of the tubular variants

As in the case of spherical virus particles, the tiling approach not only predicts the protein stoichiometry but also the location of the inter-subunit bonds that stabilize the structure.

Theorem 3.10 *The bonding structure of the tubular variants in the papilloma-polyoma family corresponds to the bonding structure shown in Fig. 12. In particular, arrows are schematic representations of the C-terminal arm extensions.*

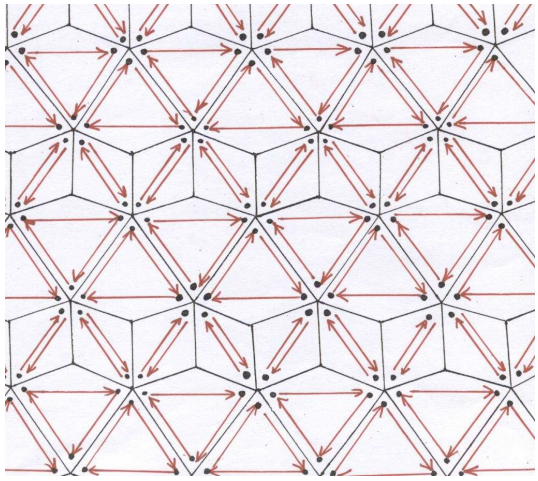

Figure 12: Inter-subunit bonds superimposed on the planar surface lattice.

PROOF: It is an immediate consequence of Theorem 3.8, using the interpretation of the individual tiles as representing dimer, respectively trimer, interactions as in the case of the spherical virus particles. □

4 Comparison with experimental results

In [6] tubular variants of viruses in the papilloma-polyoma family are considered. In particular, two types of narrow tubes have been observed, which are called zero-start and one-start tube respectively. Both are built from unit cells of about $200\mathring{A} \times 80\mathring{A}$, that contain two of the $90\mathring{A}$ pentamers. While zero-start

tubes are ring-like structures with 3 pairs of pentamers per turn of the basic helix, one-start tubes are helical in nature with about 3.3 pairs of pentamers per turn. Both structures are contained in the classification presented here, and correspond to the cylindrical tilings $T(6,0)$ and $T(2,3)$, respectively. It would be interesting to investigate if further tubular structures contained in the classification can be detected experimentally.

Moreover, the tiling approach implies that only all-pentamer structures, that is surface lattices formed from clusters of five protein subunits throughout, are possible in the papilloma-polyoma family due to the measures of the tiles in Fig. 1. This coincides with the experimental observations[1].

In order to compare the tiling in Fig. 3 with the lattice suggested in [6], it is convenient to represent the tiling via its dual tiling, that is the tiling with vertices at the centres of the original tiling. A patch of the dual tiling is shown superimposed on the original tiling in Fig. 13 on the left. For comparison, the surface lattice suggested in [6] is shown on the right. Both lattices share a lot

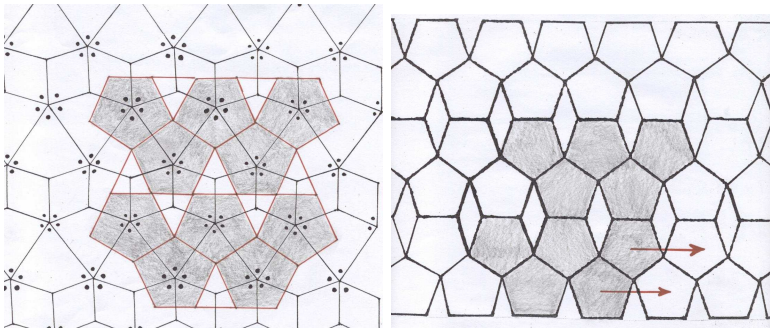

Figure 13: The dual lattice compared with the surface lattice in [6].

of properties, for example the presence of four types of 2-fold symmetry axes, but they are shifted with respect to each other as indicated by the arrows in the figure.

