# Does life compute?

Govind Menon
Division of Applied Mathematics
Brown University

Experiments

Shivendra Pandey
David Gracias (lab)

Johns Hopkins University

Computations

Ryan Kaplan, Joe Klobusicky,
Daniel Johnson-Chyzhykov, Emily Russell

Brown University

Funding: NSF, Simons Foundation, Charles Simonyi Foundation, IAS

# Overview

This talk consists of:

(a) Mathematical models for self-assembly. In particular, it includes a mathematical framework for synthetic self-assembly that was developed with the Gracias lab. (Work inspired by viruses from 2009-2016).

(b) Speculation: Can we reconcile different (mathematical) ways of thinking about geometry within the context of self-assembly? In particular, can we understand (with our current mathematical knowledge) the internal workings of AlphaFold?

# Part 1: Self-assembly (with a lot of symmetry)

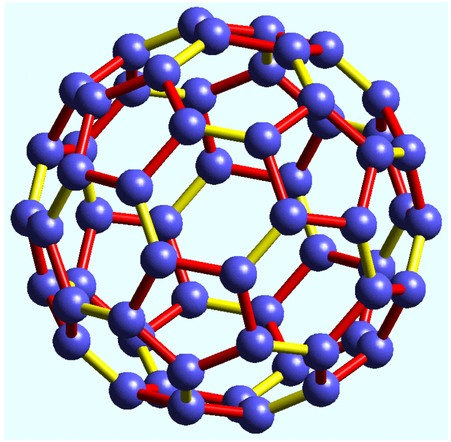
# Self-assembly in molecular biology

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:

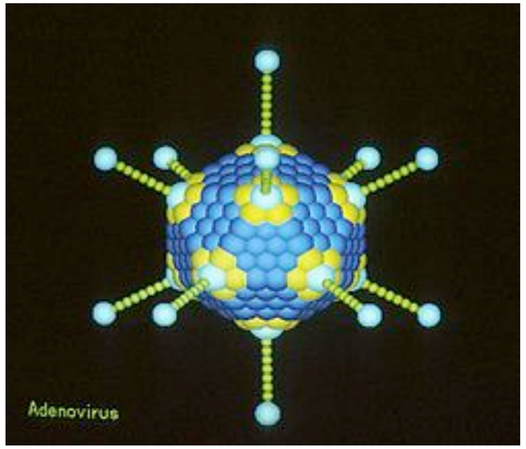
Self-assembly (of a virus) is a process akin to crystallization and is governed by the laws of statistical mechanics. The protein subunits and the nucleic acid chain spontaneously come together to form a simple virus particle because this is their lowest free energy state ."

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

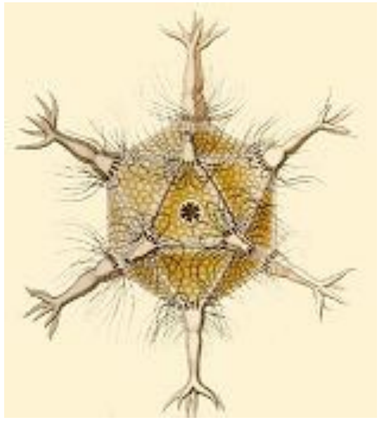
# Examples of icosahedral symmetry in nature



$C_{60}$ **molecule, 0.7 nm**




**Adenovirus, 90 nm**



**Radioalarian 10** $\mu$m

Widely different self-assembly mechanisms at different scales.

Mathematical structure: the "coding of symmetry" in the genome, and the interplay between symmetry and the pathways of self-assembly.

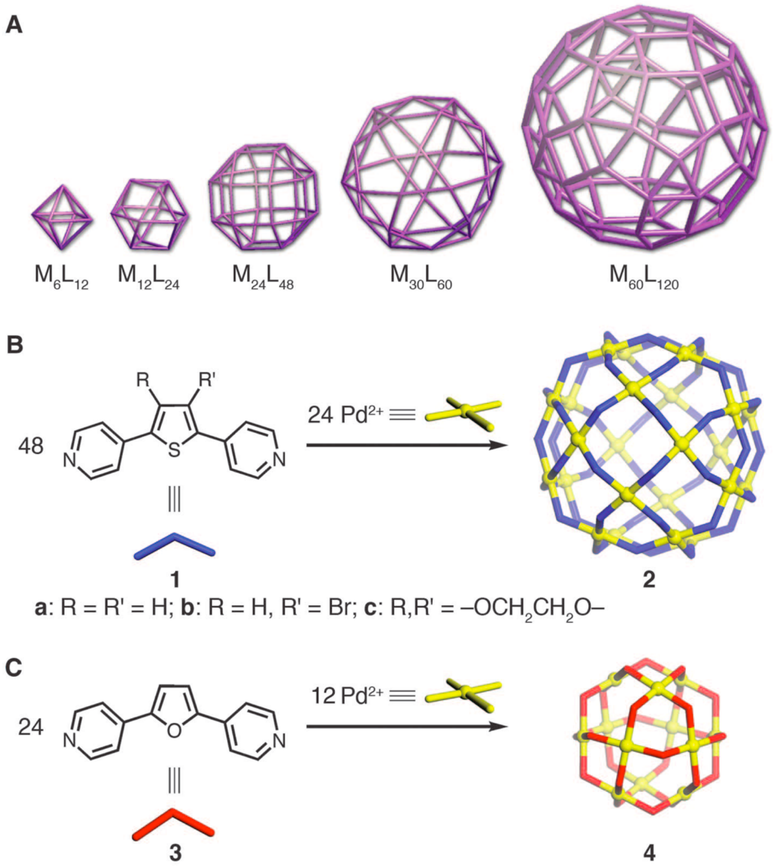
# Synthetic self-assembly

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:

