Building polyhedra by self-assembly

Govind Menon

Experiments

Shivendra Pandey
David Gracias (lab)

Johns Hopkins University

Computations

Ryan Kaplan, Daniel Johnson

Brown University

Part 1. Biology, technology, and a little math…

This talk is a description of a common mathematical framework to describe self-assembly of polyhedra. I will contrast two problems:
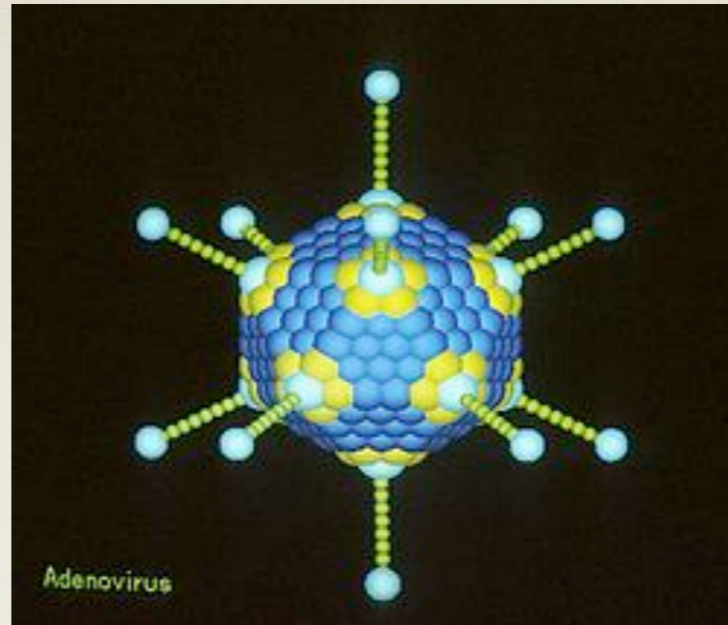
The self-assembly of the bacteriophage MS2 (Reidun Twarock's team)

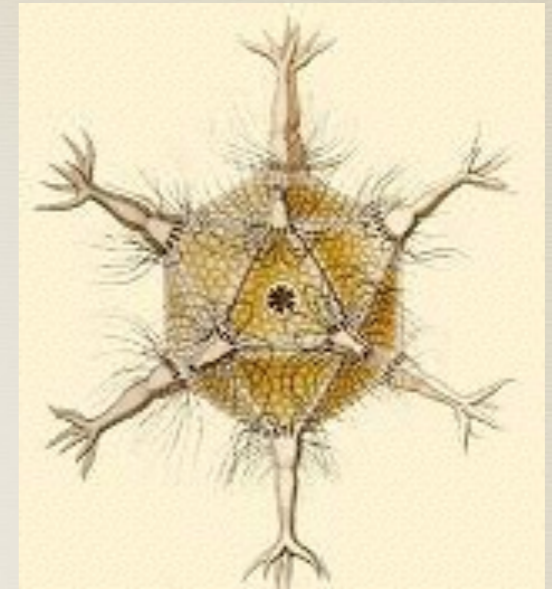Surface tension driven self-folding polyhedra (our work).

Examples of icosahedral symmetry in nature
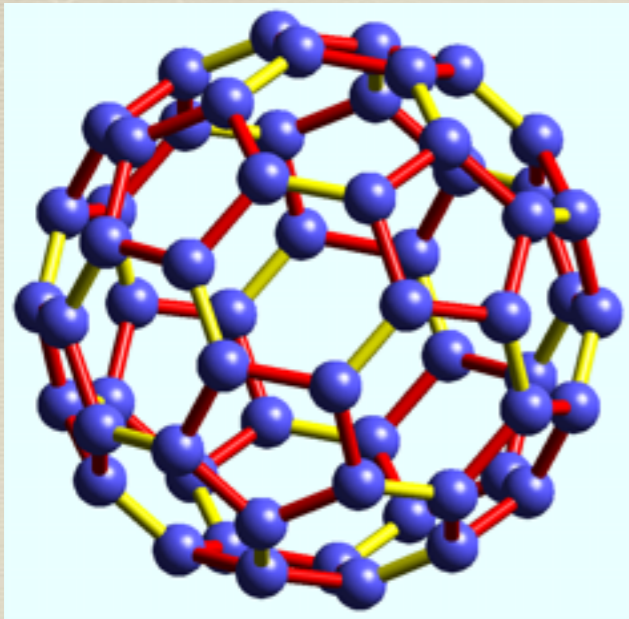
C 60 molecule, 0.7 nm   Adenovirus, 90 nm   Radioalarian   10 μm
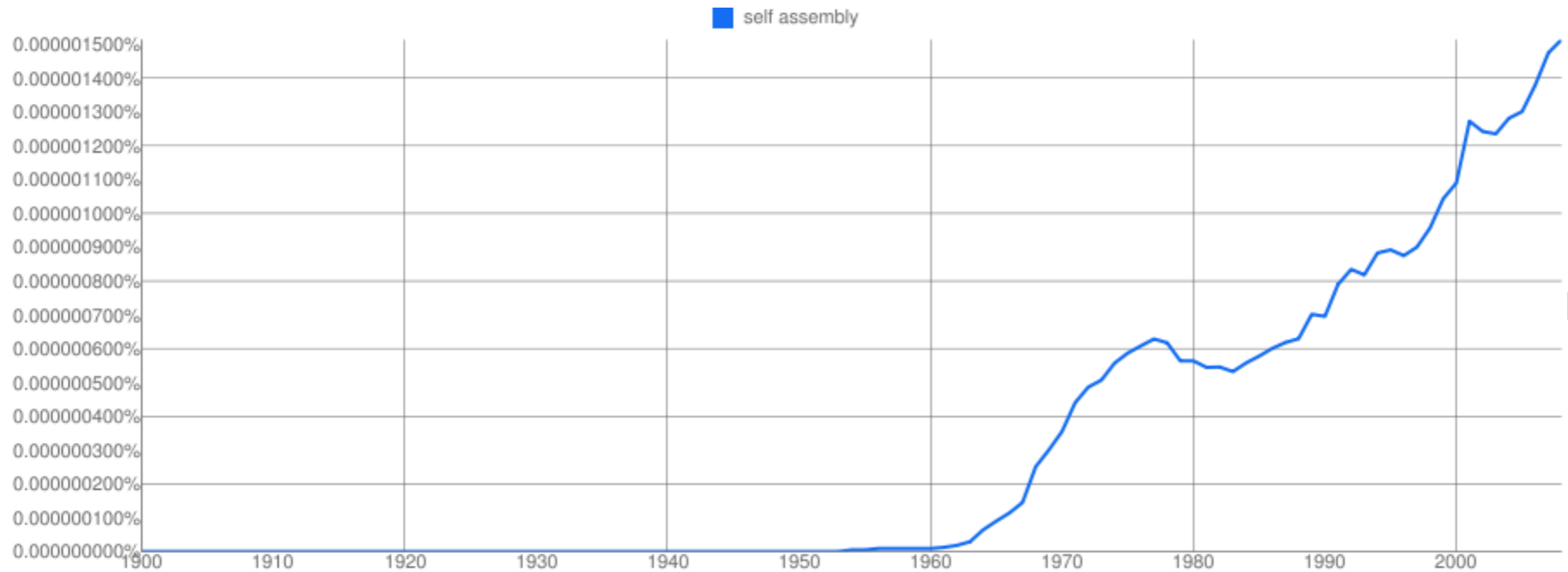
Widely different self-assembly mechanisms at different scales.

It is a very interesting to understand the types of symmetry, the "coding
of symmetry" in the genome, and the interplay between symmetry and
the pathways of self-assembly.

One of the first uses of the phrase "self assembly" is by Caspar and Klug in their work on the structure of viruses. They distinguish grades of organization in a cell as sub-assembly and self-assembly and write:

Self-assembly (of a virus) is a process akin to crystallization and is governed by the laws of statistical mechanics. The protein subunits and the nucleic acid chain spontaneously come together to form a simple virus particle because this is their lowest free energy state ."

Caspar and Klug; Cold Spring Harbor Symposium, (1962)

Our work is on <u>synthetic</u> self-assembly.

We want biology to inspire the design of devices and materials. In turn, we hope that synthetic models will shed light on biological self-assembly.

Typical themes:

stripped down interactions (e.g. one dominant energy scale),
simple shapes built out of a few simpler motifs,
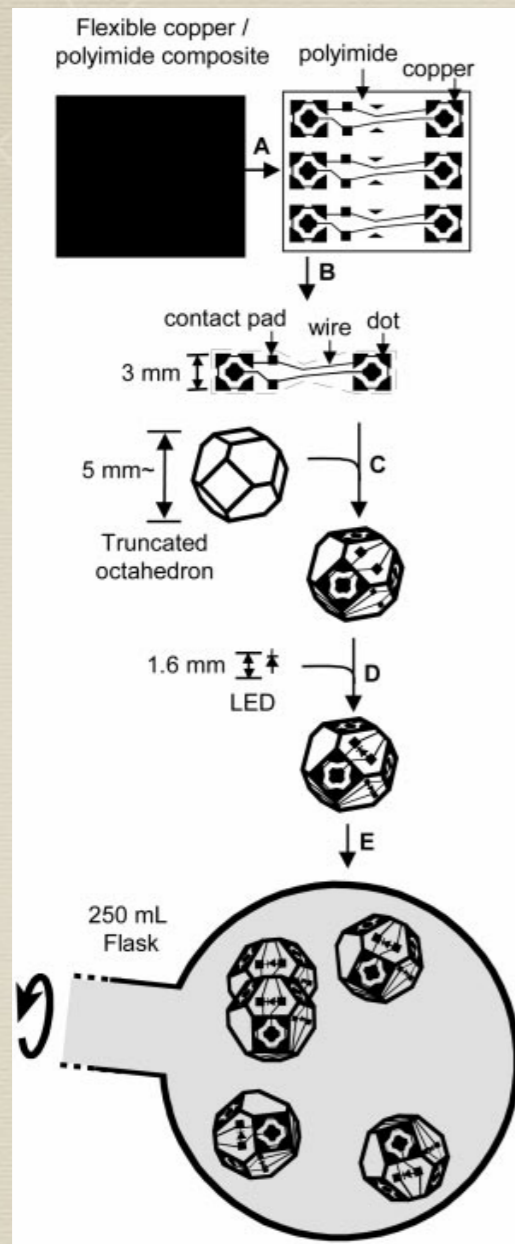some randomness.

A self-assembling circuit

Fig. 1. The procedure used to form electrical networks in 3D by self-assembly (12). (A) An array of the basic pattern of copper dots, contact pads, and wires was defined on a flexible copper-polyimide sheet using photolithography and etching. (B) These pattern elements were cut out along the dotted line, (C) glued on the faces of the polyhedron, and (D) LEDs were soldered manually onto the contact pads. (E) The copper dots and wires on the TOs were coated with solder, and self-assembly occurred in hot, isodense, aqueous KBr solution.
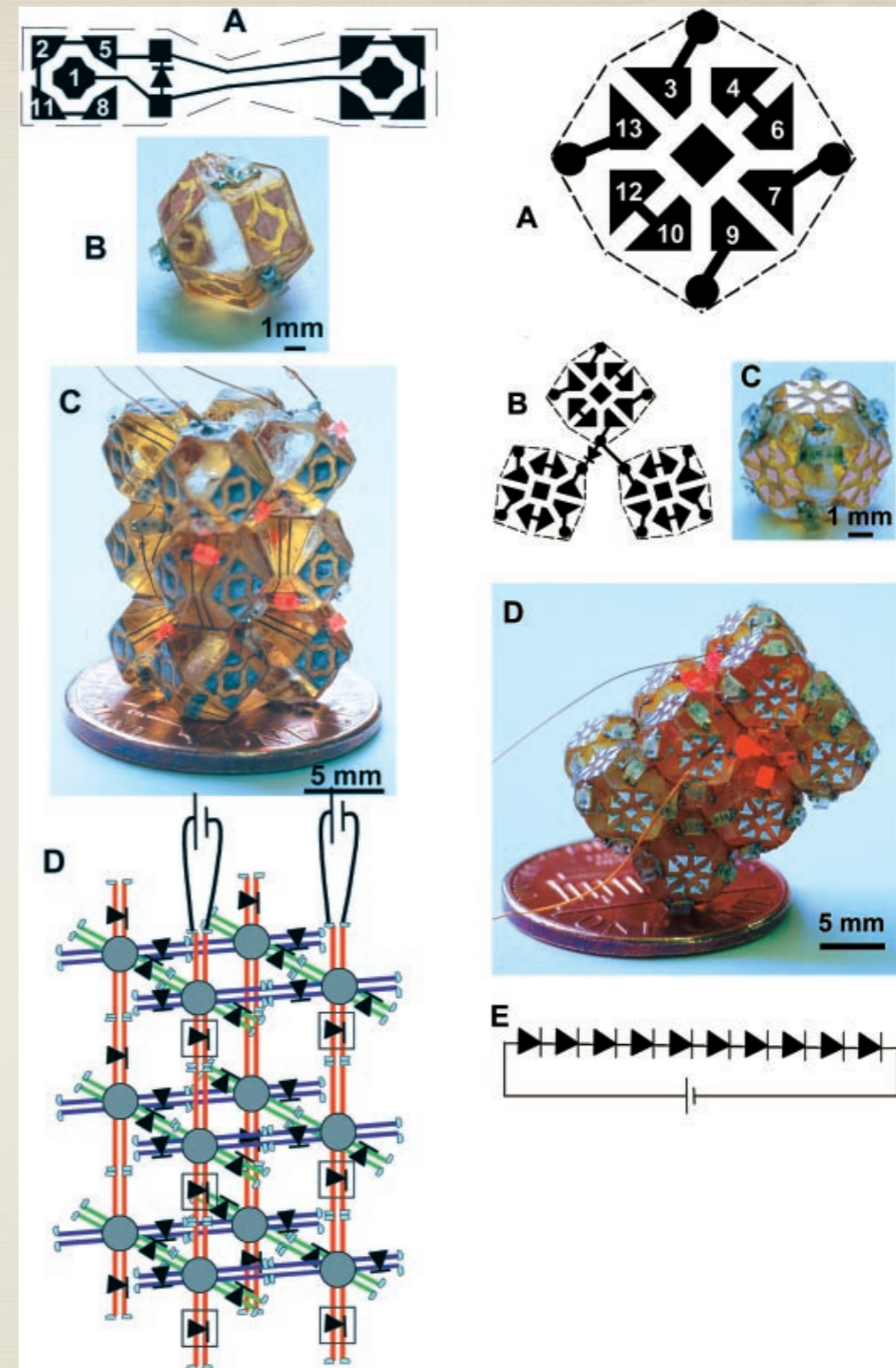
Gracias, Tien, Breen, Hsu, Whitesides, (Science, 2000).

From "proof of concept" to design principles

Many initial laboratory experiments provide "proof of concept". In the past ten years, there have been significant improvements in self-assembly across many scales. For example, polyhedra can now be self-assembled in several distinct laboratory experiments.
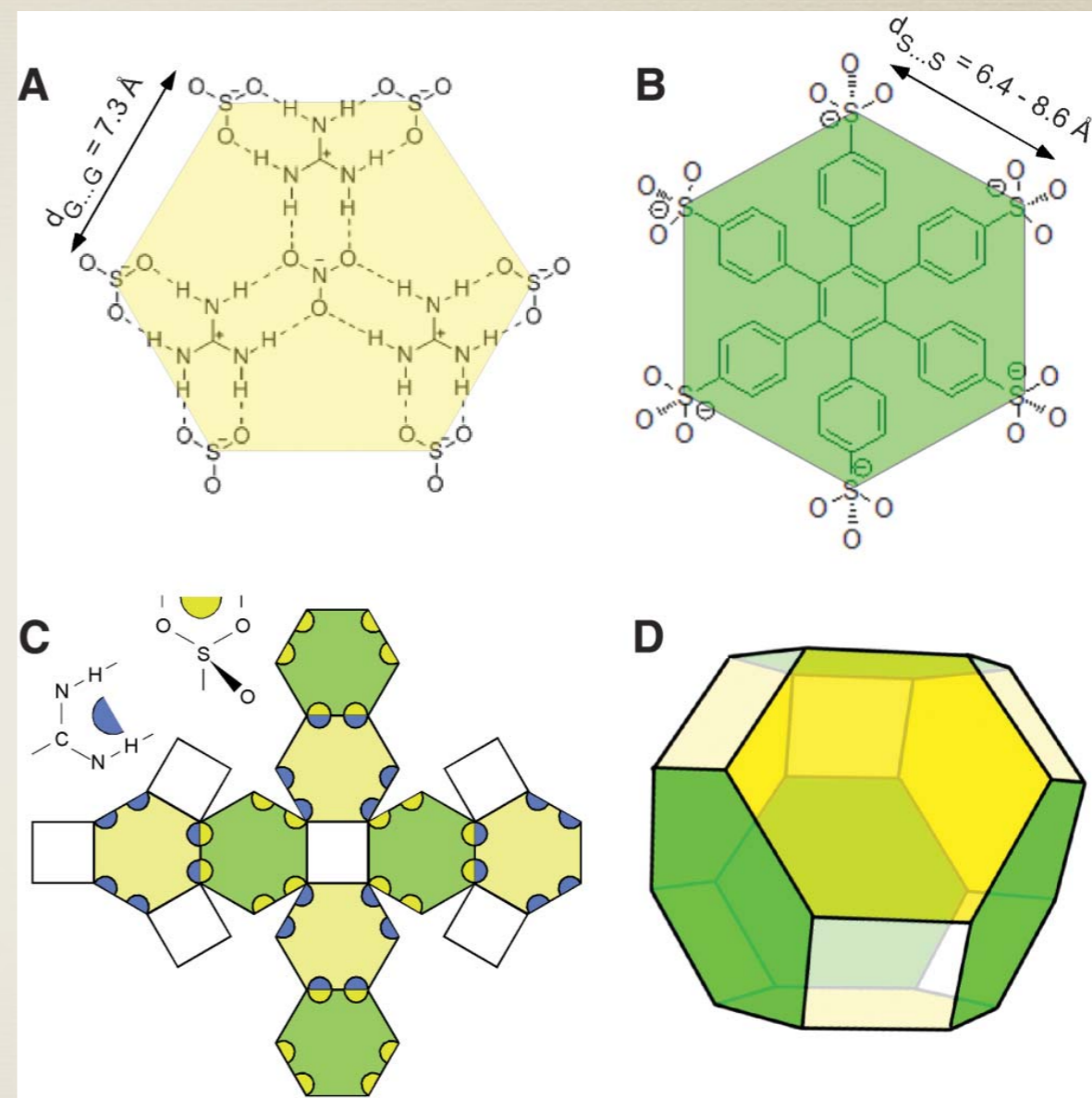
Our interest lies in:

(1) mathematical modeling of self-assembly, especially geometric principles.

(2) connections with biological self-assembly.

(3) principles for efficient design.

