UNIVERSITY OF PLYMOUTH

Mathematical and computational modelling in Mathematical Virology

Plymouth, December 3, 2020

Pierre-Philippe Dechant
work with R Twarock and Y-H He

School of Science, Technology & Health, York St John University
York Cross-disciplinary Centre for Systems Analysis, University of York
Department of Mathematics, University of York

Overview

Mathematical Modelling

- Watson & Crick: Icosahedral and helical symmetry
- Caspar & Klug: Triangulations
- Twarock: More general surface tilings
- Affine symmetry: genome and capsid

Computational Modelling

- Bioinformatics: packaging signals
- Gillespie stochastic simulations: epidemiological, infection, assembly
- Machine learning: fitness landscape

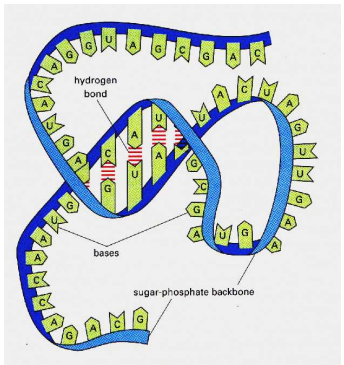
Main references

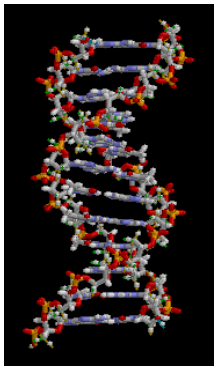
Models of Viral Capsid Symmetry as a Driver of Discovery in Virology and Nanotechnology
P-P Dechant, R Twarock, The Biochemist, 2021

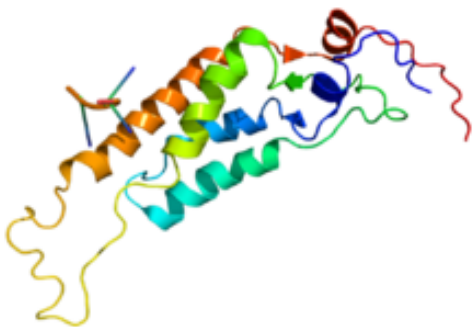
Machine-learning a virus assembly fitness landscape
P-P Dechant, Y-H He, PLOS One, arXiv preprint arXiv:1901.05051, 2021

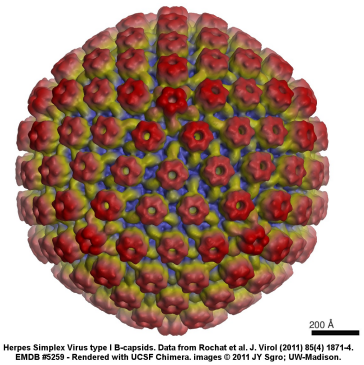
What is a Virus?

- Transported piece of genetic information that e.g. can run a programme in a host cell
- Genome: RNA or DNA – single- or double-stranded
- Fragile – needs to be protected by a protein shell: capsid
- Gene $\rightarrow$ mRNA $\rightarrow$ protein (transcription and translation)
- Each protein = amino acid chain folds into a 3D shape: one geometric building block

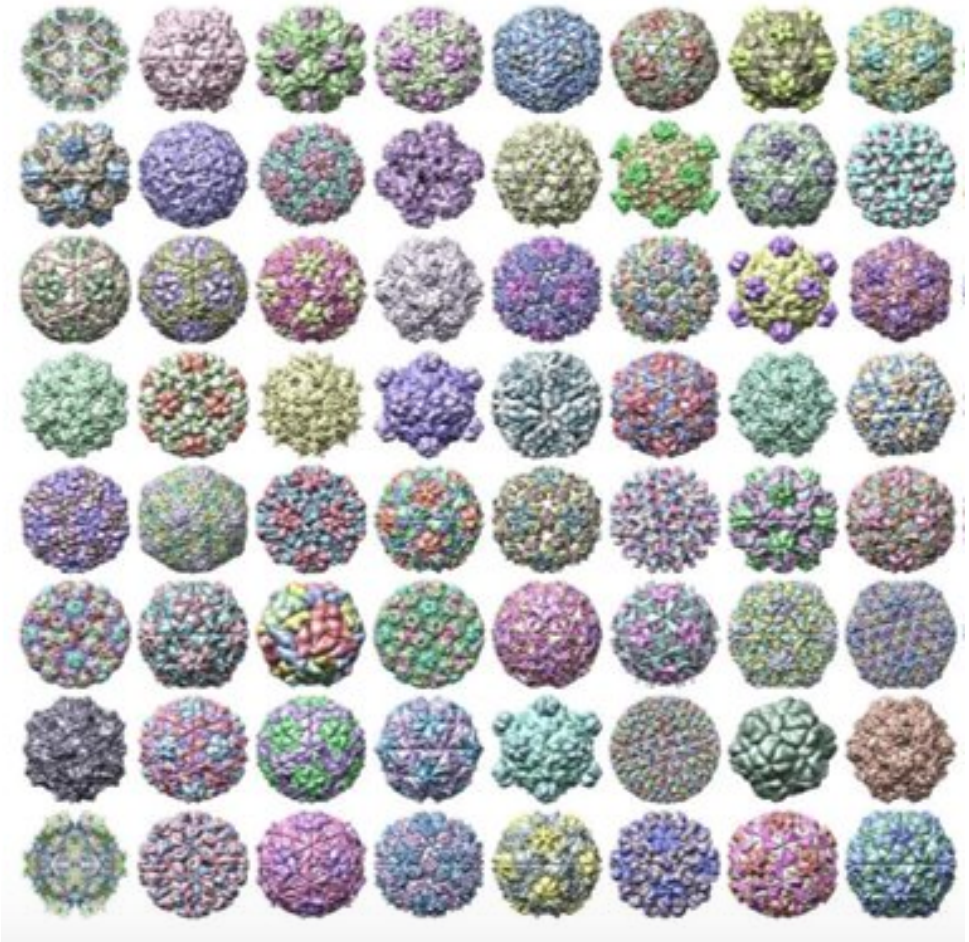


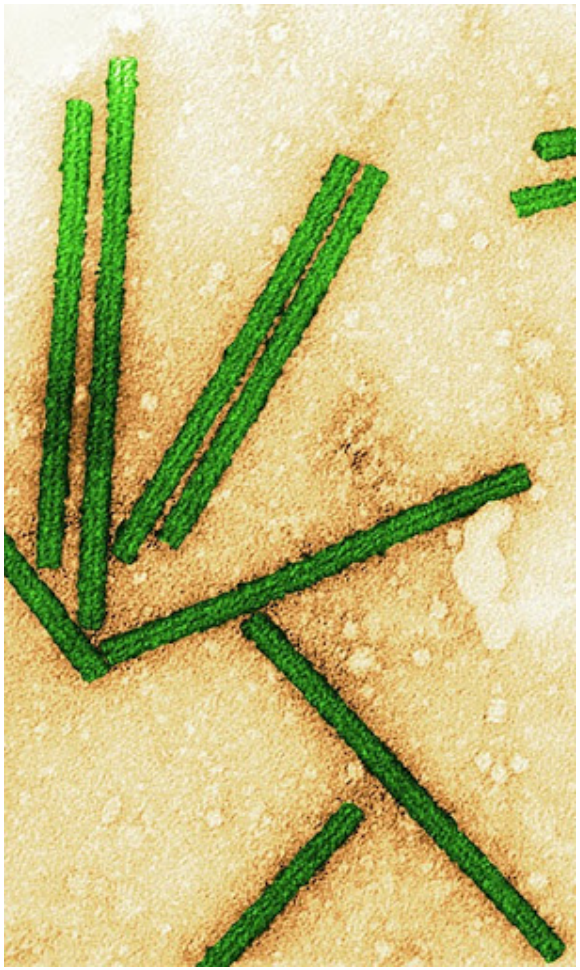




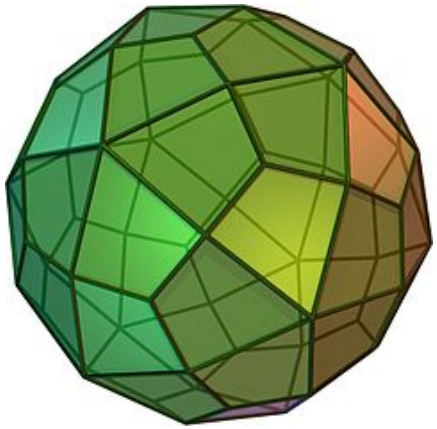
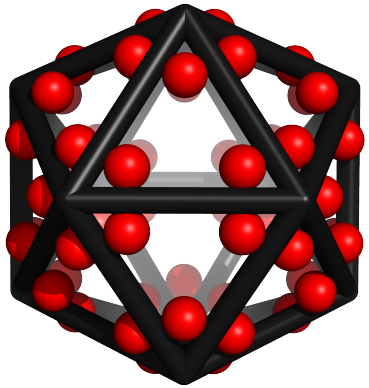
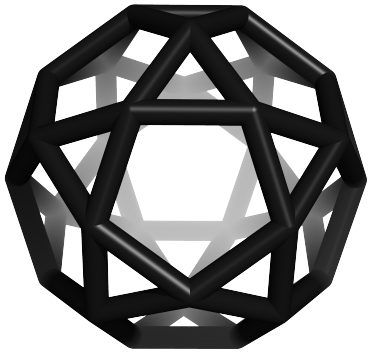


Many viruses are icosahedral – others helical



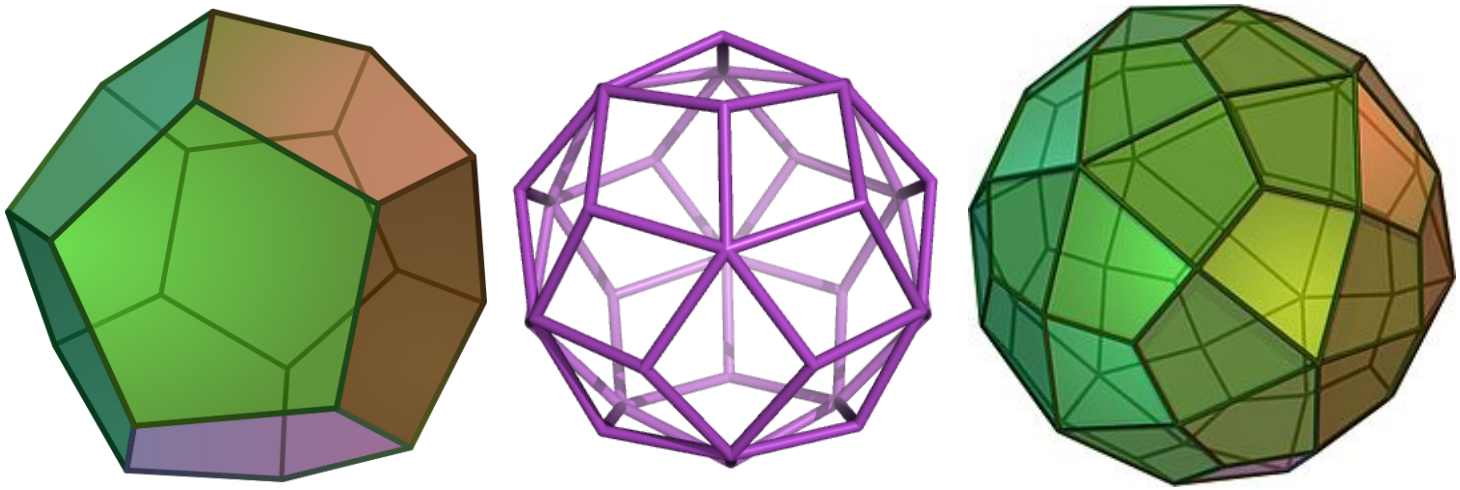


Watson and Crick: Principle of Genetic Economy

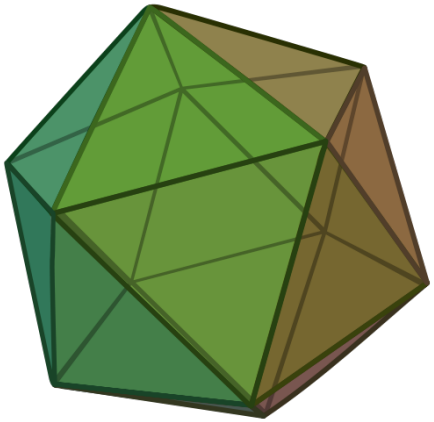
- Watson & Crick: Genetic economy $\rightarrow$ symmetry $\rightarrow$ icosahedral is largest
- Rotational icosahedral group is $I = A_5$ of order 60
- Full icosahedral group is the Coxeter group H_3 of order 120 (including reflections/inversion); generated by the root system icosidodecahedron

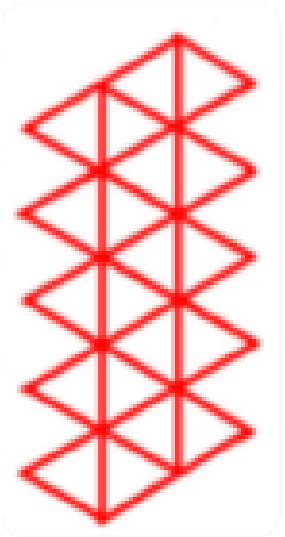
Icosahedral solids

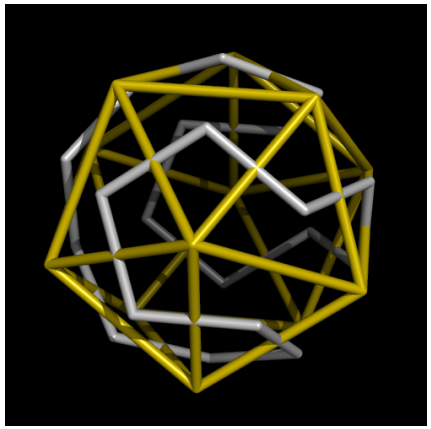


Other tile shapes can also give icosahedral tilings: pentagons (dodecahedron), rhombuses (rhombic triacontahedron), kites (deltoidal hexecontahedron)

Assembling an Icosahedron

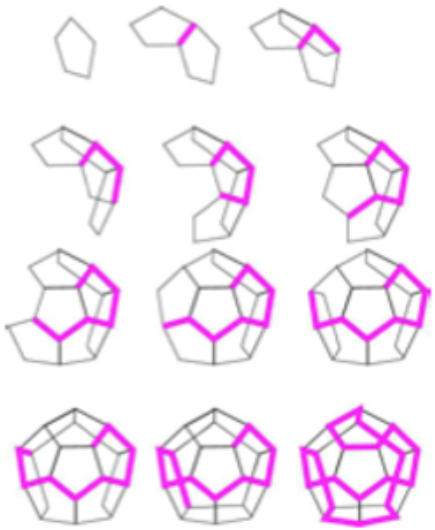


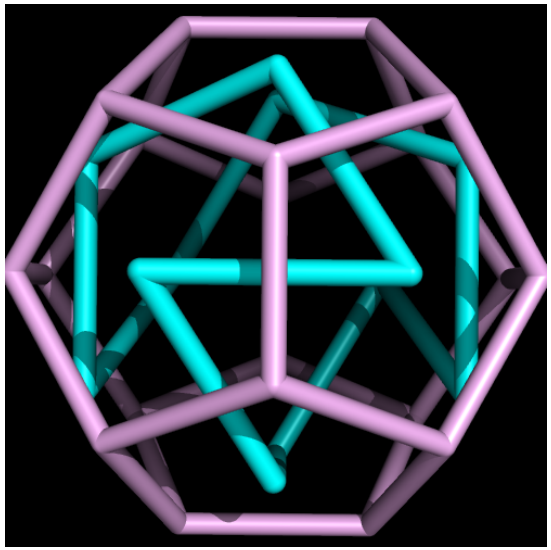


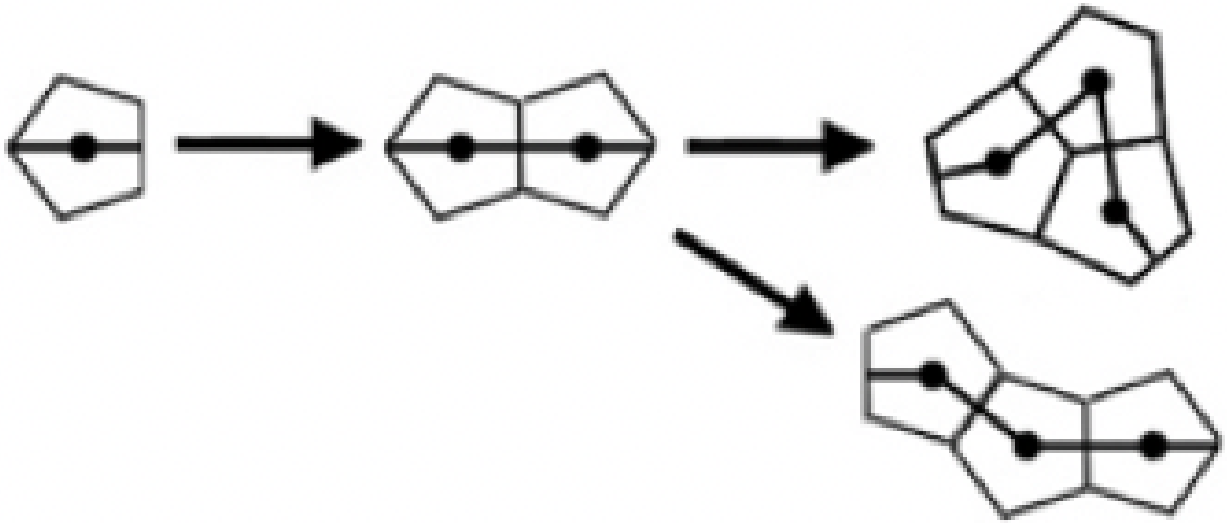


- Assemble from 20 identical triangular building blocks
- The order of addition gives a Hamiltonian path on the dual dodecahedron

Assembly and thermodynamics – Hamiltonian paths



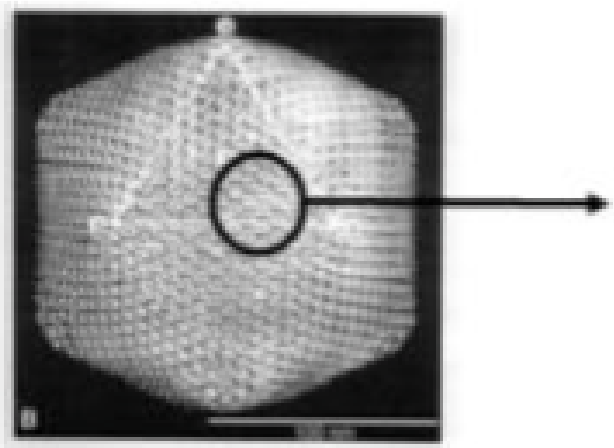




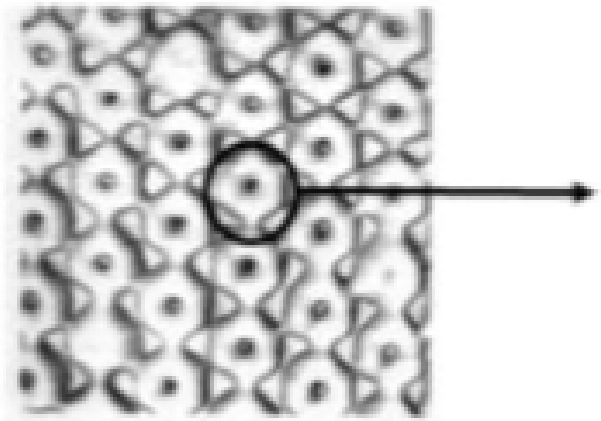
More than just icosahedral symmetry?

Solved the original problem

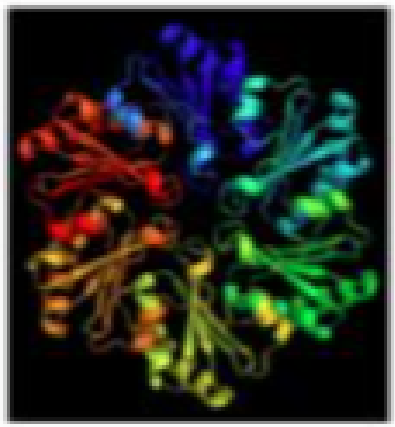
But with that solution (triangular building blocks), can viruses do better?



Protein shell
(viral capsid)



Protein clusters
(capsomeres)

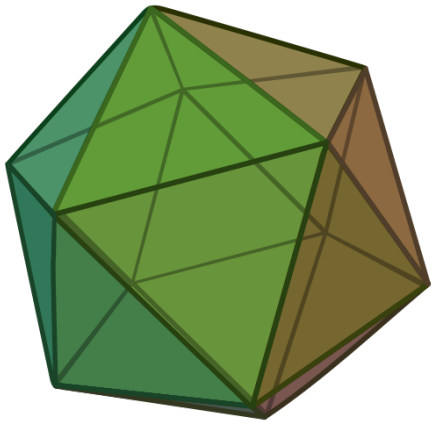


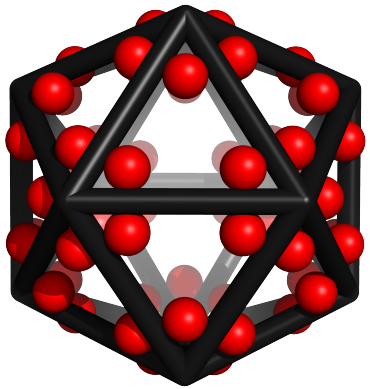
a cluster of 6
proteins
(hexamer)

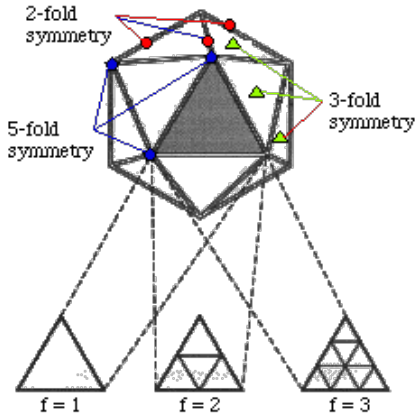
Caspar and Klug: Triangulations

A compromise between mathematics & biology: quasi-equivalence

- Mathematical upper limit of 60 for equivalent subunits, but biologically want to do better!
- Gene $\rightarrow$ can already make a triangle $\rightarrow$ might as well make many! Triangles are distinguished in that they can be decomposed into smaller triangles.
- Caspar-Klug ideas of quasi-equivalence and triangulations



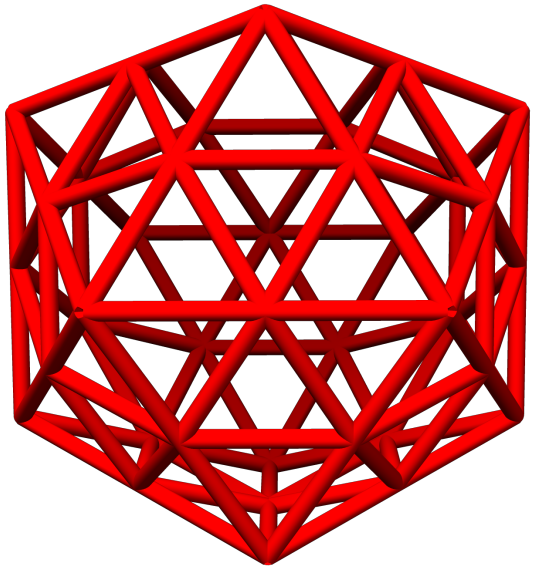


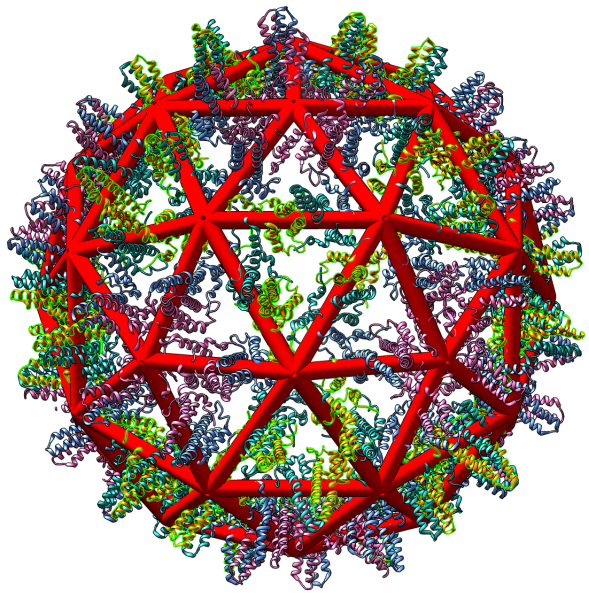


Caspar-Klug Triangulations

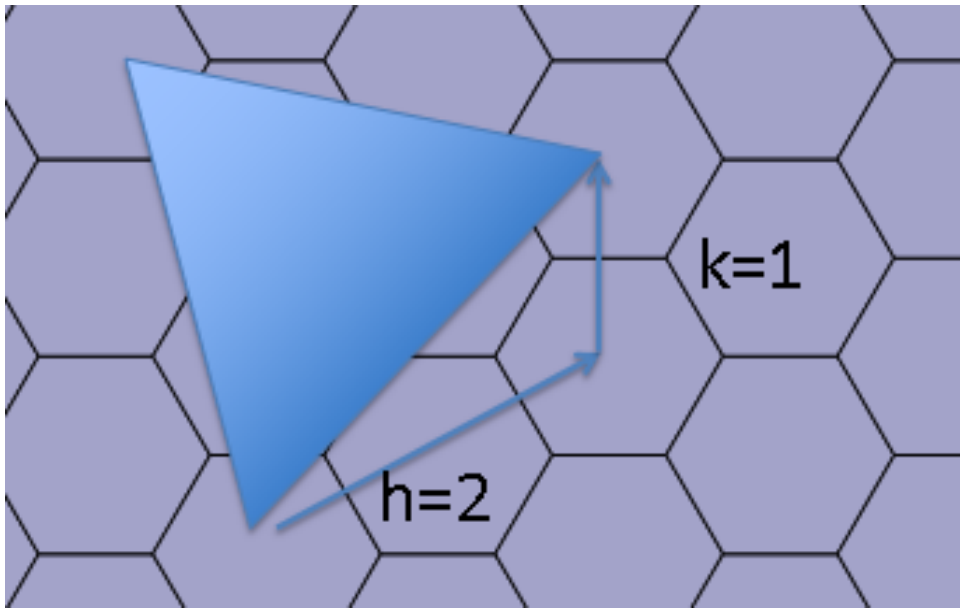
Triangulation number T

- Counts the number of small triangles per icosahedral face
- E,g. Hepatitis B virus (only one structural gene) has $T = 4$