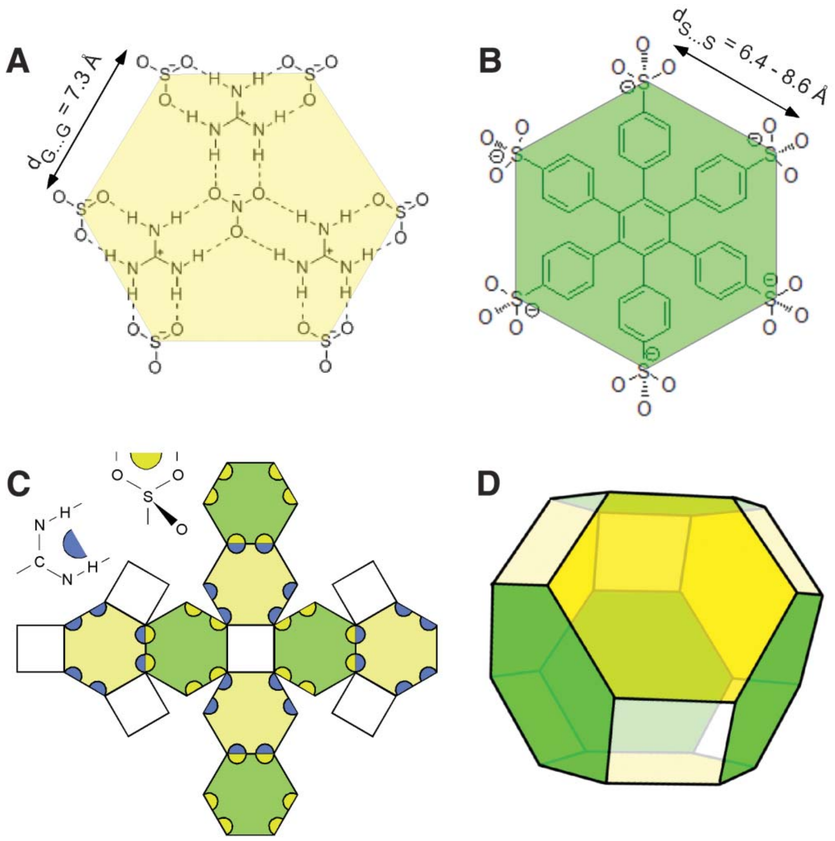
stripped down interactions (e.g. one dominant energy scale) ,
simple shapes built out of a few simpler motifs,
some randomness.

# Examples from supramolecular chemistry: 1 nm scale




Organometallic supermolecules built by covalent bonds.

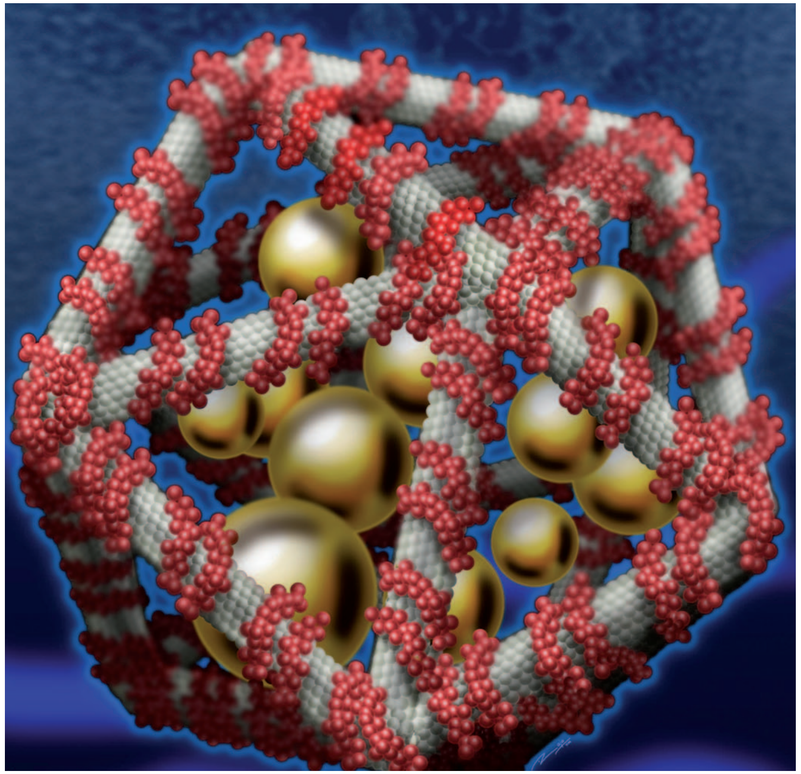
Fujita lab (Tokyo), Science (2010).

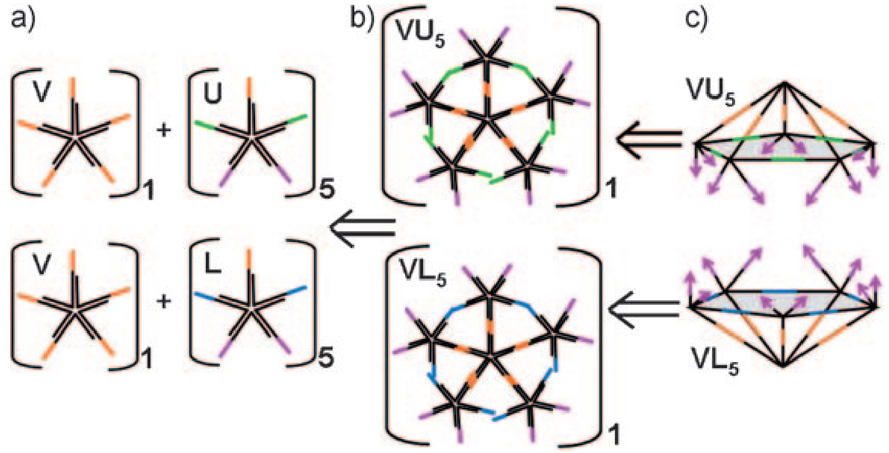



Archimedean cage built by hydrogen bonds.

Ward lab (NYU), Science (2011).

# Examples of synthetic polyhedra: 10 nm scale





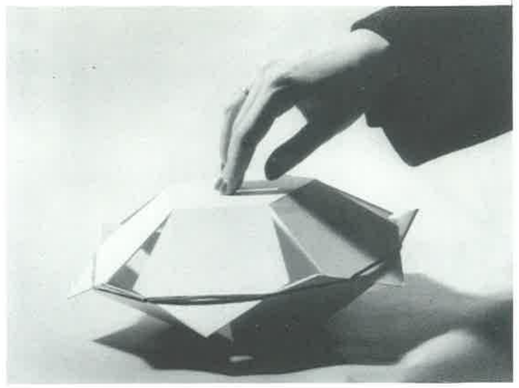

Icosahedral DNA cage built in modules.
Yamuna Krishnan's lab (NCBS), Ang. Chemie. (2009).