Examples of synthetic polyhedra: 1 nm scale
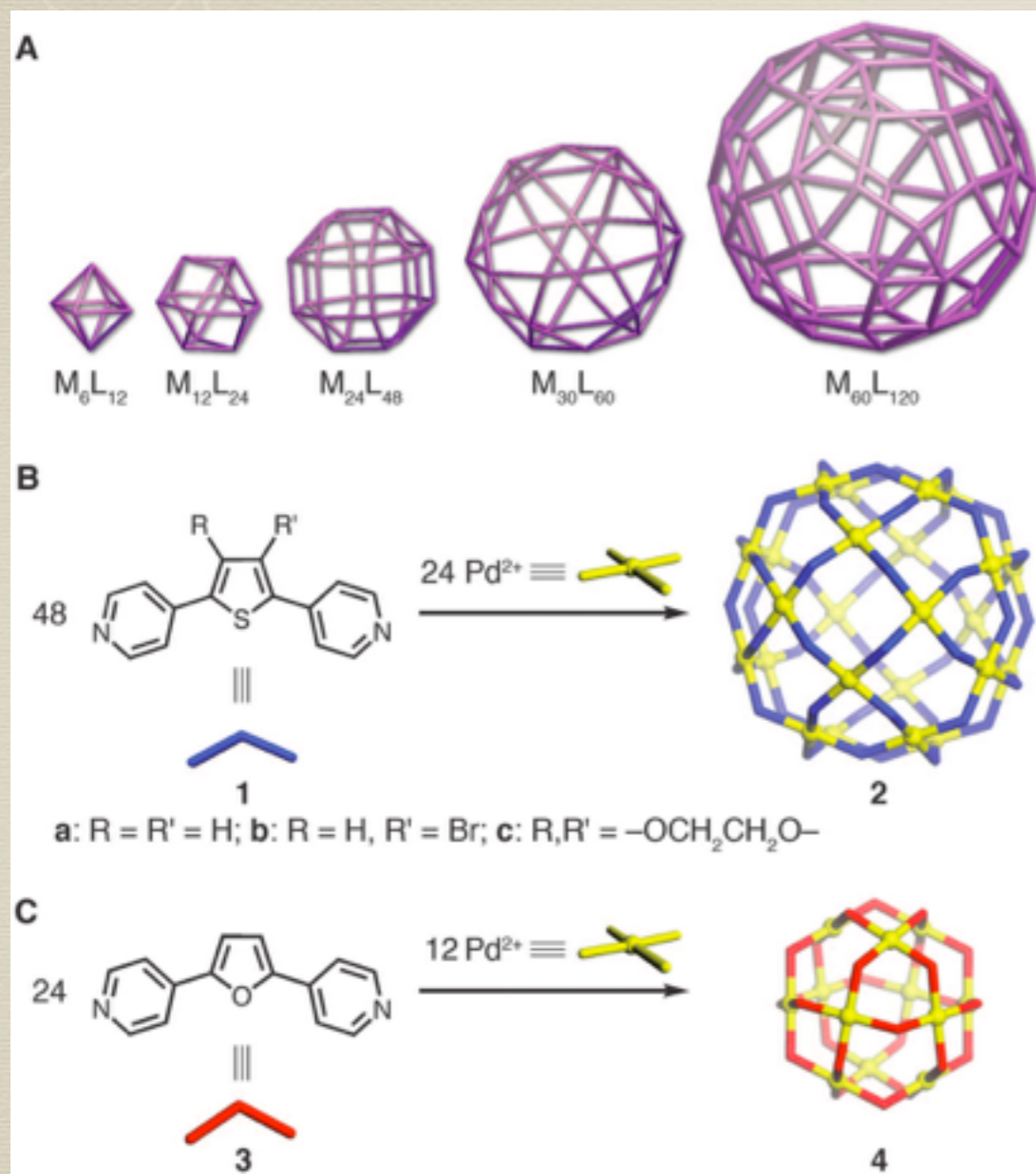
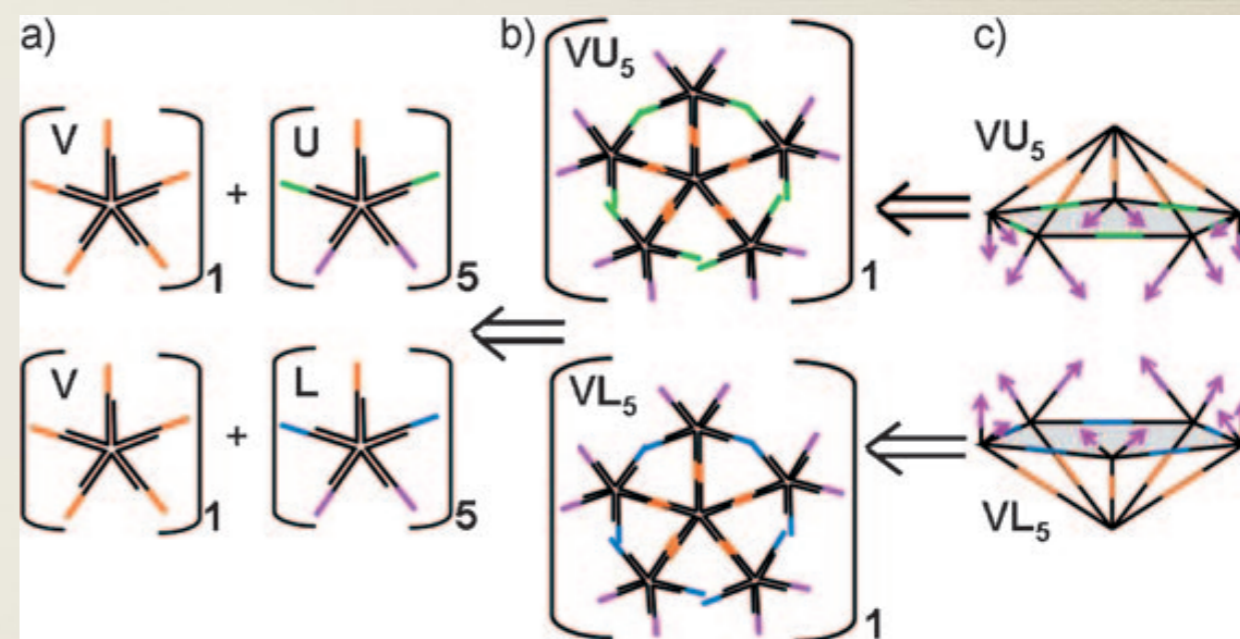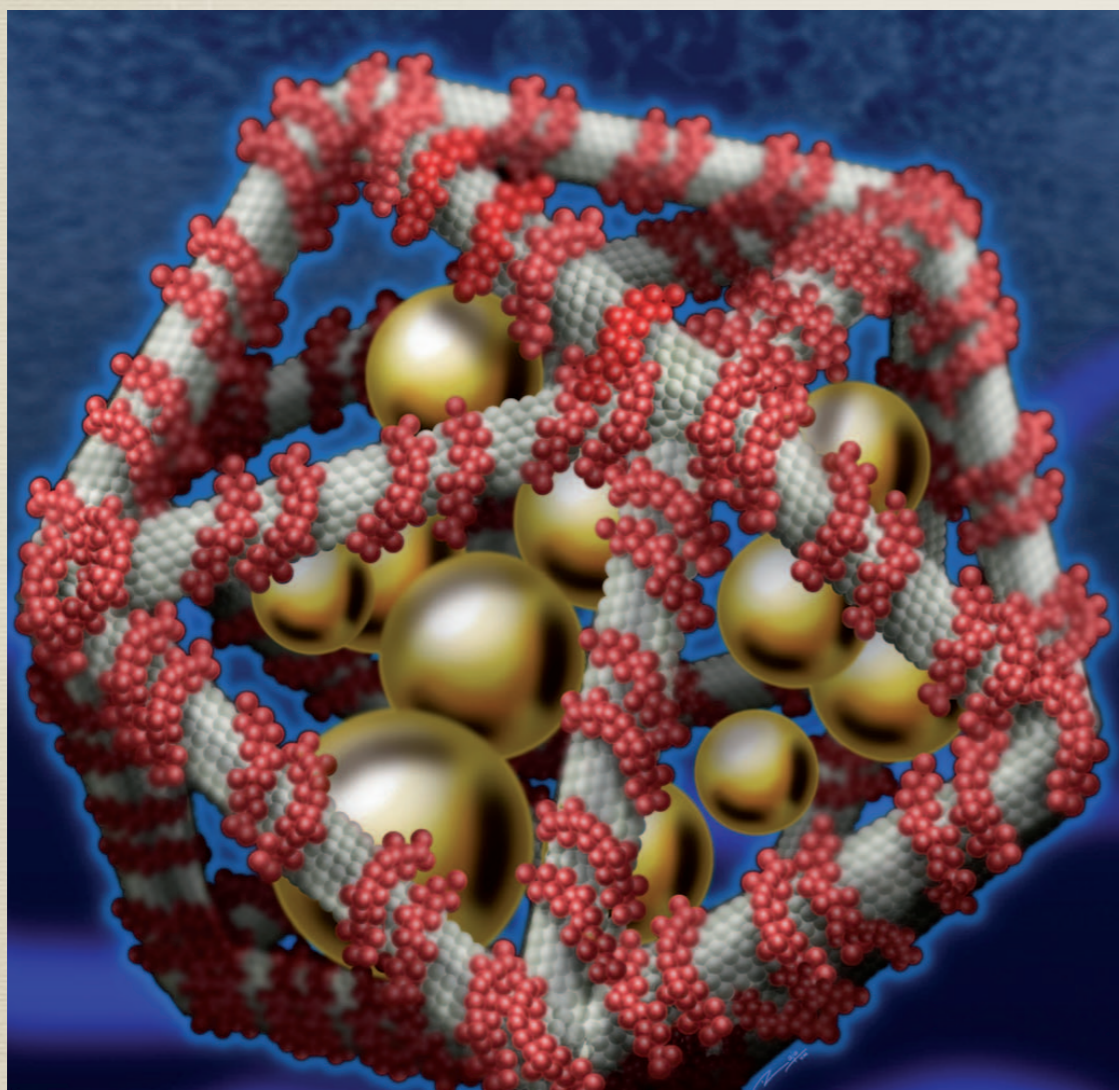Organometallic supermolecules built by covalent bonds.

Fujita lab (Tokyo), Science (2010).

Archimedean cage built by hydrogen bonds.

Ward lab (NYU), Science (2011).

Examples of synthetic polyhedra: 10 nm scale

Icosahedral DNA cage built in modules.

Yamuna Krishnan's lab (NCBS), Ang. Chemie. (2009).

passing an elastic thread alternately above and beneath the corners of the double star, holding the model flat with the other hand. Removing the hand (230), we see the dodecahedron rising (231) as a perfect model. To paint its faces so that adjoining faces have different colors, no less than four colors are sufficient. Choosing four colors, e. g. red, green, blue, and yellow, we may distribute them on a dodecahedron either as illustrated by the sketch (227) or in an essentially different way (228). (By rotation and reflection can we transform one of these models into the other?)

We can inscribe a cube into a regular dodecahedron in such a manner that every edge of the cube becomes a diagonal of a
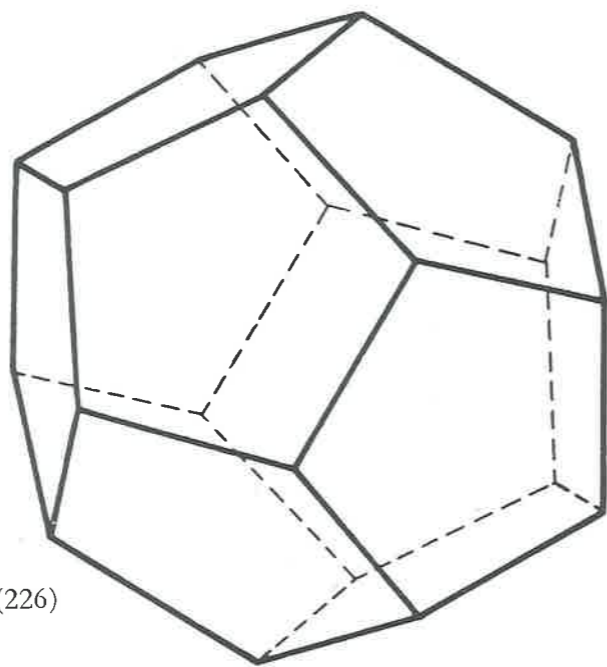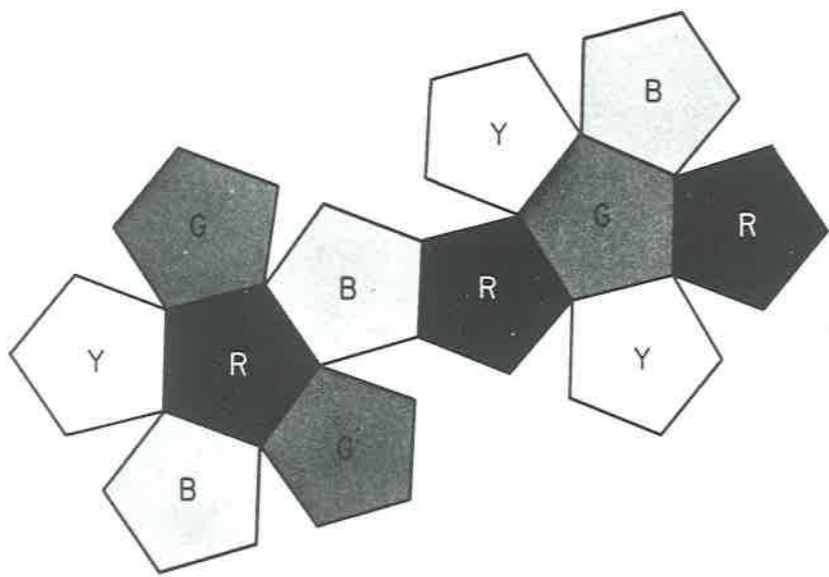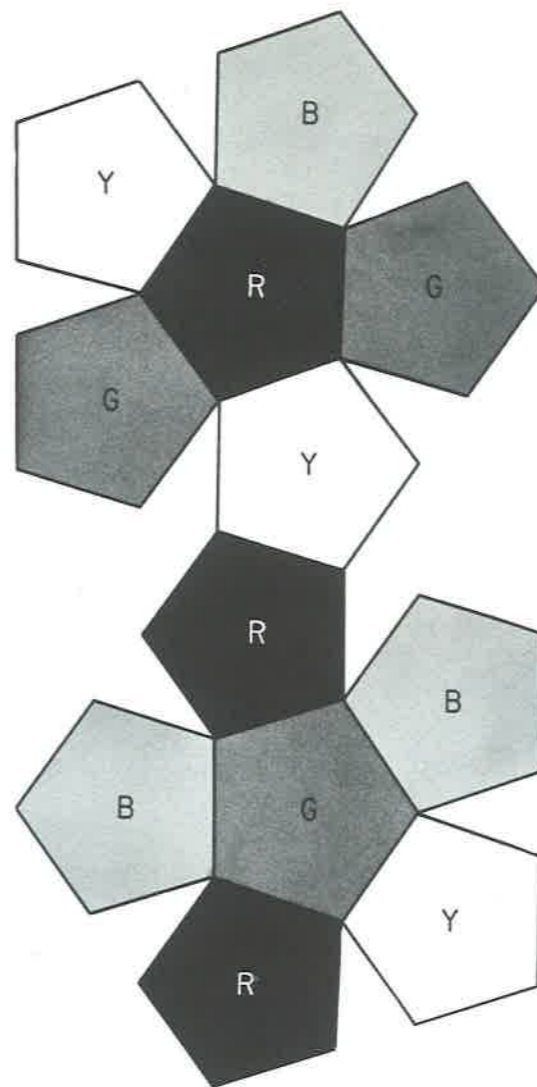
(229)

(230)

(231)

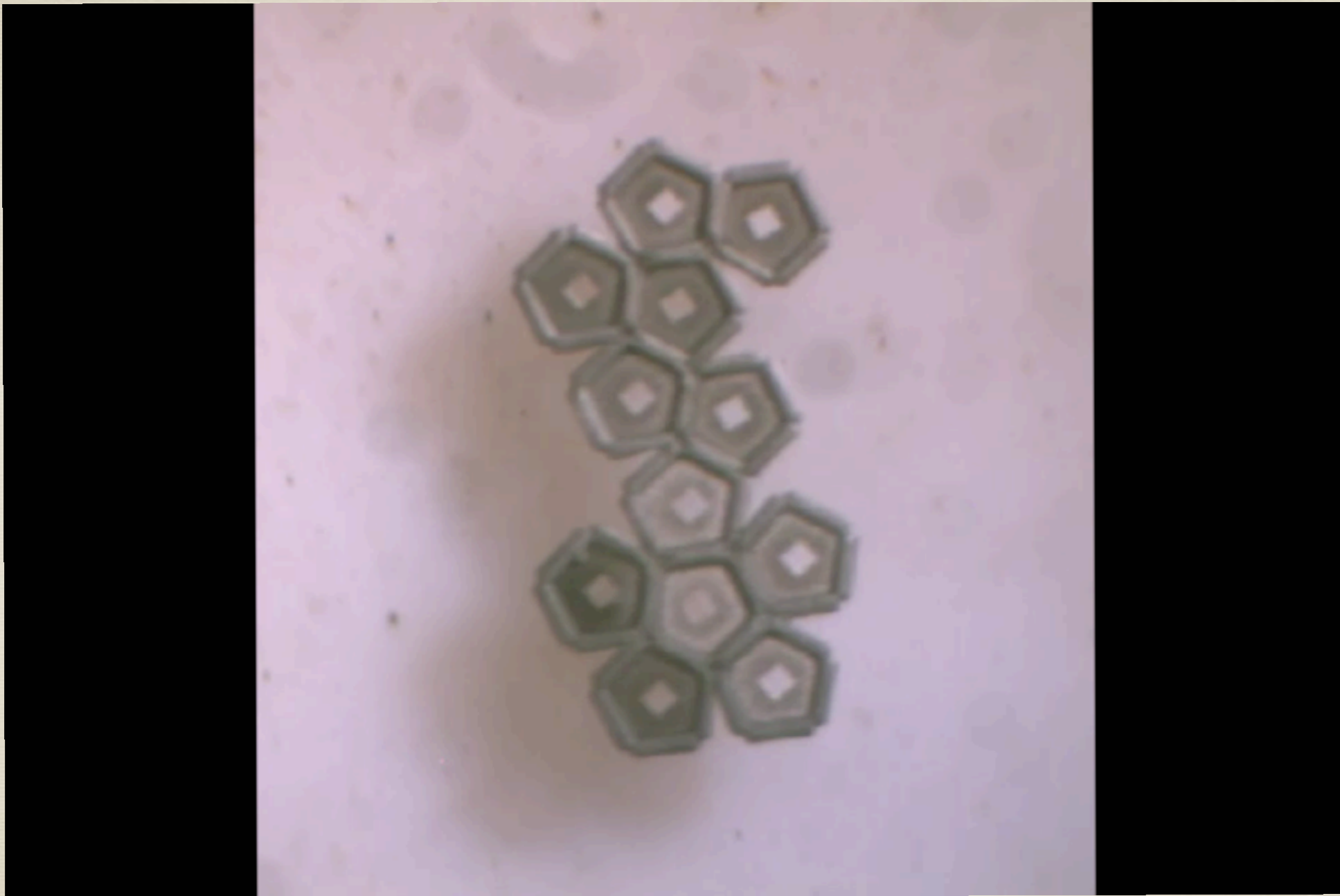Hugo Steinhaus, Mathematical Snapshots.
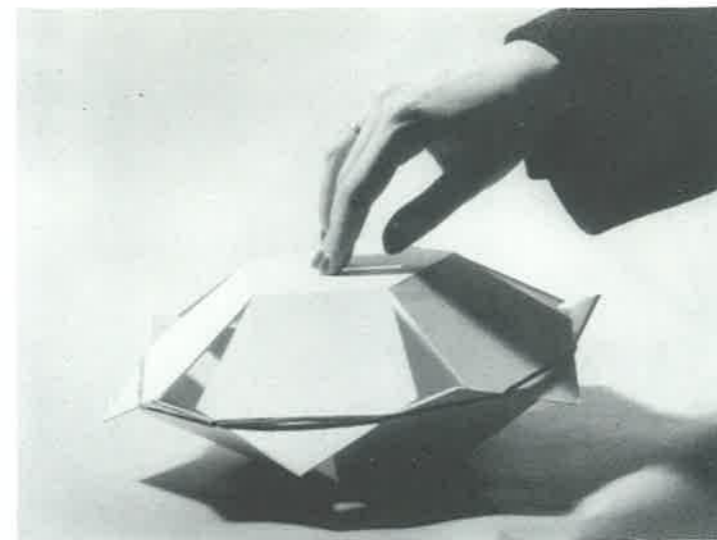
(226)

(227)

(228)

Examples of synthetic polyhedra: micron scale

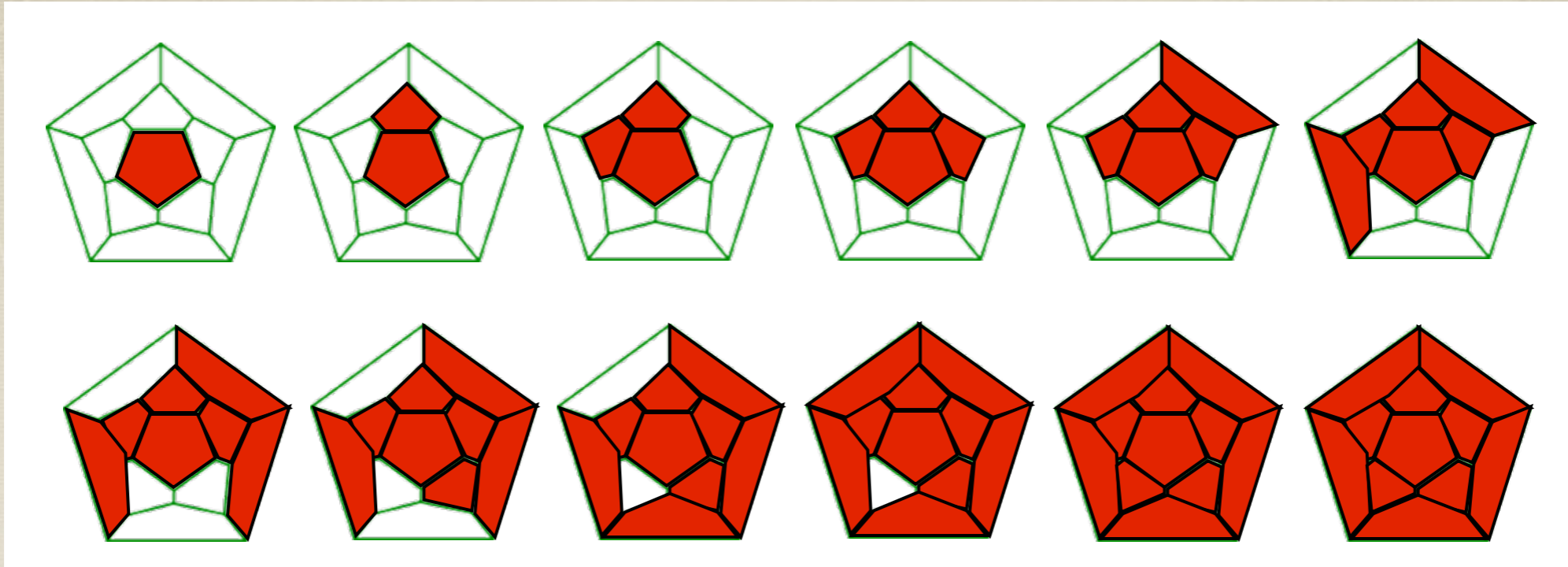Truncated octahedra built by self-folding.

Gracias, Menon + students, PNAS, (2011).

The main question

What is common to biological and synthetic self-assembly?

Our approach: use discrete geometry to model the intermediates and pathways of assembly. Try to understand the most robust features of simple models.

(1)   Discretize the assembly process into intermediate states.

(2) Model attachment/detachment kinetics using experimental data (or more often simple-minded physics) to determine rates.

