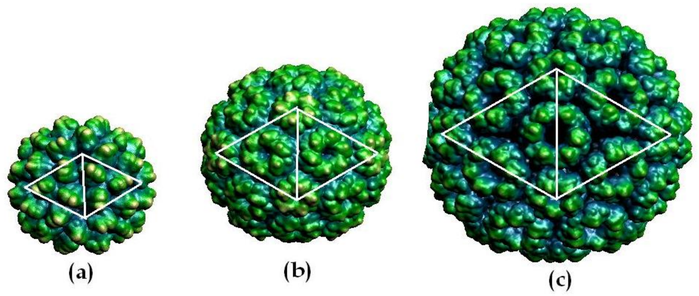
Self-Assembly of Viral Capsids via a Hamiltonian Paths Approach: The Case of Bacteriophage MS2

Thomas Keef, Department of Mathematics, University of York
tk506@york.ac.uk

Abstract: For large classes of viruses the assembly of the viral capsids that encapsulate and hence provide protection for the viral genome can be modeled as tile assembly, where tiles represent are schematic representations of suitable protein building blocks of the capsid. However, for certain classes of RNA viruses, where interactions between the capsid proteins and the RNA are crucial for assembly, this is not sufficient. We therefore developed a new method that incorporates these interactions as boundary conditions into tile assembly.

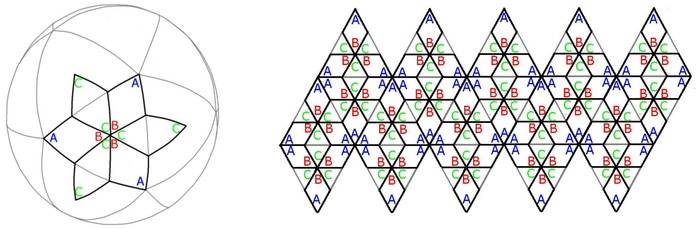
Introduction:

Viruses are fascinating micro-organisms, consisting of a very compact genome and a protective protein shell that hijack host cells typically between one hundred or one thousand times their size. Viral capsids are shells constructed from many copies of one, or a few, identical protein subunits, normally with icosahedral symmetry. The first model for predicting the location of these proteins and the resulting capsid sizes was provided by Capsar and Klug over forty years ago [1]. It involved the sub-triangulating of icosahedral faces into smaller triangular facets, giving rise to a series of numbers, called *the Caspar-Klug series* with values $T = h^2 + hk + k^2$ where $h, k \in \mathbb{Z}^{\geq 0}$. This sub-triangulation of icosahedra gives rise to more protein subunits on the new vertices as shown below, producing a capsid with a total of $60T$ protein subunits.



The first three members of the Caspar-Klug series are (a) $T = 1$, (b) $T = 3$ and (c) $T = 4$ with two of the icosahedral triangles sketched to show the introduction of more protein subunits. All figures have been adapted from [2].

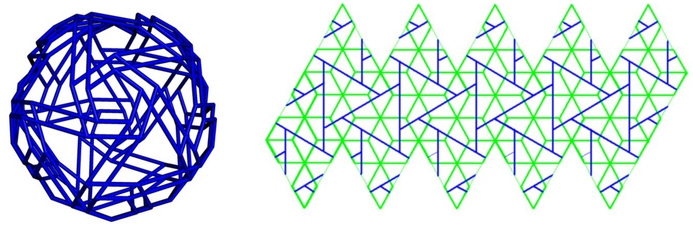
However, some viruses cannot be described by triangulations of icosahedra, and Viral Tiling Theory (VTT) gives a classification for these viruses relaxing the requirement for tiling with triangles, allowing the use of different shapes such as rhombs [3]. VTT also encodes the locations of inter-subunit bonds and and contains Caspar-Klug theory as a special case.
MS2 is an example of a virus that has the correct numbers of proteins in the locations predicted by VTT, but incorrect orientations. The Virus is a $T = 3$ particle with proteins organised with clusters (capsomeres) of five proteins on the global five-fold symmetry axes and clusters of six proteins elsewhere [4]. Besides the interactions of the proteins within a capsomere, each protein participates in an interaction (dimer interaction) with precisely one protein in another cluster. This is encoded by a rhomb tiling, in which each rhomb represents an interaction between the proteins that are schematically encoded by letters in the tiling.



MS2 has a capsid consisting of 180 proteins, with dimer interactions between capsomeres encoded here by rhomb tiles. The green C-C Dimers are symmetrical, while the blue/red A-B dimers have their symmetry broken by the binding of RNA stem loops.

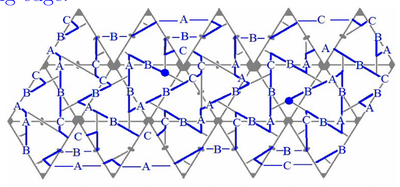
RNA Directed Assembly:

Cryo-electron microscopic data on the RNA density have shown that the first layer of RNA takes on the shape of the polyhedron shown below. Due to the facts that assembly is directed by RNA and that the RNA forms a polyhedral cage within the assembled capsid, we suggest that capsid assembly follows pathways on the cage structure. During the assembly process capsid proteins attach to the partly formed capsid with the aid of the RNA, which follows Hamiltonian pathways, i.e. pathways meeting each vertex exactly once. The vertices can be thought of as the locations of the RNA stem loops binding into the non-symmetrical A-B dimers producing a breaking of symmetry from the symmetrical C-C dimers.