Due to this shift, the basic unit of the tiling differs from the one in [6]. But a comparison of the basic unit of the tiling as shown on the left in Fig. 14 with experimental data corresponding to the near side view (left) and far side view (right) of a zero-start pentamer tube (adapted from [6]) shows that the basic unit agrees with the surface structure of the tube.

Moreover, there are differences in the bonding structure implied by the two approaches. While the maximally achievable bondings in [6] are three bonds per pentamer for every pentamer in the surface lattice, we observe here that the inter-subunit bonds are located both along all of the 2-fold symmetry axes (in fact there are two C-terminal arm extensions along each of the two-fold symmetry axes, see Fig. 12) as well as one further C-terminal arm extension

[1]We remark in this respect that also the wide tubes in [6] were subsequently found to be all-pentamer structures.

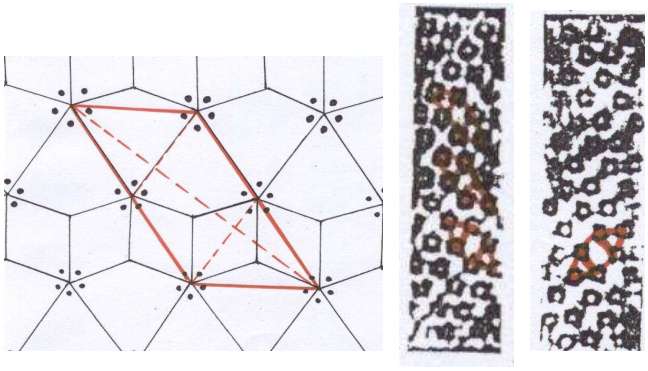

Figure 14: The basic unit compared with experimental data adapted from [6].

per kite tile. The lattice based on the tiling approach hence describes a more stable structure.

5 Discussion

It has been suggested in [6] that the pentagonal surface lattice of the tubular structures in the papilloma-polyoma family of viruses should be rooted in the specific bonding properties of the pentameric units of which they are composed, and that the action of specific bonds between pentamers should be important. The tiling approach implemented here shows that this is indeed the case, and that geometric objects, that is the tiles representing dimer and trimer interactions, are the key to uncovering the structures of the surface lattices of the tubes.

A crucial feature about the tiling approach is that it is able to predict besides the structure of the tubular surface lattices also the location of the inter-subunit bonds, and it would be interesting to probe the tiling approach experimetally by validating these predictions.

Mathematical models are an important tool in the analysis of experimental data since the large width of the diffraction spots compared with the small number of unit cells contained in the width of the tubular particles leads to a relatively poor quality of the optically filtered images. Optical masks for filtering operations are hence crucial, and suitable masks can be constructed based on the mathematical models. It is hence expected that the results presented here should have direct impact on experiments concerning tubular variants of viruses in the papilloma-polyoma family.

6 Acknowledgements

Financial support via an EPSRC Advanced Research Fellowship is gratefully acknowledged.

References

[1] D.L.D. Caspar and A. Klug, Cold Spring Harbor Symp. Quant. Biol. 27, 1 (1962).

[2] I. Rayment, T.S. Baker, D.L.D. Caspar and W.T. Murakami, Nature 295, 110 (1982)

[3] R.C. Liddington, Y. Yan, J. Moulai, R. Sahli, T.L. Benjamin and S.C. Harrison, Nature 354, 278 (1991).

[4] S. Casjens, "Virus structure and assembly, Jones and Bartlett, Boston , Massachusets (1985).

[5] R. Twarock, J. Theor. Biol. 226, 477 (2004).

[6] N.A. Kiselev and A. Klug, J. Mol. Biol. 40, 155 (1969).

[7] R. Twarock, *Protein stoichiometry and bonding structure for viral capsids based on tiling theory*, in preparation.

[8] Modis, Y., Trus, B.L. & Harrison, S.C. (2002) Atomic model of the papillomavirus capsid, EMBO J. 21, 4754-4762.