(3) Useful in chemistry for modeling fullerenes (Wales, 1987)

Cannot explain malformed shapes, no sequence specific information. Many other models exist (Berger, Shor, 1994; Bruinsma (2005).
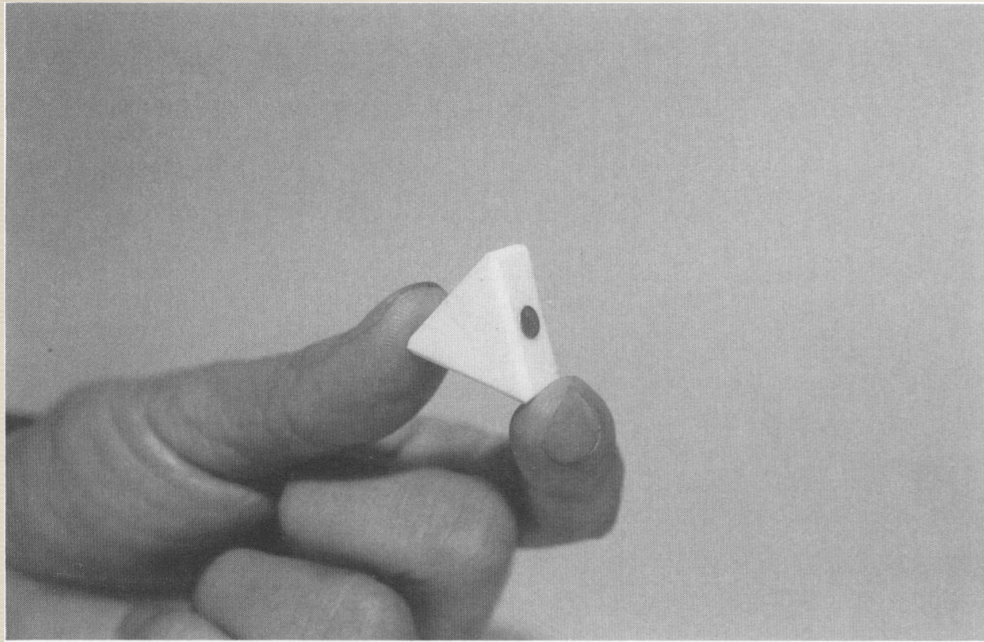
Macroscopic "chemical reactions"
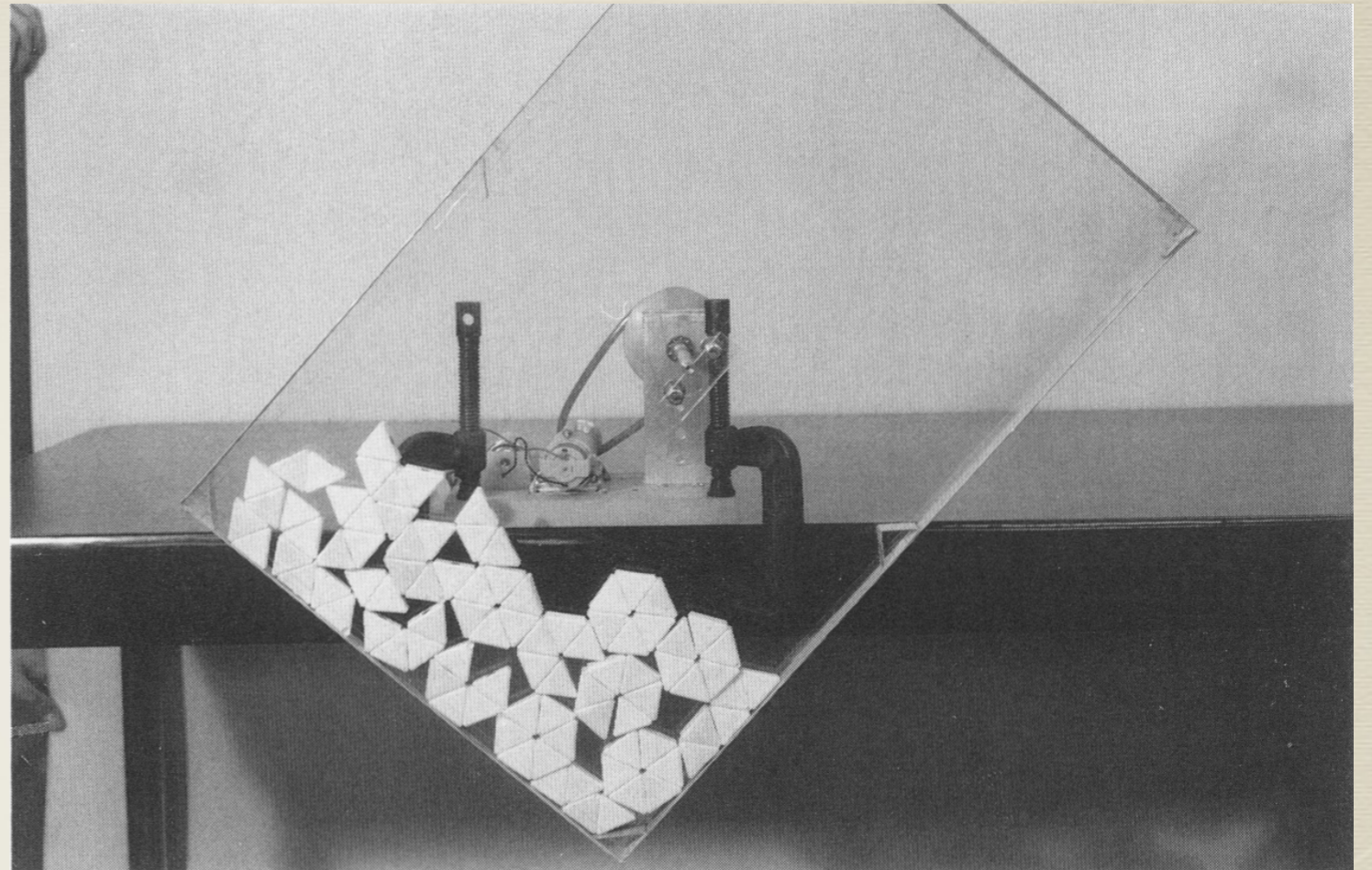

Figure 2. Photograph of the basic unit.

Figure 4. The experimental apparatus.

Hosokawa, Shimoyama, Miura, Artificial Life, (1996).

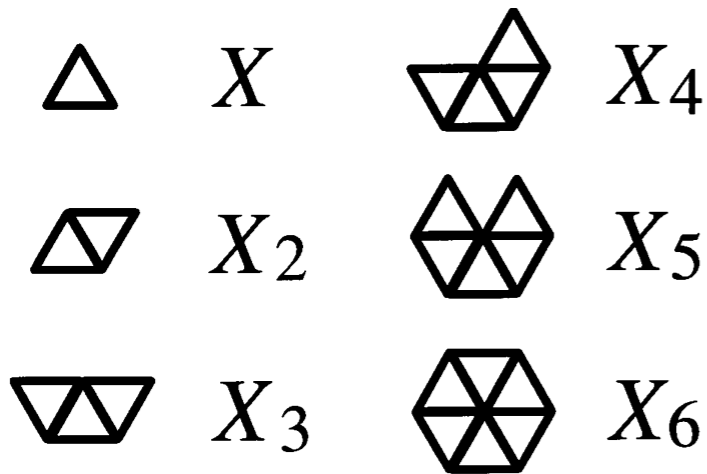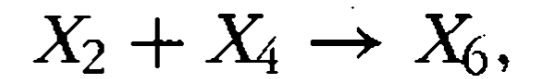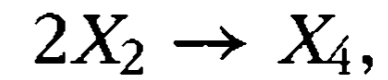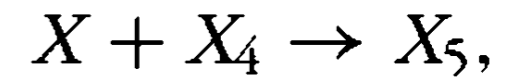# Chemical reactions theory: states, reactions, rates.

Figure 6. Initial, intermediate, and final products of the system.

$2X \rightarrow X_2,$ $X + X_2 \rightarrow X_3,$

$X + X_3 \rightarrow X_4,$ $X + X_4 \rightarrow X_5,$

$X + X_5 \rightarrow X_6,$ $2X_2 \rightarrow X_4,$
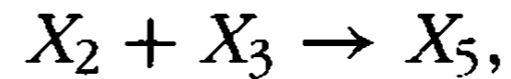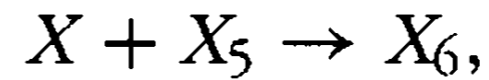
$X_2 + X_3 \rightarrow X_5,$ $X_2 + X_4 \rightarrow X_6,$

$2X_3 \rightarrow X_6.$
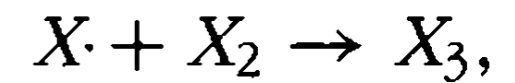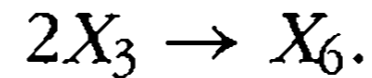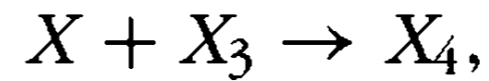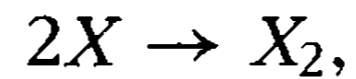
States   Reactions

Figure 12. Divisions of the plane using the units for the basis.

Rates: compute probabilities of collisions and bonding.
Must include physics at this stage.

A combinatorial explosion

(1) A linear system of equations need not be easy to solve.
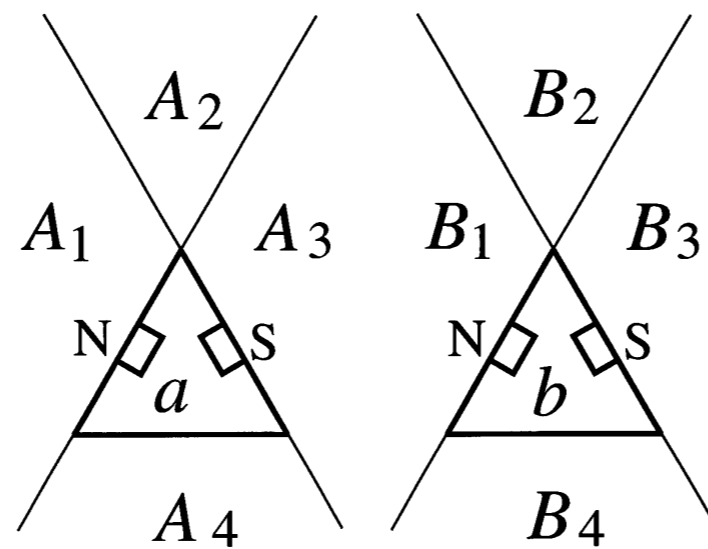(2) Rates cannot be determined from experiment (need a weight for each edge).

Polyhedron	# faces	# intermediates	# edges in $\mathcal{C}$	# assembly pathways from □ to ■
Tetrahedron	4	5 (5)	4 (4)	1 (1)
Cube	6	9 (8)	10 (8)	3 (2)
Octahedron	8	15 (12)	22 (14)	14 (4)
Dodecahedron	12	74 (53)	264 (156)	17,696 (2,166)
Icosahedron	20	2650 (468)	17242 (1984)	57,396,146,640 (10,599,738)
Truncated tetrahedron	8	29 (22)	65 (42)	402 (171)
Cuboctahedron	14	341 (137)	1636 (470)	10,170,968(6,258)
Truncated cube	14	500 (248)	2731 (1002)	101,443,338 (5,232,294)
Truncated octahedron	14	556 (343)	3071 (1466)	68,106,377(5,704,138)

A few enumerative results from: Johnson, Menon (2015).

Part 2. Biological inspiration:

Structure and self-assembly of viruses,
and the story of MS2.

Viruses

Viruses are the most populous (10^{31}) and genetically diverse organisms (10^7 genotypes) on earth. However, they lack the biosynthetic machinery for independent existence.

The "simplest" viruses consist of a genome (RNA or DNA) contained within a protein shell (the capsid).

The (idealized) reproductive cycle

(1) Capsid disassembles when virus infects a host;

(2) Genome hijacks host's machinery to make new genome and protein;

(3) Capsid reassembles and packages the genome.

The elegant natural design of viruses

(1) Genetic economy.

(2) Structural symmetry.

(1): NIH-NCBI lists the complete genome of approximately 4000 viruses. Many of these have very short genomes. For example, ss-RNA viruses often have genomes with only 1000 to 10000 base pairs.

(2) In the mid-1950s' crystallography revealed that many viral capsids have icosahedral symmetry. As Caspar and Klug realized, structural symmetry is well matched with genetic economy -- a few basic units can go a long way...
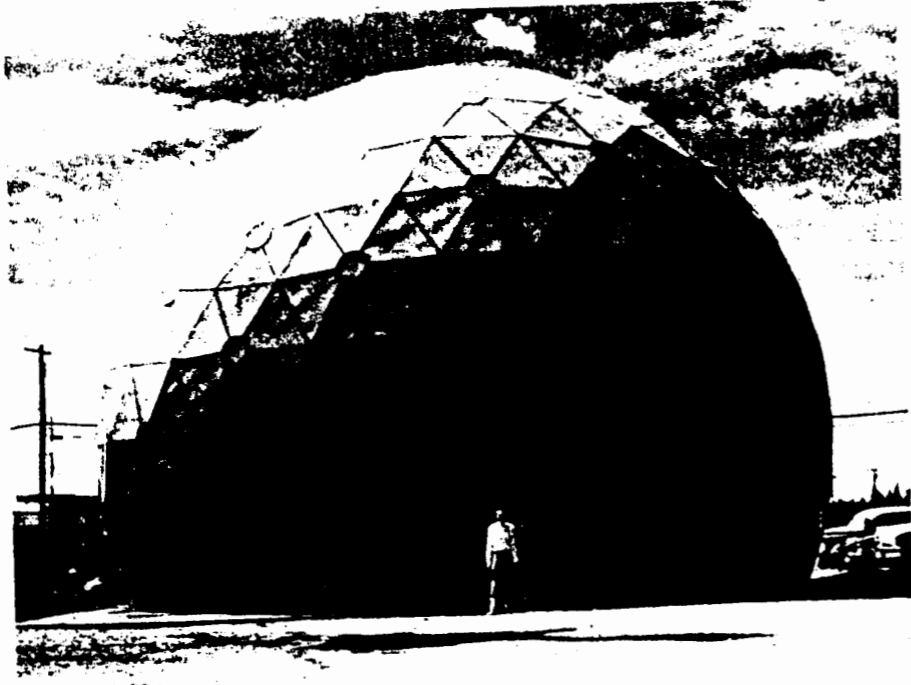
FIGURE 5. A Fuller geodesic dome (Radome designed by Geometrics, Inc., Cambridge, Mass.). Note that the surface is made up of quasi-equivalent triangles and that these are grouped in hexamers and pentamers about the small rings of the dome. (Photograph supplied by W. H. Wainwright).
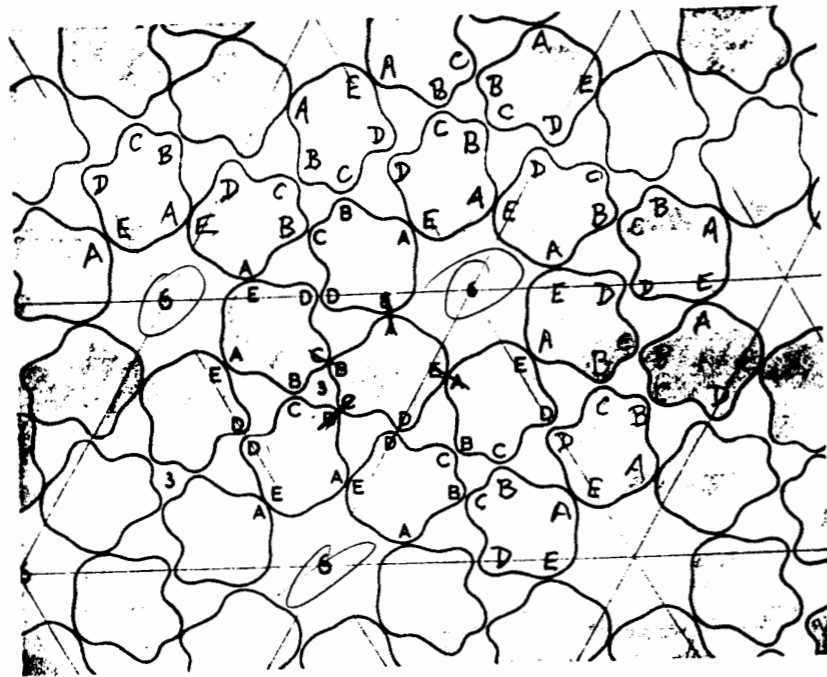

FIGURE 10. Models representing the arrangement of subunits in the icosahedral shell T = 4 (see Fig. 9 for details).
(a) Each subunit is represented by a wooden peg. There are 3 per face of the underlying deltahedron (Fig. 8b), making 240 in all.
The local 3-fold axes relating the 3 subunits per deltahedron face are shown by the black triangles.
(b) Subunits represented by pieces of rubber tubing and clustered into 30 hexamers and 12 pentamers to give 42 morphological units.
(c) Appearance of (b) at low resolution.

FIGURE 6. Asymmetric units arrayed in an equilateral-triangular plane net. Besides having translations, here a and a, the lattice has 6-fold rotational axes of symmetry. Although the asymmetric units are in 6 different orientations in space, they are all exactly equivalently related.
Each unit here is equipped with five "bond" sites, A, B, C, D, and E, forming three different "bonds", namely a hexamer bond AE, a trimer bond BC, and a dimer bond DD. (Note that only two of these bonds are absolutely essential for coherence of the array.)
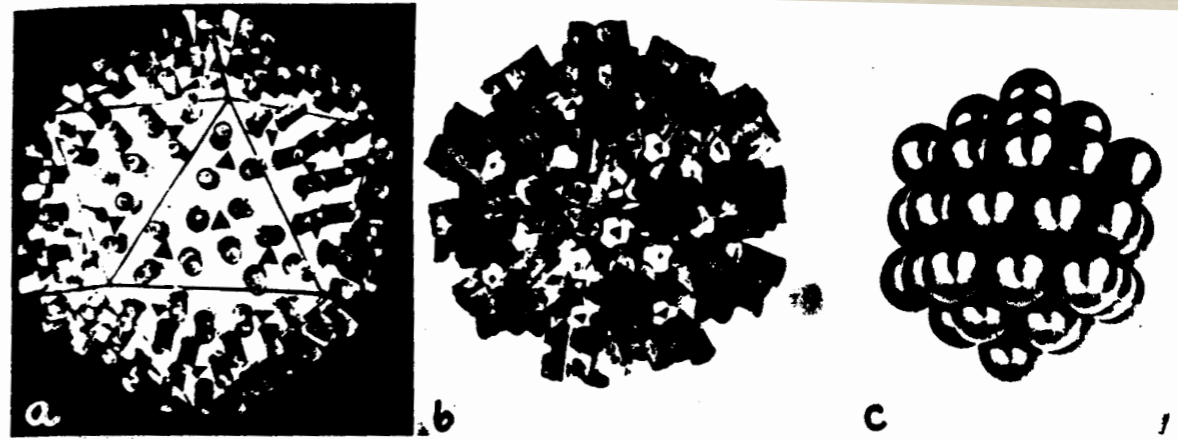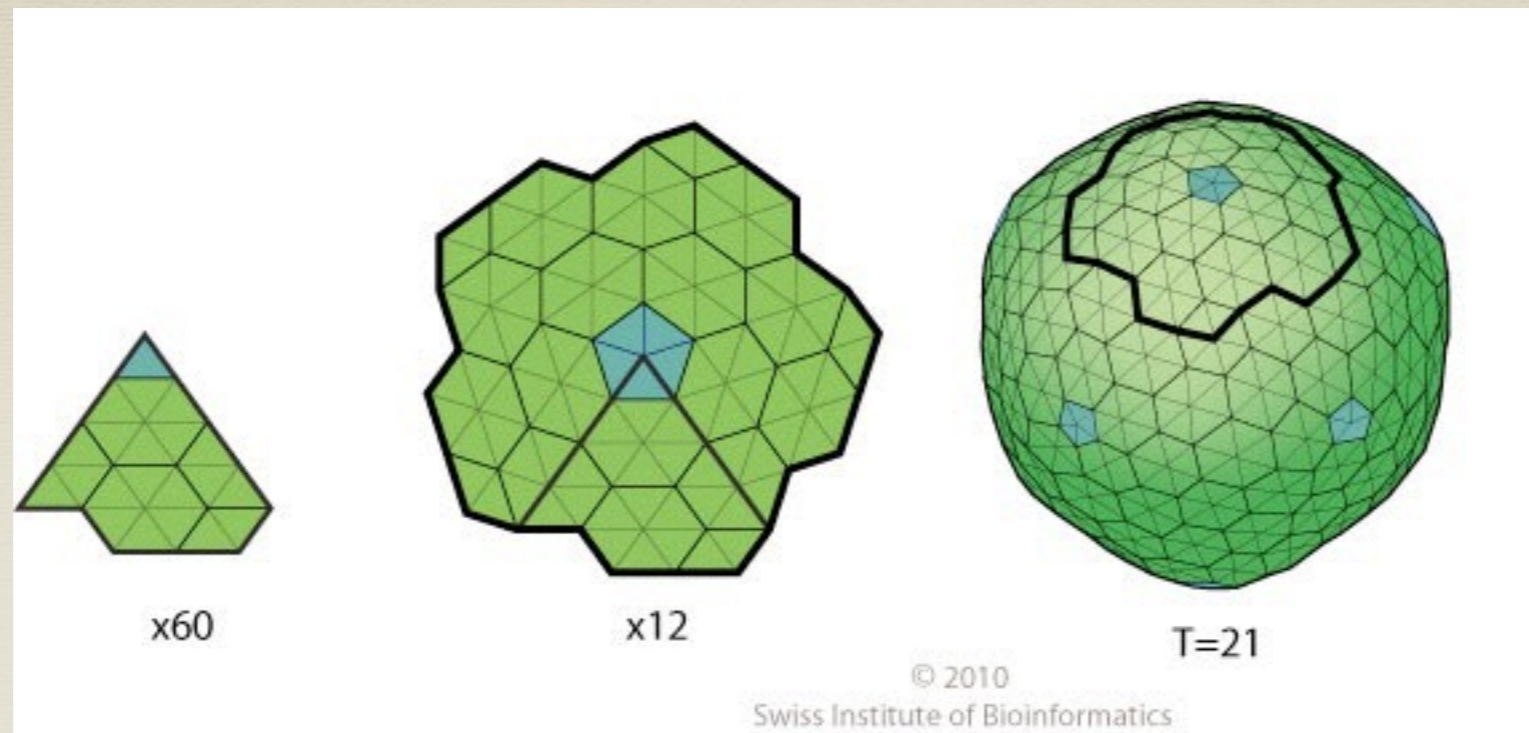
11