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

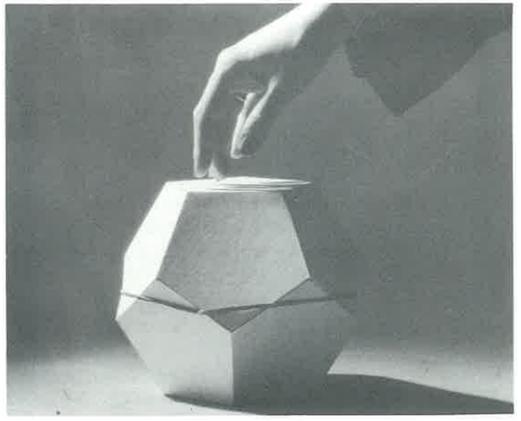
(229)





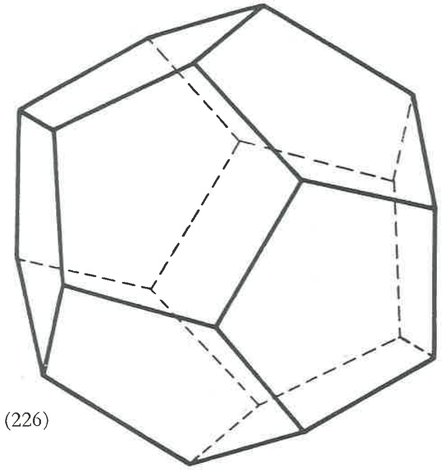
(230)

(231)

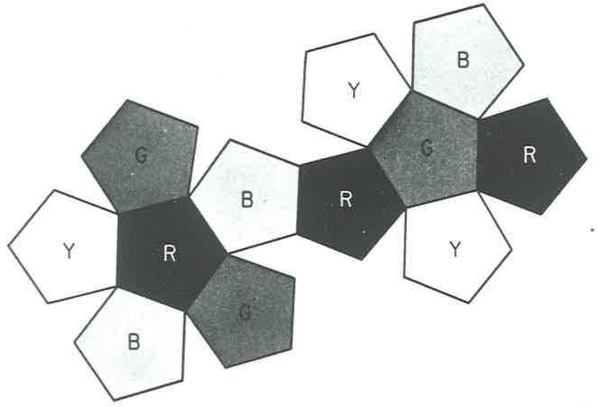




Hugo Steinhaus, Mathematical Snapshots (1938).

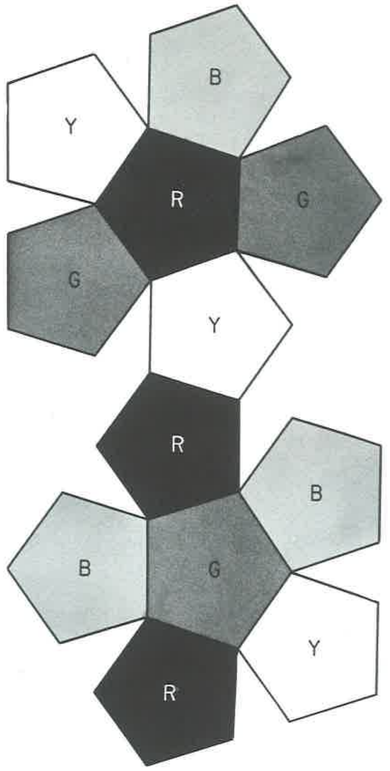


(226)



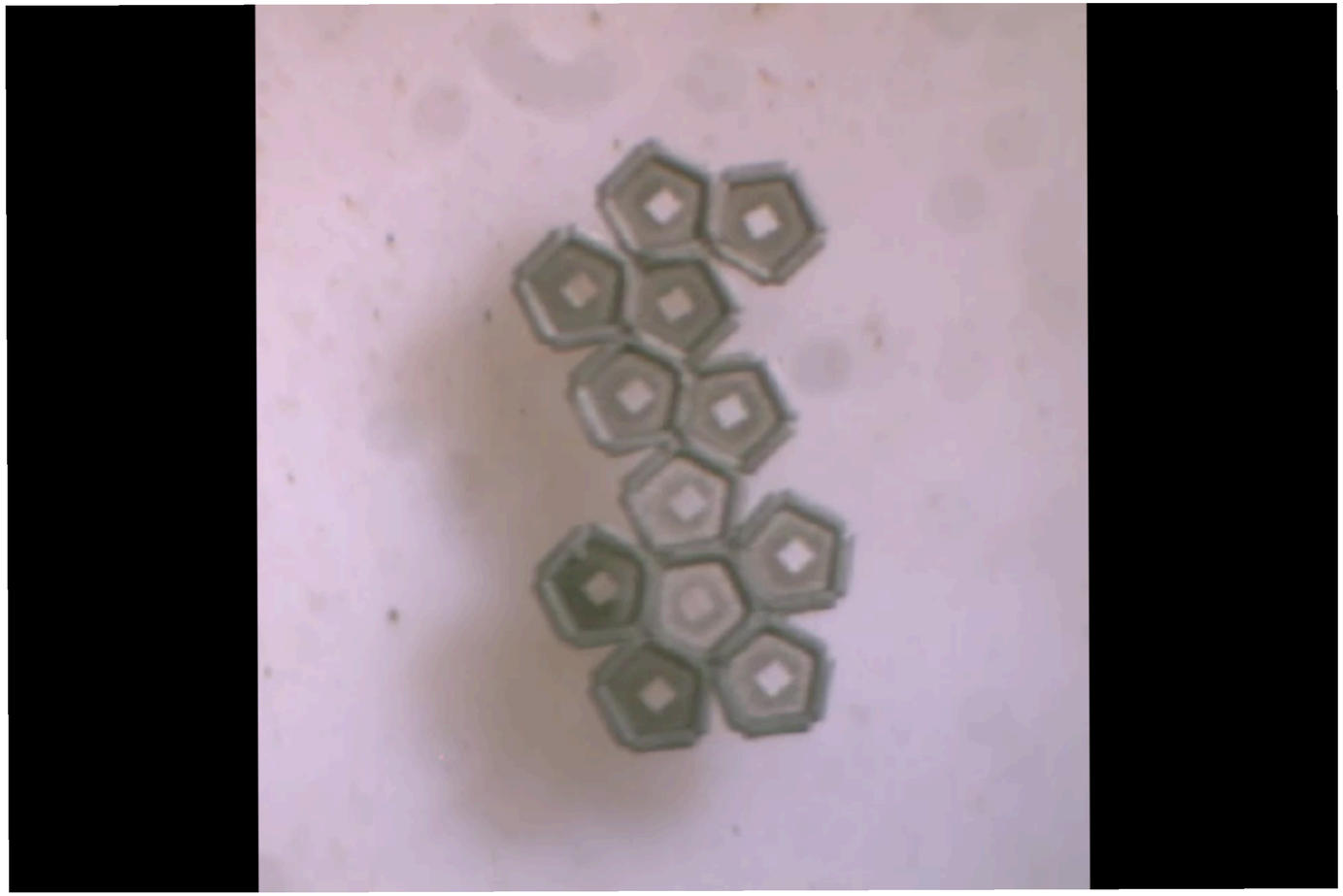

(227)






(228)