The RNA follows the lines shown in these cartoons with the RNA stem loops binding the rhombs at the vertices.

Determining the ensemble of different assembly pathways hence translates into the problem of finding Hamiltonian paths with certain properties. Each Hamiltonian path can be encoded by three letters: A labeling a clockwise move along a short edge around a five-fold axis (i.e. the centre of a pentagon formed by five short edges), C labeling an anti-clockwise move along a short edge, and B labeling a move (in any direction) along a long edge.



The route for the sequence BABABCCCBABCCBABCCBAABAABC-CBAABCBABCBAABCCBAABCBCBABAB.

Since assembly is nucleated by a hexamer of dimers, that is a cluster of six dimers [4] we start our assembly only with hamiltonian pathways starting around one of these hexameres. From these starting configurations we allow any pathways which then meet all vertices, adding dimers along the way. We then investigate the concentrations of assembly intermediates at equilibrium, following each hamiltonian pathway using this relation [5, 6]:

$$\frac{[n]}{[n-1][1]} = S_d S_n K'_n \tag{1}$$

Where $[i]$ is the concentration of pathway intermediate (i), S_i are symmetry factors and K'_i is function of the number of contacts formed in the transition from intermediate $(i-1)$ to (i). $\gamma(i)$ is the number of inter-subunit contacts, R is the gas constant and T is the temperature.

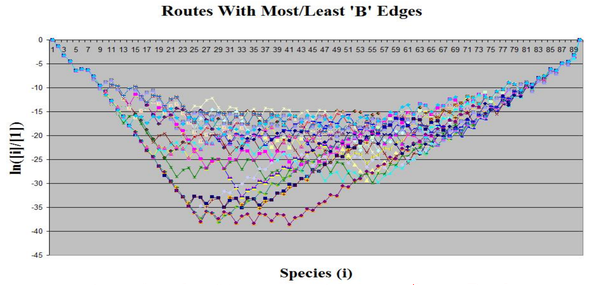
$$K'_n = e^{\left(-\frac{\gamma(n)\Delta G_{contact}}{RT}\right)} \tag{2}$$

Results:

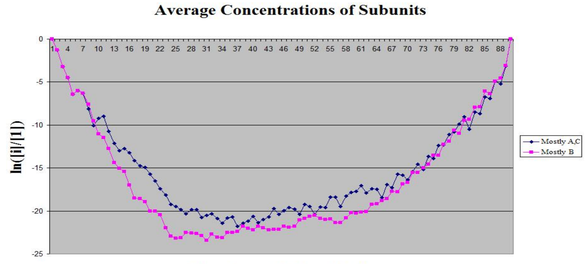
We set the association energy for contacts $\Delta G_{contact} = -1$ and compare two sets of assembly concentrations:

1. Routes preferring hexameres of dimers, i.e. following more B edges.
2. Routes preferring pentameres of dimers, i.e. following more A and C edges.

Below is a graph of relative concentrations $[i]/[1]$ for the 32 routes with the most B edges and the 32 routes with the least B edges. This shows a wide range of relative concentrations and if we look at the averaged concentrations for the two types of routes it is clear that there is a difference in the concentrations of intermediates in the two sets. difference in the concentrations of intermediates in the two sets.



Relative concentrations for the 64 routes with most/least B edges .



Averaged concentration profiles for the routes with the most/least B edges.

We still need to incorporate energy contributions from the RNA-protein interactions, which we will do via additional boltzman factors in 2 but we expect similar differences. Once we have this, we will use this approach in collaboration with the Astbury Centre for Structural Molecular Biology at the University of Leeds to determine assembly scenarios indirectly: they measure concentrations of assembly intermediates.

We hope to deduce:

- The most likely assembly scenario (RNA following more B or A,C edges).
- The length of RNA in the first RNA cage: a ratio in lengths of RNA of 1.21:1 is predicted from the scenario with more B edges vs the one with more A and C edges.

References

[1] D. L. D. Caspar and A. Klug. *Cold Spring Harbor Symp. Quant. Biol.*, 27:1–24, 1962.
[2] V. S. Reddy, P. Natarajan, et al. *J. Virol.*, 75(24):11943–11947, 2001.
[3] R. Twarock. *J. Theor. Biol.*, 226(4):477–482, 2004.
[4] P.G. Stockley et al. A simple, rna-mediated allosteric switch controls the pathway to formation of a t=3 capsid. page submitted.
[5] A. Zlotnick. *J. Mol. Biol.*, 241(1):59–67, 1994.
[6] T. Keef, A. Taormina, and R. Twarock. *J. Phys. Biol.*, (2):175–188, 2005.

Acknowledgments: I would like to thank my collaborators Nick Grayson, Simone Severini and Reidun Twarock, my PhD supervisor, for helpful discussions. This work has been funded by the EPSRC LSI programme, grant number GR/T26979/01.