Geodesic domes, planar tilings, and models of an icosahedral virus.
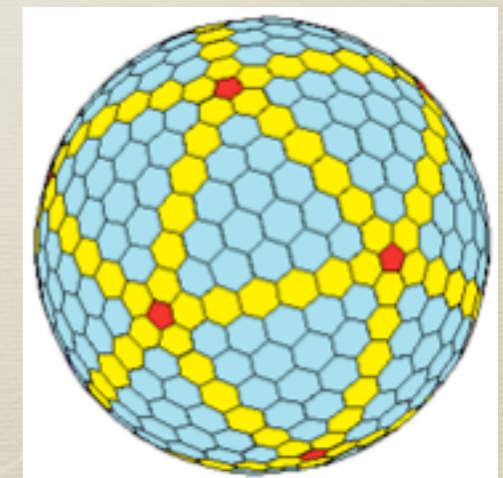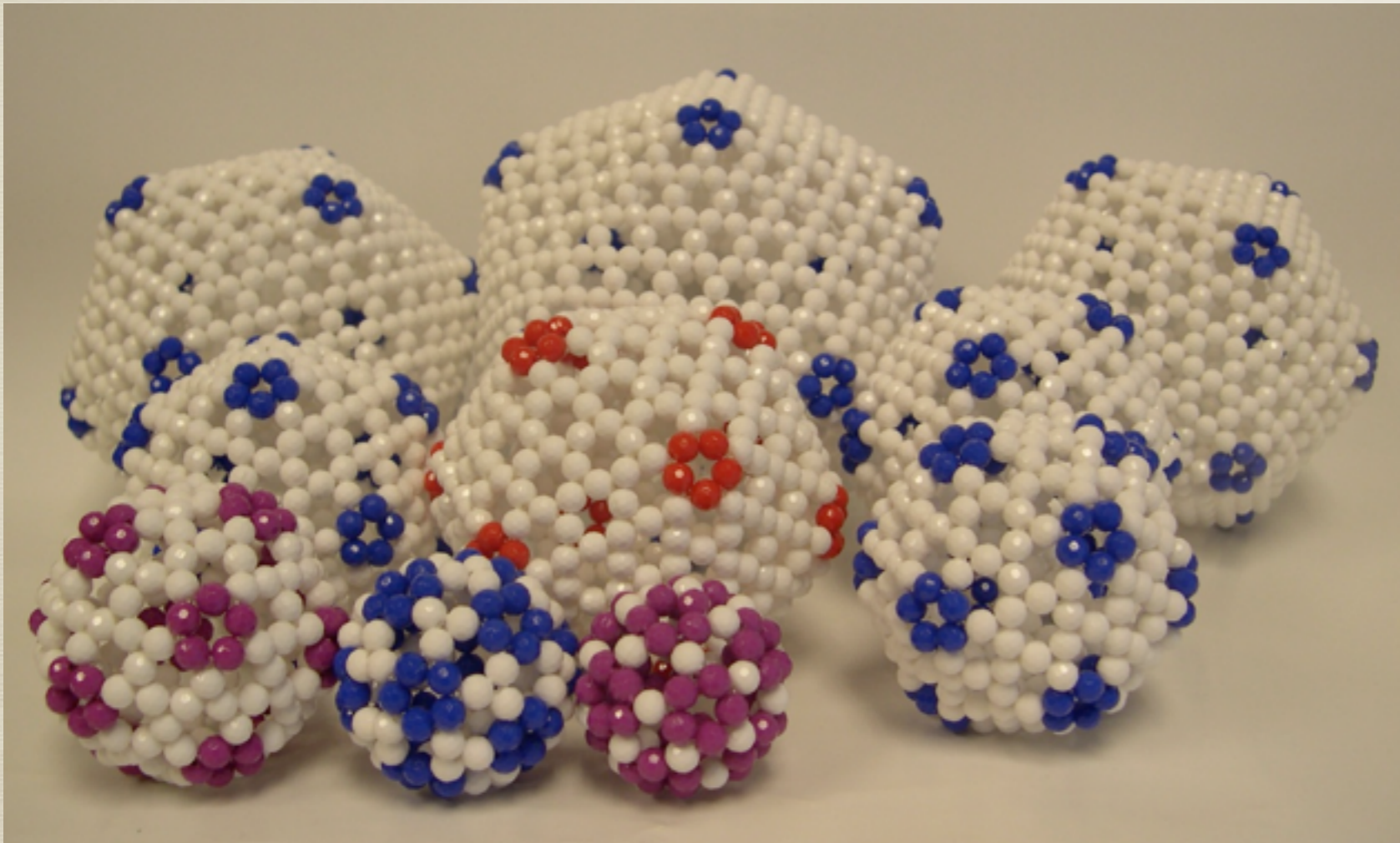
Caspar, Klug (1962)

The Caspar-Klug "quantization"

The number of subunits in the viral capsid must be 60T where

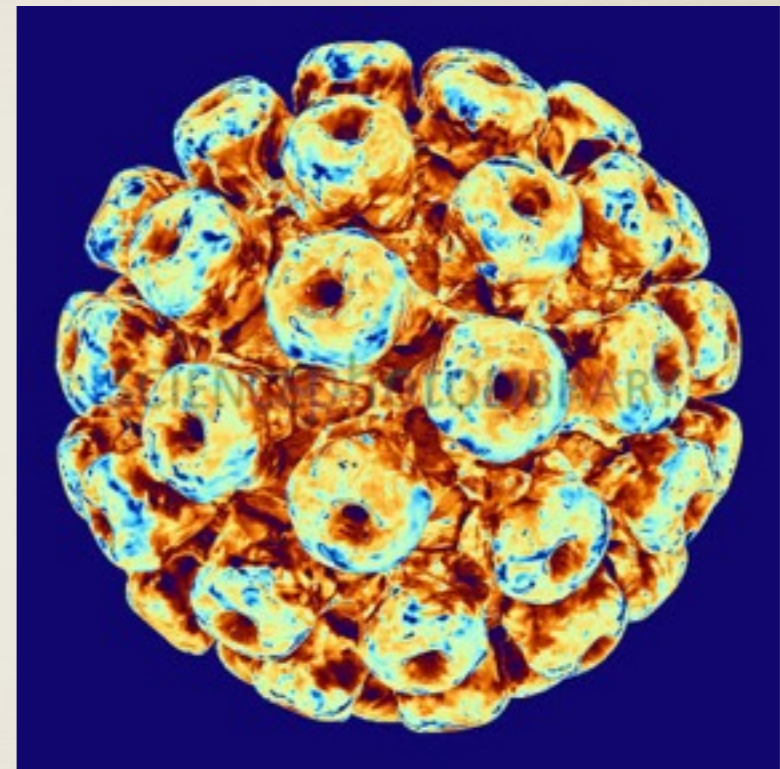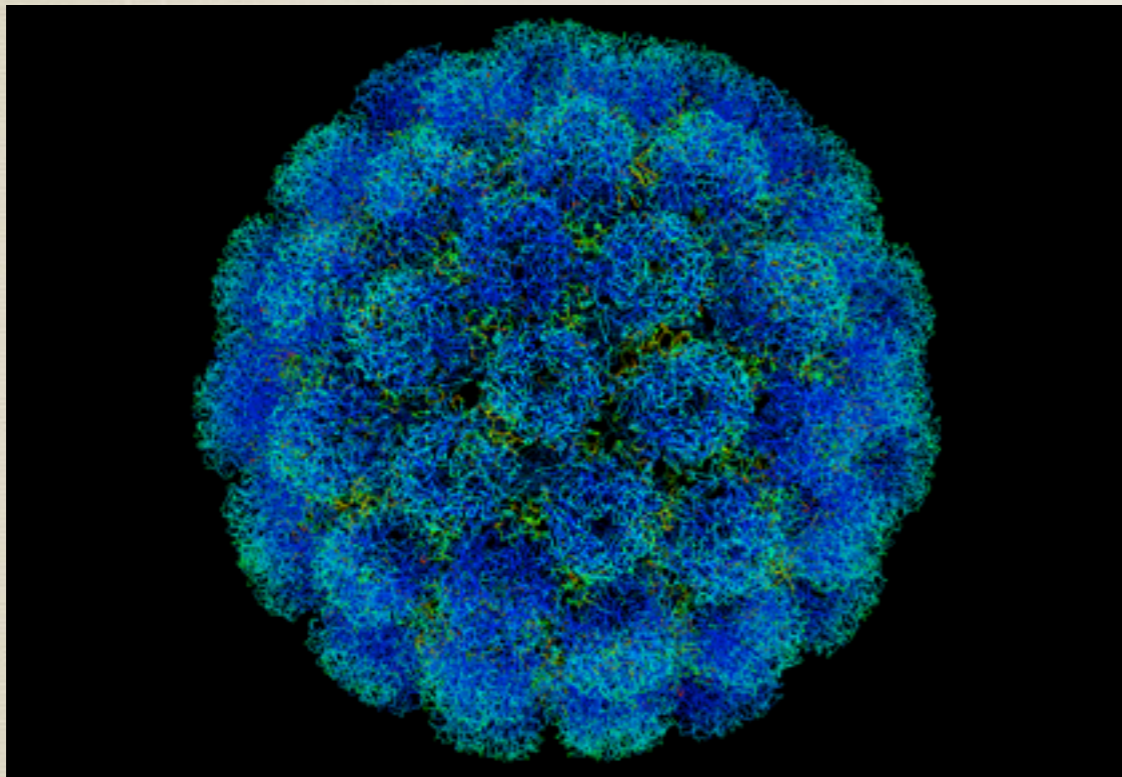$$T = h^2 + hk + k^2, \quad h, k = 0, 1, 2, \ldots$$

This restriction had been found by Goldberg in 1937 in his study of discrete isoperimetric problems.

Caspar-Klug virus= Goldberg polyhedron.

"Forbidden" structures

T=6 is impossible, thus a virus structure based on Caspar-Klug theory cannot have 360 subunits. But such virus structures were found in 1991.
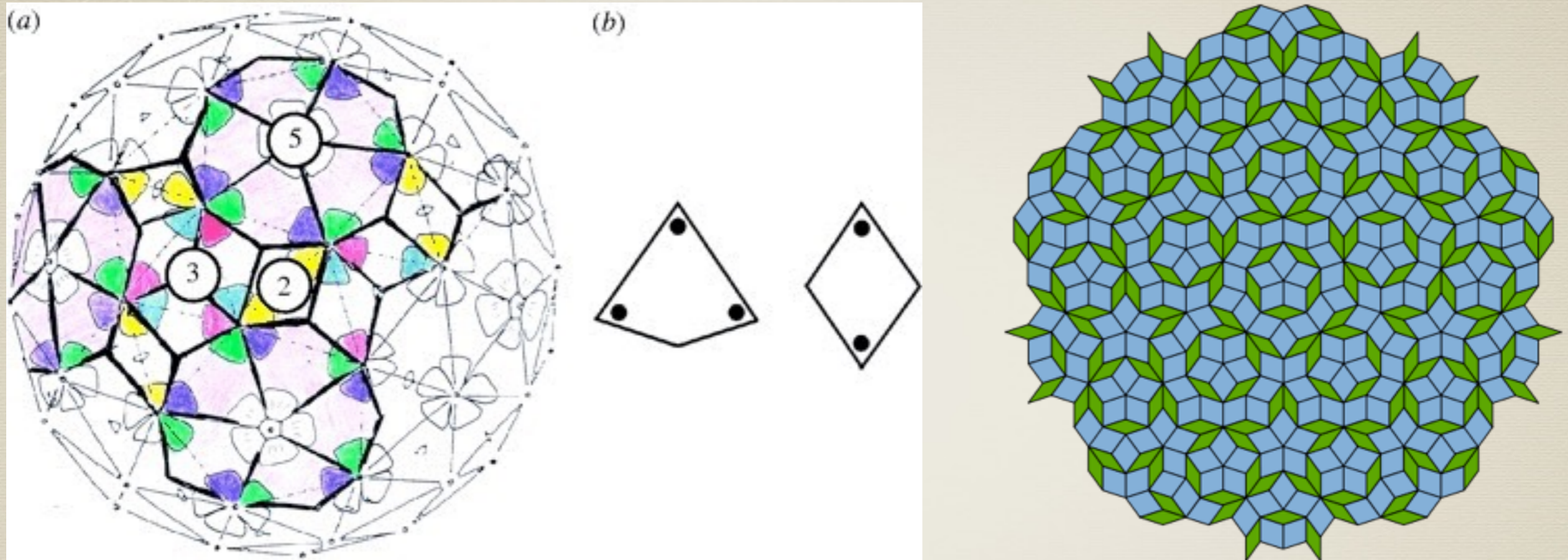

Both SV40 and polyomavirus have 72 pentamers, thus 360 subunits.

Liddington et al (Nature , Nov. 28, 1991).

Aperiodic tilings and SV40

The Sv40 structure is simple and beautiful (Twarock, 2004):

Certain aperiodic tilings (e.g. the Penrose tiling) can be "folded" into a closed polyhedron, analogous to the Goldberg polyhedra.

Thus, SV40 discovered quasicrystals two billion years before us!

The story of MS2

MS2 is an icosahedral virus with a single-stranded RNA. It infects the bacteria e.coli, and other enterobacteria. The MS2 genome was the first to be completely sequenced (Fiers 1972-1976).

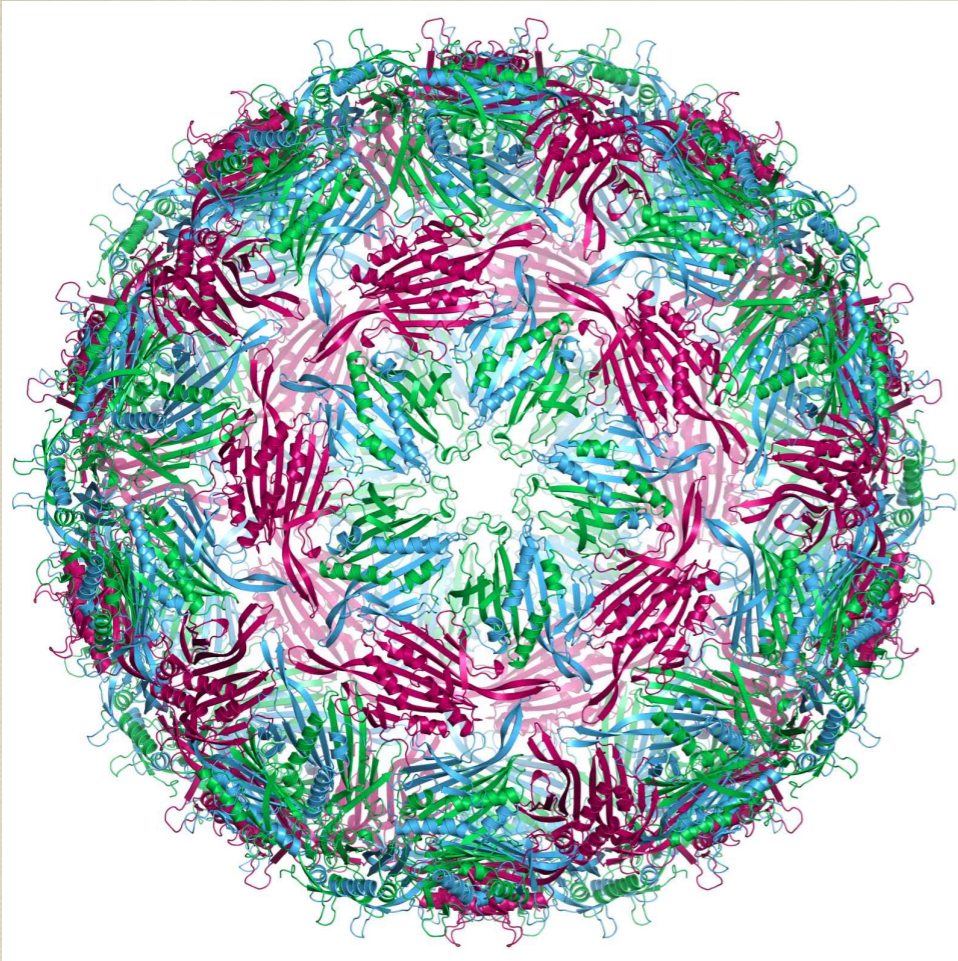The genome has 3569 nucleotides that code for four proteins:

(1) coat protein; (2) maturation protein; (3) lysis enzyme; (4) replicase enzyme.

The MS2 genome
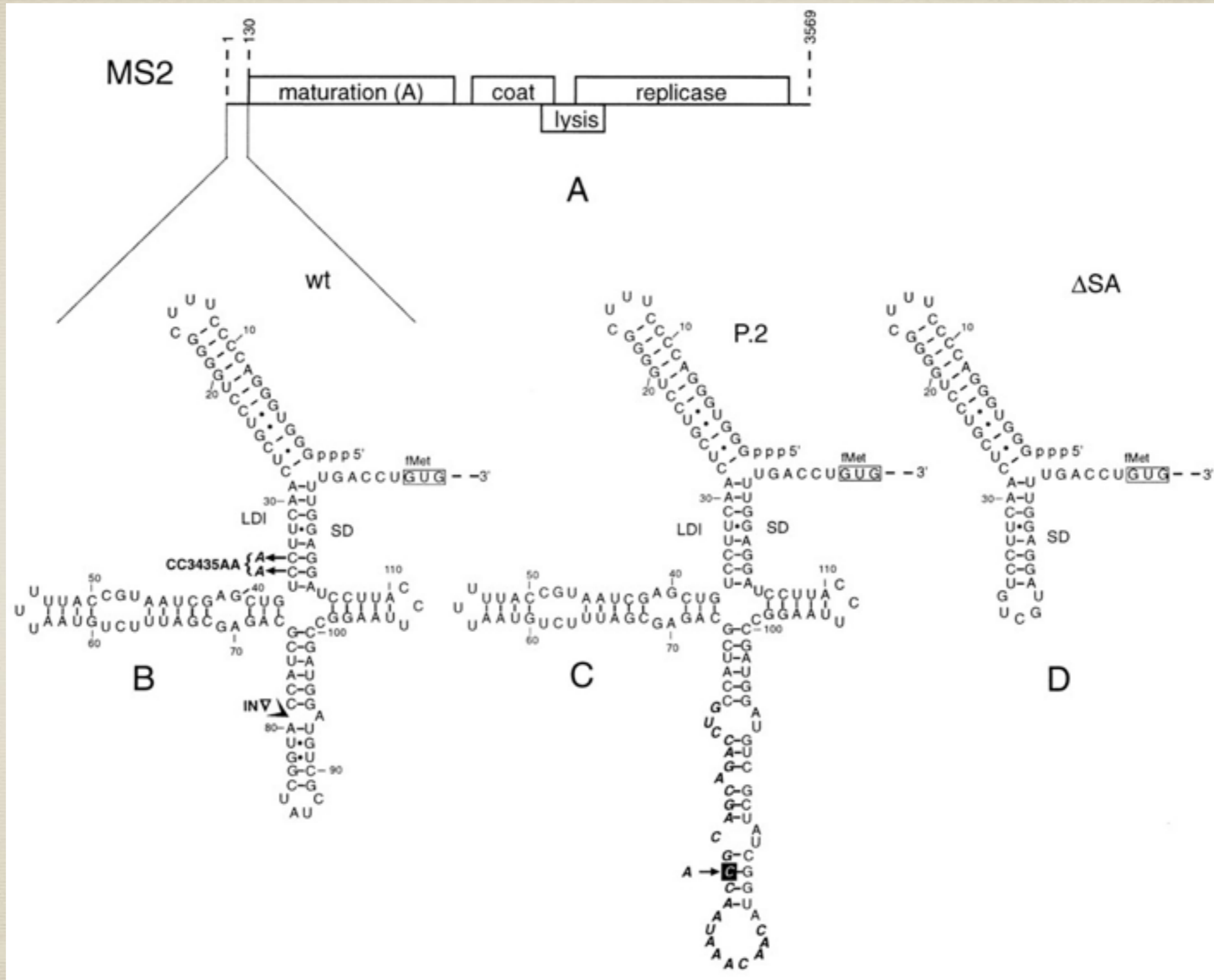
The equilibrium structure of MS2
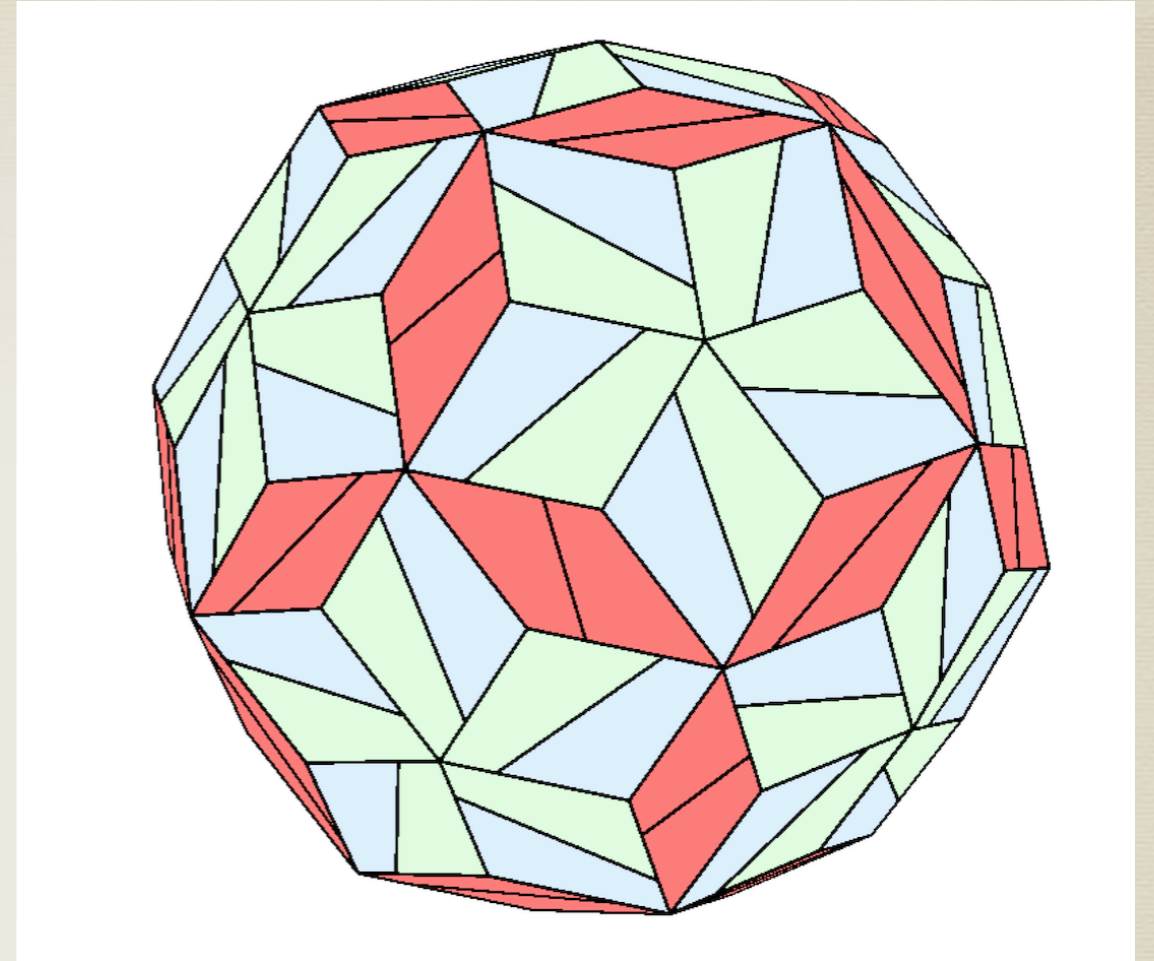
The capsid consists of 180 copies of the coat protein. However, this protein exists in three configurations (A,B,C) which bind into two dimers (A/B and C/C).