Viruses: Caspar-Klug triangulations $T = h^2 + hk + k^2$



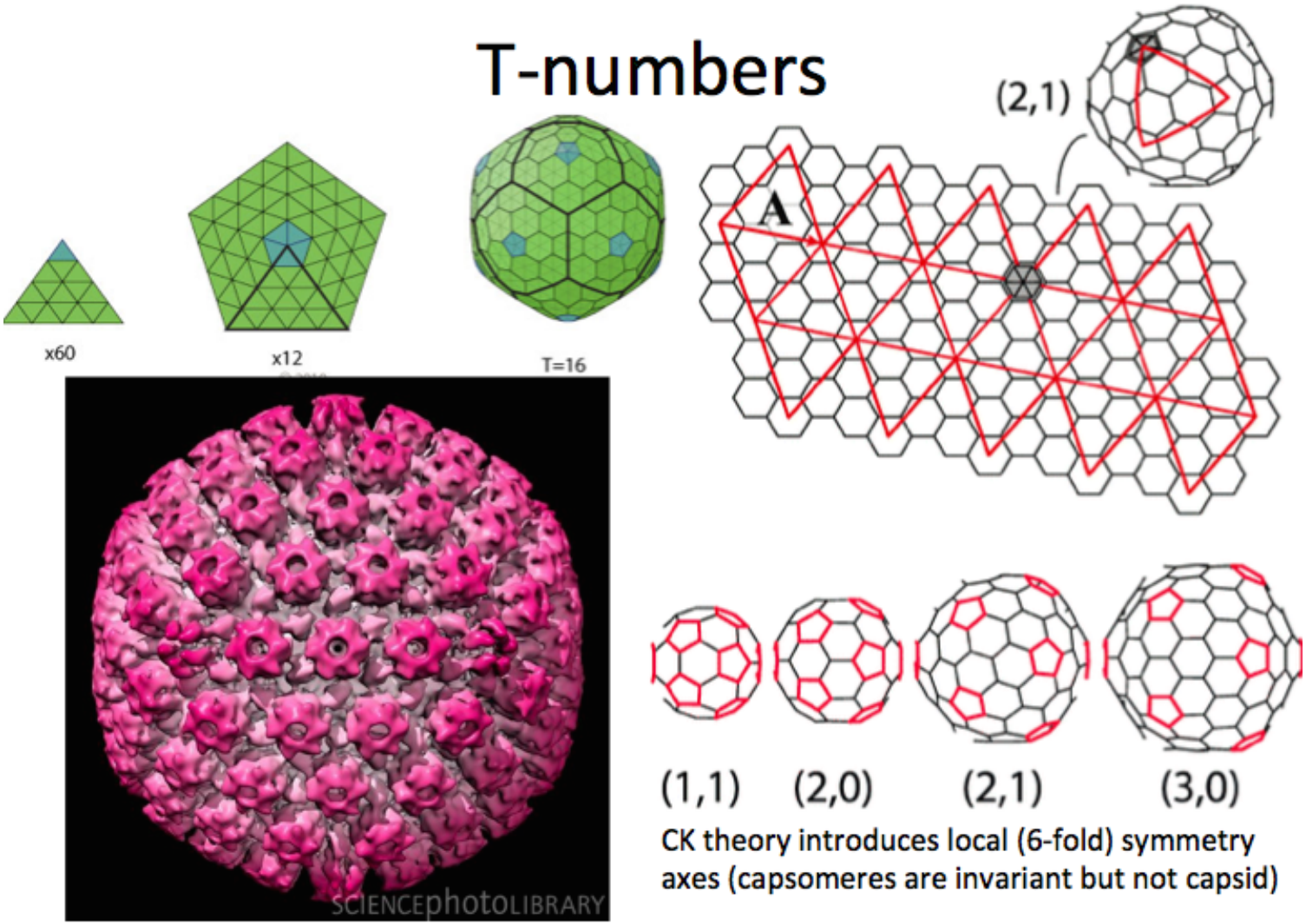
Integer steps h and k in hexagonal directions

give allowed triangulation numbers $T = h^2 + hk + k^2$.

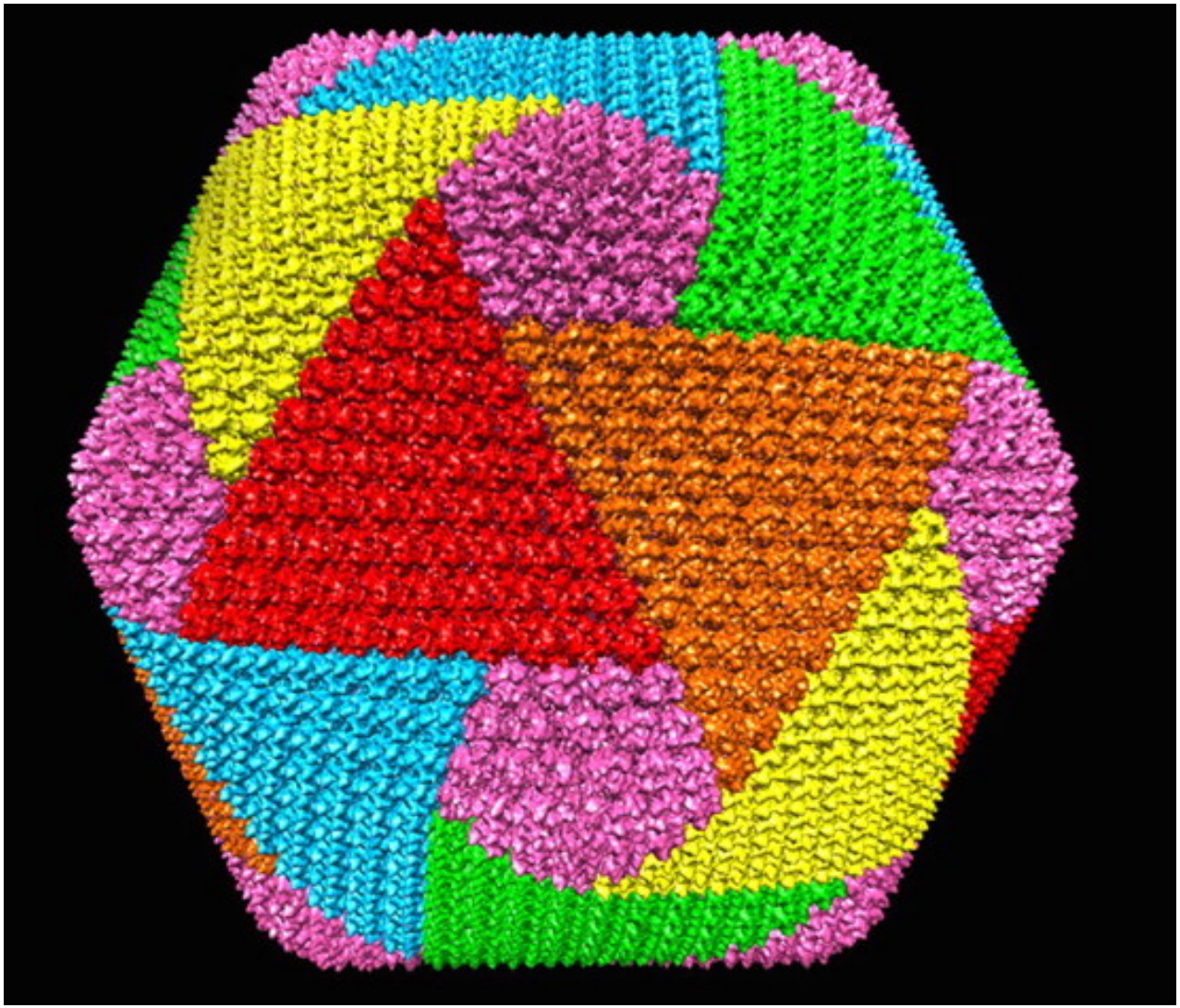
T orbits so $60T$ proteins

60 of which form 12 pentamers, and $60(T-1)$ form $10(T-1)$ hexamers.

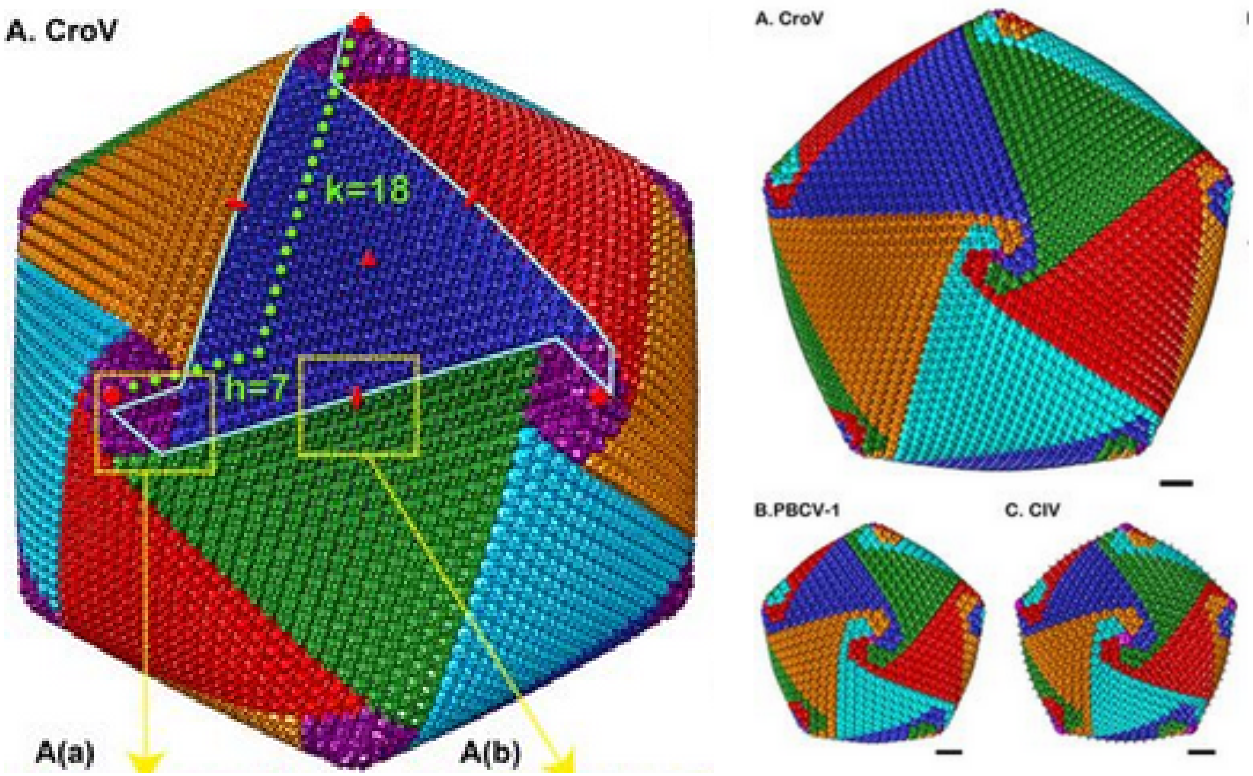
Viruses: Caspar-Klug $T = h^2 + hk + h^2$ triangulations



A recent discovery: Giant viruses

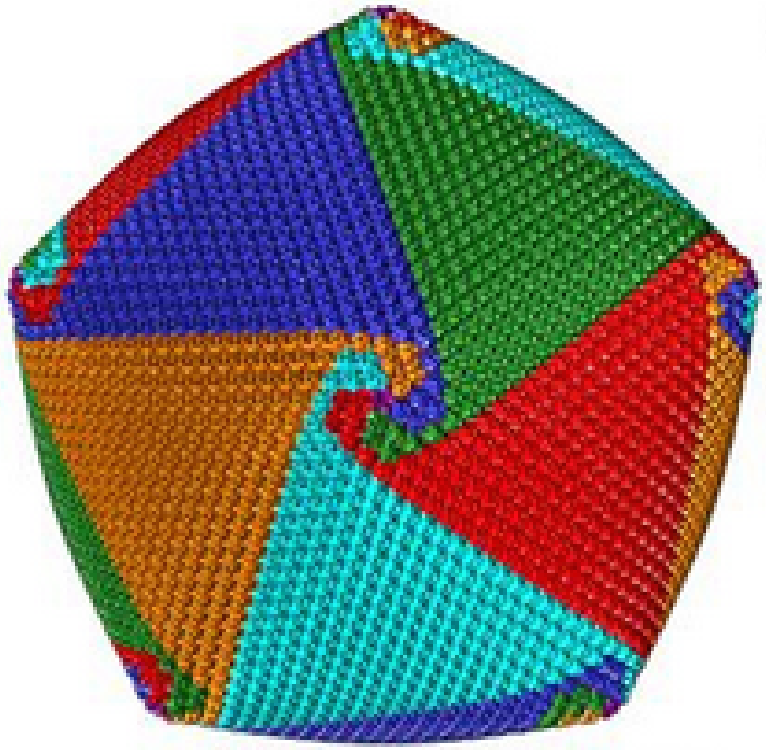


A family of cages with a common approach – little hooks



Pentasymmetrons and trisymmetrons

A common approach – little hooks

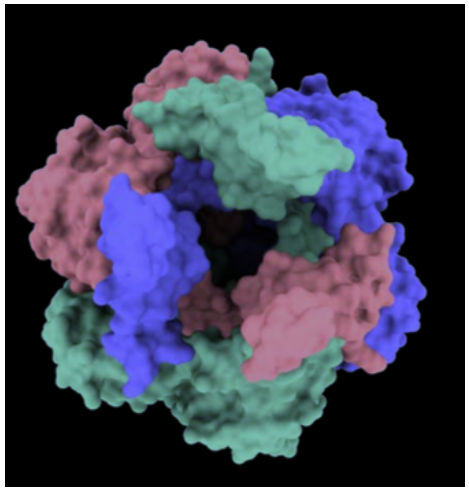


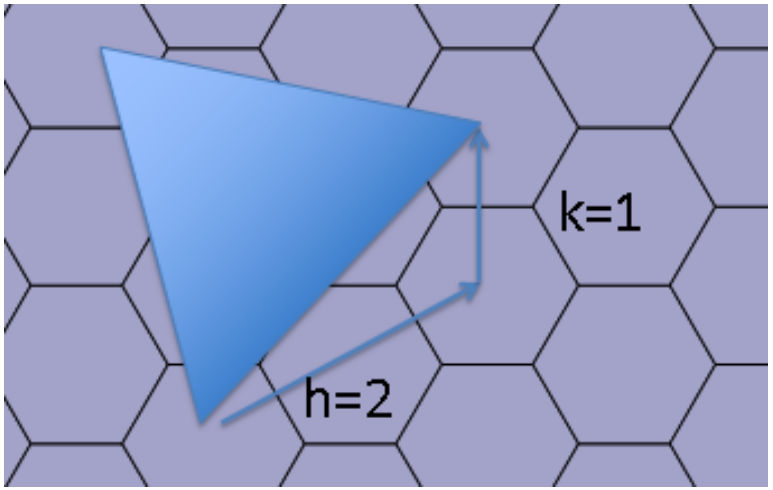
Pentasymmetrons and trisymmetrons

A family of solutions: $h = 7$ – and some gaps

- Chilo iridescent virus: $T = 147$, $h = 7$ and $k = 7$
- Paramecium bursaria Chlorella virus 1: $T = 169$, $h = 7$ and $k = 8$
- Phaeocystis pouchetti virus: $T = 219$, $h = 7$ and $k = 10$
- Faustovirus: $T = 277$, $h = 7$ and $k = 12$
- Pacman virus: $T = 309$, $h = 7$ and $k = 13$
- Cafeteria roenbergensis: $T = 499$, $h = 7$ and $k = 18$

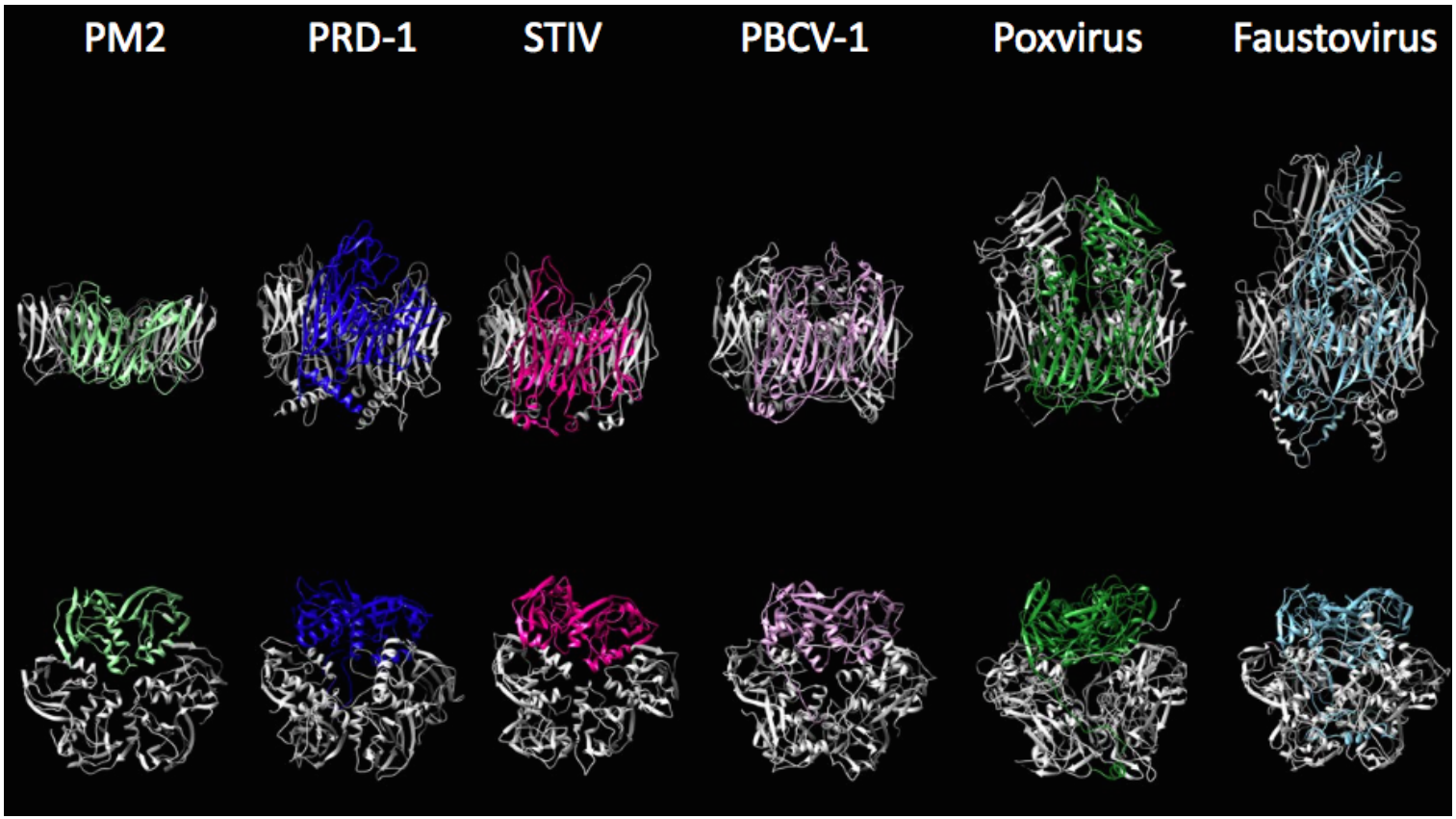
Major capsid protein





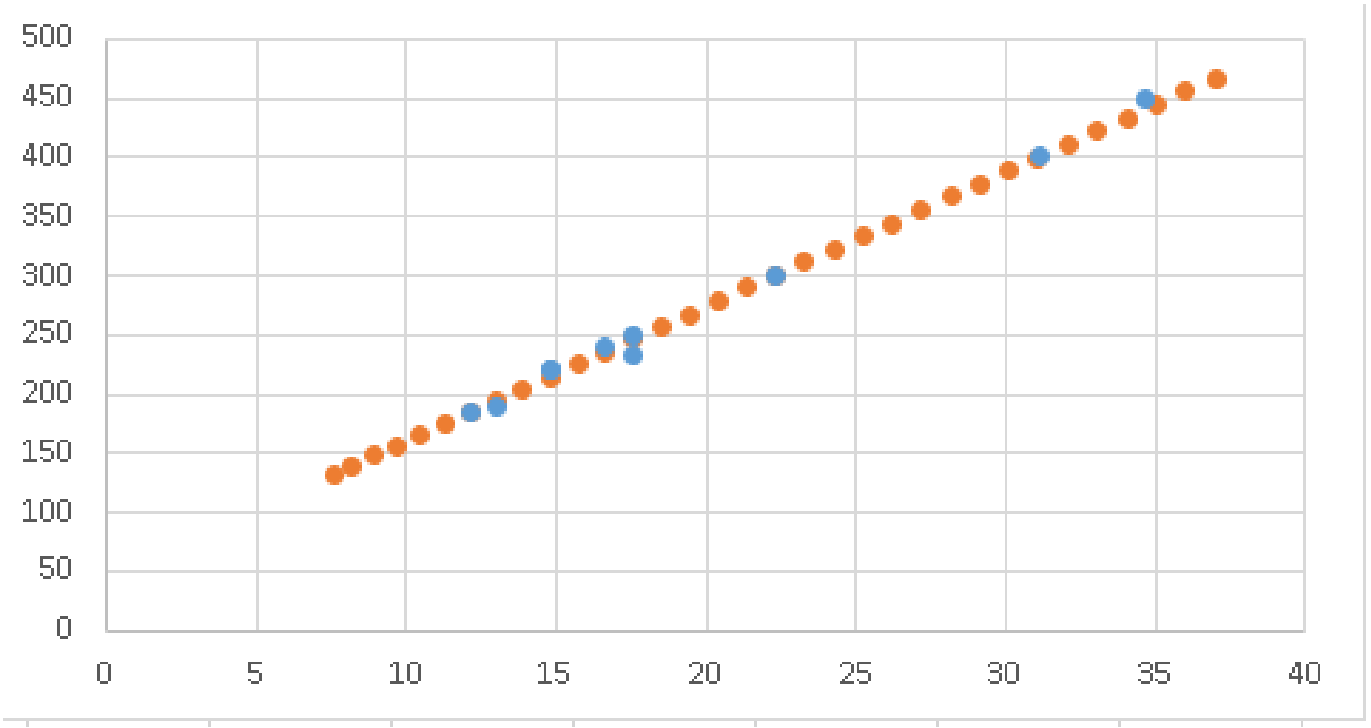
T is an area, so $\sqrt{T}$ gives size of triangle and thus also particle diameter