# Surface tension driven self-folding:  micron scale



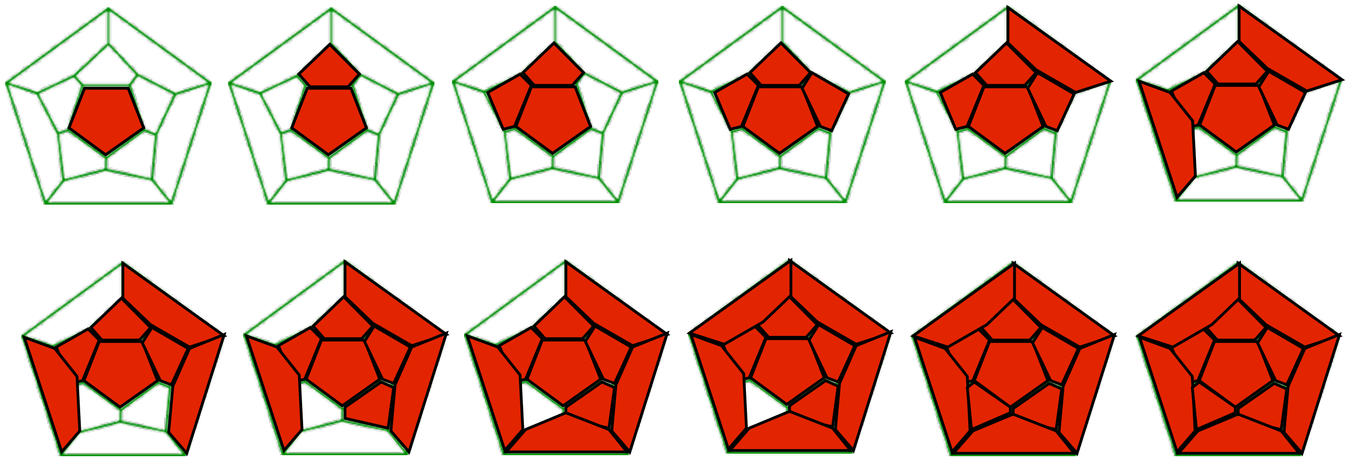
Truncated octahedra built by self-folding.
Pandey et al, PNAS,  (2011).

# The main theoretical  question

Can we develop common frameworks  to understand 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.

# Zlotnick's model (1994) and the building game



(1) Discretize the assembly process into intermediate states.

(2) Next add a model for attachment/detachment kinetics.

This is useful in chemistry for modeling fullerenes (Wales, 1987). However, it does not explain malformed shapes in virus assembly and has no sequence specific information. Many other models exist (Berger, Shor (1994); Bruinsma (2005)).

# Macroscopic "chemical reactions"

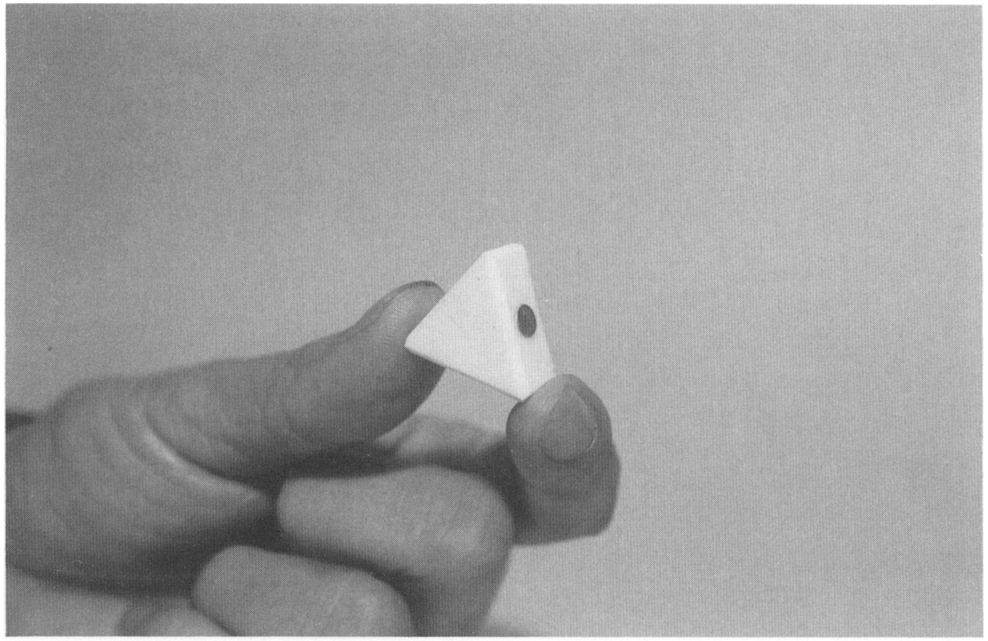


Figure 2. Photograph of the basic unit.

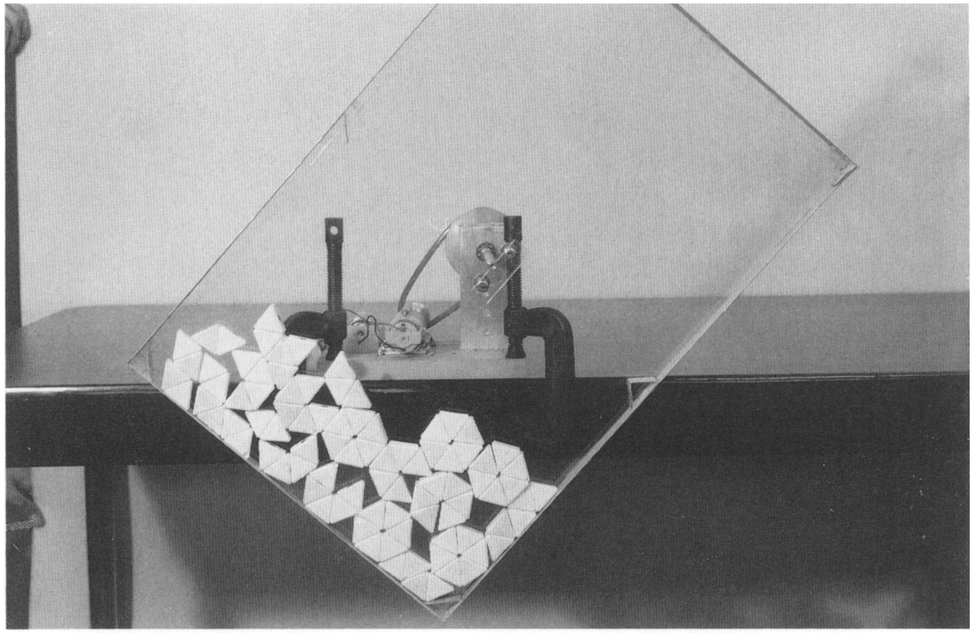


Figure 4. The experimental apparatus.

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

# Chemical reactions theory: states, reactions, rates.

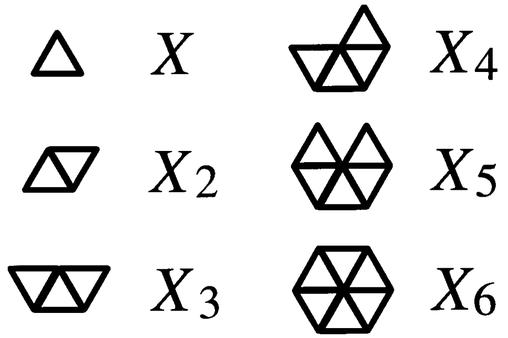



Figure 6. Initial, intermediate, and final products of the system.