The switch from one dimer to another is triggered by an RNA hairpin loop.

The early dogma on the process of self-assembly

Self-assembly (of a virus) is a process akin to crystallization and is governed by the laws of statistical mechanics. The protein subunits and the nucleic acid chain spontaneously come together to form a simple virus particle because this is their lowest free energy state ."

Caspar and Klug; Cold Spring Harbor Symposium, (1962)

The dogma soon unraveled....

(1) The time taken to reach equilibrium is too long (this is essentially the same as Levinthal's paradox for protein folding).

(2) Klug showed that RNA-driven conformation changes drive the self-assembly of Tobacco Mosaic Virus (1971). Thus, assembly is not thermodynamic, it is nucleated by RNA.

There have been many investigations of the assembly of viruses, but sequence-specific studies are very recent and use discrete geometry in an essential way.

Pariacoto virus
PDB: 1f8v
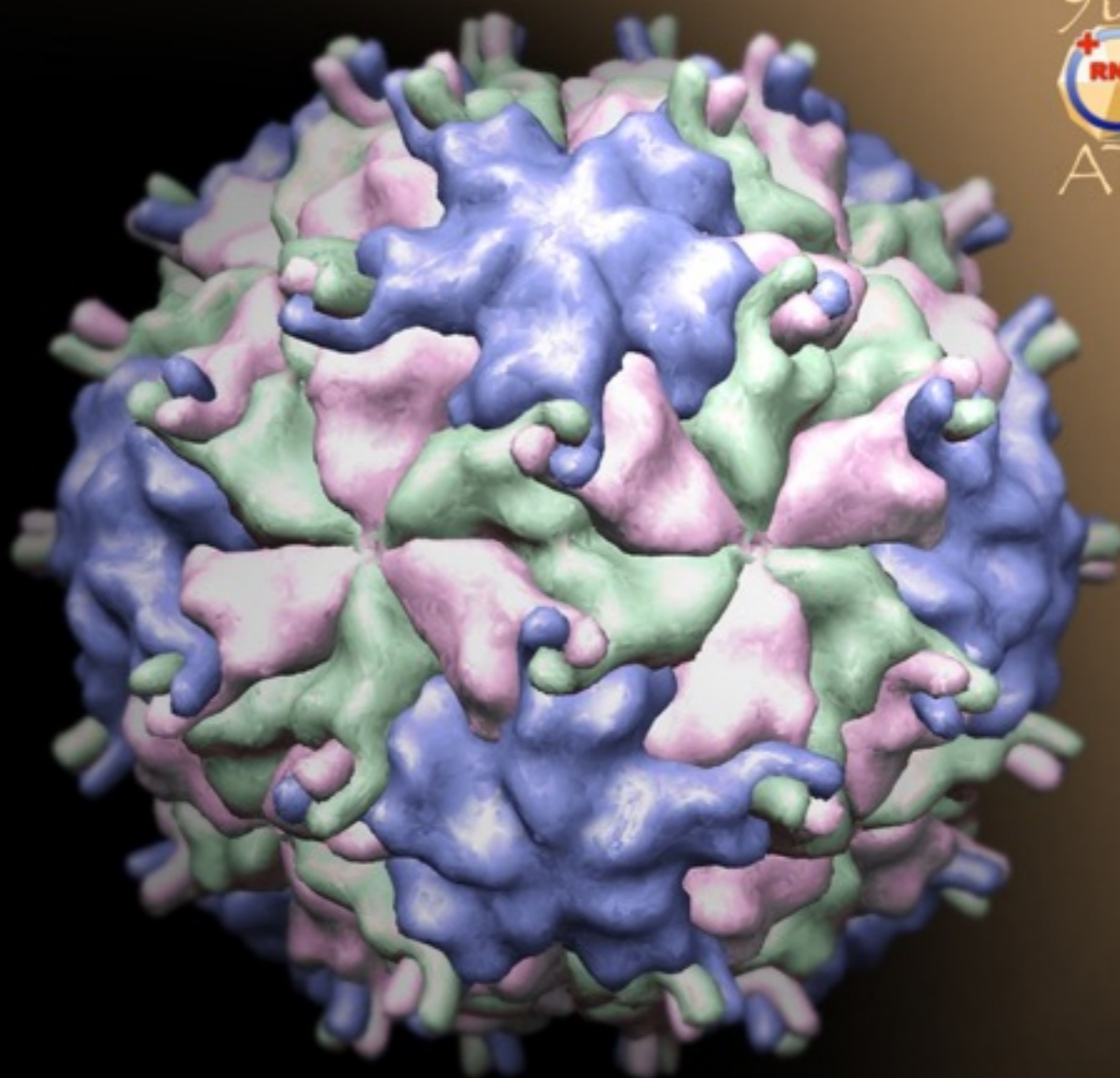
The role of RNA folding
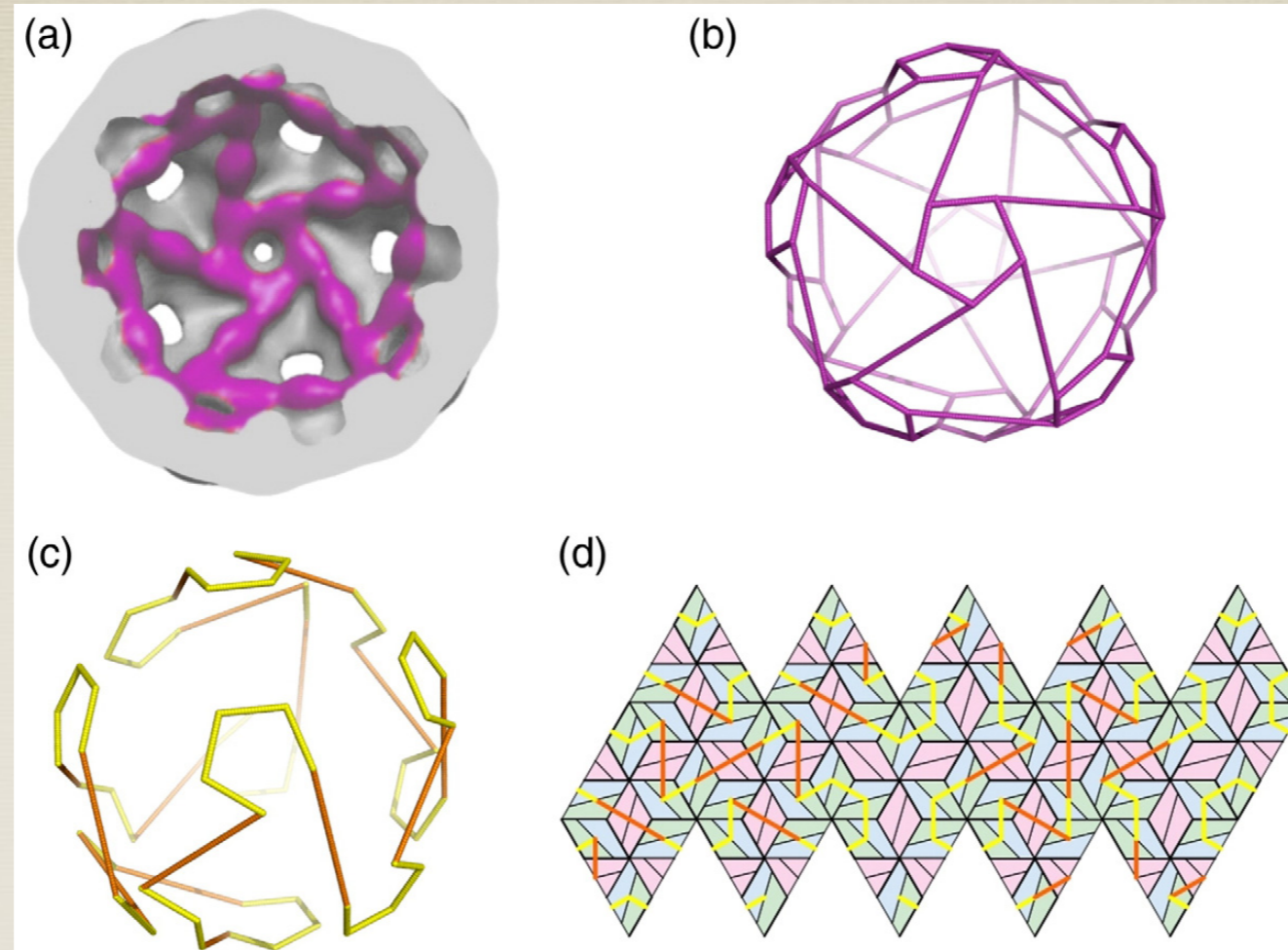
Fig. 1. The polyhedral cage of MS2 RNA density can be described as a Hamiltonian path. (a) A cryo-EM reconstruction of the outer RNA shell of bacteriophage MS2 (depicted in magenta) based on an image at ~17 Å resolution adapted from Van den Worm *et al.*[7] (b) A representation of this RNA shell as a polyhedral cage. (c) A three-dimensional view of a single Hamiltonian path, which meets every vertex of the polyhedron exactly once by moving along the short (yellow) and long (orange) edges of the polyhedral cage. (d) A planar net representation of the Hamiltonian path shown in (c) and its relation to the A (blue), B (green), and C (pink) quasi-equivalent subunits of the MS2 capsid.

Twarock's work: use a combinatorial model for RNA-capsid coassembly to locate secondary folding sites for RNA on the MS2 genome.

Dykeman, Stockley, Twarock; PNAS (2015)

Part 3. Self-folding polyhedra: computation and experiment.

computations: Ryan Kaplan, Daniel Johnson, Joe Klobusicky, (Brown)

experiments: Shivi Pandey, David Gracias (Johns Hopkins)
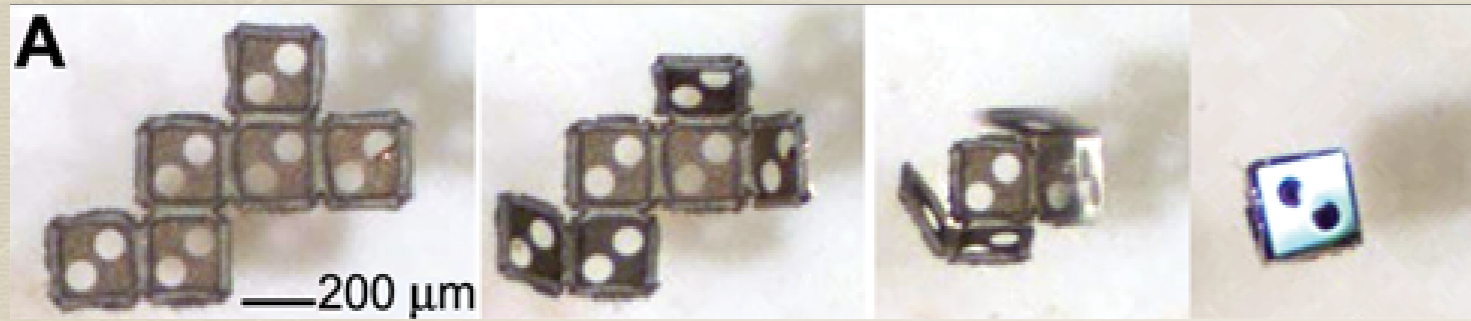
Examples of synthetic polyhedra: micron scale
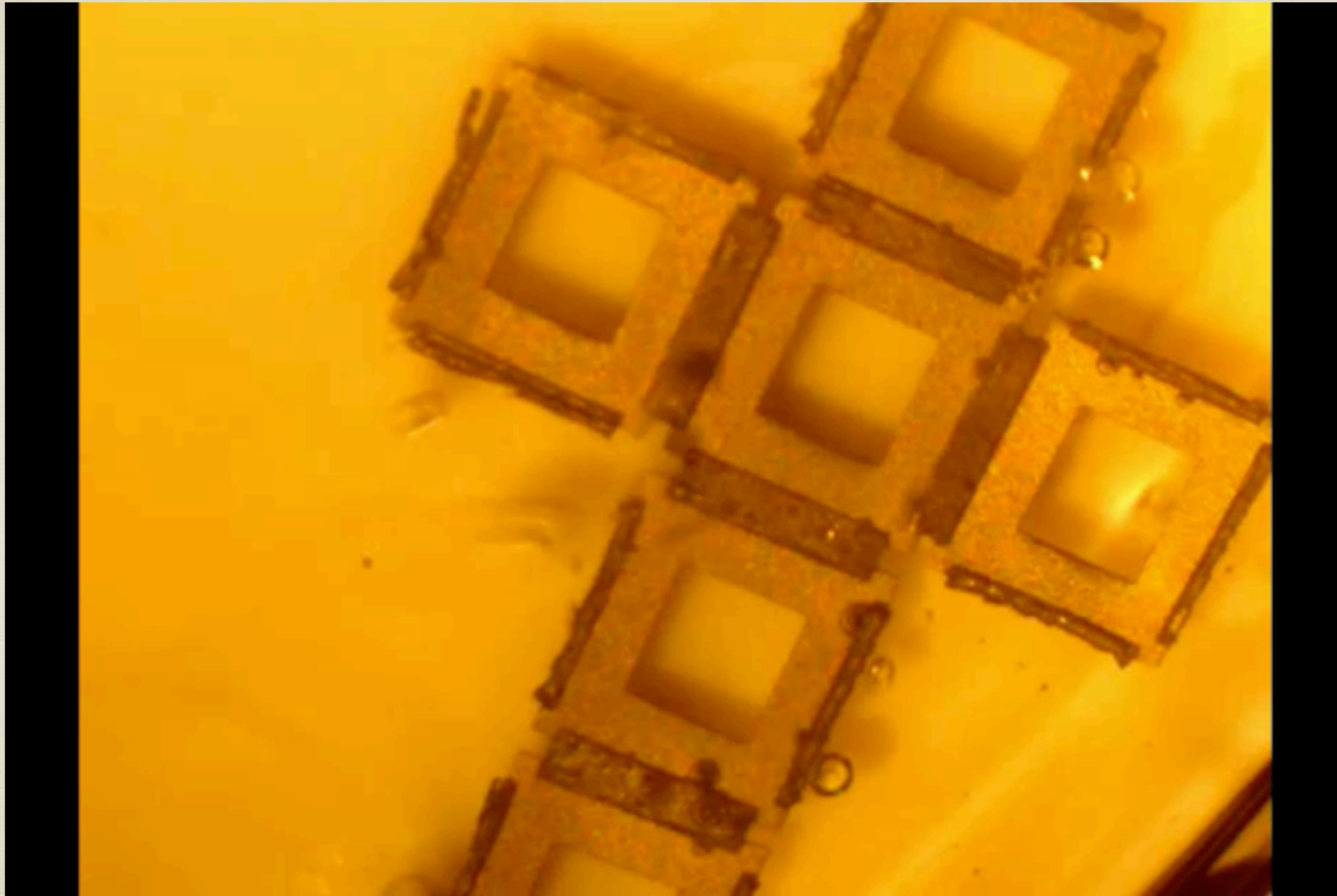
Discovering design principles
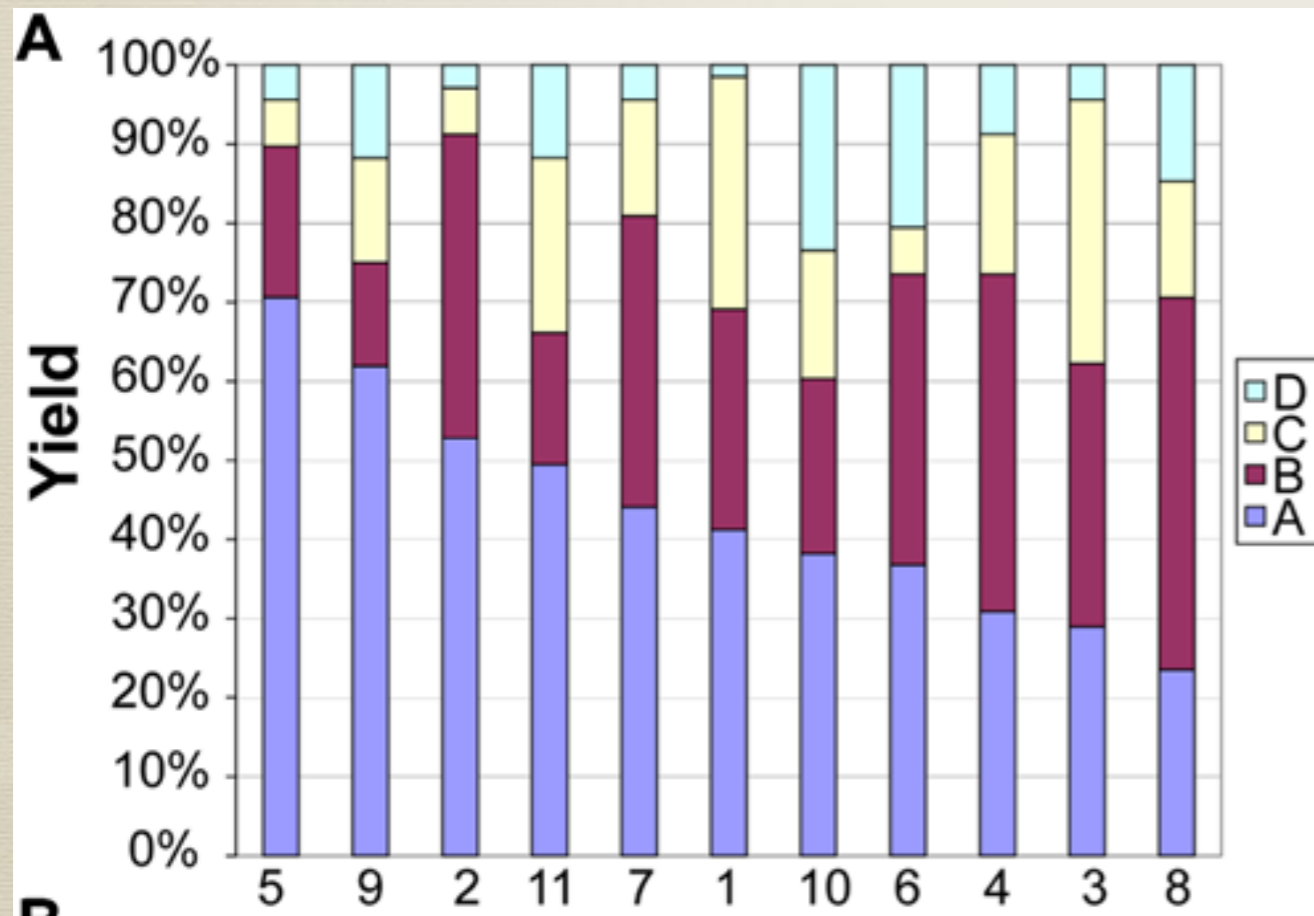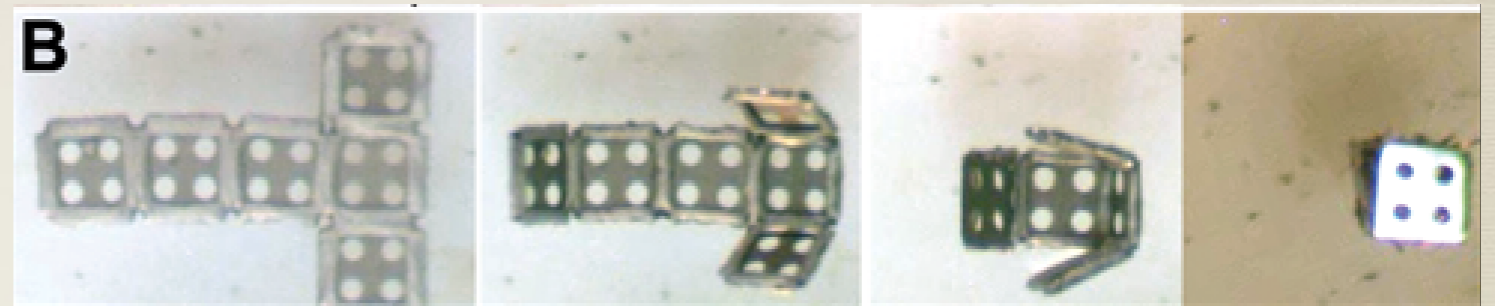
Which net self-folds with the highest yield?
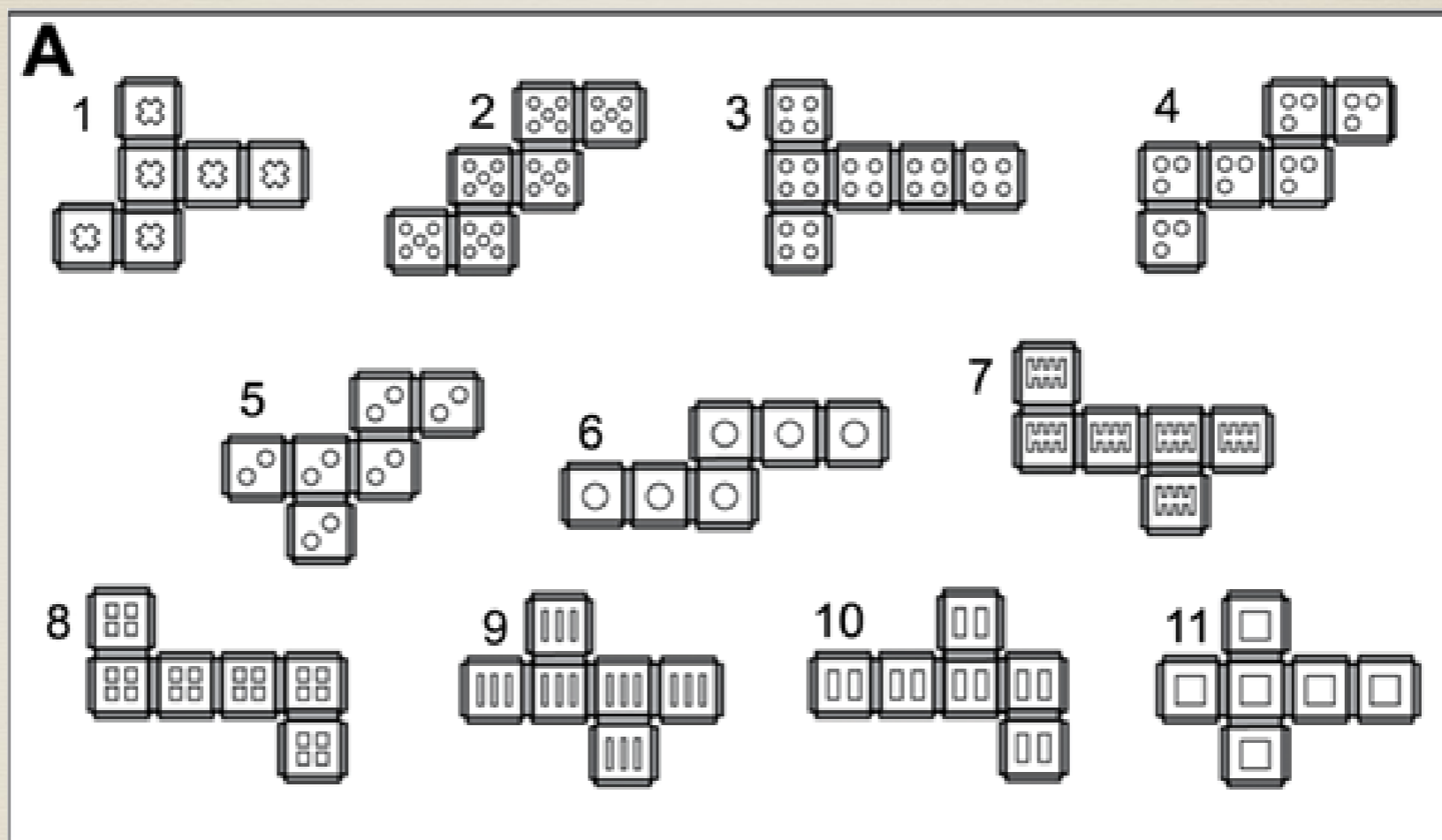
The observed pathways (net 5 vs net 11) and yield