Major capsid protein – evolutionary conservation



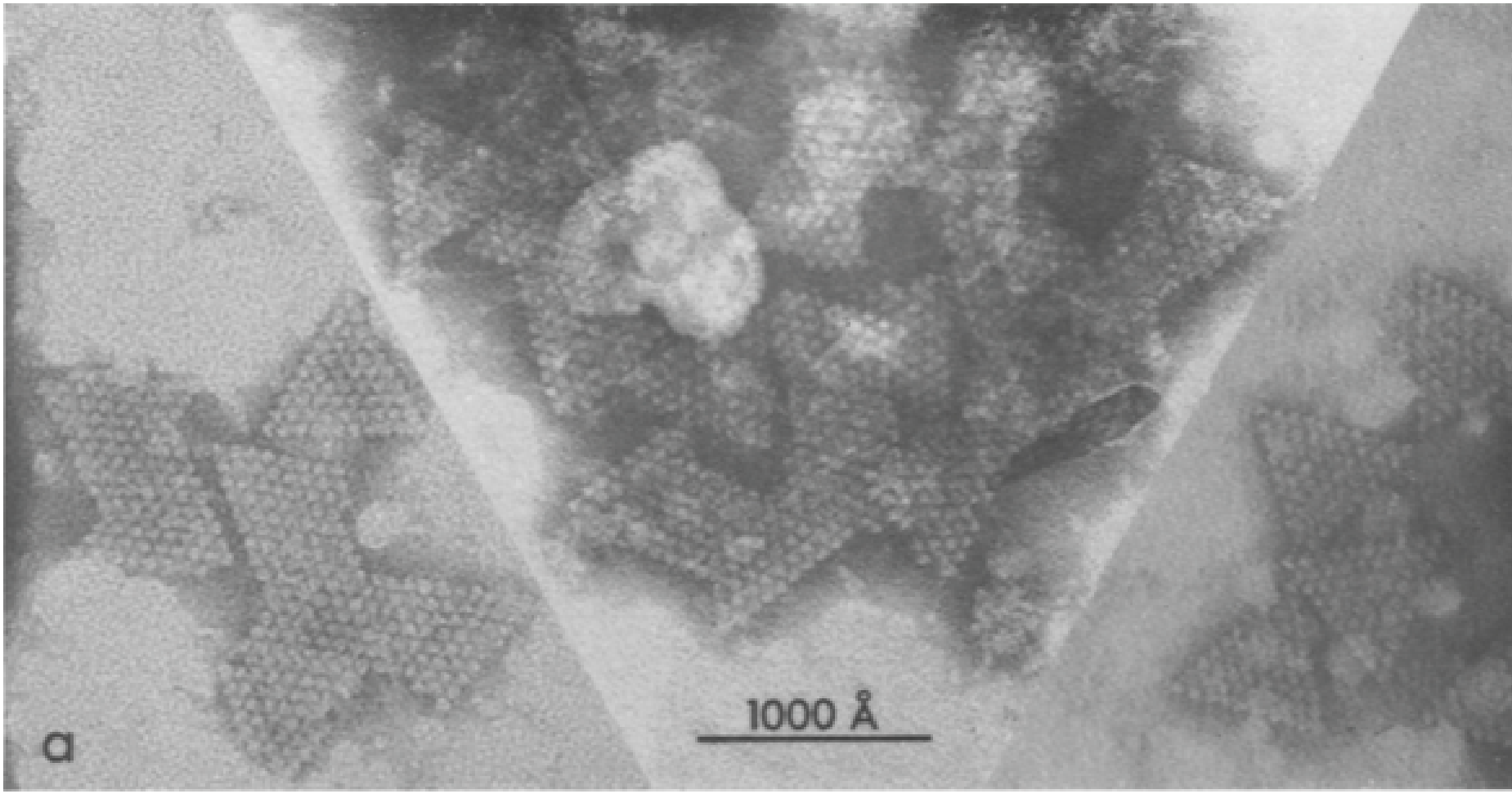
For same size tessellating unit particle size should scale as $\sqrt{T}$

Predict a scaling relation

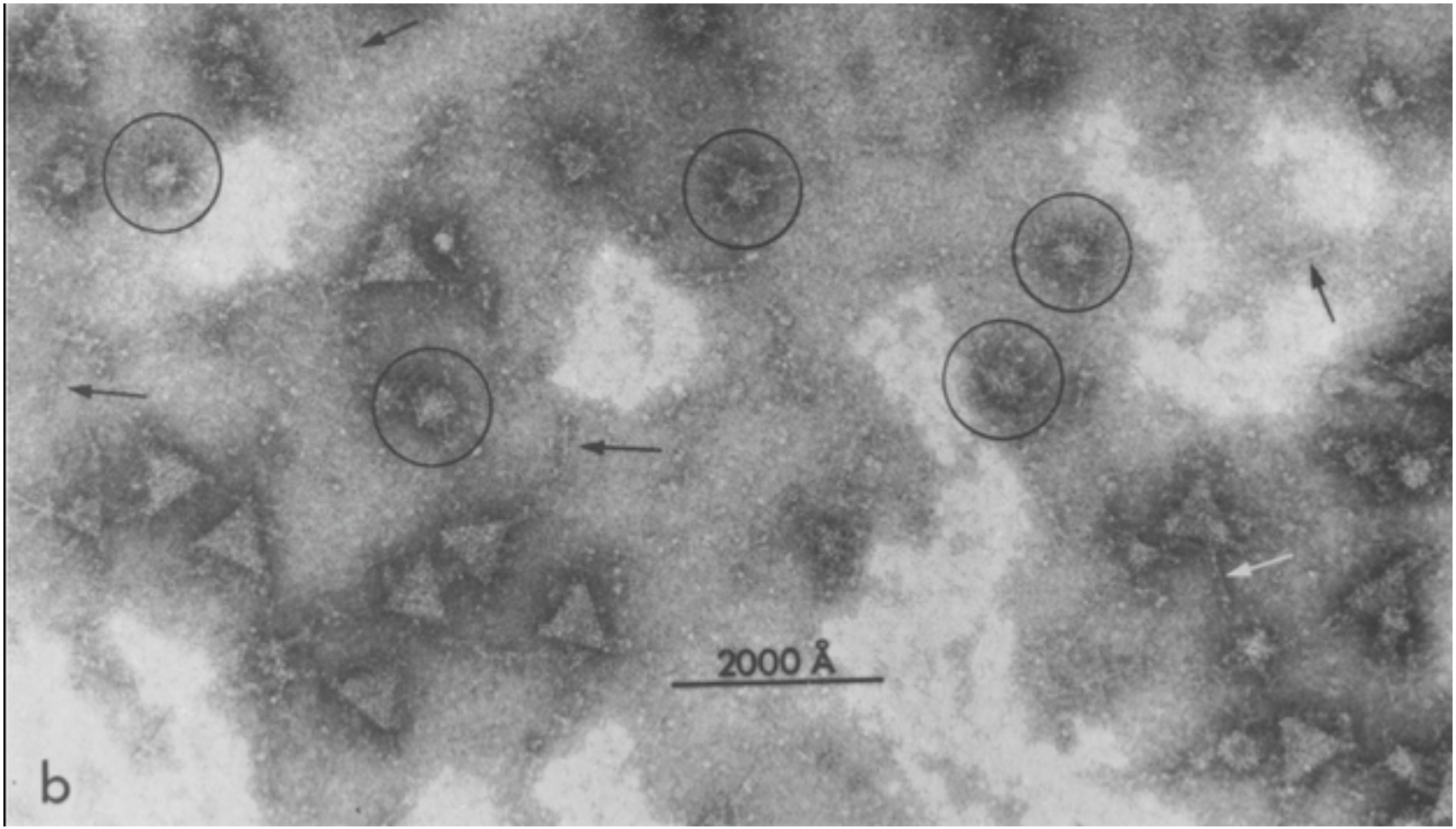


Missing points allowed geometrically but less stable? Or just not yet discovered? Predict Tetraselmis virus 1 TetV-1 of $257nm \pm 9nm$ is exactly $T = 343$. Predict holes in family exist and sizes are given by this scaling

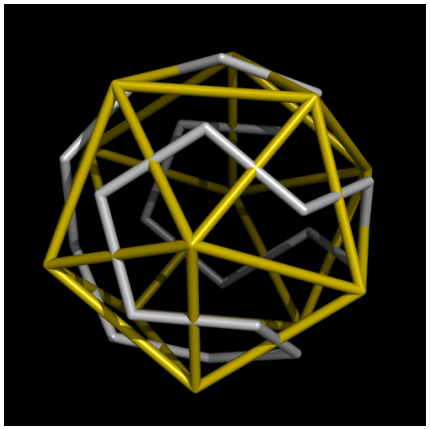
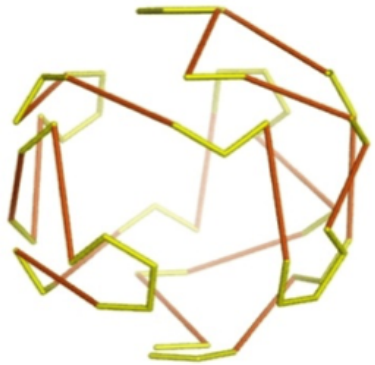
Trisymmetrons and Pentasymmetrons



Major capsid protein - trimer, pseudohexamer



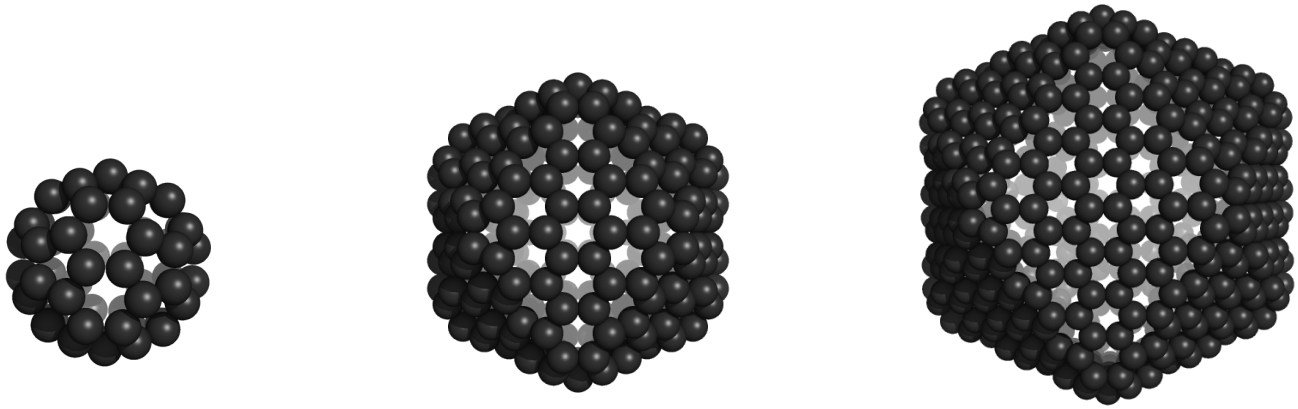
Build from prearranged blocks? Back to Hamiltonian paths


- Are the trisymmetrons and pentasymmetrons preformed? (or is that just what virions fall apart into?)
- If trisymmetrons are assembled then we're back to a Hamiltonian path for the icosahedron
- If pentasymmetrons then get a slightly new polyhedron

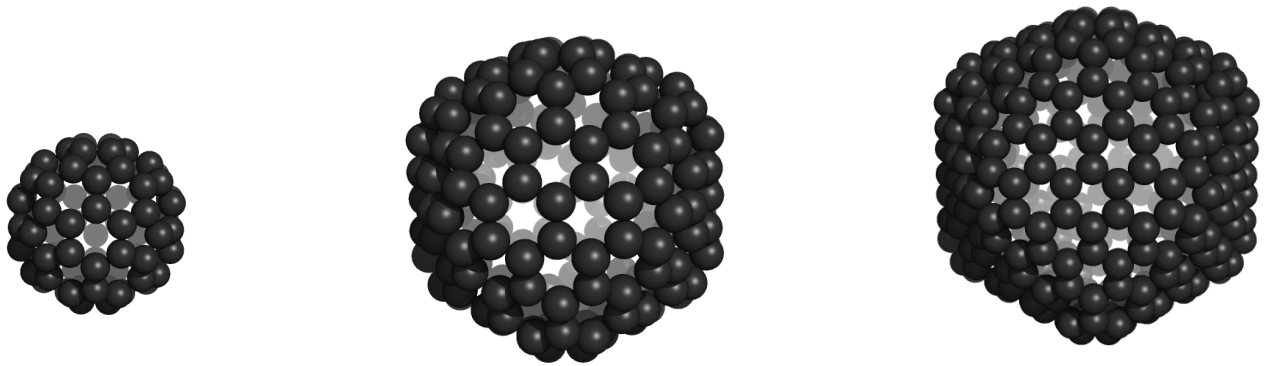
Other objects made from identical building blocks: Fullerenes

- Other icosahedral objects in nature: football-shaped fullerenes
- Different shells with icosahedral symmetry: e.g. C_{60}, C_{240}, C_{540}
- Follow Caspar-Klug-like layouts (e.g. $T = h^2$ and $T = 3h^2$ families)

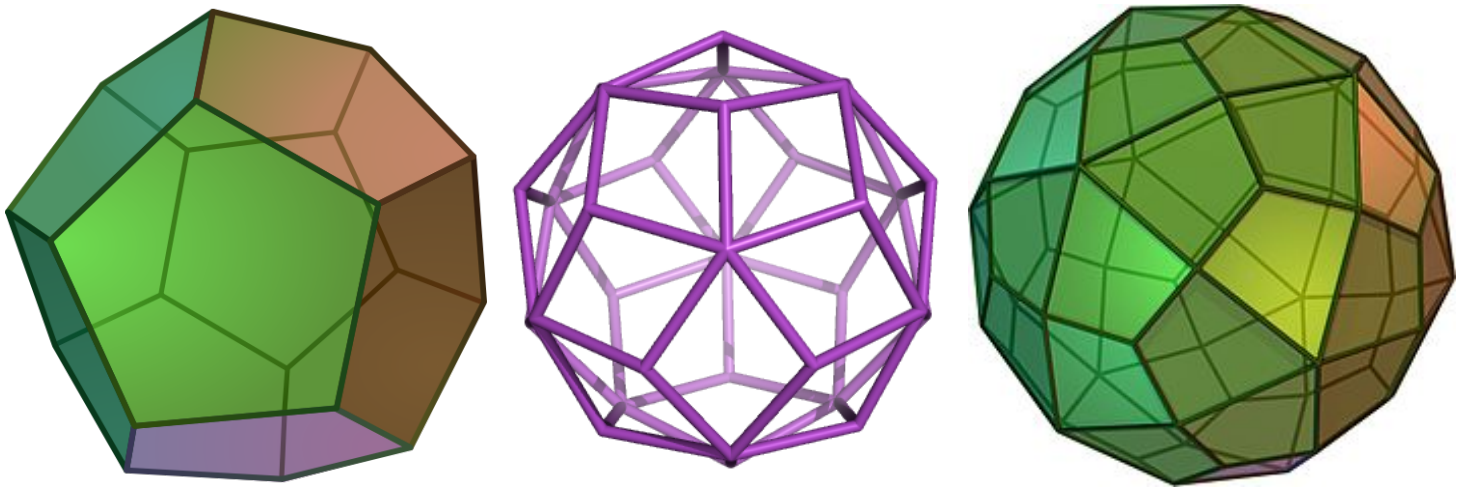


Fullerenes

- Other icosahedral objects in nature: football-shaped fullerenes
- Recover different shells with icosahedral symmetry from affine approach: carbon onions ($C_{80} - C_{180} - C_{320}$)

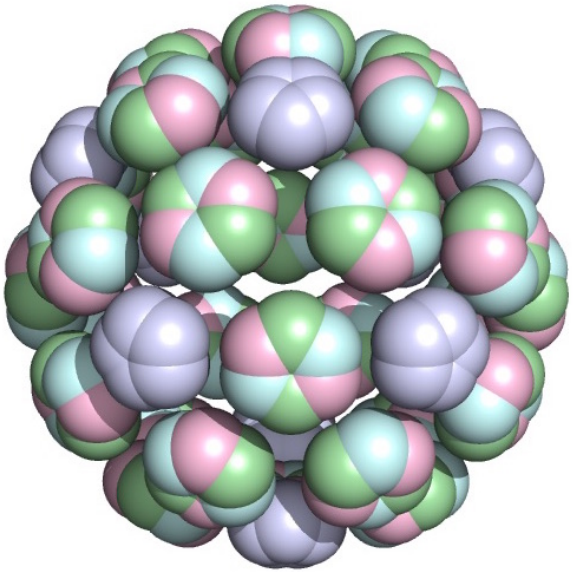
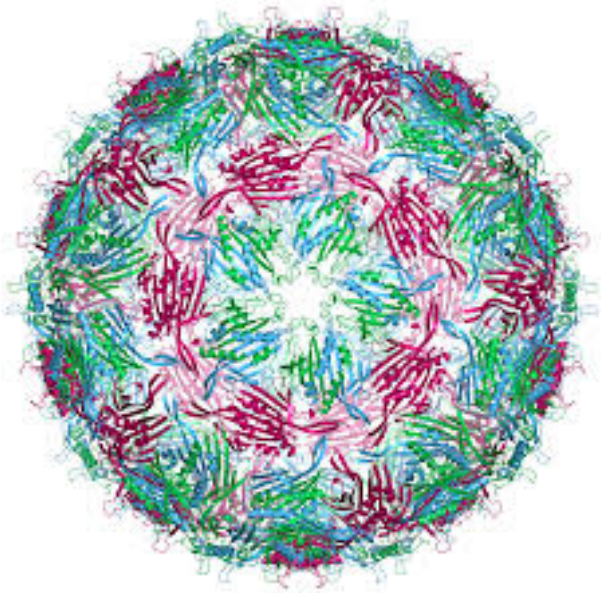
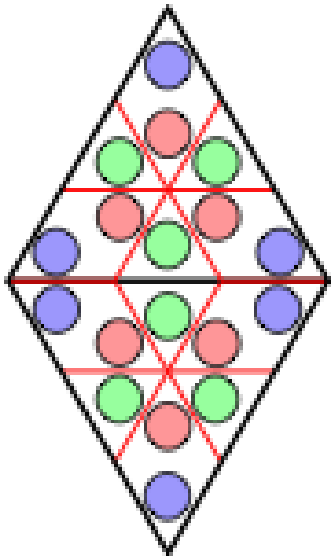


More general tile shapes from other icosahedral tilings



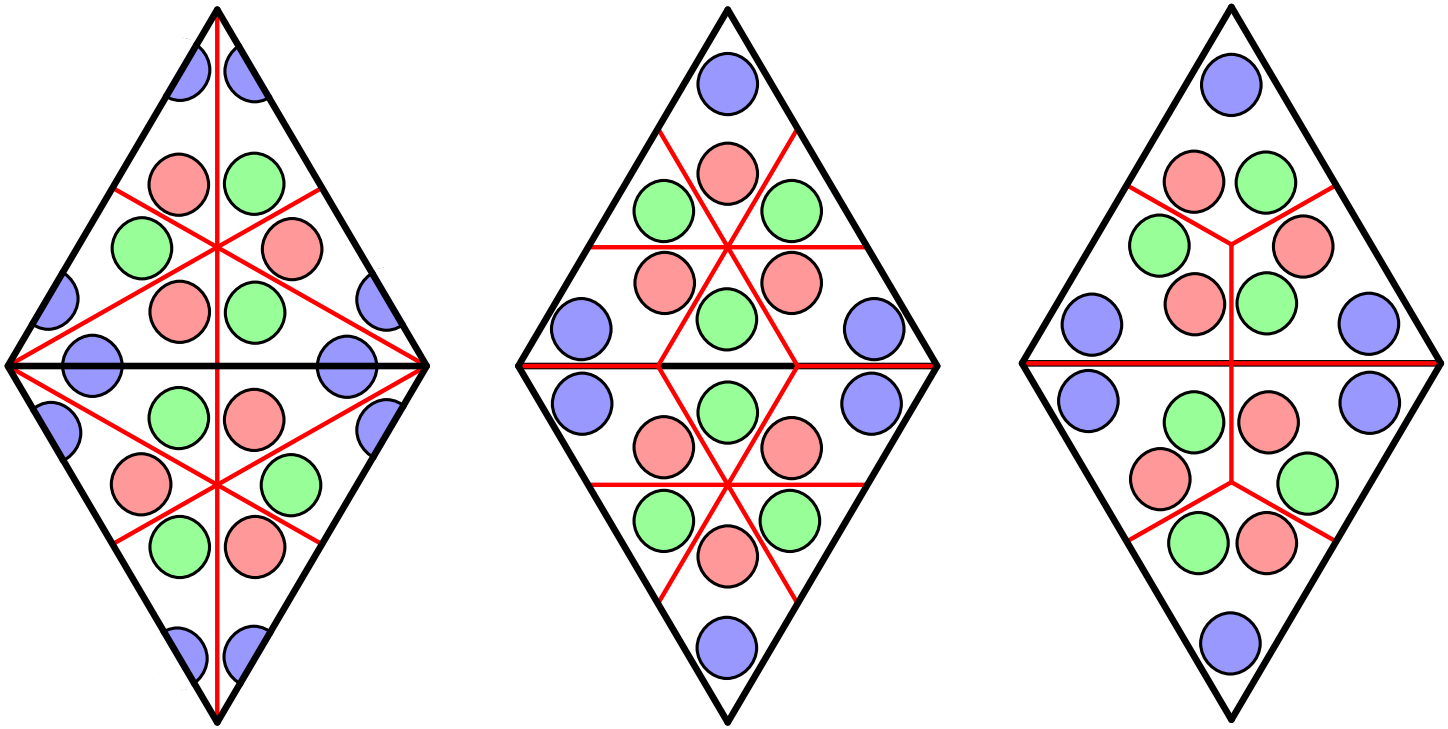
Other tile shapes can also give icosahedral tilings: pentagons (dodecahedron), rhombuses (rhombic triacontahedron), kites (deltoidal hexecontahedron)

triangulations vs other quasi-equivalent tilings

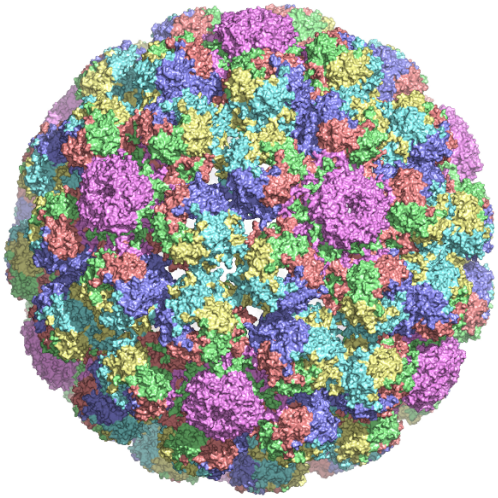
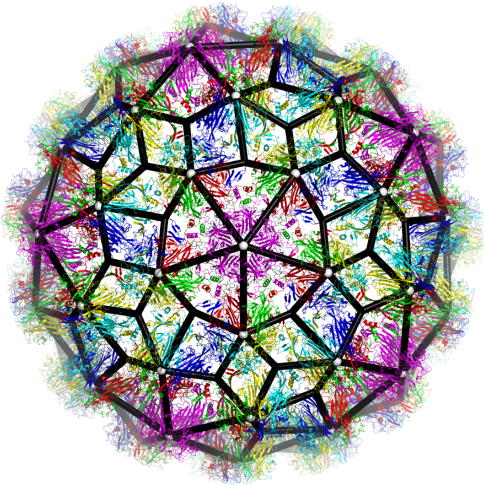
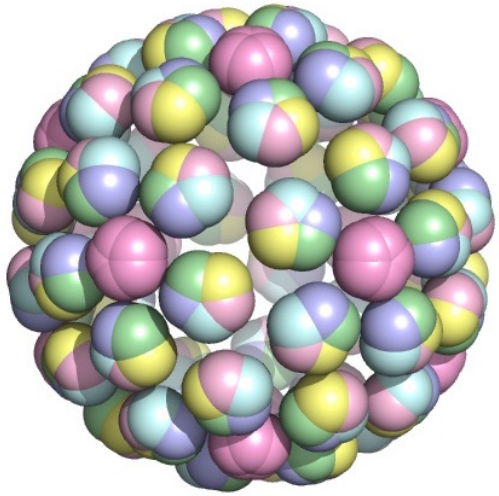
Two viral surface layouts: a $T = 4$ triangulation (e.g. HBV) and a rhombus tiling (MS2) for a pseudo $T = 3$ triangulation

Other quasi-equivalent tilings



Three (pseudo) $T = 3$ capsids: Polio, MS2 and Pariacoto.
Different building blocks depending on the underlying biology: dimer vs trimer interactions.