States

$$
\begin{aligned}
&2X \to X_2, && X\cdot + X_2 \to X_3,\\
&X + X_3 \to X_4, && X + X_4 \to X_5,\\
&X + X_5 \to X_6, && 2X_2 \to X_4,\\
&X_2 + X_3 \to X_5, && X_2 + X_4 \to X_6,\\
&2X_3 \to X_6.
\end{aligned}
$$

Reactions

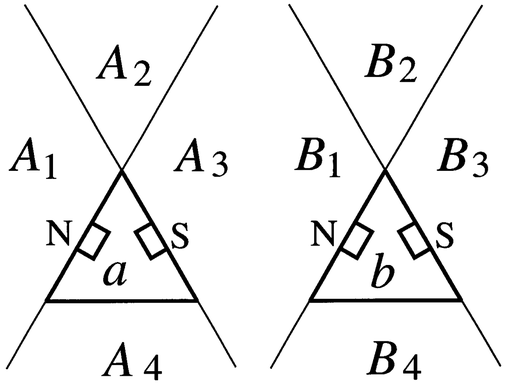
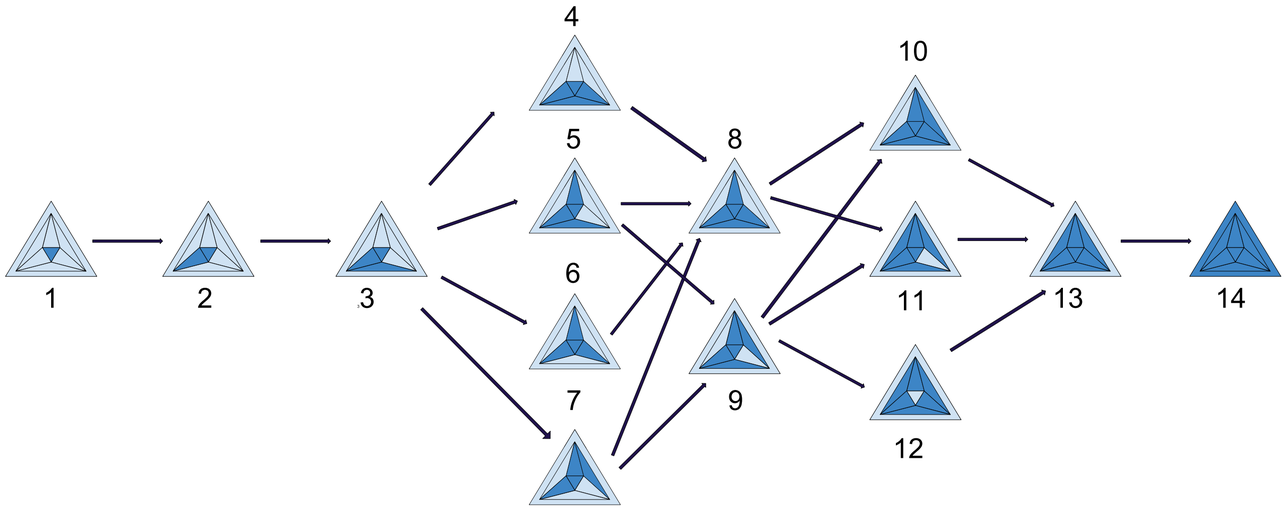



Figure 12. Divisions of the plane using the units for the basis.

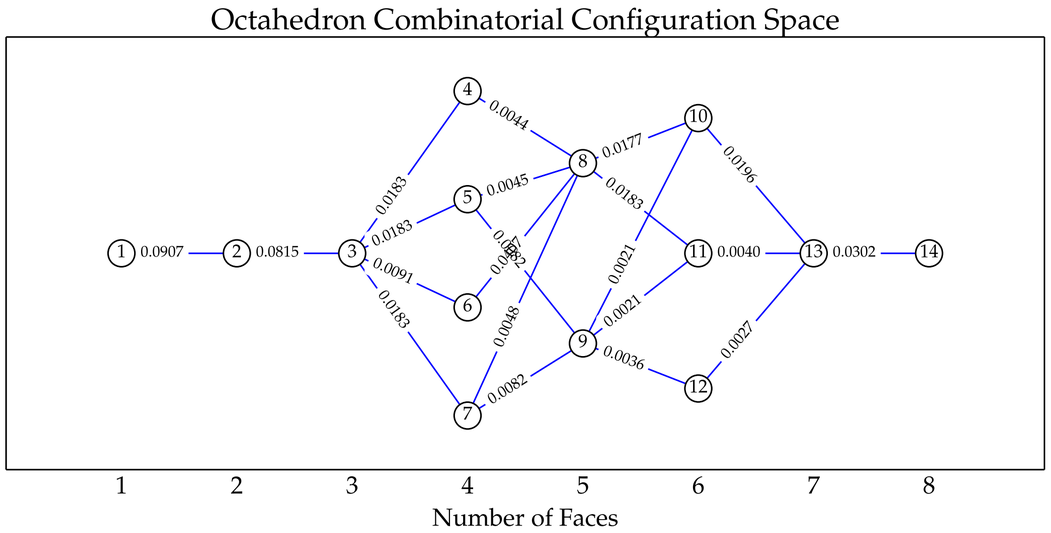
Rates: compute probabilities of collisions and bonding. Must include physics at this stage.

# The structure of a chemical reaction model of self-assembly




(a) An assembly graph

(a) Equip it with rates on edges
(unidirectional for simplicity)



Part 2. Some biological inspiration from viruses.

# The elegant natural design of viruses

(1) Genetic economy.

(2) Structural symmetry.

(1): NIH-NCBI lists the sequence of nucleotides of 13,736,078 viruses (as of 11/21/2024). Many of these have very short genomes. For example, ss-RNA viruses often have genetic sequences 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...

# The reproduction cycle of simple viruses

The “simplest” viruses consist of a genome (RNA or DNA) contained within a protein shell (the capsid). They lack the biosynthetic machinery for independent existence. Instead, they use a host to reproduce as follows.

(1) The capsid disassembles when a virus infects a host;

(2) The genome hijacks host’s machinery to make new genome and protein;

(3) Capsid reassembles and packages the new genome.