Net 5

Net 11

A: perfectly folded structures.

D: 2 or more faces did not fold.

The combinatorial explosion

Polyhedron	Number of faces	Number of nets
Tetrahedron	4	2
Cube	6	11
Octahedron	8	11
Dodecahedron	12	43,380
Icosahedron	20	43,380
Truncated octahedron	14	2.3×10^{6}
Viral capsid (T=1)	60	10^{30}

What my students did at Brown*

(1) Build models. Convert this to a labelled graph (face-edge, vertex-edge associated to a polyhedron)

(2) Construct a Monte Carlo sampler: fish for spanning trees; convert to a net; check for congruence. Repeat.

(3) Build a database of nets.

(4) Optimization: Find best/worst nets for various cost functions (radius of gyration, vertex connections).

* Maggie Ewing, Drew Kunas, Nghi Nguyen (summer 2010)

Vertex connections and radius of gyration
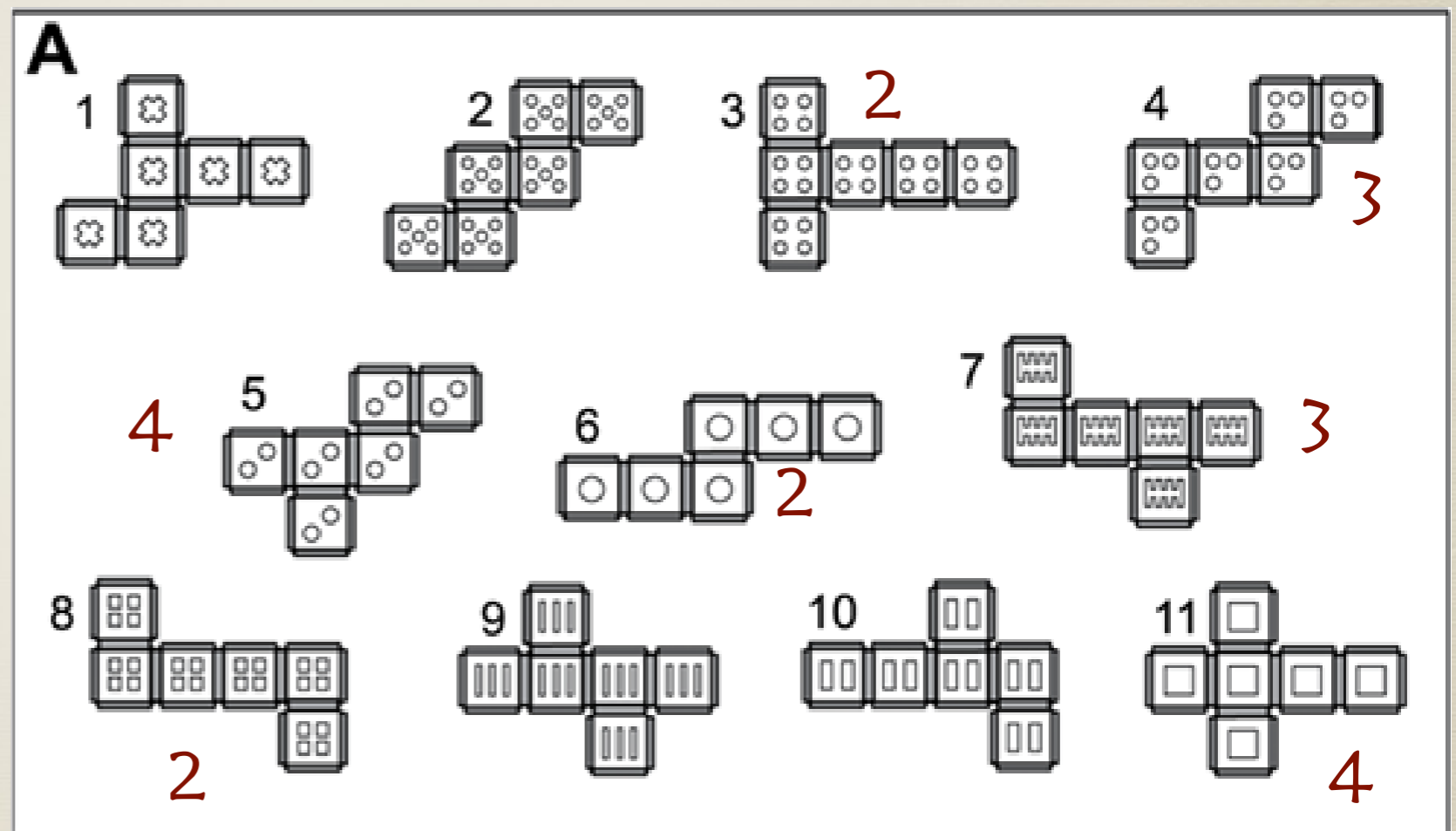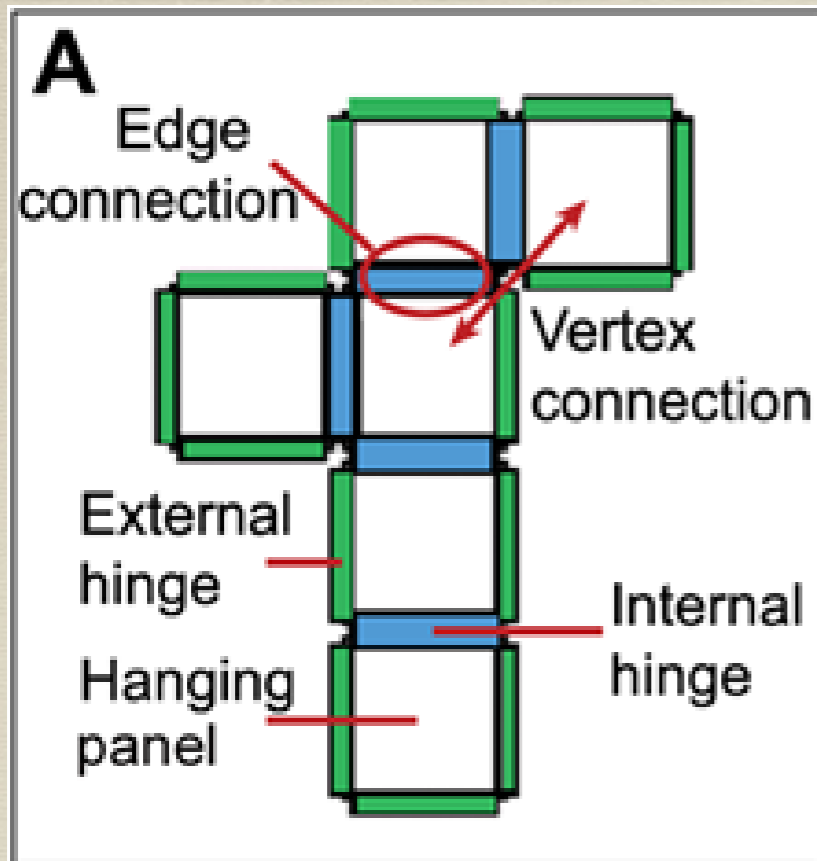
Experiments on self-folding dodecahedra (at Johns Hopkins)

C
B
A
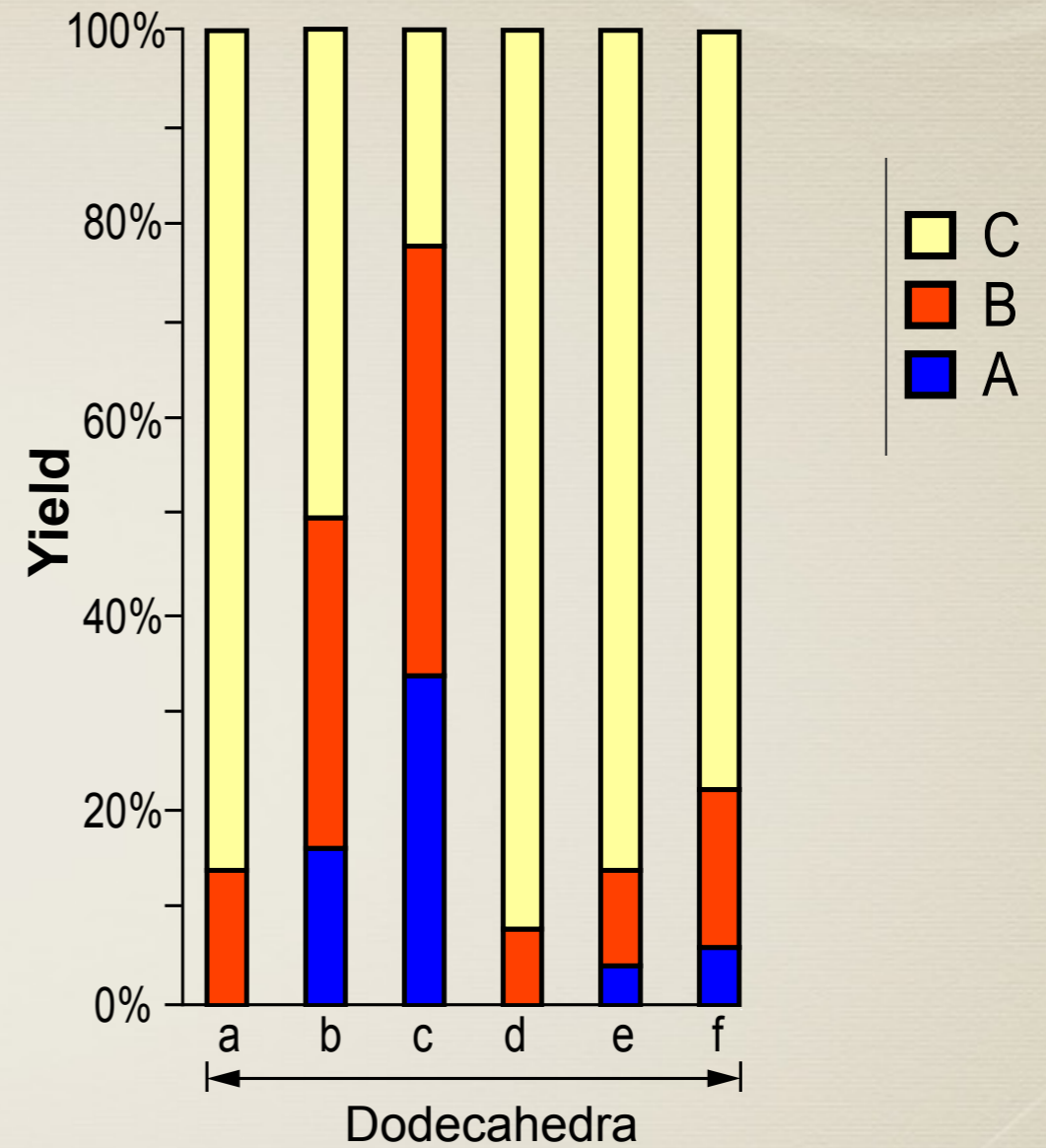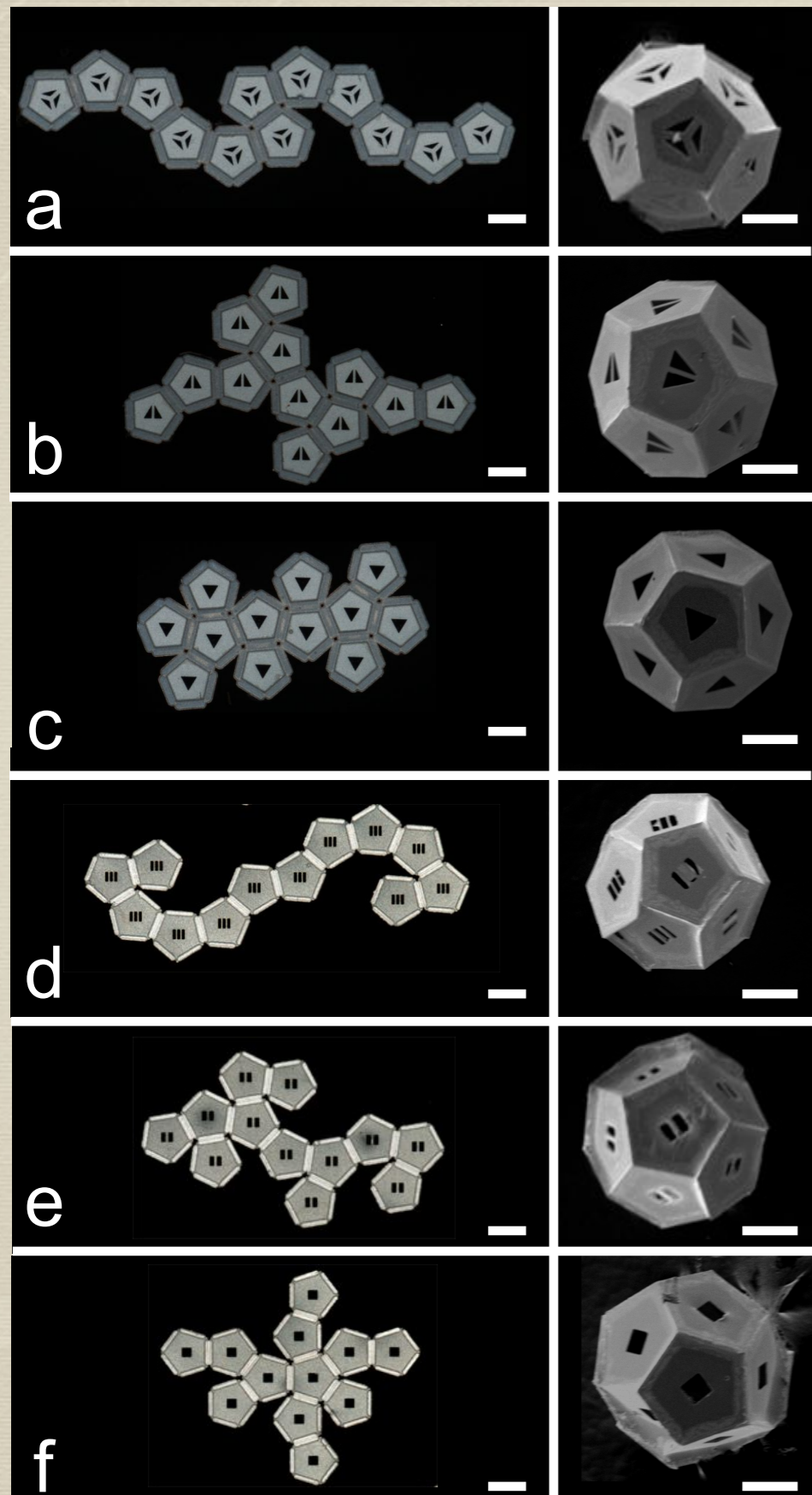
(a)--(c): ranked by vertex connections.

(d)--(f): ranked by radius of gyration

Dodecahedra

The (simple-minded) predictors work.

But the main issue is to understand the process of self-assembly.

The (simple-minded) predictors work. But the main issue is to understand the process of self-assembly. Our model involves:

(a) Construction of discrete configuration space (a graph of intermediates).

(b) Markov chains/optimization on configuration space based on "cooked up" rates and energy functions (only geometry, minimal physics).

(c) Daniel Johnson's thesis (2015): compute rates ab initio based on Brownian motion of polyhedral linkages. Impossible to know rates from experiment for simple molecules (e.g. fullerenes).