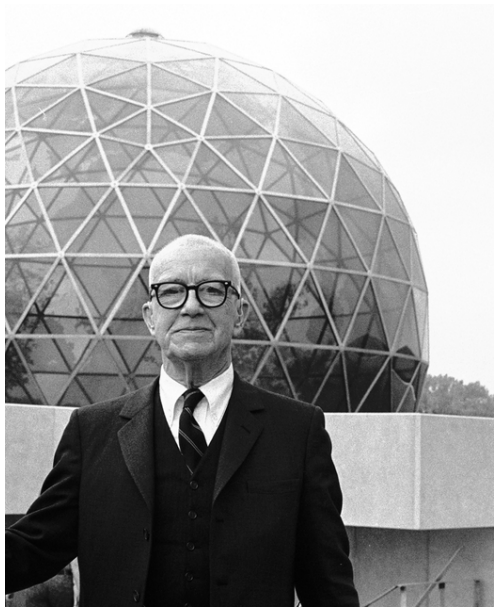
A puzzle: non-quasiequivalent tilings – Penrose

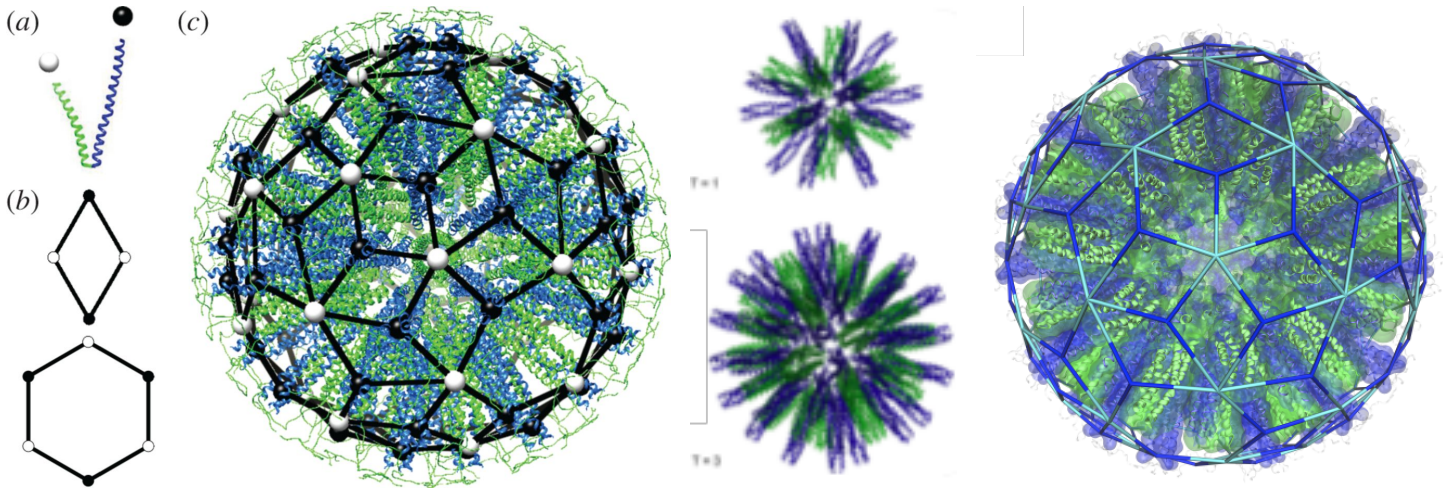
More general icosahedral tilings: Cryo-EM reconstruction of Human Papillomavirus (HPV), a kite-rhombus tiling and a pseudo $T = 7$ triangulation (but only 6 orbits).
Reidun Twarock: Viral Tiling theory

Architecture

- Triangulations: Buckminster Fuller geodesic domes
- Kite-rhombus tiling: the new Amazon HQ



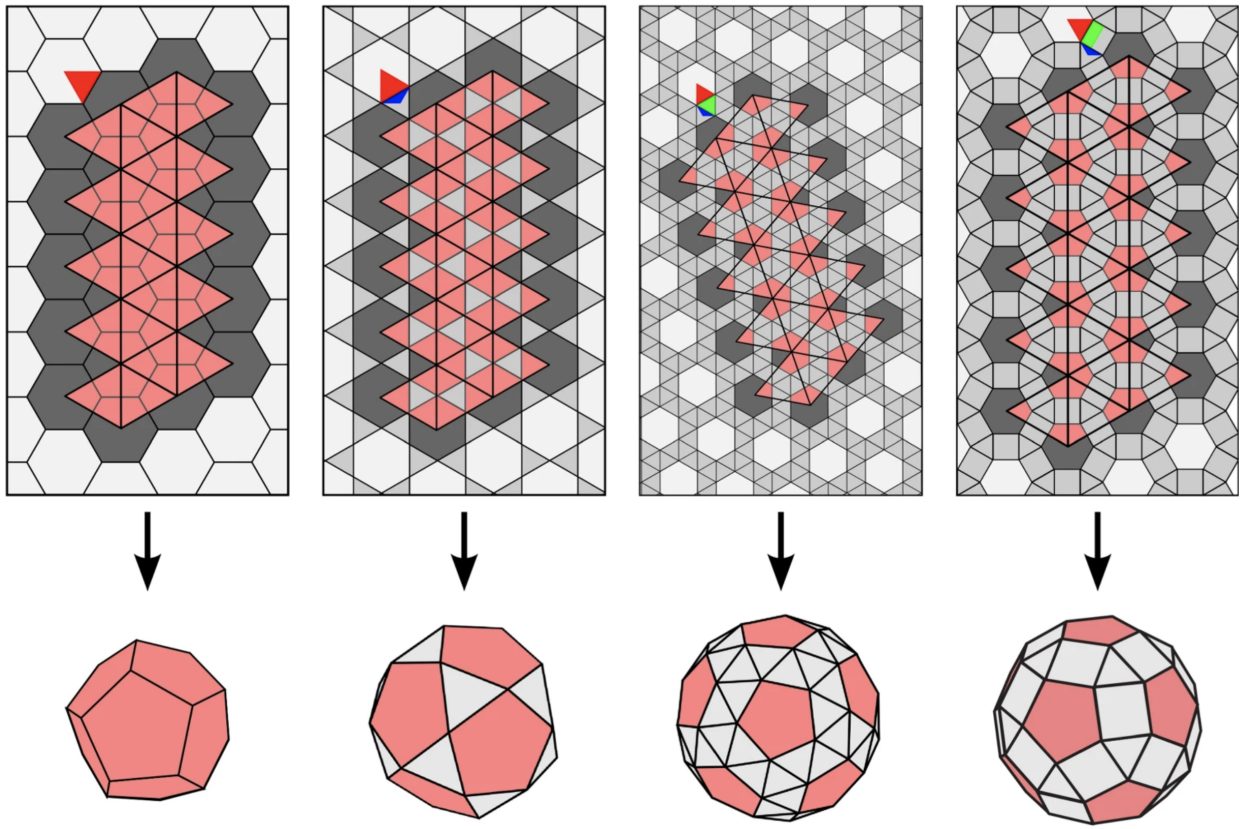
Nanotech 1: Self-assembling protein nanoparticles



De novo design of nanoparticles from identical building blocks.

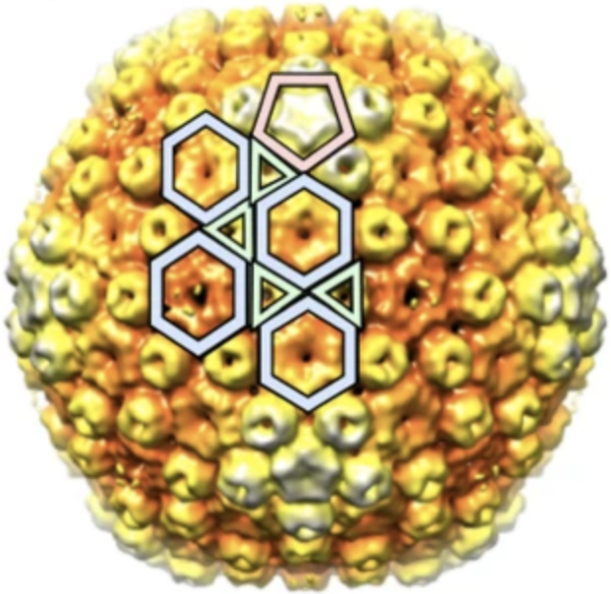
Quantised number of building blocks (e.g. in mass spec) - mathematically predict structure and properties. Particles used for vaccine design (malaria).

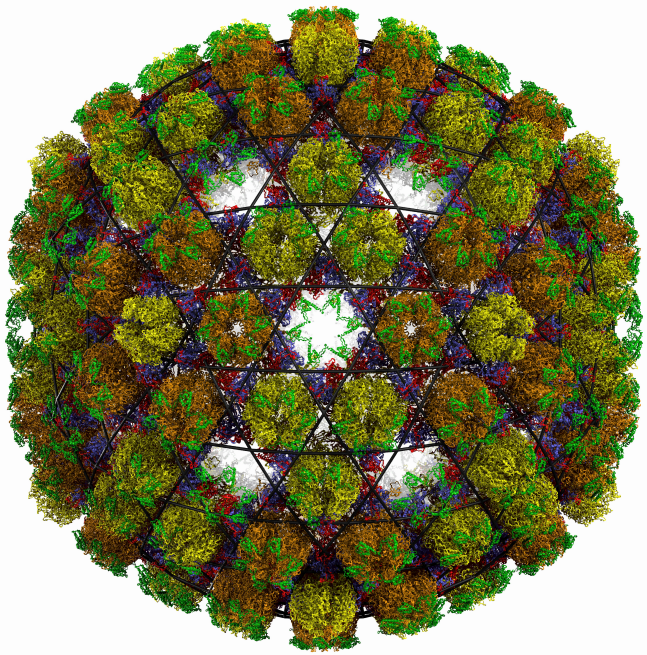
More generalised: Archimedean tilings



Reidun Twarock & Antoni Luque: Put an icosahedral net on more general hexagonal tilings.

Beyond quasi-equivalence: Archimedean tilings

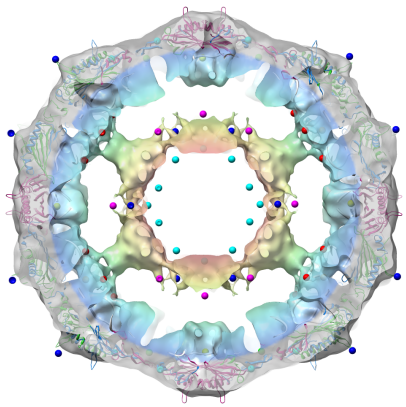


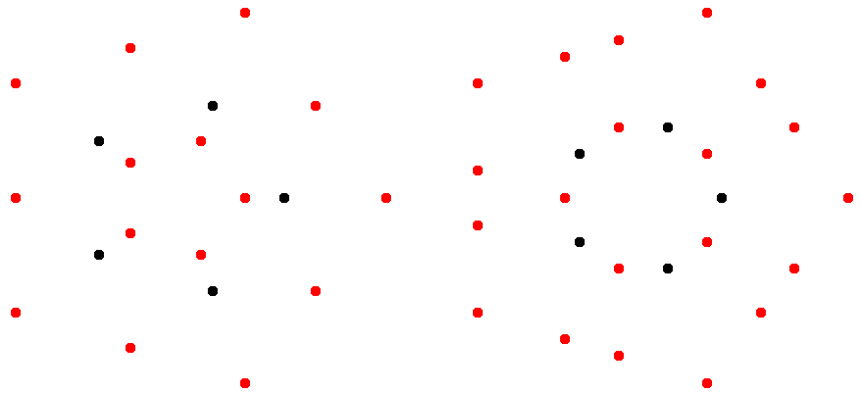


More general surface tilings e.g. for phage Basilisk and Herpes Simplex Virus.

More general symmetry still? Affine symmetry ideas.

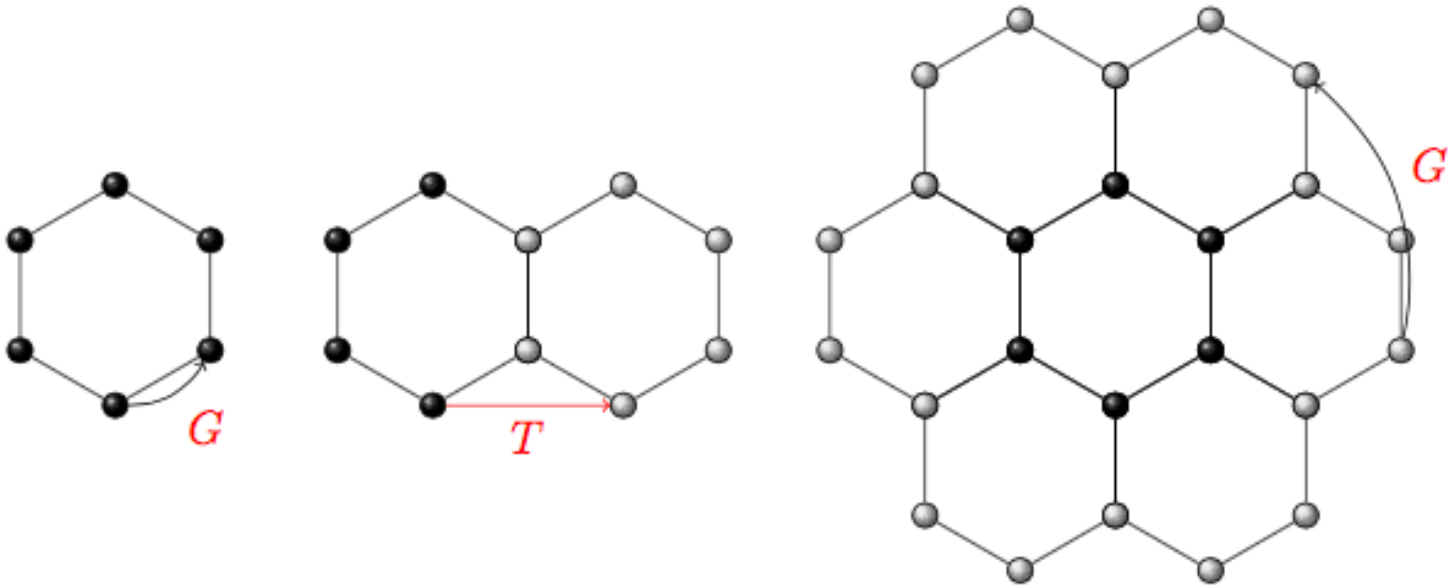
- Description only in terms of surface structures.
- Making the symmetry non-compact might allow more general symmetry, simultaneously constraining different 'radial levels'
- Non-compact generator is a translation – motivates looking into affine extensions of icosahedral symmetry
- There is an inherent length scale in the problem – given by size of nucleic acid/protein molecules





Affine extensions - A_2

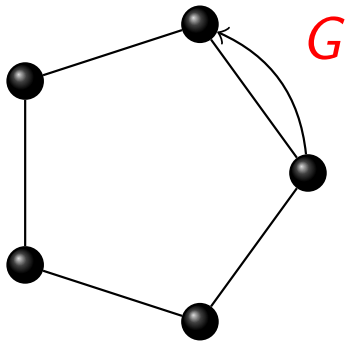
Unit translation of a unit hexagon



A random translation would give 6 secondary hexagons, i.e. 36 points. Here we have degeneracies due to 'coinciding points', and building up the hexagonal lattice.

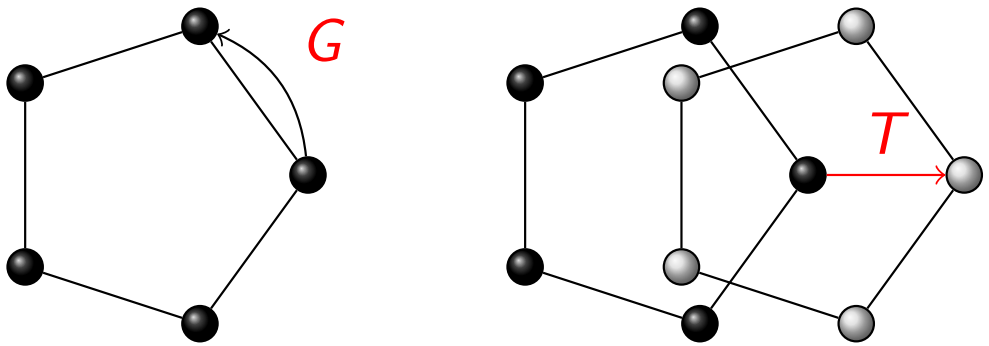
Affine extensions of non-crystallographic groups?

Unit translation along a vertex of a unit pentagon



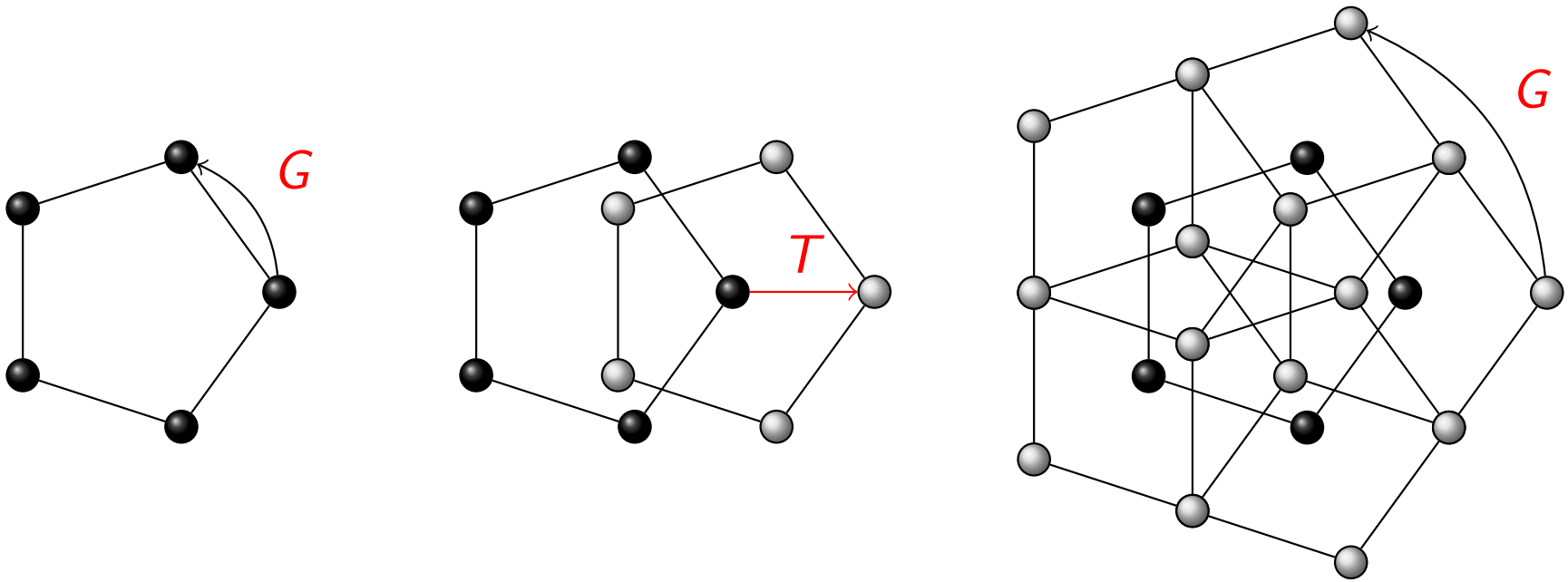
Affine extensions of non-crystallographic groups?

Unit translation along a vertex of a unit pentagon



Affine extensions of non-crystallographic groups?

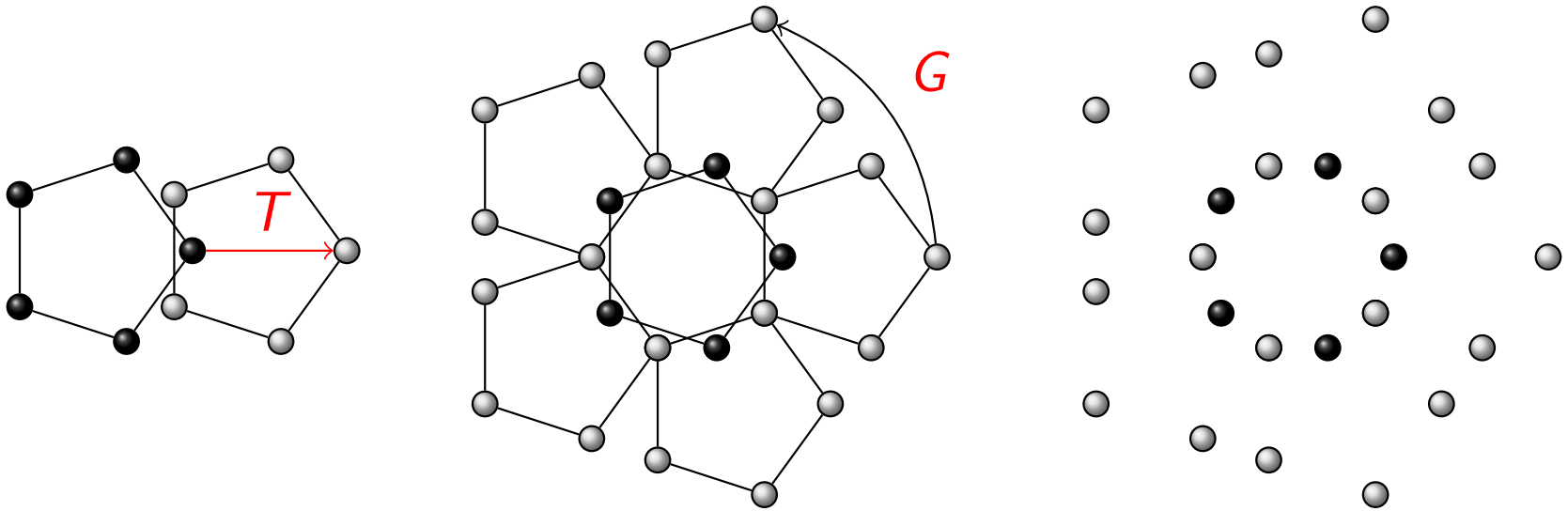
Unit translation along a vertex of a unit pentagon



A random translation would give 5 secondary pentagons, i.e. 25 points. Here we have degeneracies due to 'coinciding points'.

Affine extensions of non-crystallographic root systems?

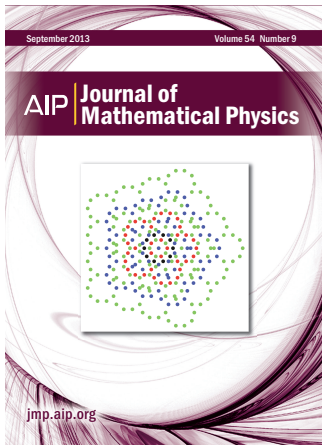
Translation of length $\tau = \frac{1}{2}(1+\sqrt{5}) \approx 1.618$ (golden ratio)

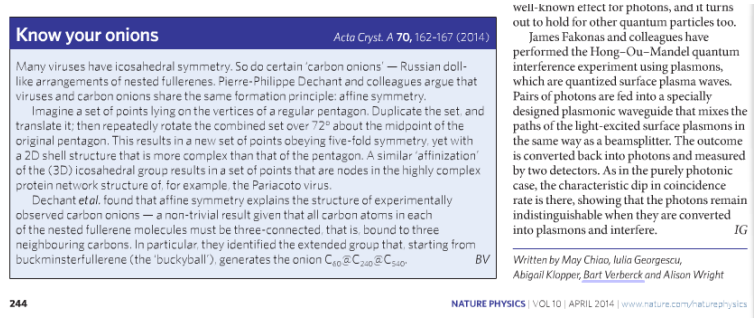


Cartoon version of a virus or carbon onion. Would there be a biological benefit to have more than just compact symmetry? The problem has an intrinsic length scale.

Affine extensions of non-crystallographic Coxeter groups

- 2D and 3D point arrays for applications to viruses, fullerenes, quasicrystals, proteins etc
- Two complementary ways to construct these





Know your onions *Acta Cryst. A* **70,** 162–167 (2014)

Many viruses have icosahedral symmetry. So do certain 'carbon onions' — Russian doll-like arrangements of nested fullerenes. Pierre-Philippe Dechant and colleagues argue that viruses and carbon onions share the same formation principle: affine symmetry.

Imagine a set of points lying on the vertices of a regular pentagon. Duplicate the set, and translate it; then repeatedly rotate the combined set over 72° about the midpoint of the original pentagon. This results in a new set of points obeying five-fold symmetry, yet with a 2D shell structure that is more complex than that of the pentagon. A similar 'affinization' of the (3D) icosahedral group results in a set of points that are nodes in the highly complex protein network structure of, for example, the Pariacoto virus.