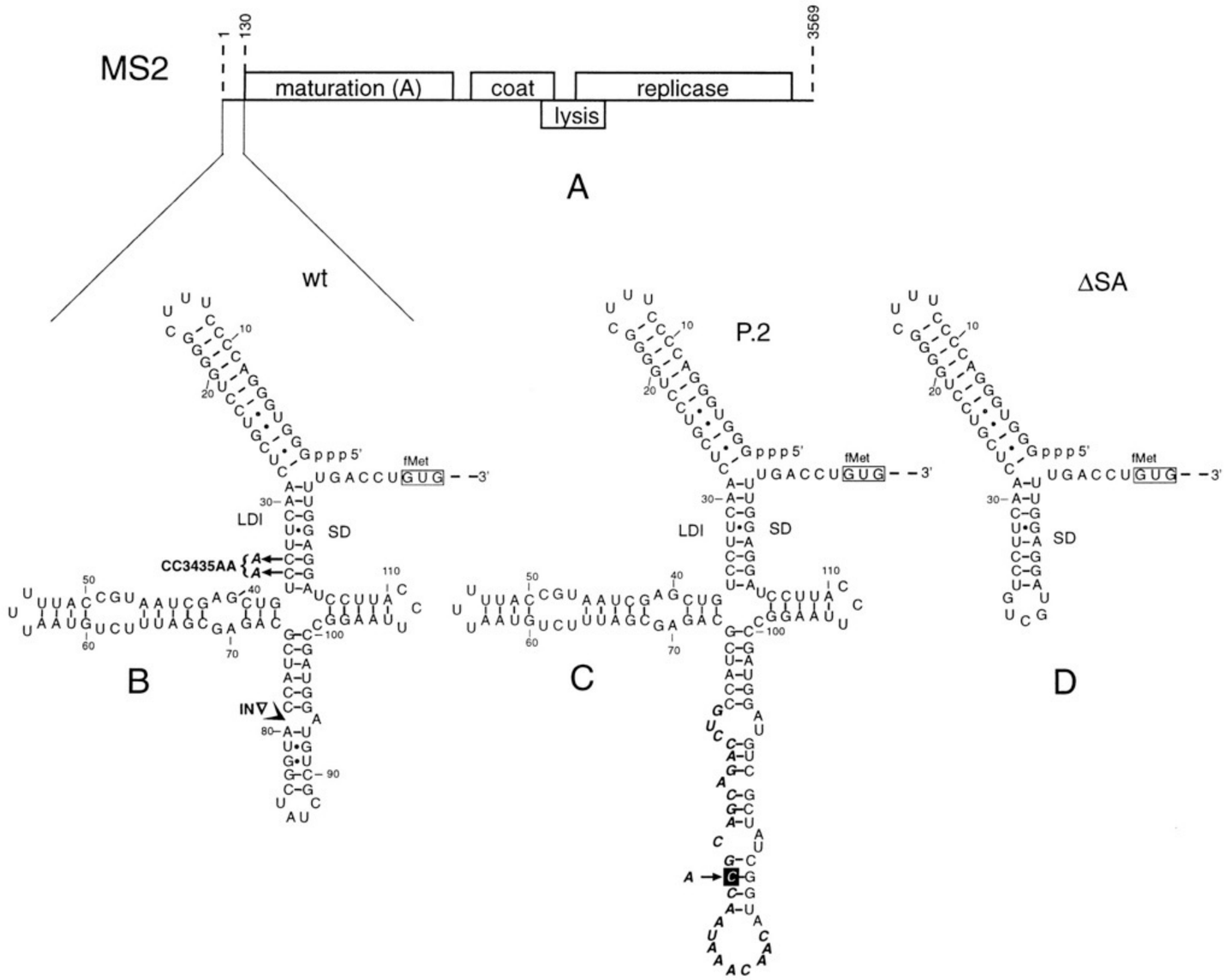
# The story of MS2

MS2 is an icosahedral virus with a single-stranded RNA. It infects the bacteria e.coli and other enterobacteria. The MS2 genome was isolated in 1961 and is the first genome to be completely sequenced (Fiers 1972-1976).

The genome has 3569 nucleotides that code for four proteins:

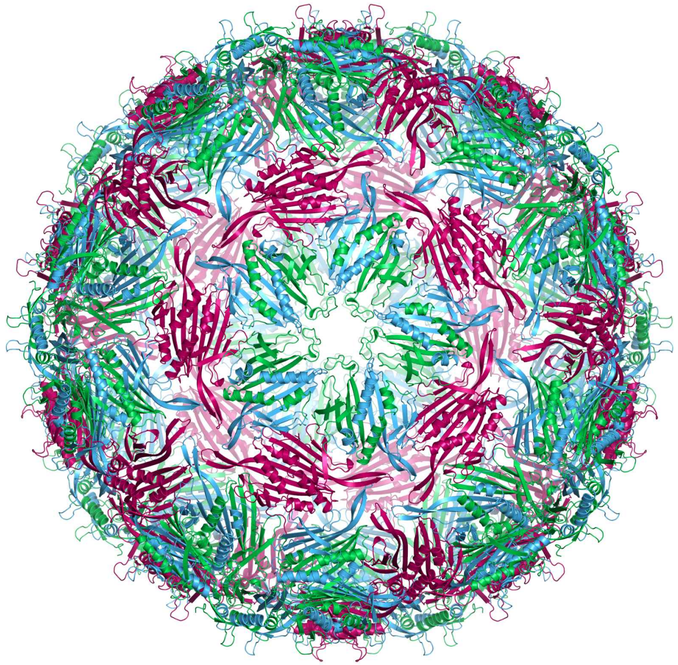
(1) coat protein; (2) maturation protein; (3) lysis enzyme; (4) replicase enzyme.

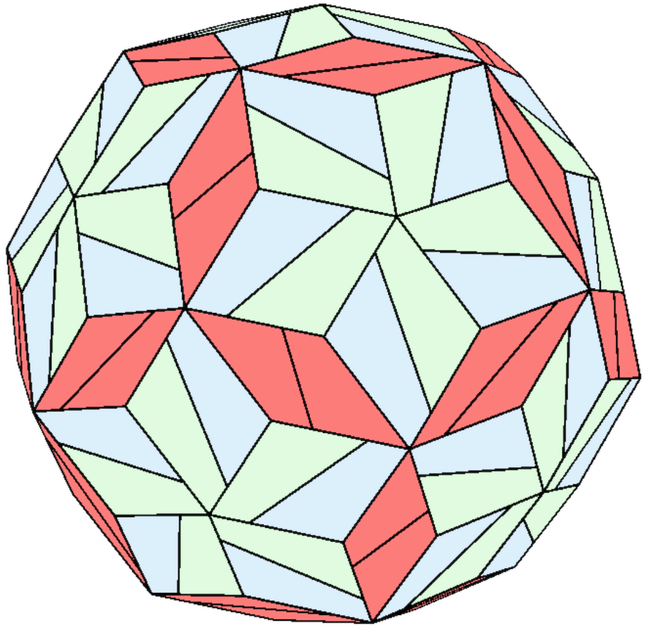
# The MS2 genome




Fiers et al; Nature (1972)

# 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 just based on the capsid protein, 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.