H

P
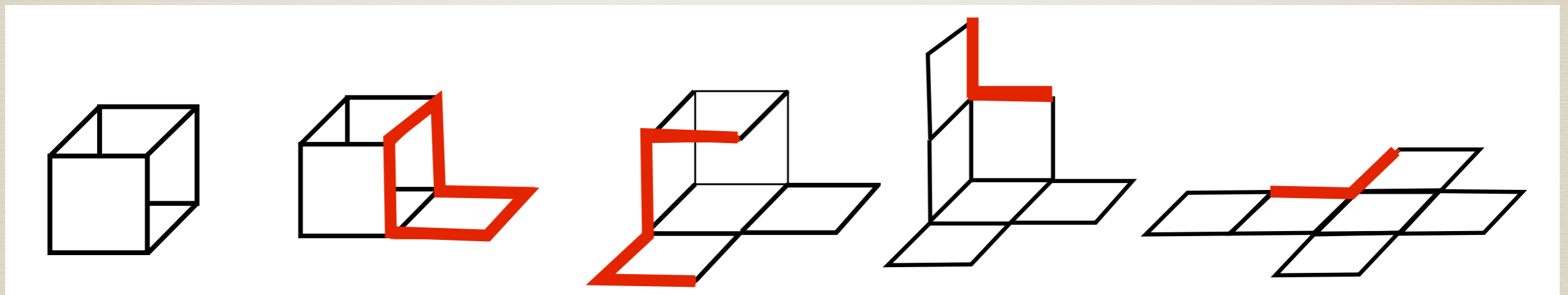
trees on the vertex-edge graph by the removal of edges, or the
transformation of an initial spanning tree on the face-edge graph by
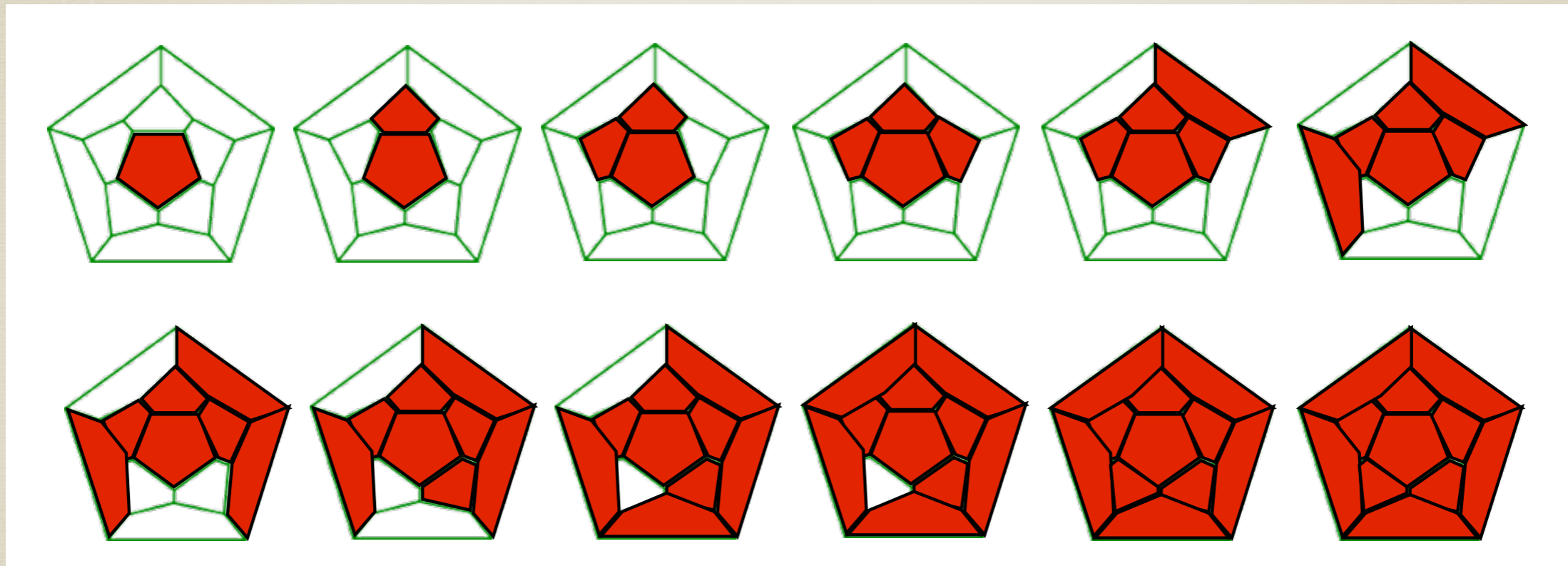the addition of cycles.

Pathways for net 3
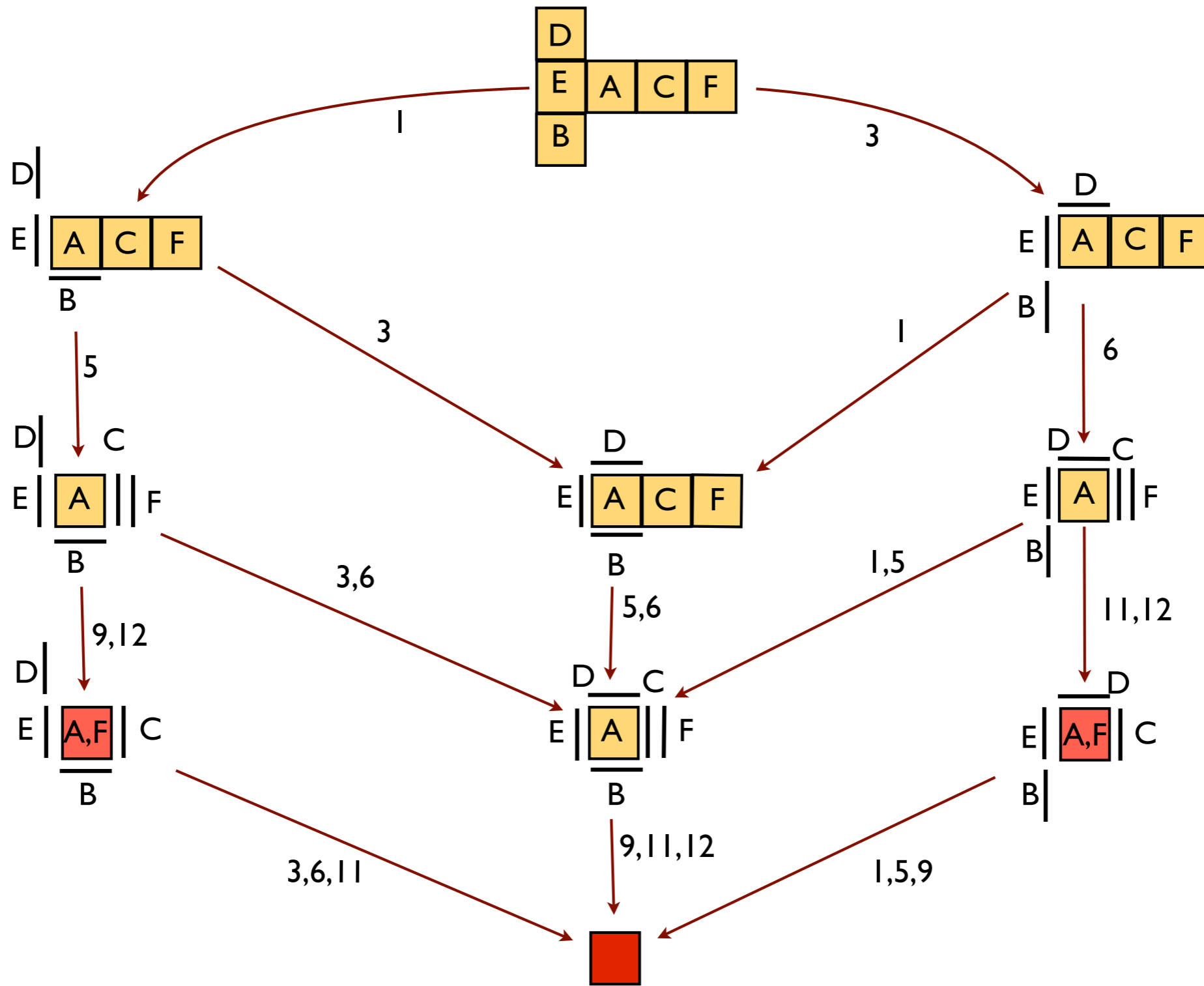
Pathways for net 5
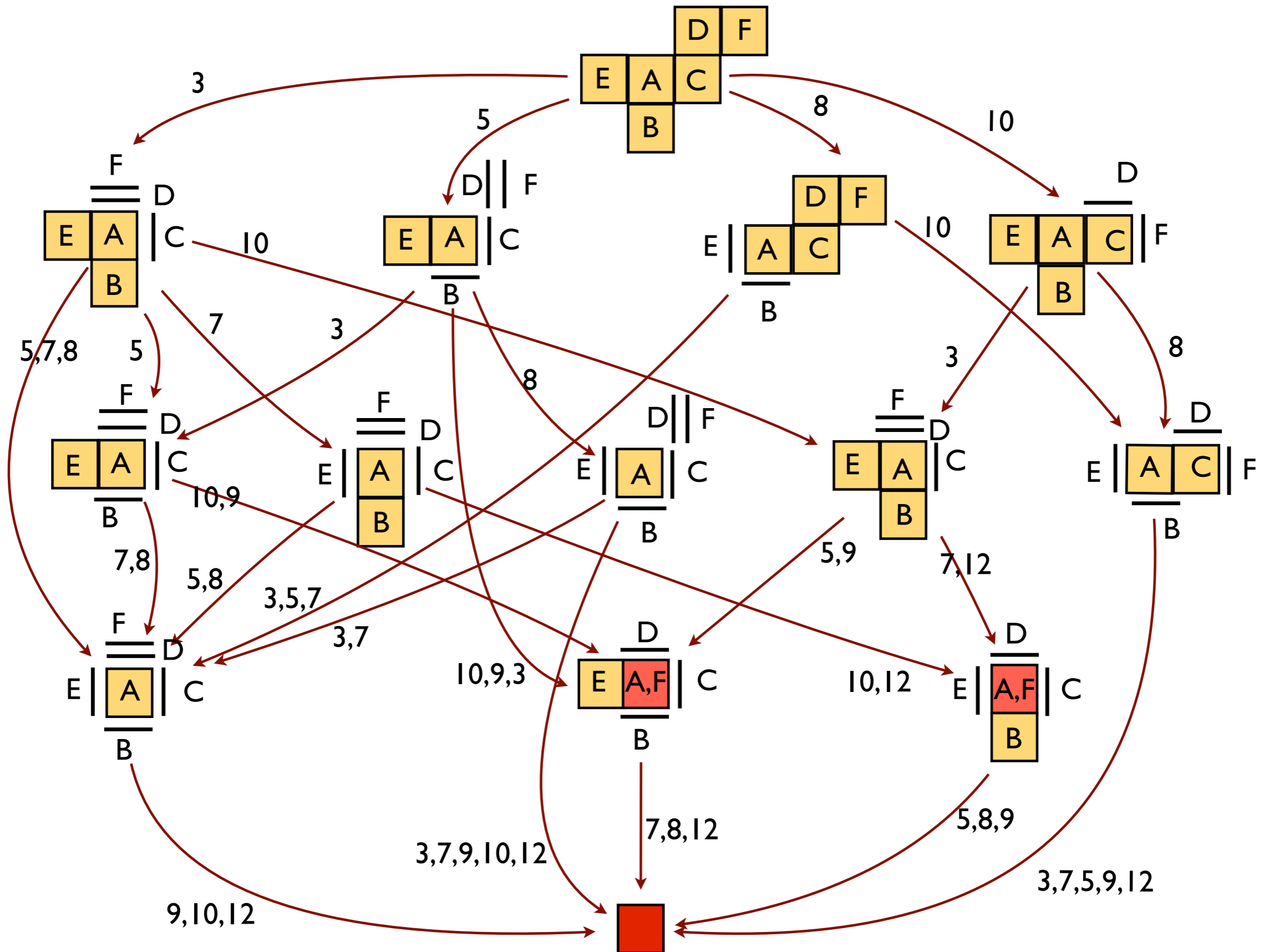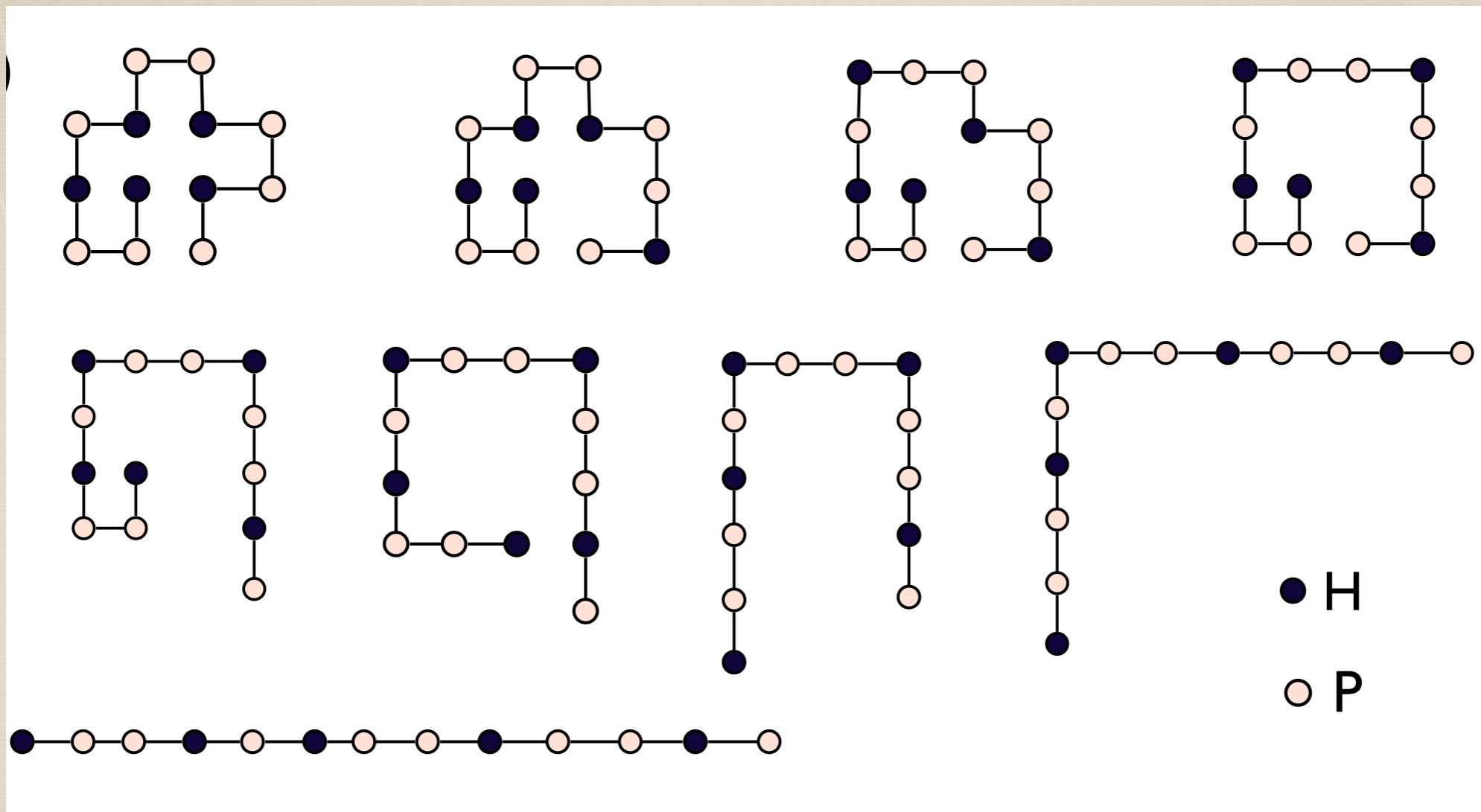
Unfold an HP chain by breaking secondary HH bonds. Initial configuration from Chan and Dill (1997). Crankshaft moves adapted from Verdier and Stockmayer (1962).

● H

(c)