Dechant *et al.* found that affine symmetry explains the structure of experimentally observed carbon onions — a non-trivial result given that all carbon atoms in each of the nested fullerene molecules must be three-connected, that is, bound to three neighbouring carbons. In particular, they identified the extended group that, starting from buckminsterfullerene (the 'buckyball'), generates the onion $C_{60} \subset C_{240} \subset C_{540}$. *BV*

well-known effect for photons, and it turns out to hold for other quantum particles too.

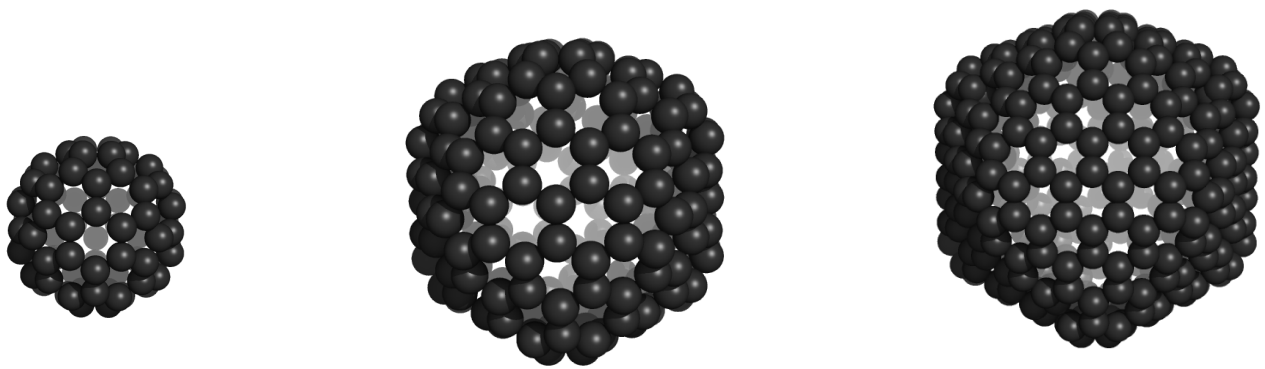
James Fakonas and colleagues have performed the Hong–Ou–Mandel quantum interference experiment using plasmons, which are quantized surface plasma waves. Pairs of photons are fed into a specially designed plasmonic waveguide that mixes the paths of the light-excited surface plasmons in the same way as a beamsplitter. The outcome is converted back into photons and measured by two detectors. As in the purely photonic case, the characteristic dip in coincidence rate is there, showing that the photons remain indistinguishable when they are converted into plasmons and interfere. *IG*

Written by May Chiao, Iulia Georgescu, Abigail Klopper, Bart Verberck and Alison Wright

244 NATURE PHYSICS | VOL 10 | APRIL 2014 | www.nature.com/naturephysics

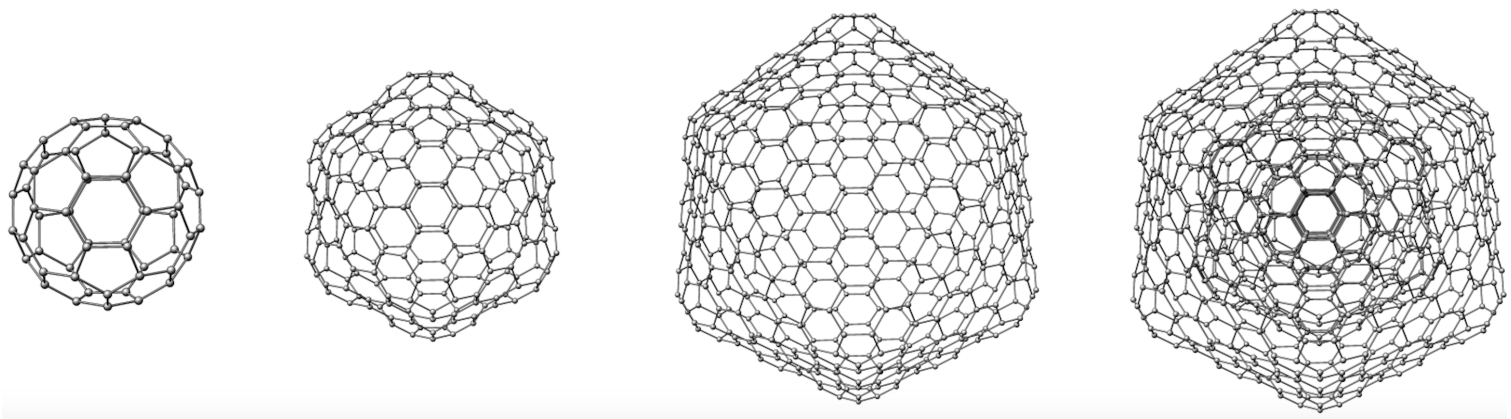
Extension to fullerenes: carbon onions

- Extend idea of affine symmetry to other icosahedral objects in nature: football-shaped fullerenes
- Recover different shells with icosahedral symmetry from affine approach: carbon onions ($C_{80} - C_{180} - C_{320}$)

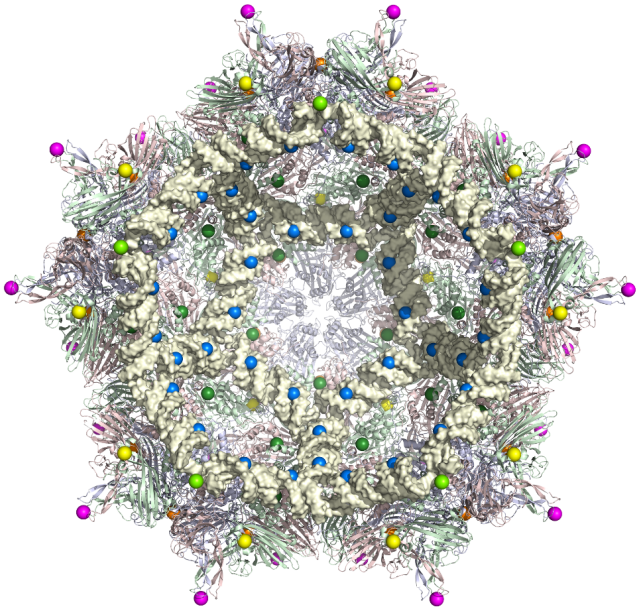


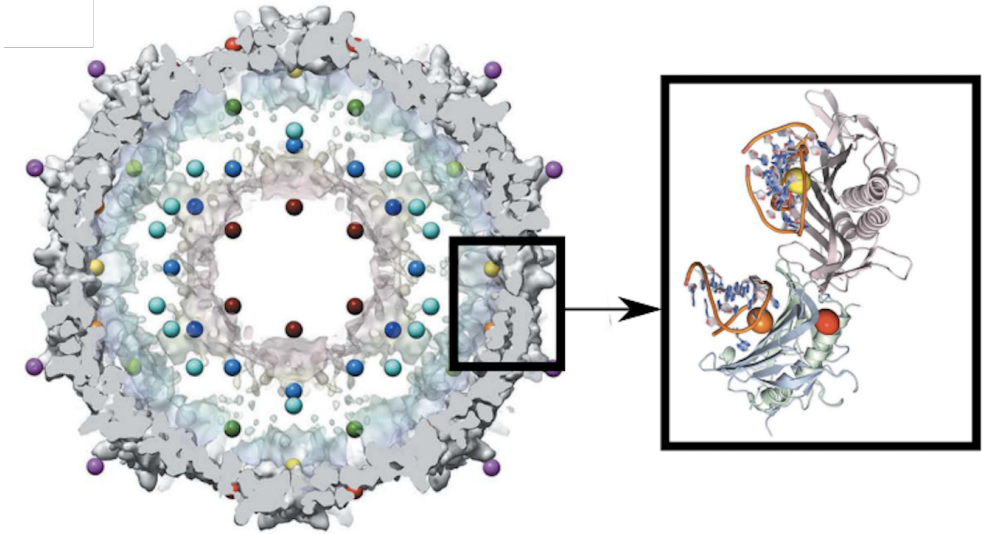
Extension to fullerenes: carbon onions

- Extend idea of affine symmetry to other icosahedral objects in nature: football-shaped fullerenes
- Recover different shells with icosahedral symmetry from affine approach: carbon onions ($C_{60} - C_{240} - C_{540}$)

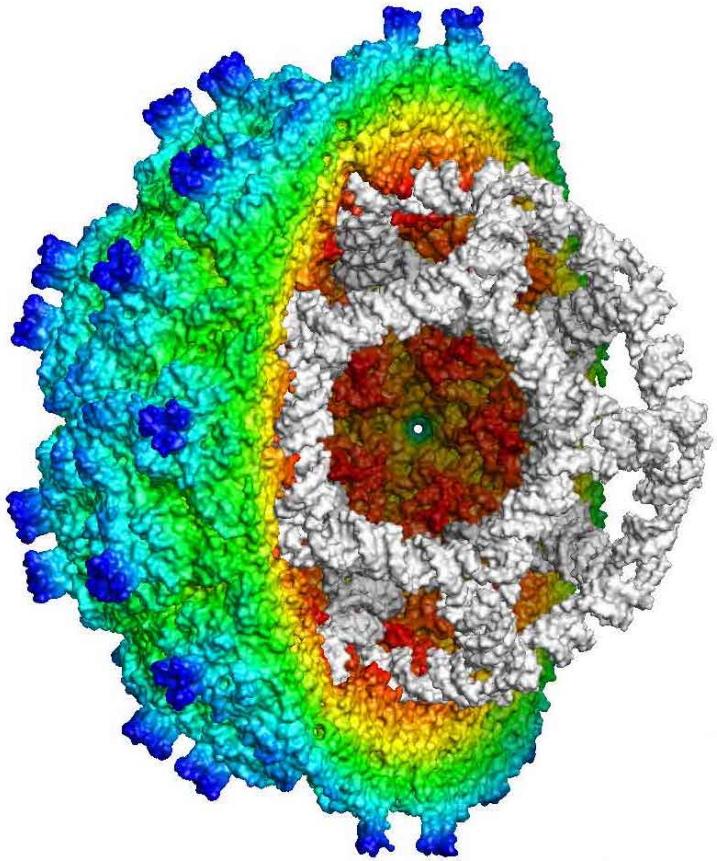


Use in Mathematical Virology



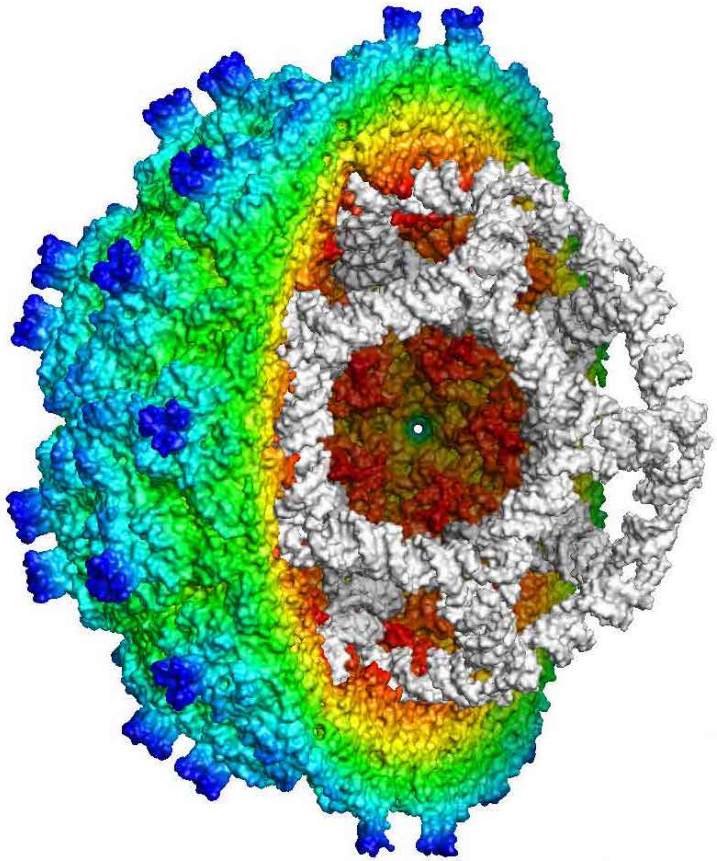


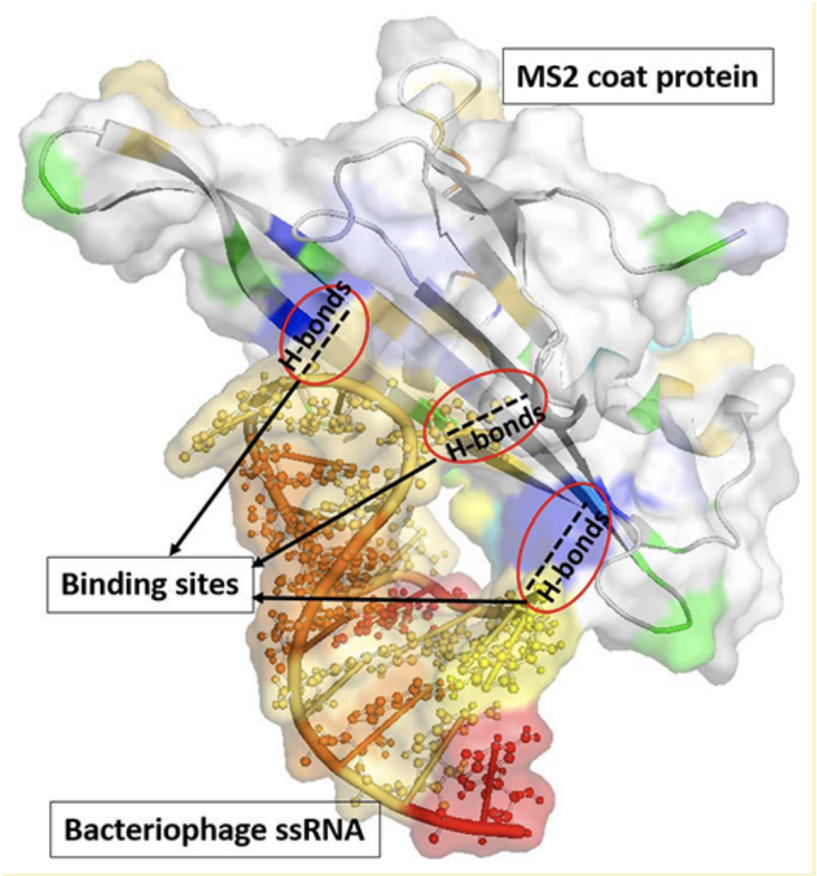
3D distribution: RNA-CP contacts



There are specific interactions between RNA and coat protein (CP) given by icosahedral symmetry axes

3D distribution: RNA-CP contacts

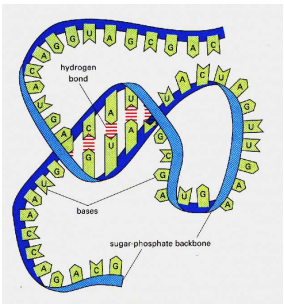


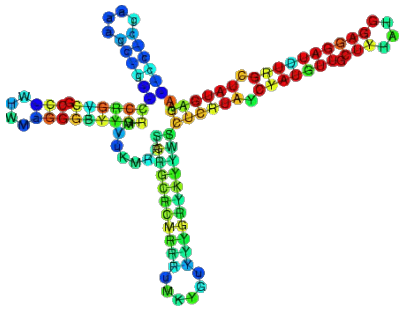


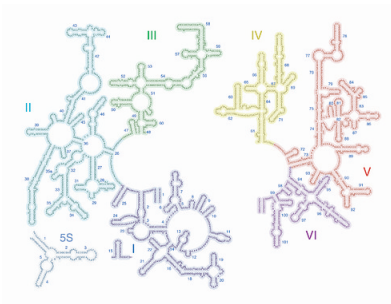
There are specific interactions between RNA and coat protein (CP) given by icosahedral symmetry axes

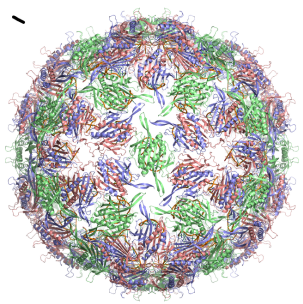
New insight into RNA virus assembly

- There are specific interactions between RNA and and inner (capsid) surface
- Essential for (co-)assembly, as only this RNA-CP interaction turns CP into right geometric shape for capsid formation for MS2
- Hamiltonian cycle visiting each RNA-CP contact once – dictated by symmetry
- Even the RNA has an icosahedrally ordered component









RNA is involved in a co-assembly process through packaging signals

These packaging signals help recruit coat protein (CP) in a co-assembly process.

Overview

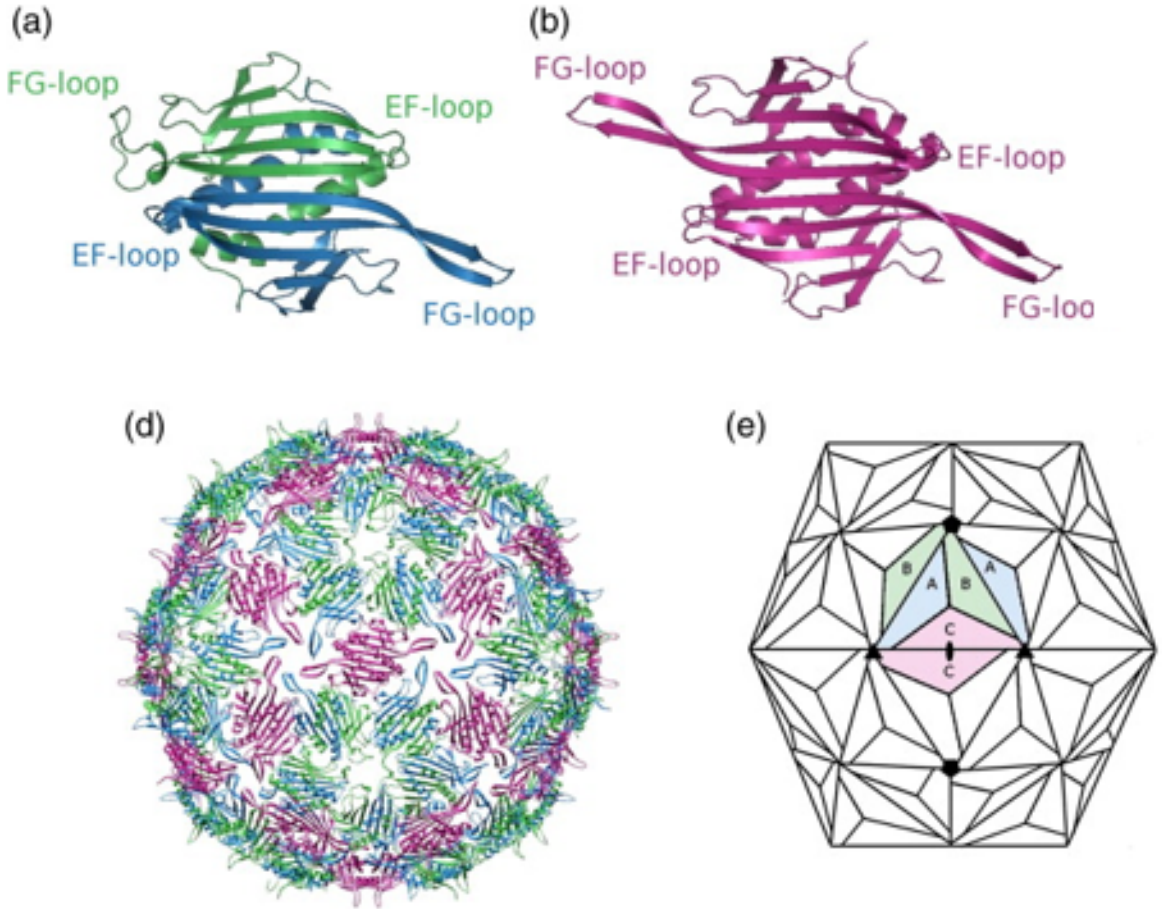
Mathematical Modelling

- Watson & Crick: Icosahedral and helical symmetry
- Caspar & Klug: Triangulations
- Twarock: More general surface tilings
- Affine symmetry: genome and capsid

Computational Modelling

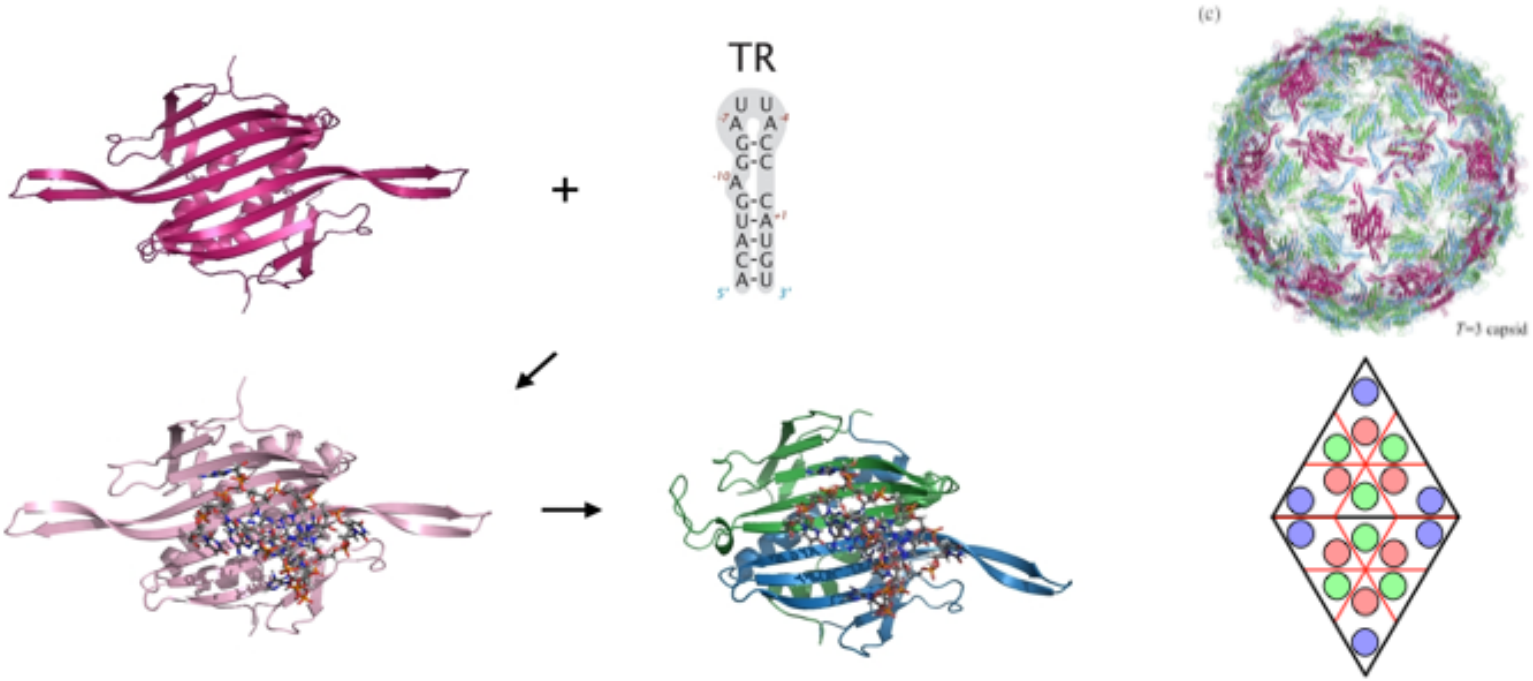
- Bioinformatics: packaging signals
- Gillespie stochastic simulations: epidemiological, infection, assembly
- Machine learning: fitness landscape

MS2 tiling and dimeric building blocks: A/B and C/C



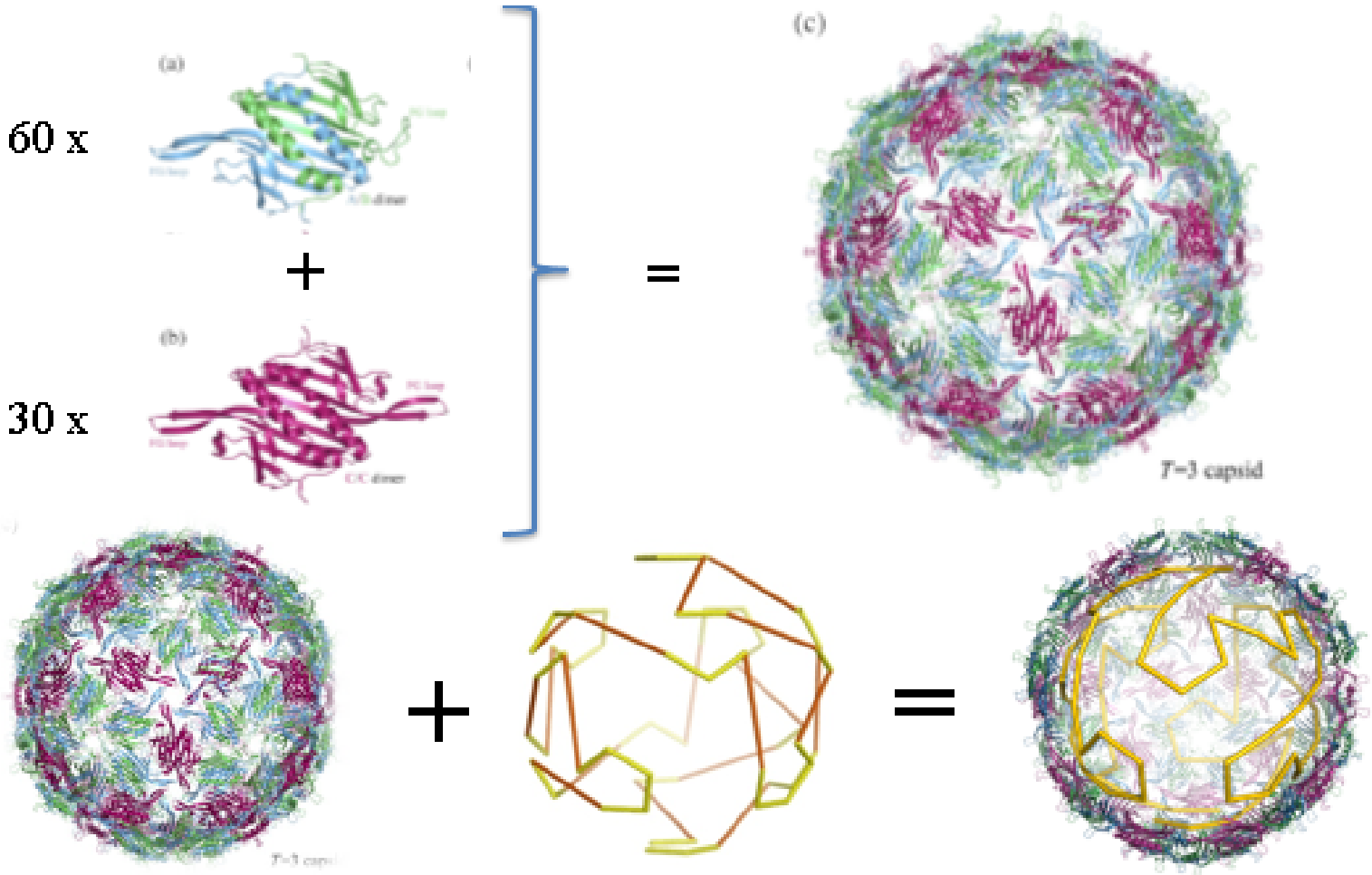
Need to bind RNA in 60 places

The TR sequence is known to initiate assembly by associating with the maturation protein. It forms a stemloop and it has been shown that the stemloop changes the conformation of the symmetric C/C dimer to the asymmetric A/B dimer (allosteric switch).