# Co-assembly with RNA folding

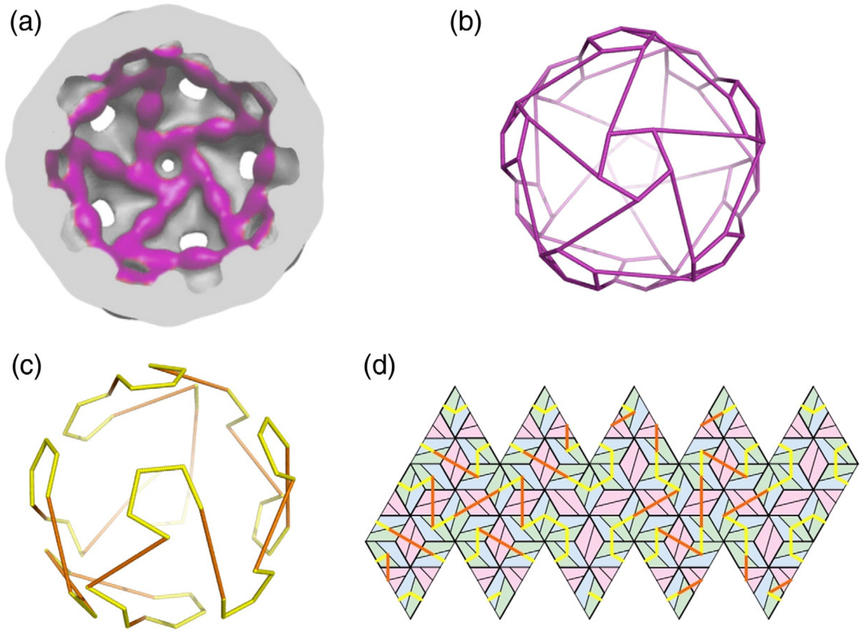



**Fig. 1.** The polyhedral cage of MS2 RNA density can be described as a Hamiltonian path. (a) A cryo-EM reconstruction of the outer RNA shell of bacteriophage MS2 (depicted in magenta) based on an image at ~17 Å resolution adapted from Van den Worm *et al.*[7] (b) A representation of this RNA shell 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.

The RNA folds on a complementary Hamiltonian path. The folding process creates secondary folding sites that nucleate the dimer transition.

Dykeman, Stockley, Twarock; PNAS (2015)

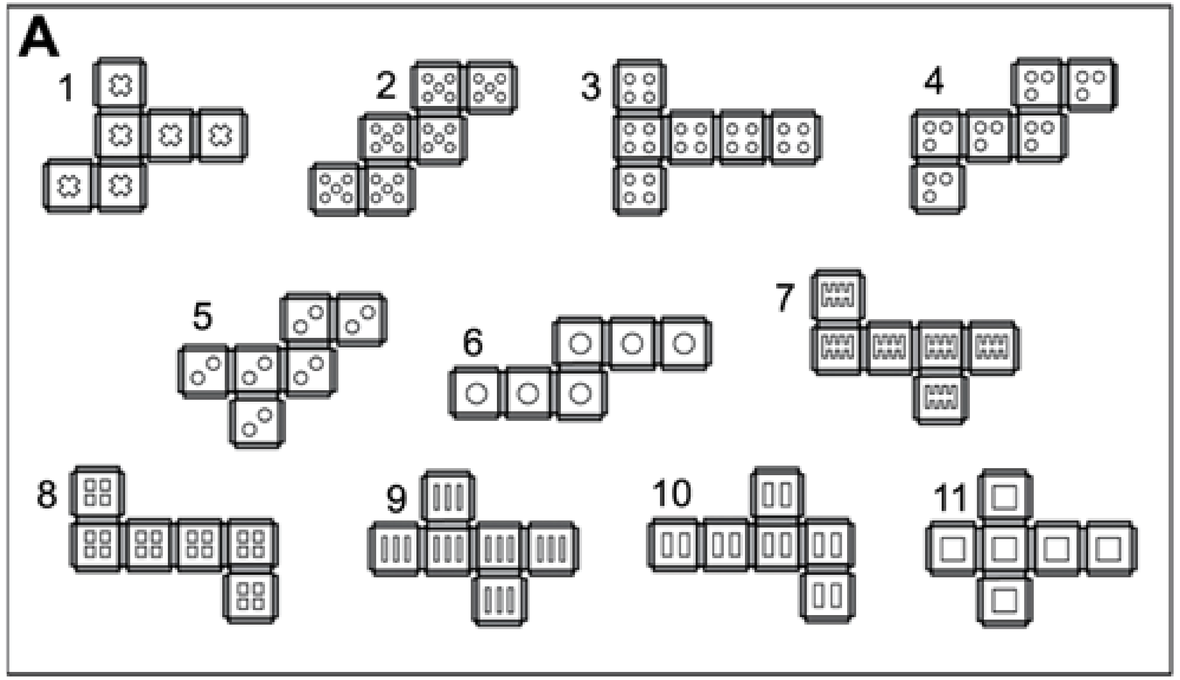
# Part 3. Self-folding polyhedra: computation and experiment.

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

experiments: Shivendra Pandey, David Gracias (Johns Hopkins University)

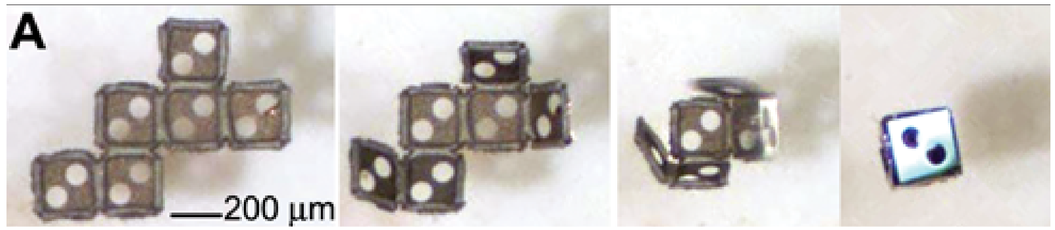
# Discovering design principles

Which net self-folds with the highest yield?