Some interesting mathematical structure

The discrete configuration space explodes combinatorially. However, there seems to be a collapse onto a few dominant intermediates that is robust for various choices of metrics and rates on the configuration space.
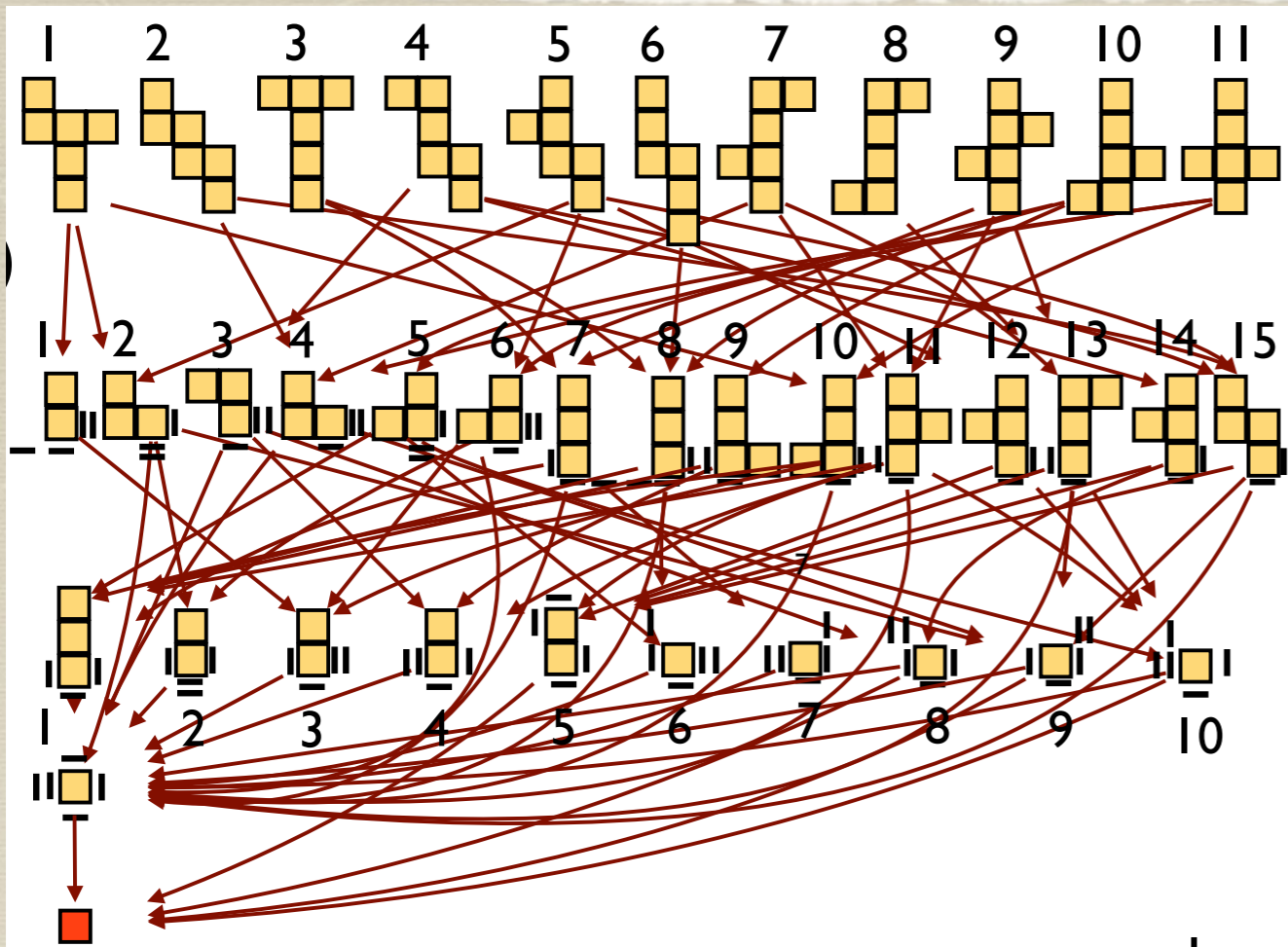
We illustrate this for folding. Similar results have been reported for various models of virus self-assembly.
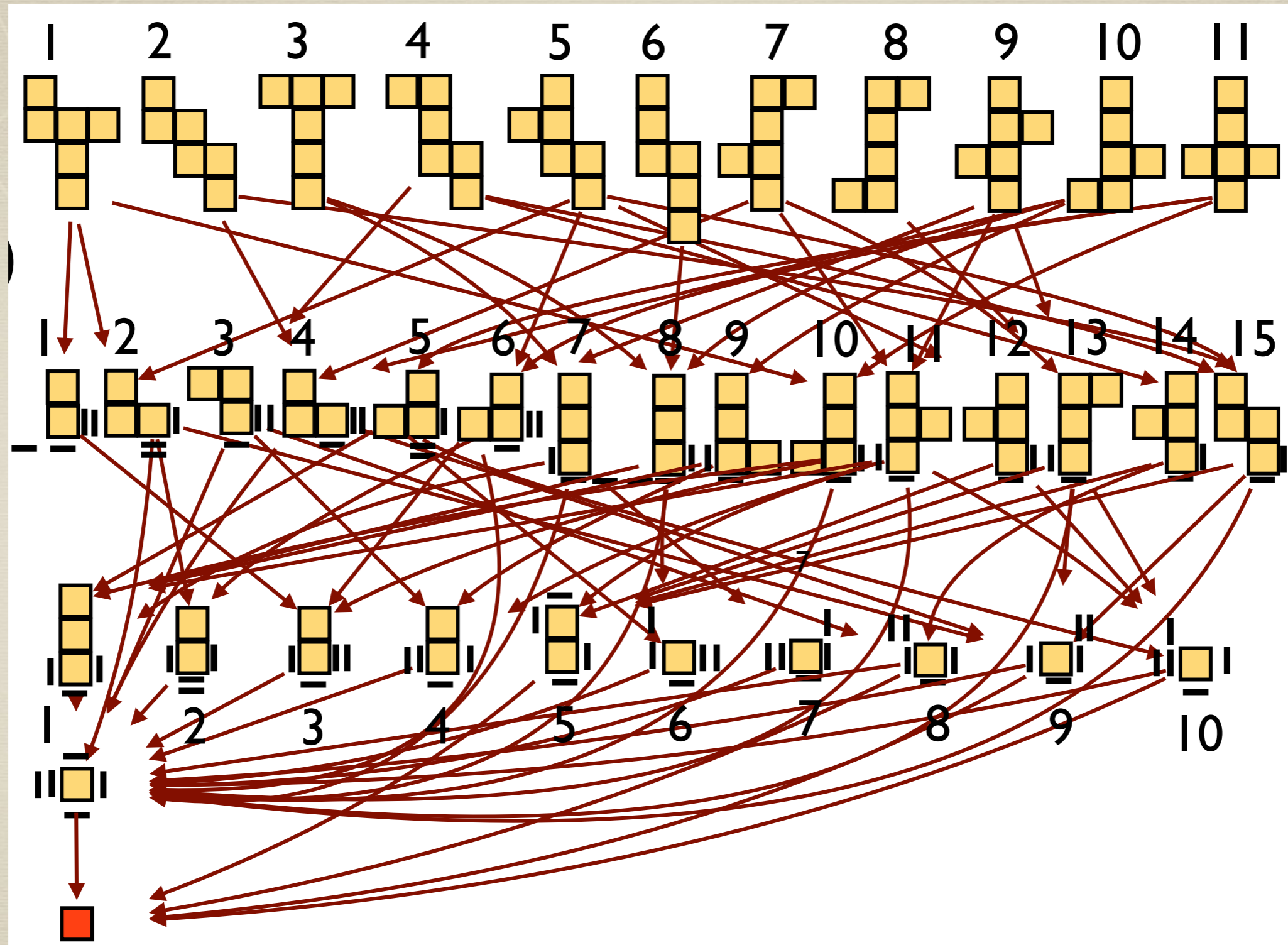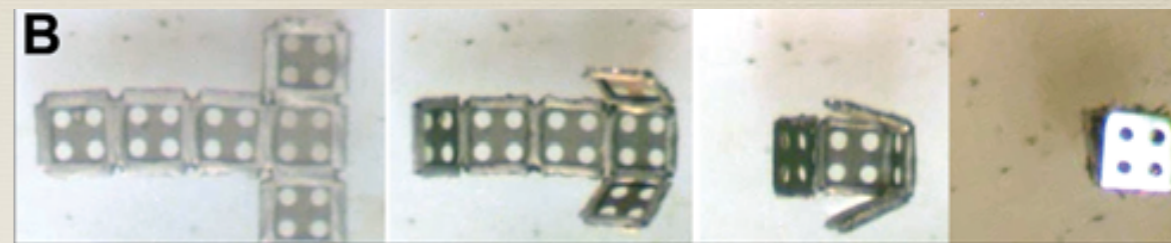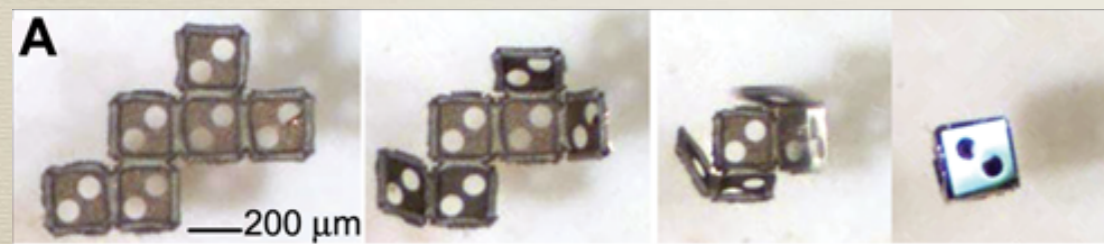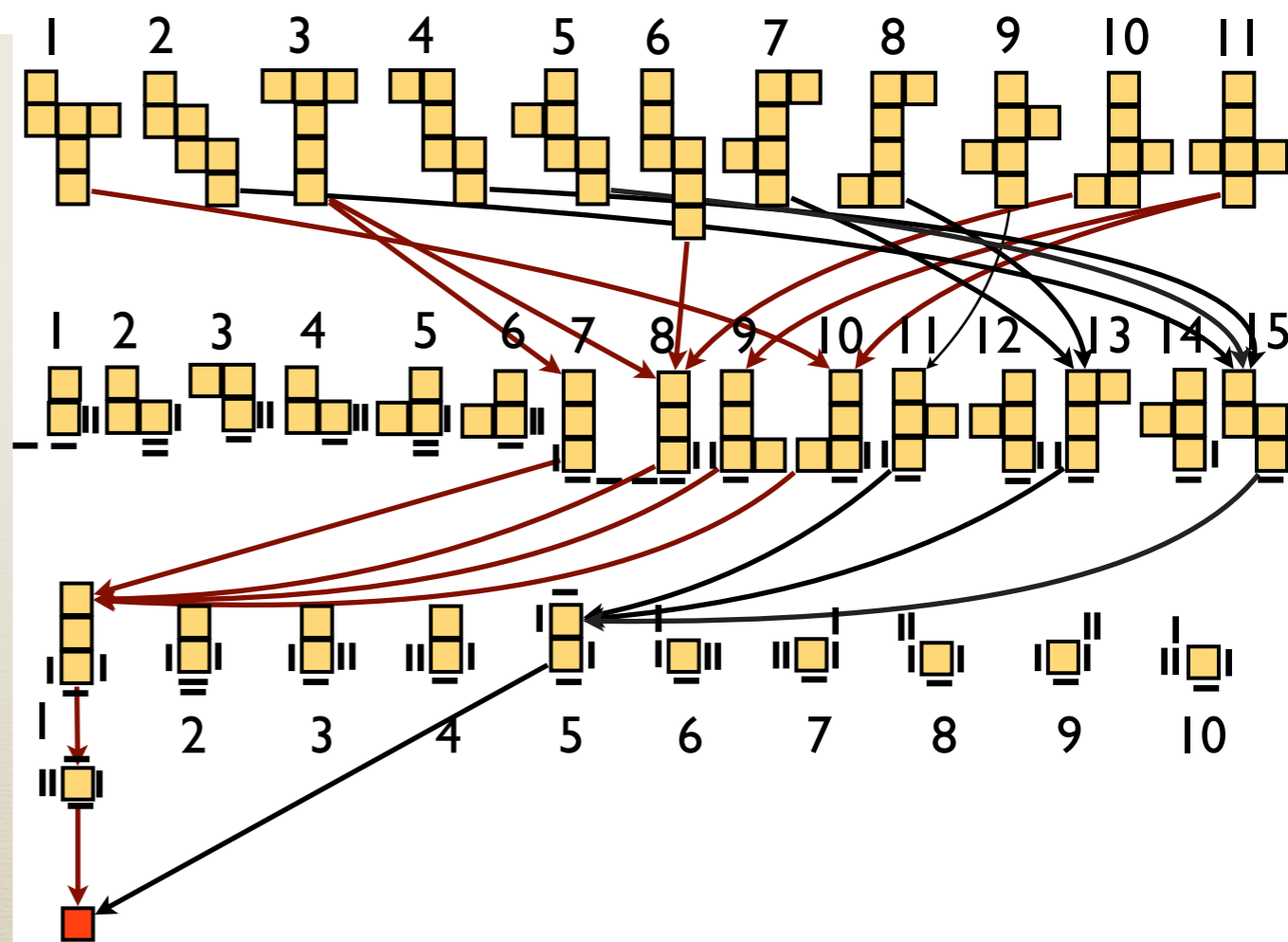
All folding pathways for the cube.
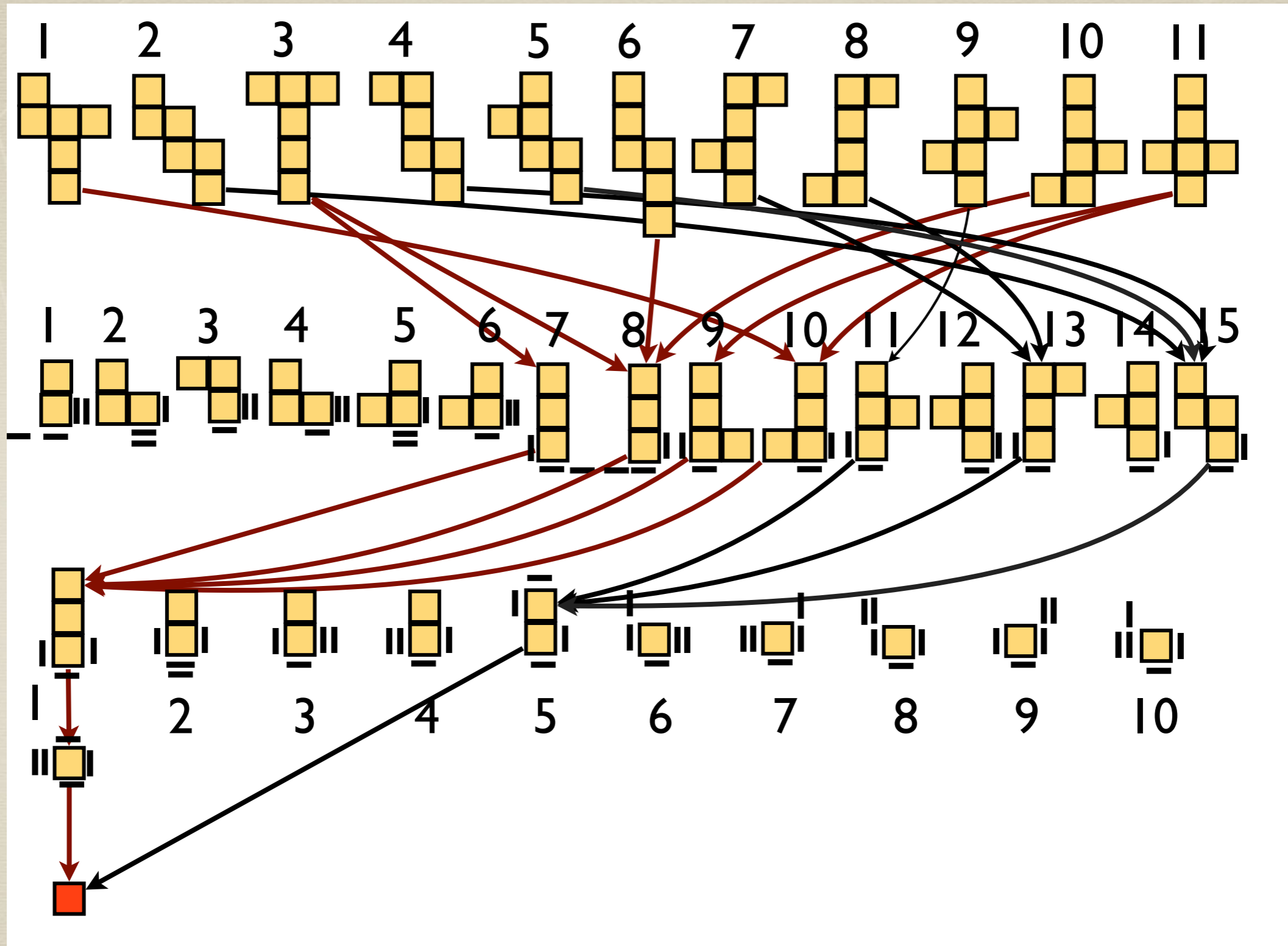
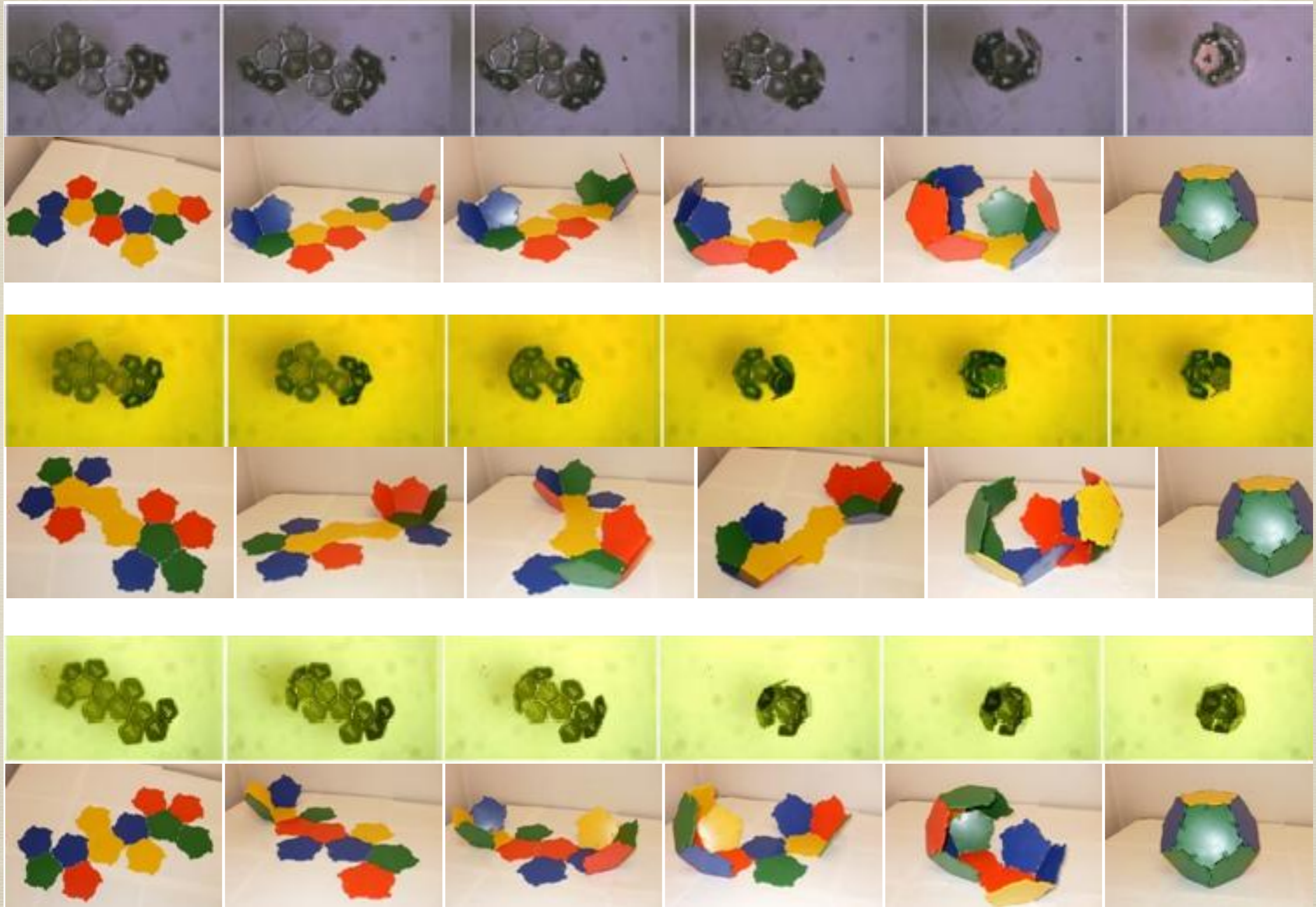Geodesic folding pathways for the cube.

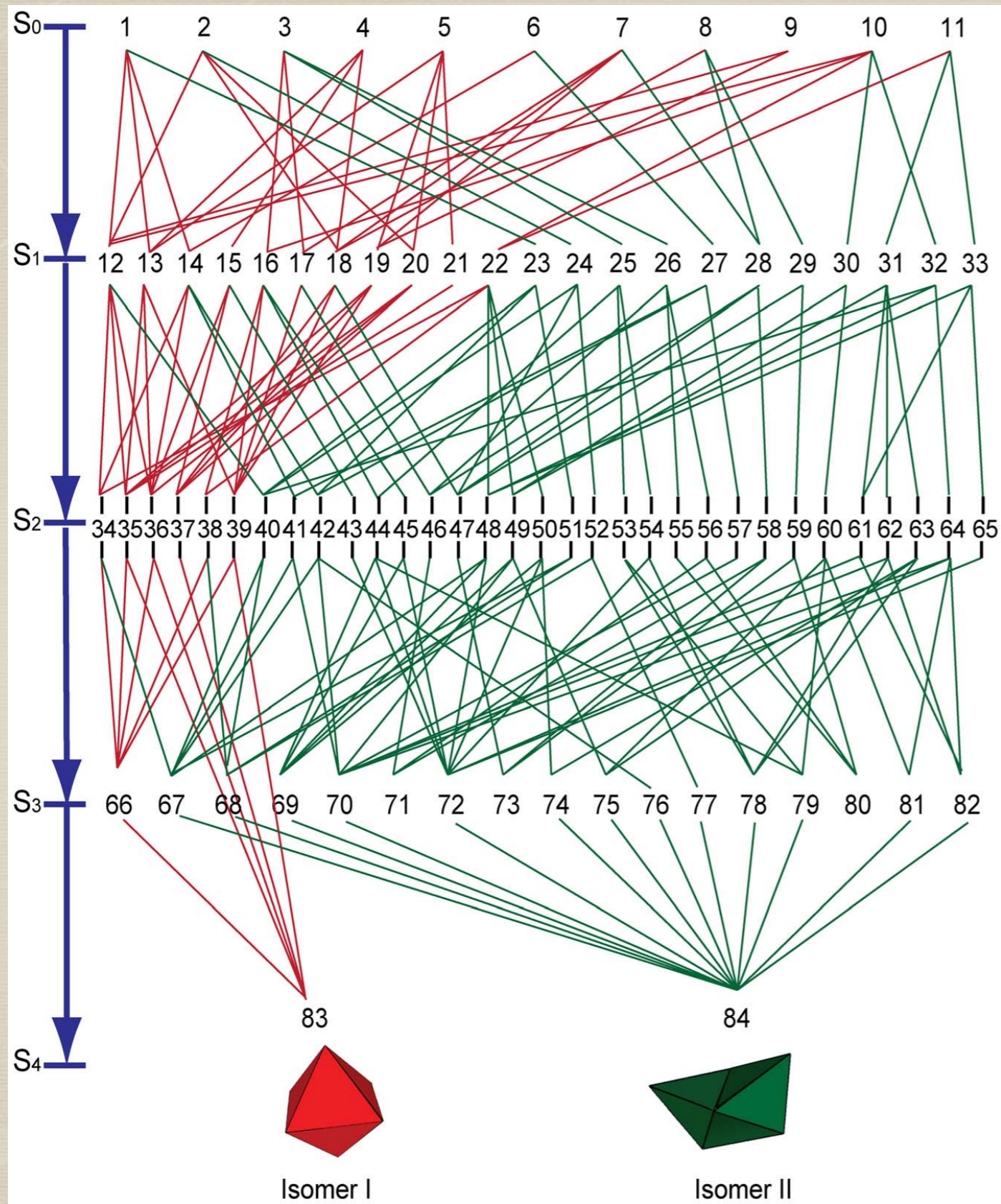All folding pathways.

Optimal assembly pathways.
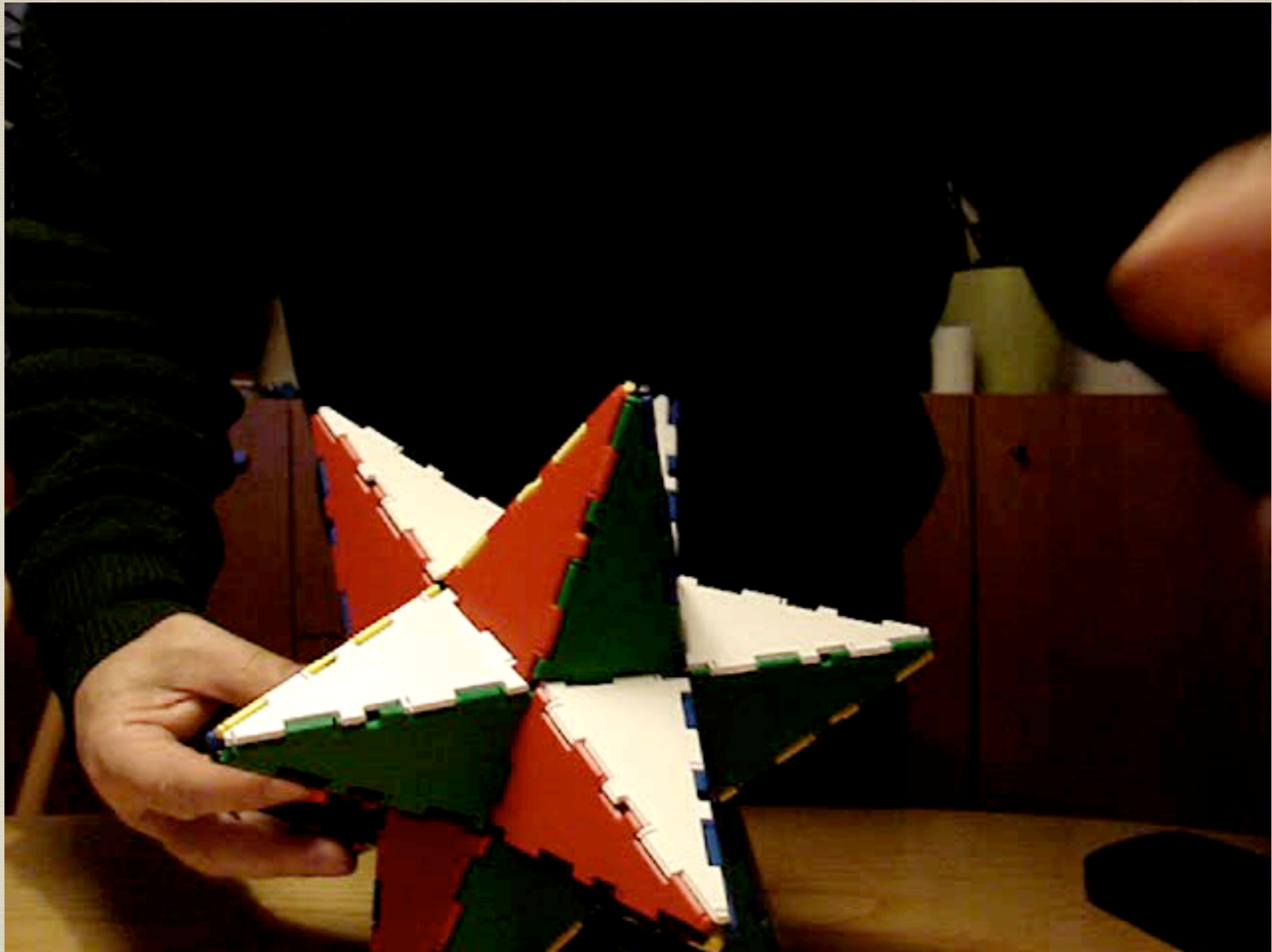
Comparison of computation and experiment

<u>Where is the math?</u>

(1) Configuration spaces, numerical schemes, computation of rates for linkages executing Brownian motion.

(2) What can we build?

The octahedron and an "isomer"

Izidor Hafner's "bellows" based on Connelly's polyhedron.