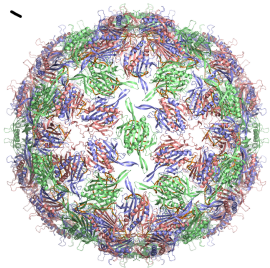
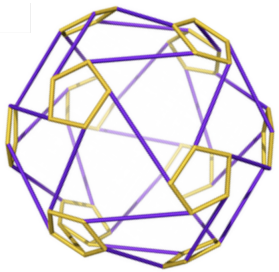
Peter Stockley (Leeds), Neil Ranson (Leeds), Eric Dykeman (York) et al

MS2 Hamiltonian path

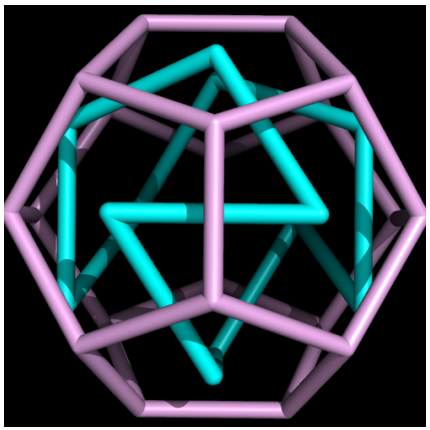
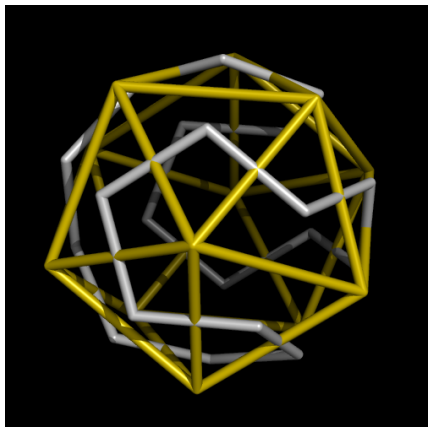
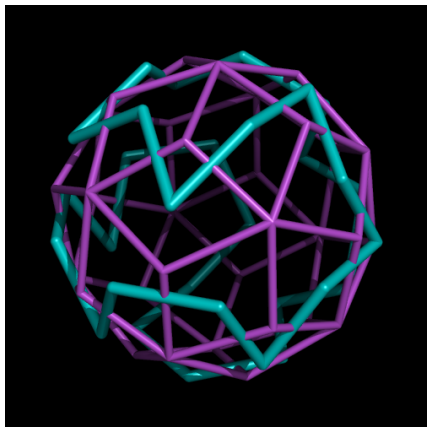


New insight into RNA virus assembly

- Example of MS has 60 vertices with 41,000 paths
- The RNA is actually circularised by Maturation Protein: only 66 cycles
- With thermodynamical assembly kinetics and 5-fold averaging experiments uniquely idenfied an evolutionarily conserved cycle
- Patents for new antiviral strategies and virus-like nanoparticles e.g. for drug delivery (Twarock group)

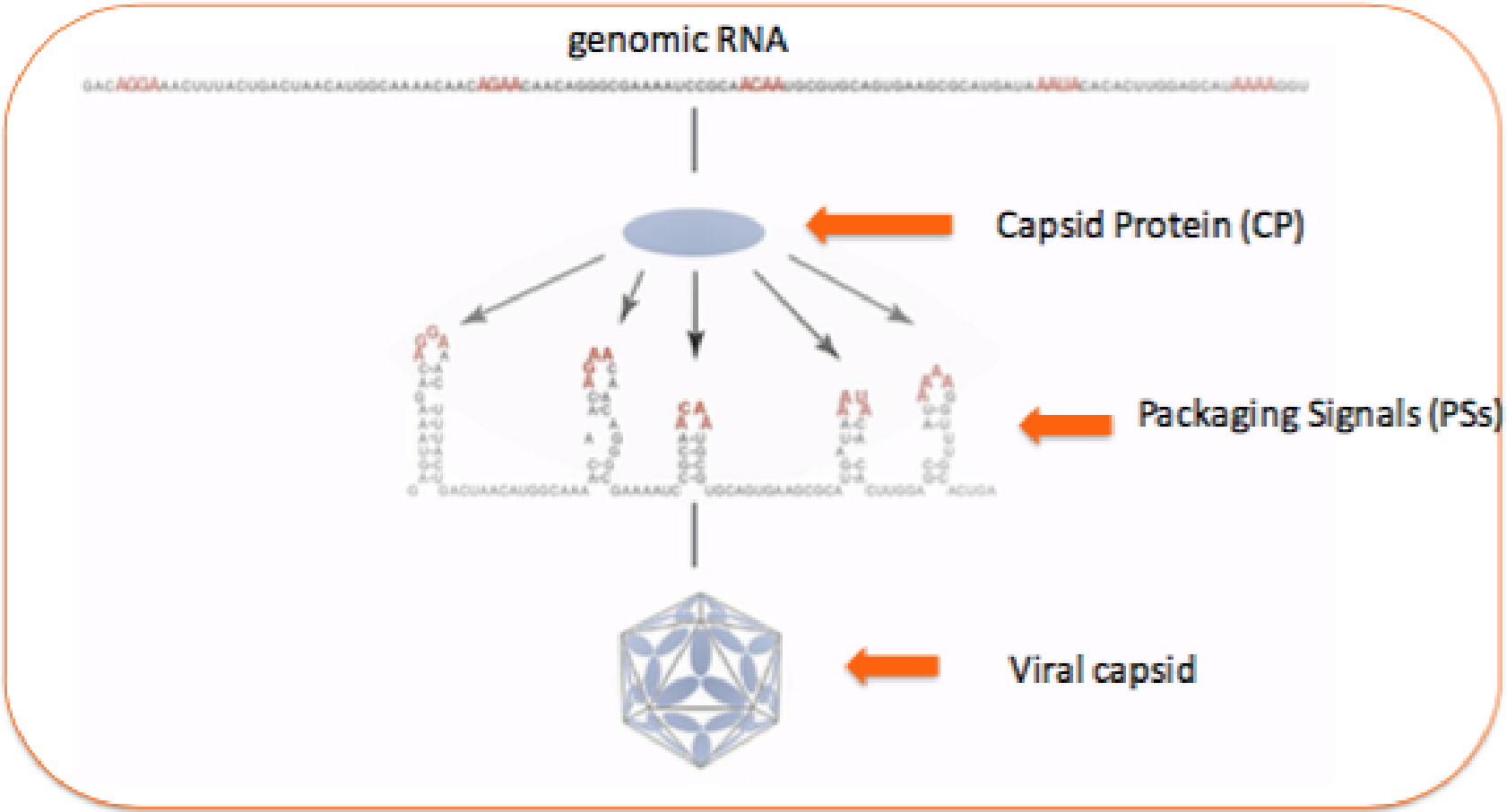





Hamiltonian cycles on icosahedral solids

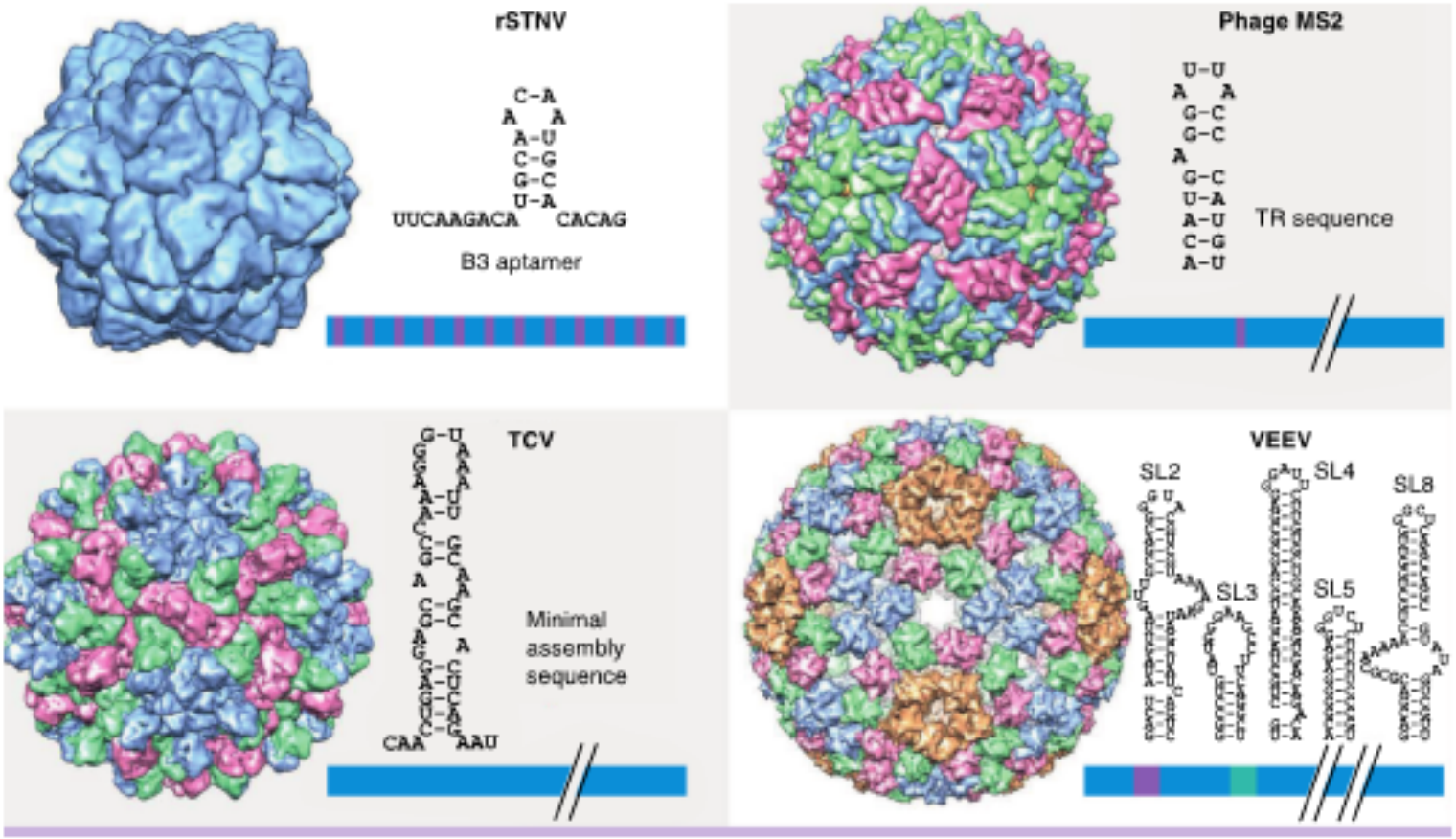




- So interaction contacts are given by the symmetry
- Orbits of the interaction points have to be visited by the RNA exactly once
- Even the RNA has an icosahedrally ordered component
- Hamiltonian cycles for dodecahedron, icosahedron and rhombic triacontahedron

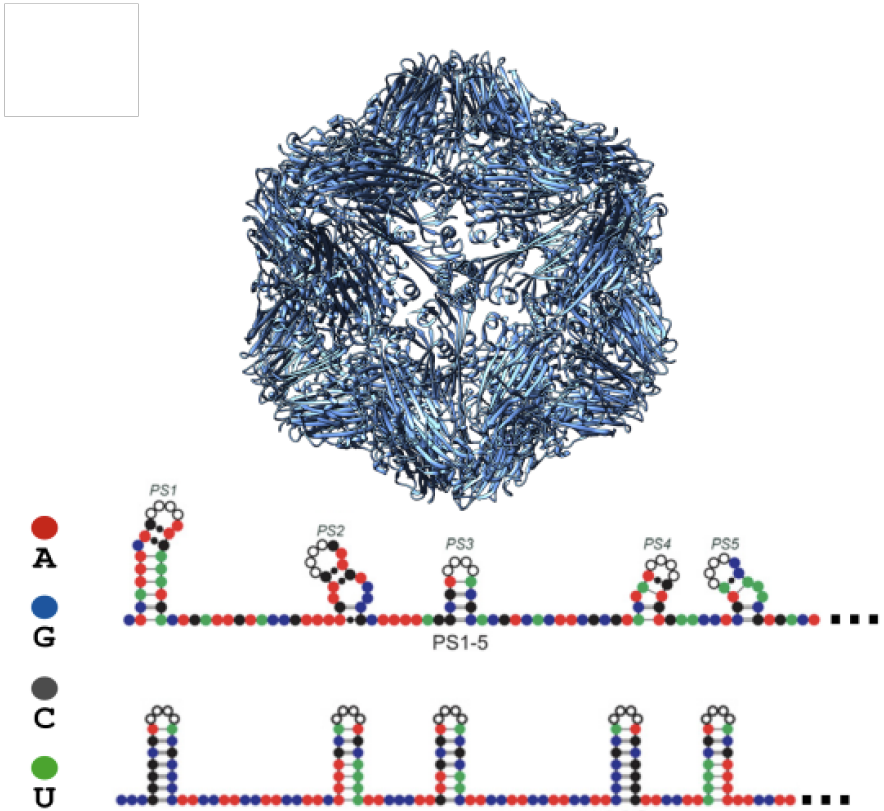
Assembly via Multiple dispersed Packaging Signals



Common Mechanism across groups of viruses

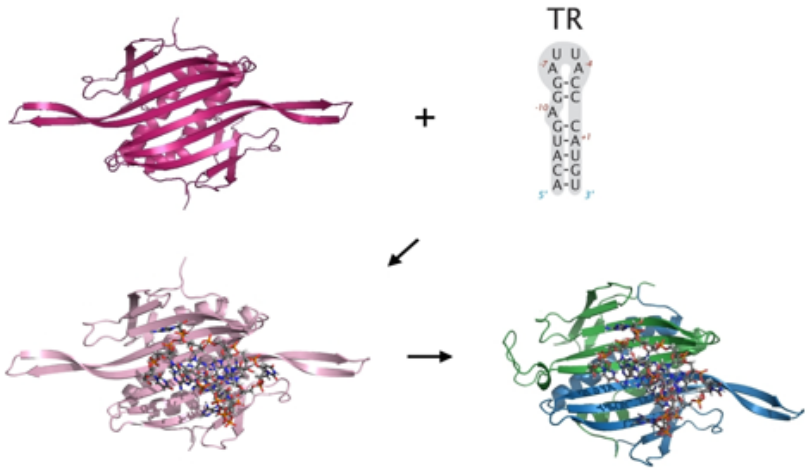


Nanotech 2: Engineering Packaging Signals to make VLPs



Repurpose: Improved sequences optimised for assembly efficiency (e.g. STNV). Potential applications to vaccines or drug delivery.

Antivirals: Understanding assembly allows one to interfere



- target RNA
- target CP
- introduces competitors
- this might still drive evolution due to exerting selection pressures but less so than biochemical antivirals

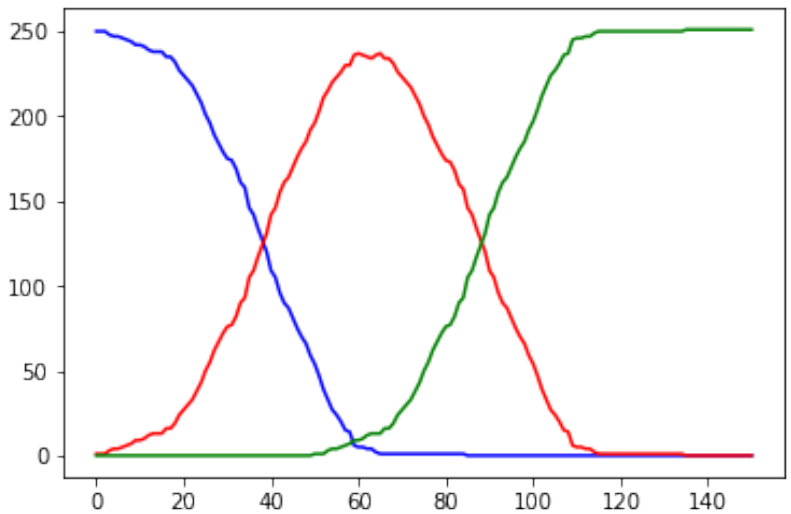
Simulations

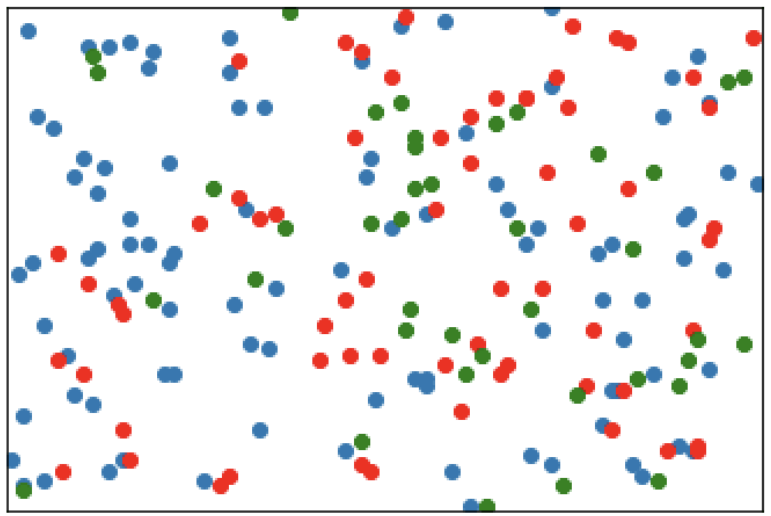
Stochastic simulations

rather than ODE models because of discrete nature and low numbers: Gillespie-type algorithms that select a random reaction to occur

- SIR and spatial modelling (epidemiology)
- Multi-scale model coupling of an intracellular model with an immune system (infection model)
- Assembly toy model simulations

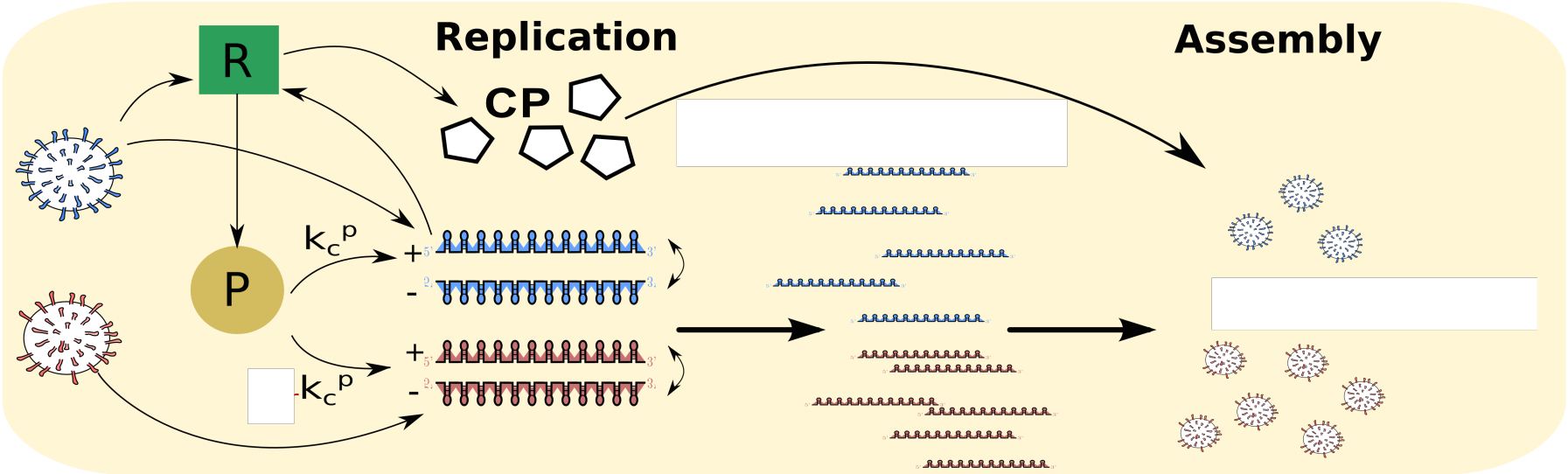
Basic epidemiological simulations





Gillespie SIR model and spatial/movement model.

Infection model: intracellular model



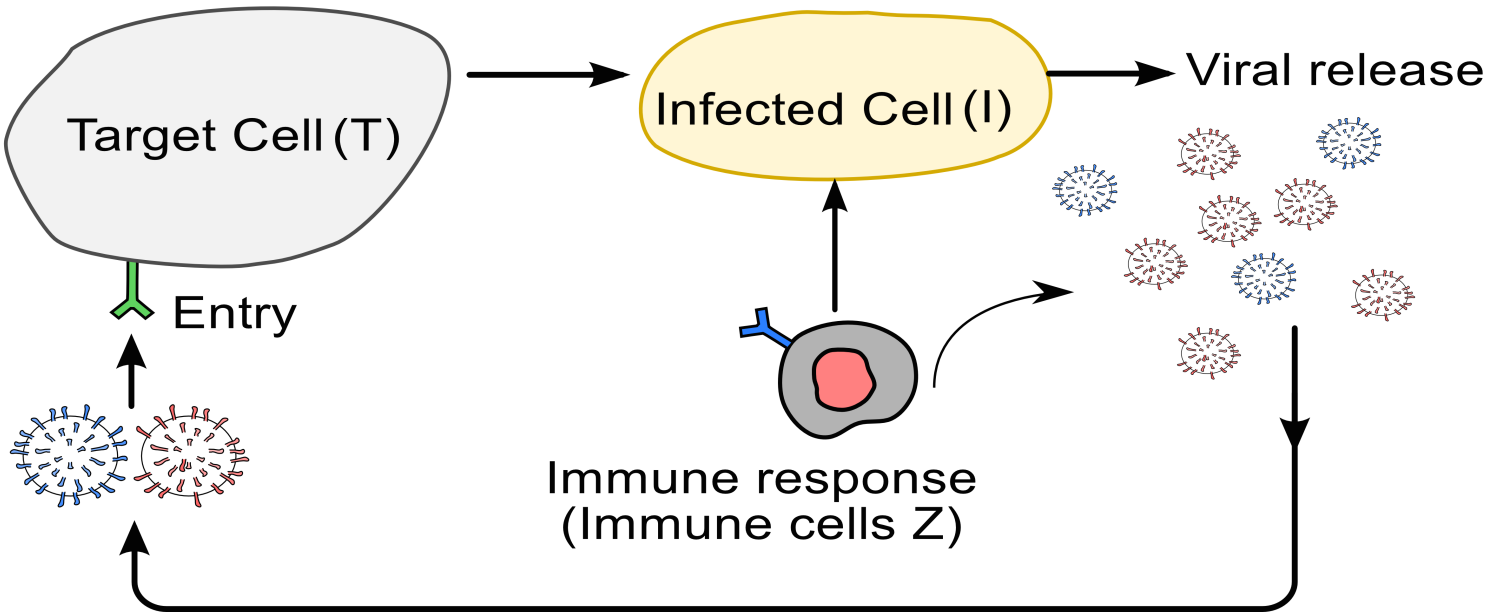
$$p_V^+ + R \underset{k_r^{off}}{\overset{k_r^{on}}{\rightleftarrows}} (p_V^+, R) \text{(Ribosome positive strand virus binding/unbinding)}$$

$$(p_V^+, R) \xrightarrow{k_r^c} p_V^+ + R + P \text{ (Genome translation - makes P and abundant CP)}$$

$$p_{V/S}^{\pm} + P \underset{k_p^{\pm off}}{\overset{k_p^{\pm on}}{\rightleftarrows}} (p_{V/S}^{\pm}, P) \text{(Polymerase positive/negative strand virus binding/unbinding)}$$

$$(p_{V/S}^{\pm}, P) \xrightarrow{k_p^c} p_{V/S}^{\pm} + p_{V/S}^{\mp} + P \text{ (complementary strand production)}$$

Infection model: immune system



$T \xrightarrow{\lambda} 2T$ (Target cell birth)

$T \xrightarrow{d_T} 0$ (Target cell death)

$T + pV \xrightarrow{\beta} I$ (Infection of target cell)

$I \xrightarrow{a} rV$ (Infected cell death/lysis)

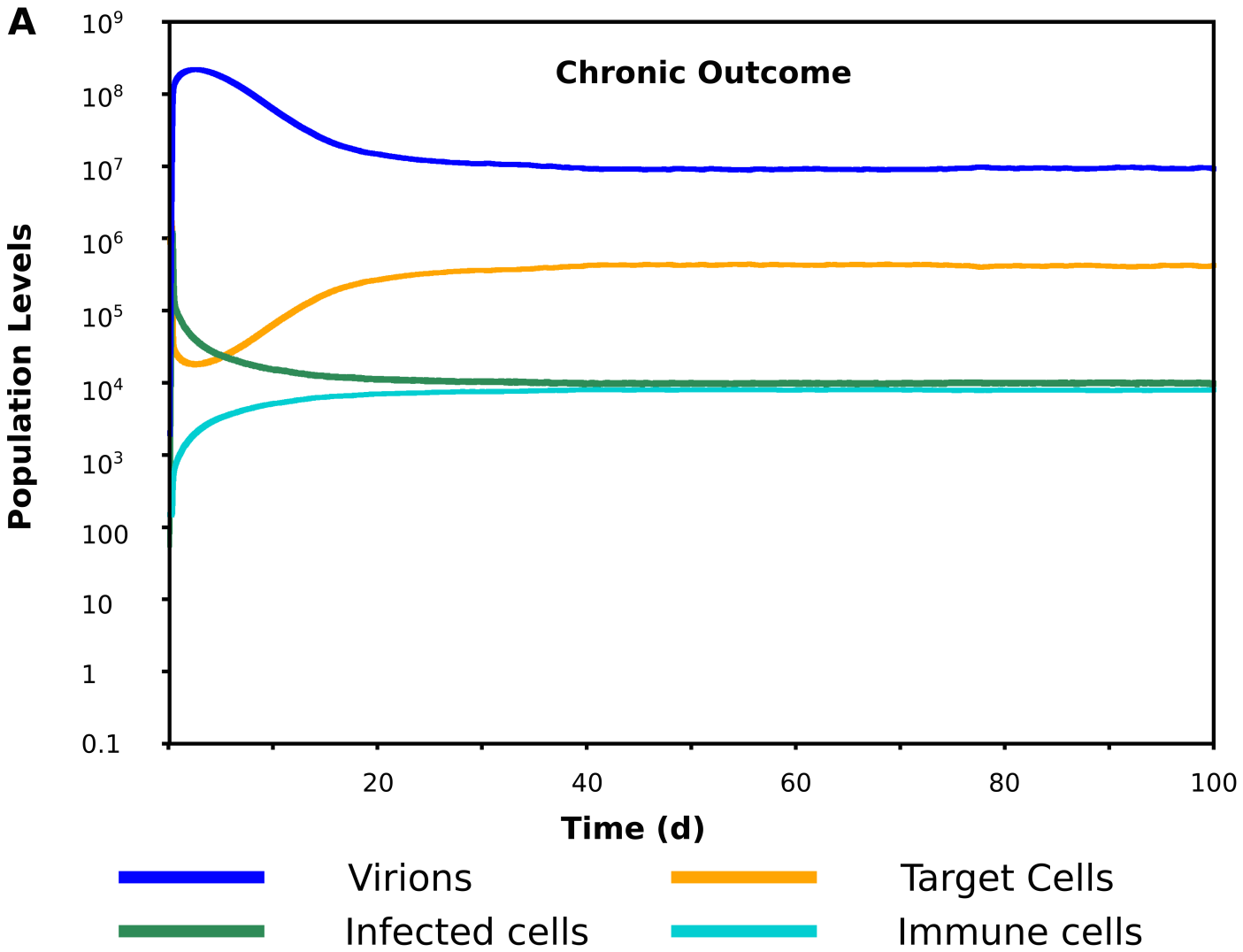
$I + Z \xrightarrow{\pi} Z$ (Infected cell removal by immune system)

$V + Z \xrightarrow{u} Z$ (Virion removal by immune system)

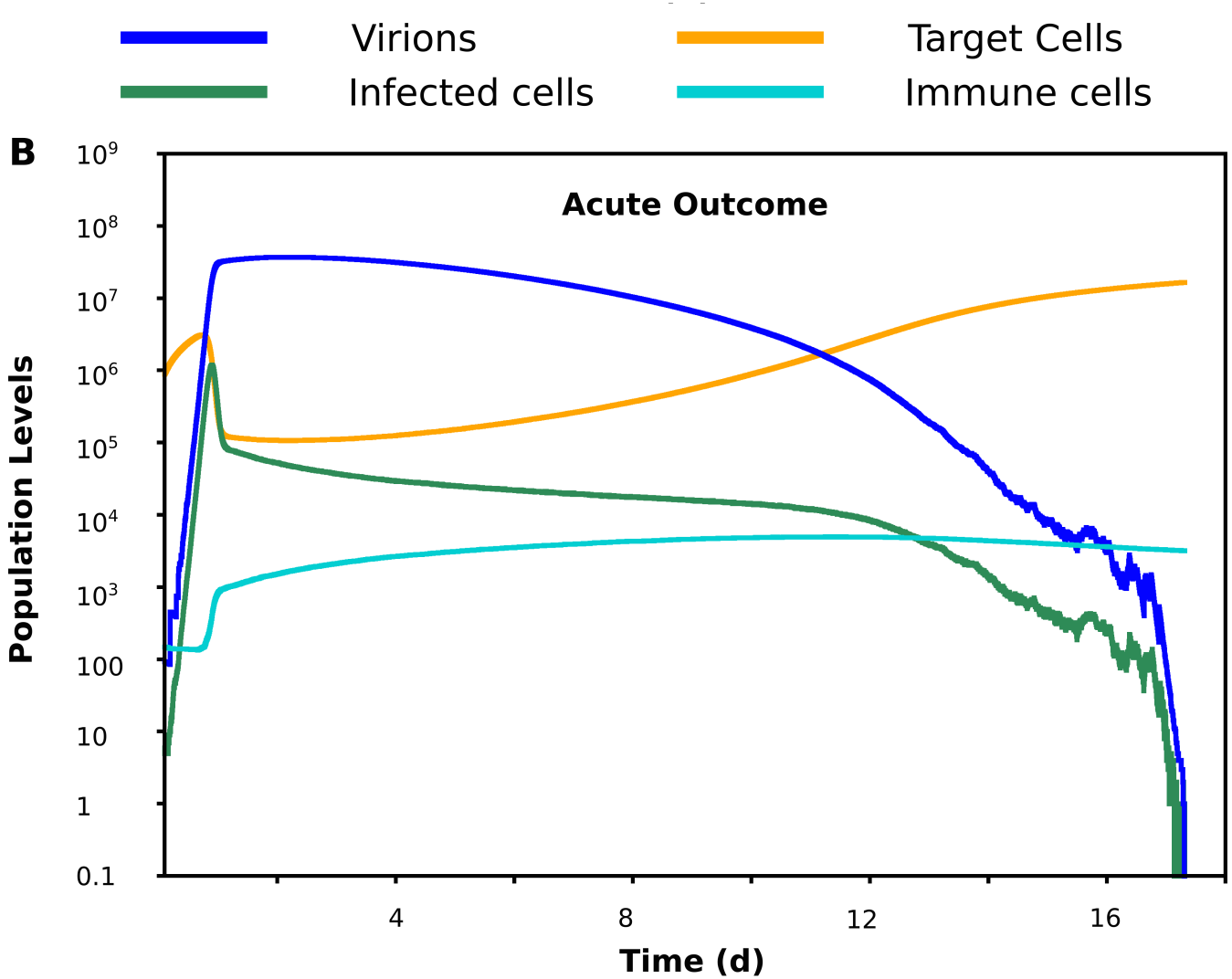
$I + Z \xrightarrow{c} I + 2Z$ (Immune cell birth)

$Z \xrightarrow{b} 0$ (Immune cell death)

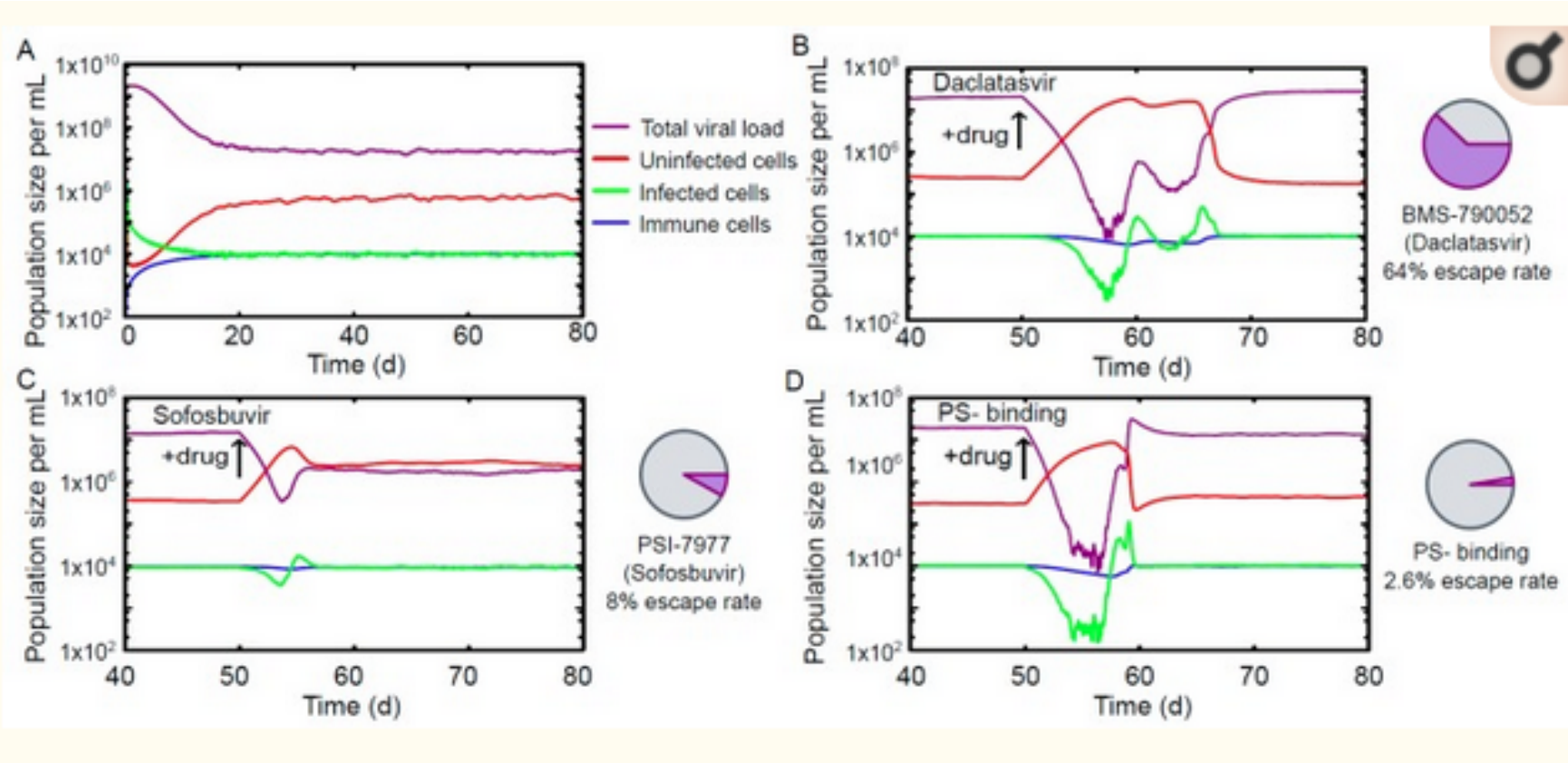
Chronic infections



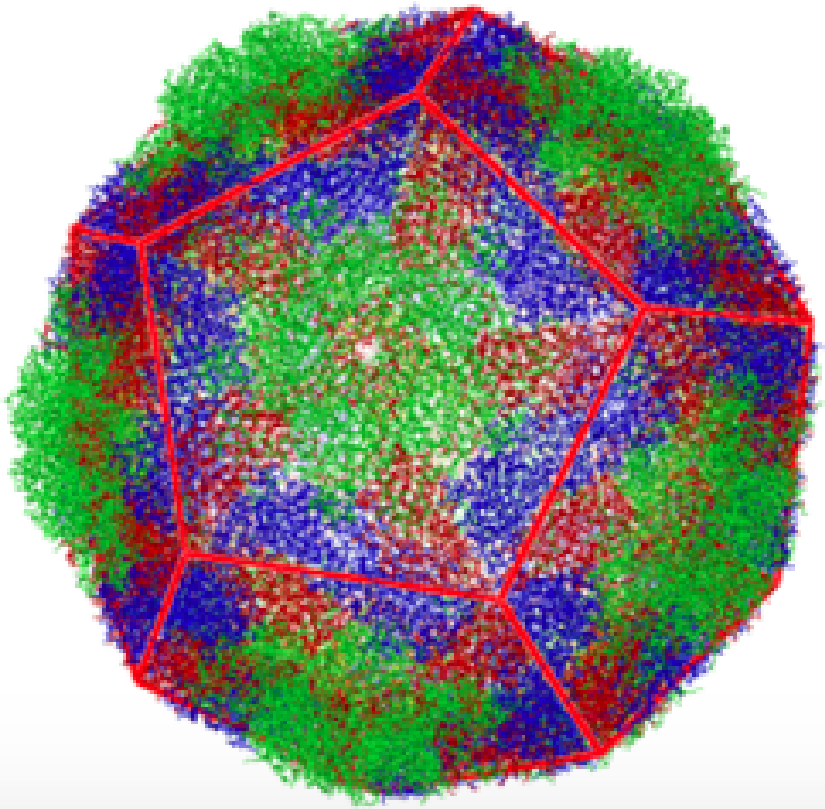
Acute infections

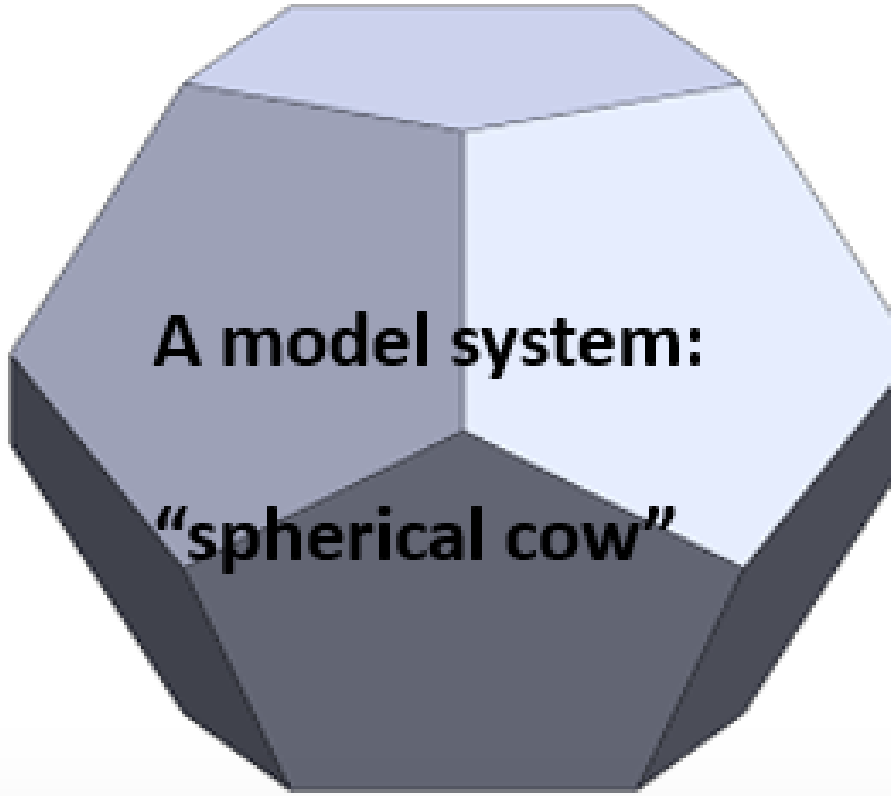


Antivirals: Evolutionarily stable drugs

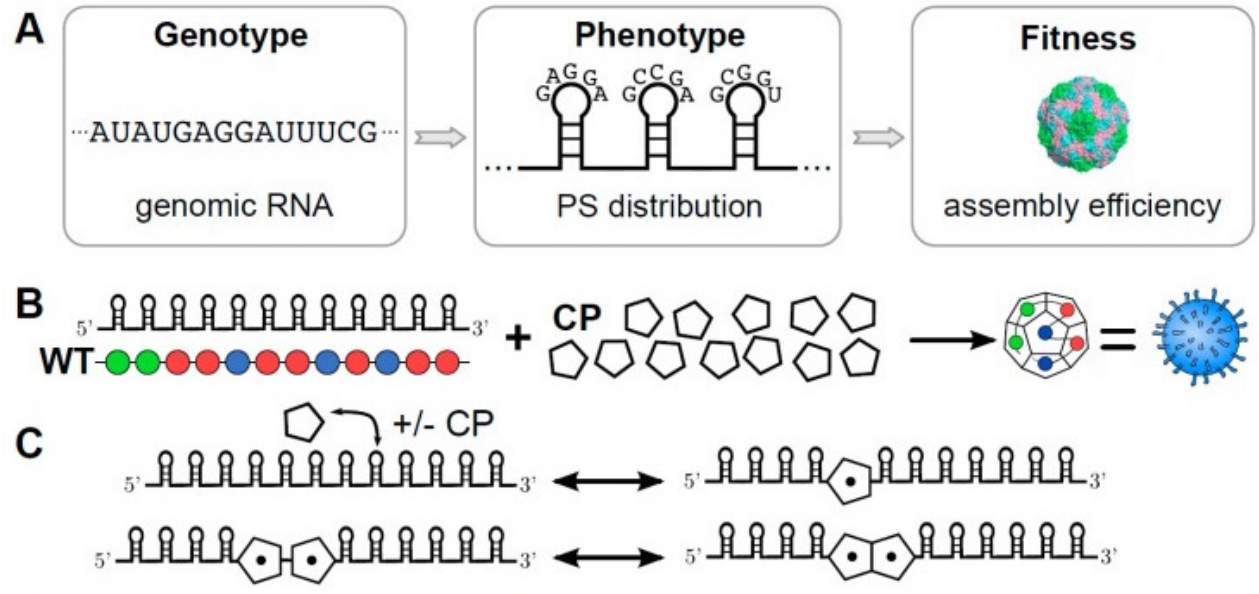


Simulating an assembly toy model: Dodecahedral cow



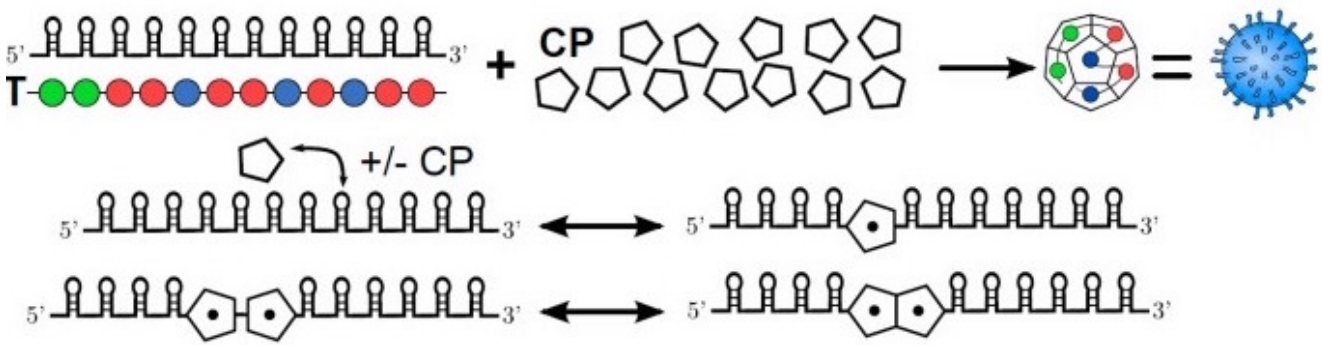


A mathematical and biophysical assembly toy model



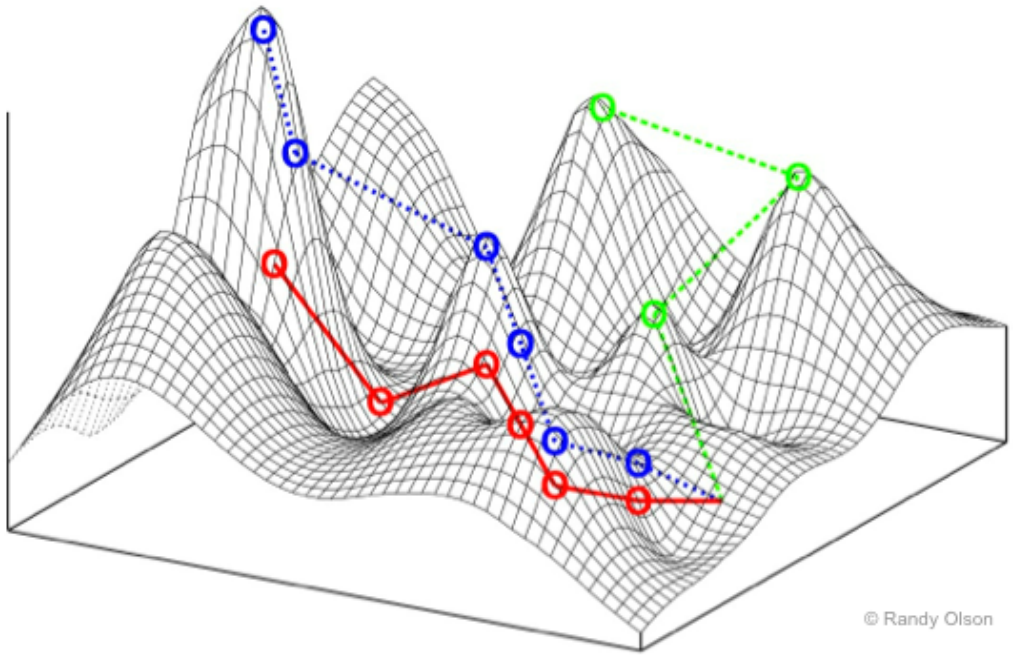
A phenomenological genome space of 12 packaging signals with 3 binding affinity bands (weak, medium, strong). Can compute the whole space explicitly in terms of assembly efficiency.

Simplest model: the dodecahedron



- 12 PSs in 3 bands (strong/intermediate/weak, 3/2/1, green/blue/red)
- Gillespie molecular simulation: stochastically select one possible reaction at a time
- Enough resources for 2000 virus capsids
- Computed fitness landscape in 4 weeks on a supercomputer

Fitness Landscape



Generally messy (many contributions) and difficult to quantify.
Here capture the assembly contribution for the phenotype space of 3^{12} points with (stochastic) assembly efficiency (< 2000).

A prime machine learning example

- Input vector: Genotype/Phenotype of length 12 (packaging signal strengths in 3 bands): 12D vector
- Output vector: Assembly efficiency (out of 2000 possible capsids): scalar
- Black box: Expensive map: Molecular dynamics simulations (computationally very costly)

A prime machine learning example

- Input vector: Genotype/Phenotype of length 12 (packaging signal strengths in 3 bands): 12D vector
- Output vector: Assembly efficiency (out of 2000 possible capsids): scalar

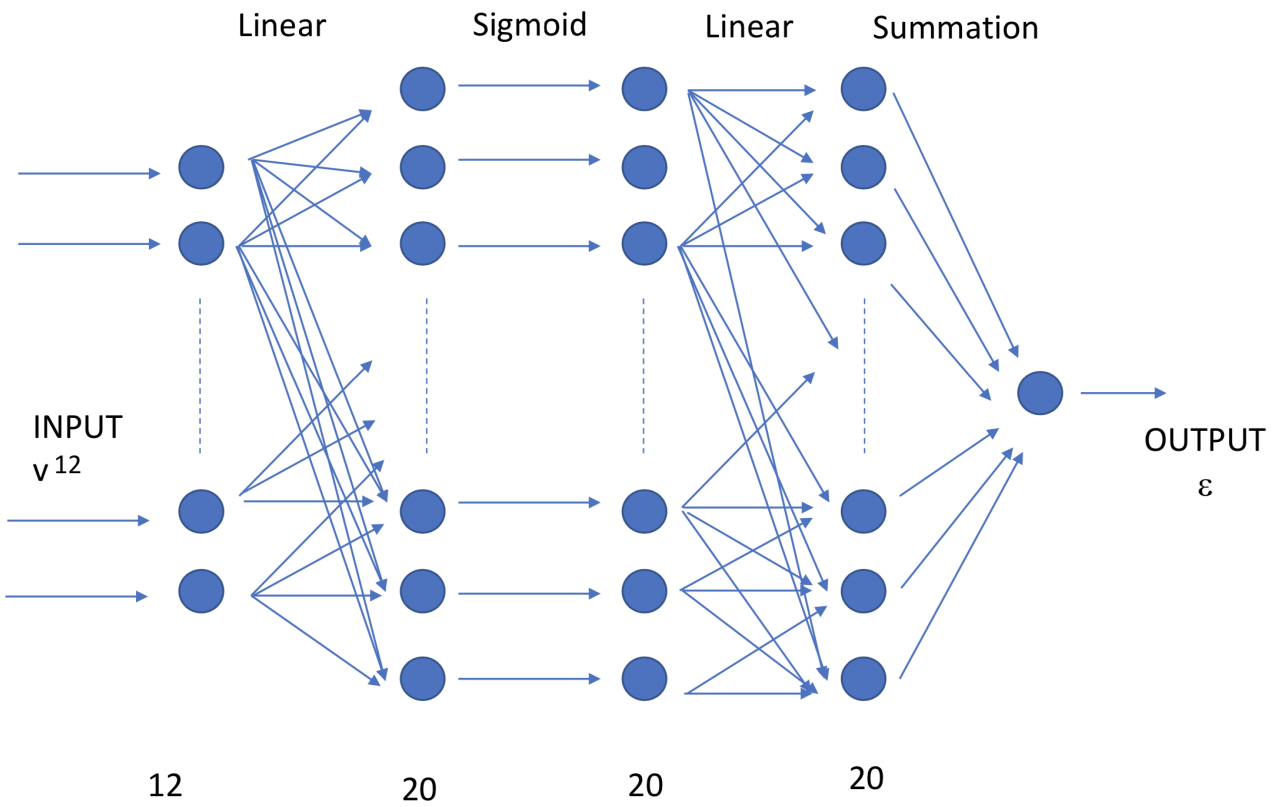
- Black box: Cheap approximation of map: Machine learning via a neural network

Input and Output dataset

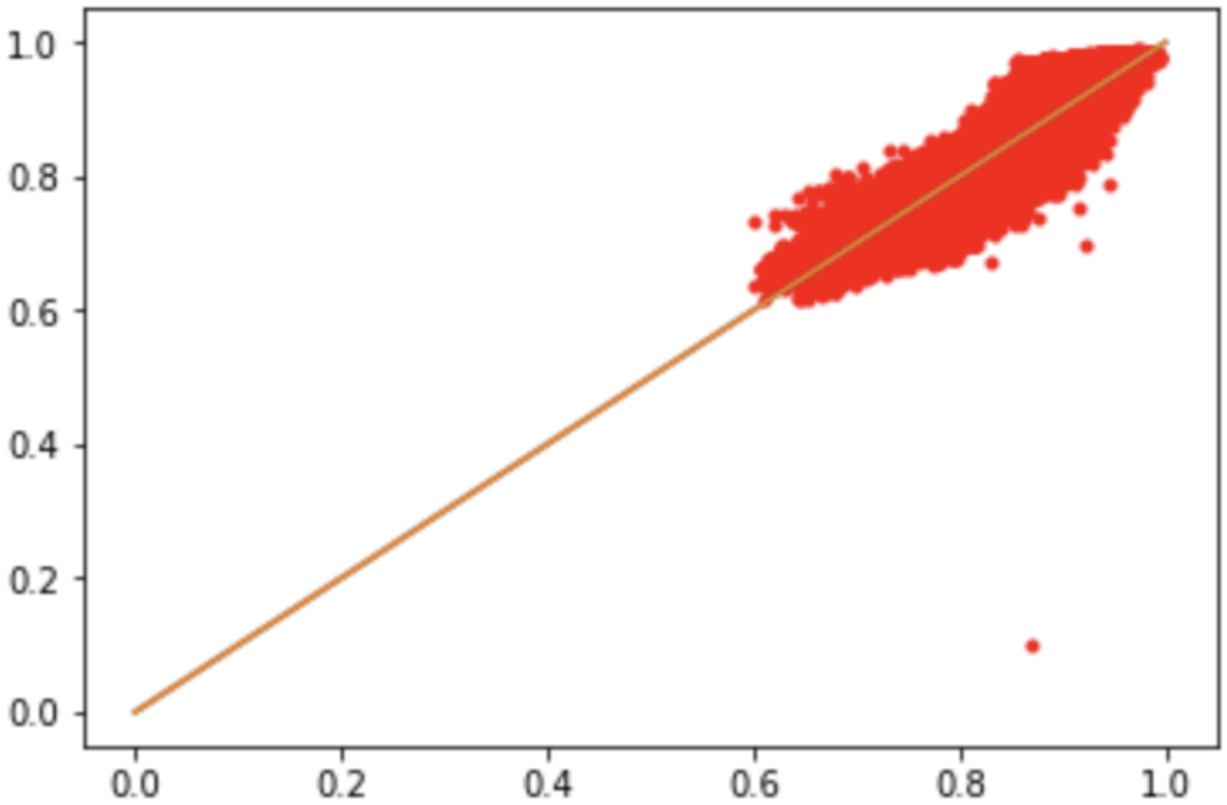
	Genome	Fitness
0	111111111111	200
1	111111111112	1393
2	111111111113	1869
3	111111111121	1597
4	111111111122	1896
5	111111111123	1960
6	111111111131	1875
7	111111111132	1959
8	111111111133	1961
9	111111111211	1639
10	111111111212	1683
11	111111111213	1895
12	111111111221	1848
13	111111111222	1904
14	111111111223	1964
15	111111111231	1904
16	111111111232	1949
17	111111111233	1959
18	111111111311	1852
19	111111111312	1858

$3^{12} \sim \frac{1}{2}$ Million data points

Machine Learning with a Neural Network

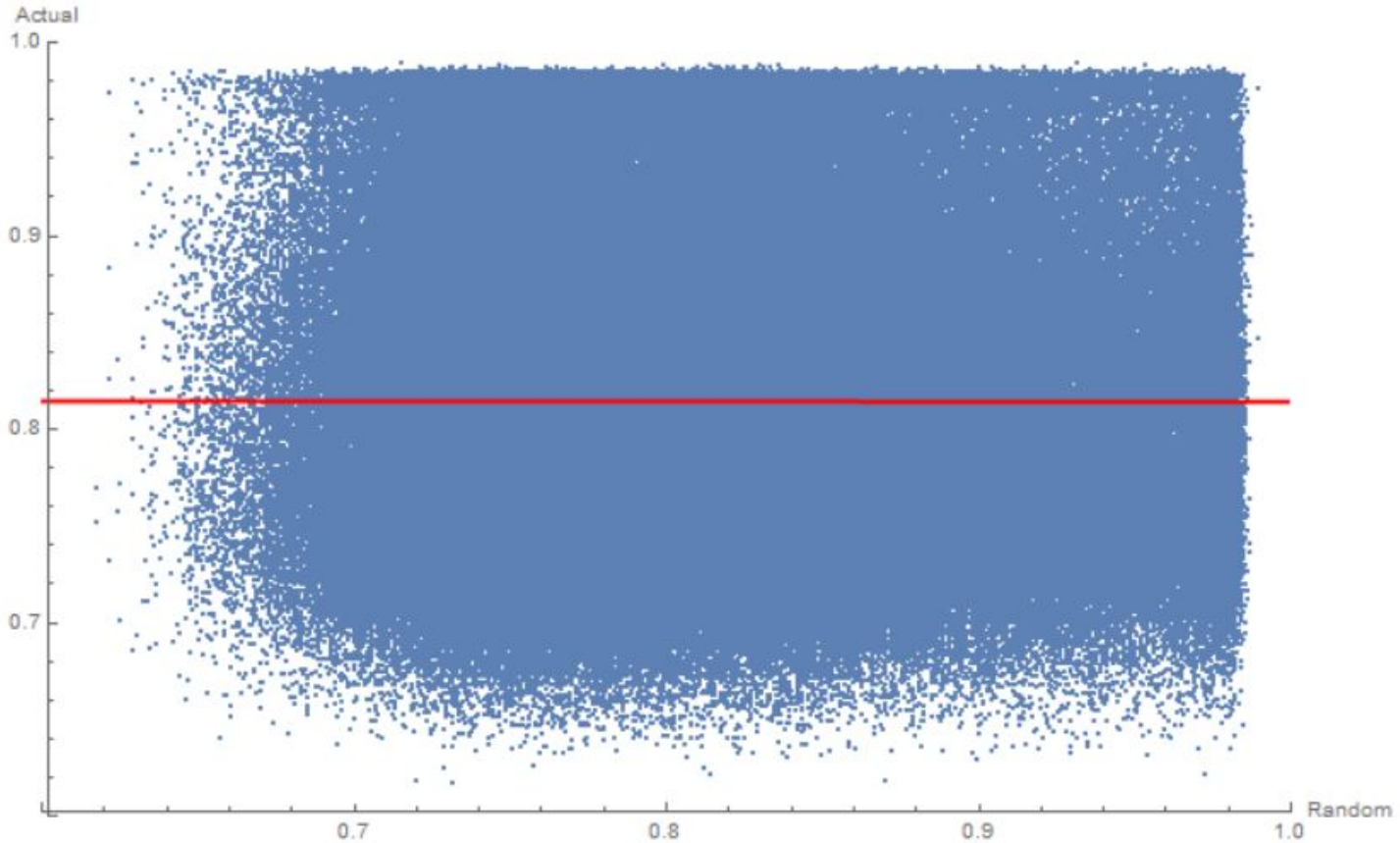


Predictions



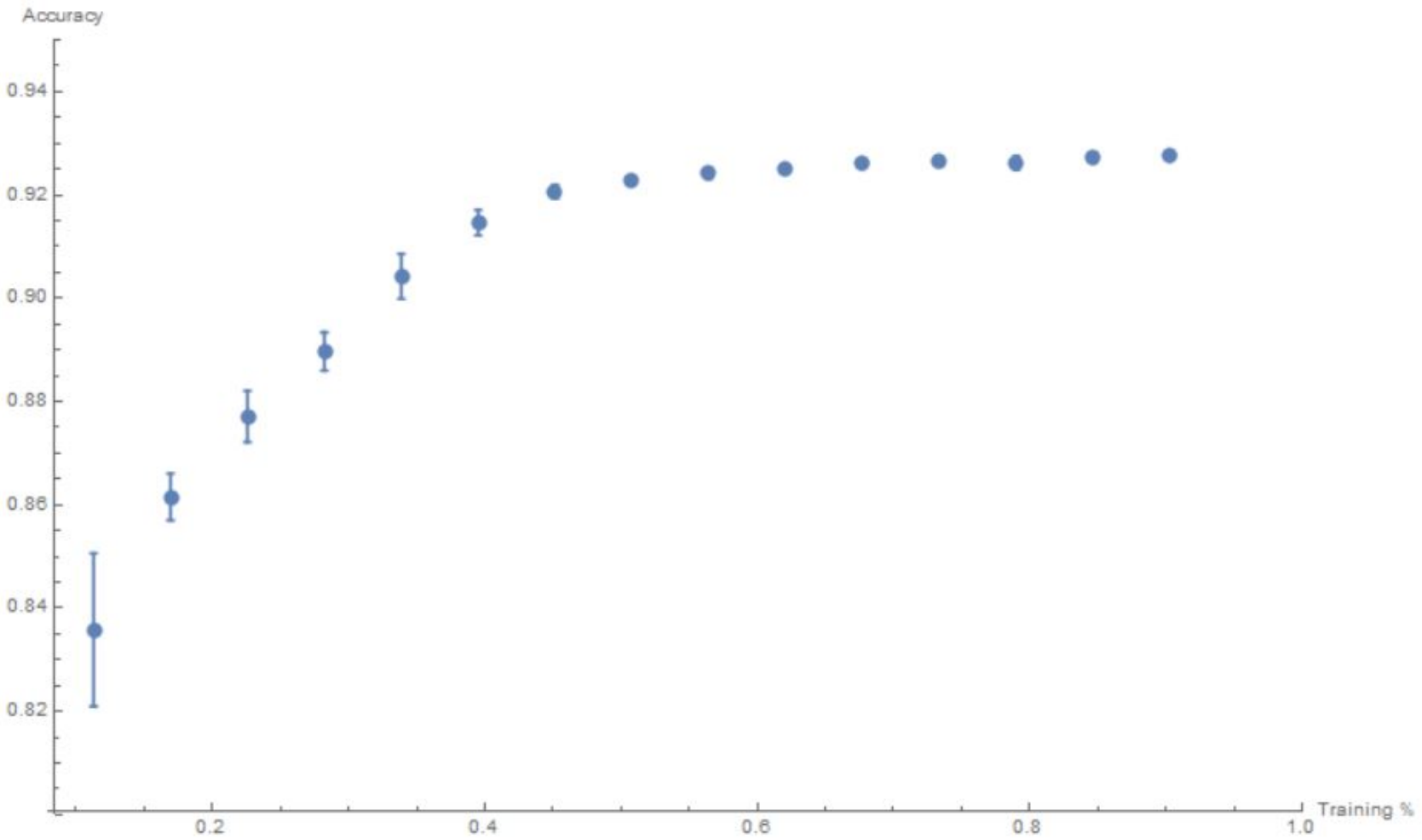
predicted vs actual value of assembly efficiency

Predictions

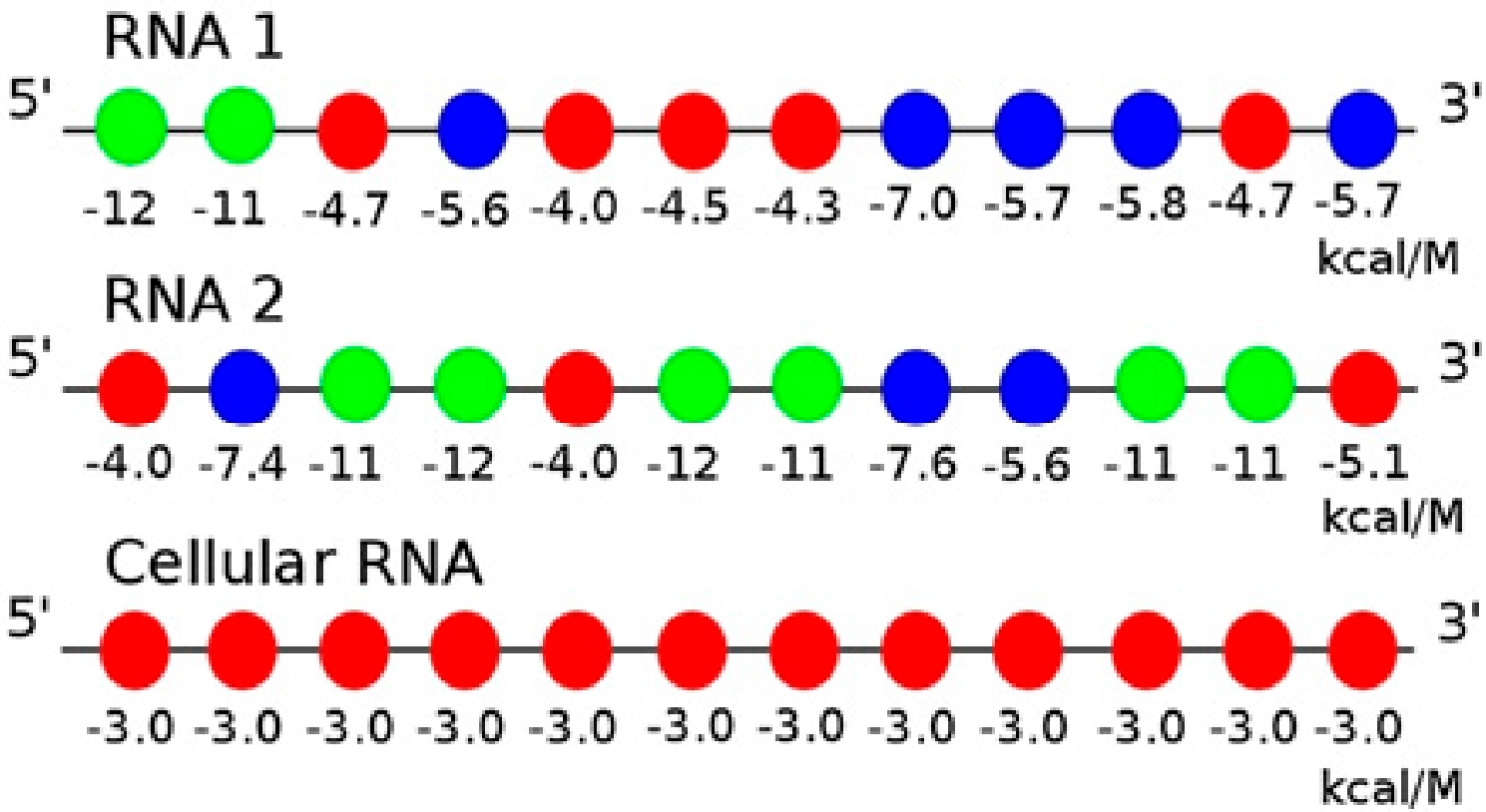


vs random assignments of assembly efficiency

Learning Curve

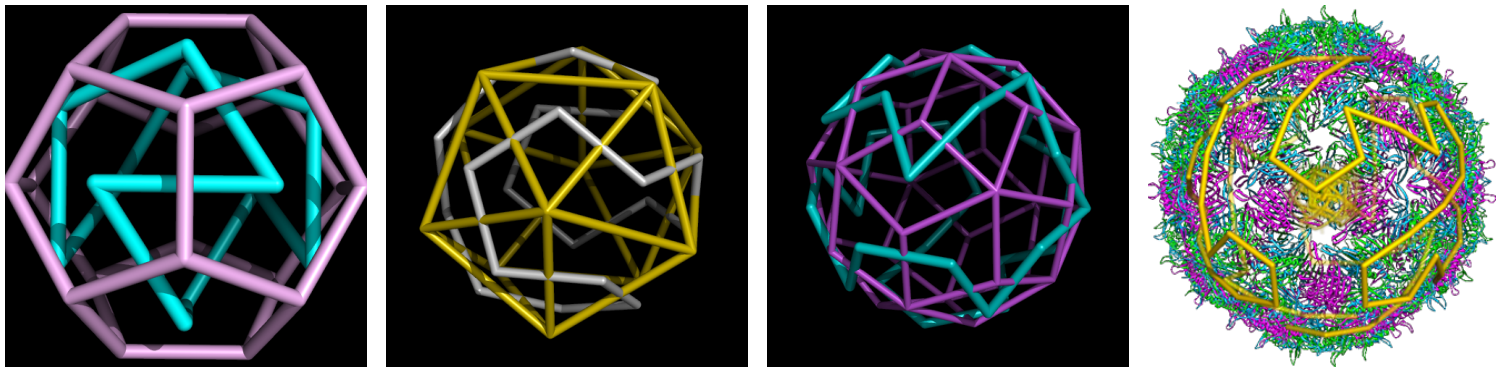


Not just random, intrinsic features?



Definite starting point with strong binding, then weaker binding in an error-correcting bit, driven to completion by thermodynamics

Conclusions: Mathematical and Computational Virology



ML might allow us to do more realistic models in future – geometry, binding strength gradation. Partially explore the landscape and predict the rest (procedurally)?

Insights into mathematical and biophysical design principles open up novel directions for biomedicine and nanotechnology.

Overview

Mathematical Modelling

- Watson & Crick: Icosahedral and helical symmetry
- Caspar & Klug: Triangulations
- Twarock: More general surface tilings
- Affine symmetry: genome and capsid

Computational Modelling

- Bioinformatics: packaging signals
- Gillespie stochastic simulations: epidemiological, infection, assembly
- Machine learning: fitness landscape

Further references

[1] F. Crick, J. Watson, Nature 177, 1956, 473.

[2] D. Caspar, A. Klug, Cold Spring Harb. Symp. Quant. Biol. 27, 1962, 1.

[3] R. Twarock, A tiling approach to virus capsid assembly explaining a structural puzzle in virology, J. Theor. Biol. 226, 477-482 (2004)

[4] R. Twarock & A. Luque, Structural puzzles in virology solved with an overarching icosahedral design principle, Nature Comms 10.1 (2019): 1-9

[5] P-P. Dechant, C. Boehm & R. Twarock, Novel Kac-Moody-type affine extensions of non-crystallographic Coxeter groups, J. Phys A, 285202 (2012)

[6] T. Keef, J.P. Wardman, N.A. Ranson, P.G. Stockley & R. Twarock, Structural constraints on the three-dimensional geometry of simple viruses: case studies of a new predictive tool. Acta Cryst A. 69, 2013, 140-50

[7] P. Dechant, J. Wardman, T. Keef & R. Twarock, Viruses and fullerenes - symmetry as a common thread? Acta Cryst A 70, 2014, 162-7

[8] R. Twarock, G. Leonov & P.G. Stockley, Hamiltonian Path Analysis of Viral Genomes, Nature Communications 9, 2018, 2021

[9] R. Twarock & P.G. Stockley, RNA-Mediated Virus Assembly: Mechanisms and Consequences for Viral Evolution and Therapy, Annu Rev Biophys. 48, 2019, 495-514

[10] G. Indelicato, N. Wahome, P. Ringler, S. A. Müller, M.-P. Nieh, P. Burkhard & R. Twarock, Principles Governing the Self-Assembly of Coiled-Coil Protein Nanoparticles, Biophys J. 110, 2016, 646-60

Main references

Models of Viral Capsid Symmetry as a Driver of Discovery in Virology and Nanotechnology
P-P Dechant, R Twarock, The Biochemist, 2021

Machine-learning a virus assembly fitness landscape
P-P Dechant, Y-H He, PLOS One, arXiv preprint arXiv:1901.05051, 2021

Topical Collection: Machine-learning mathematical structures
Editors: Y-H He, P-P Dechant, A Kasprzik, A Lukas
Advances in Applied Clifford Algebras, August 2021

Algebraic interests

Thank you!

- Exceptional root systems/geometries (H_4, E_8 etc)
- Clifford algebras
- ADE correspondences
- (Reflexive) polytopes