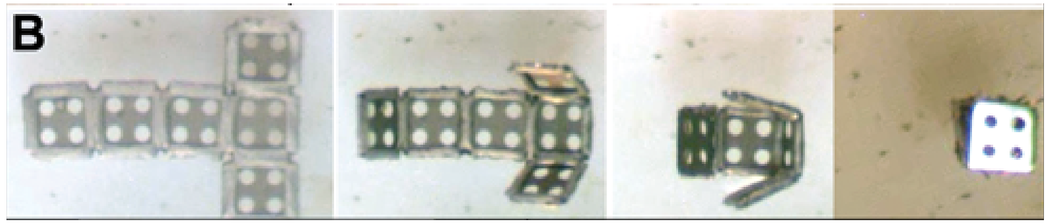

Azam, Leong, Zarafshar, Gracias (PloS One 2010)

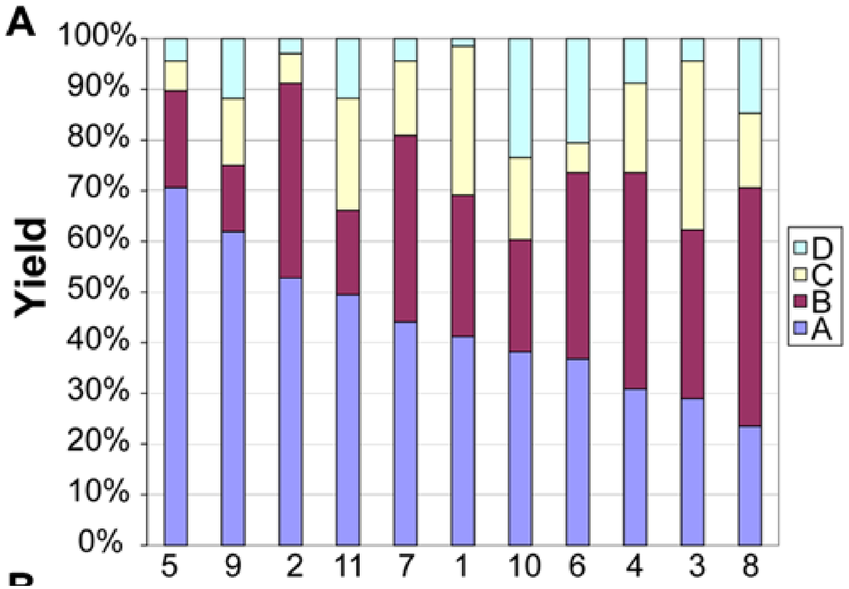
# The observed pathways (net 5 vs net 11) and yield




Net 5

Net 11



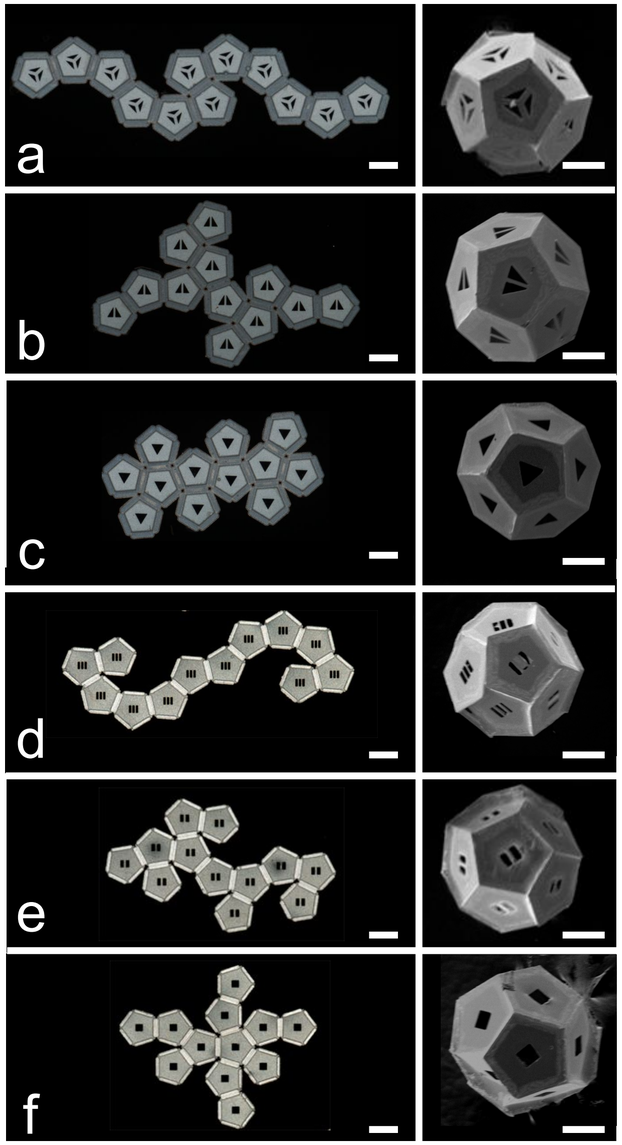

A: perfectly folded structures.

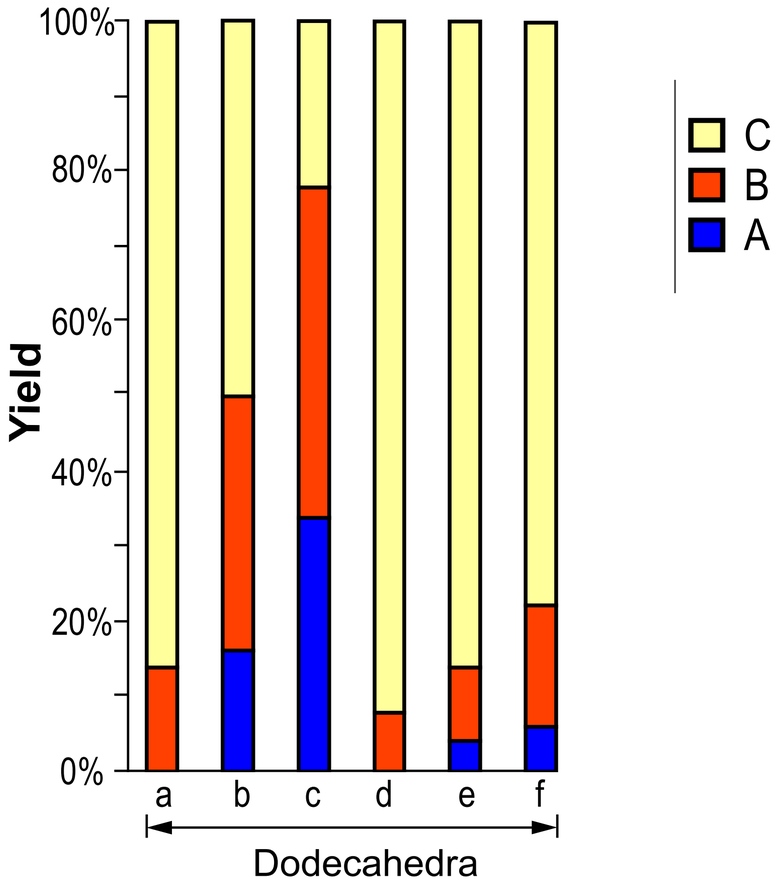
D: 2 or more faces did not fold.

# 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 | $2.3 \times 10^{6}$ |
| Viral capsid (T=1) | 60 | $10^{30}$ |

# Experiments on self-folding dodecahedra (at Johns Hopkins)

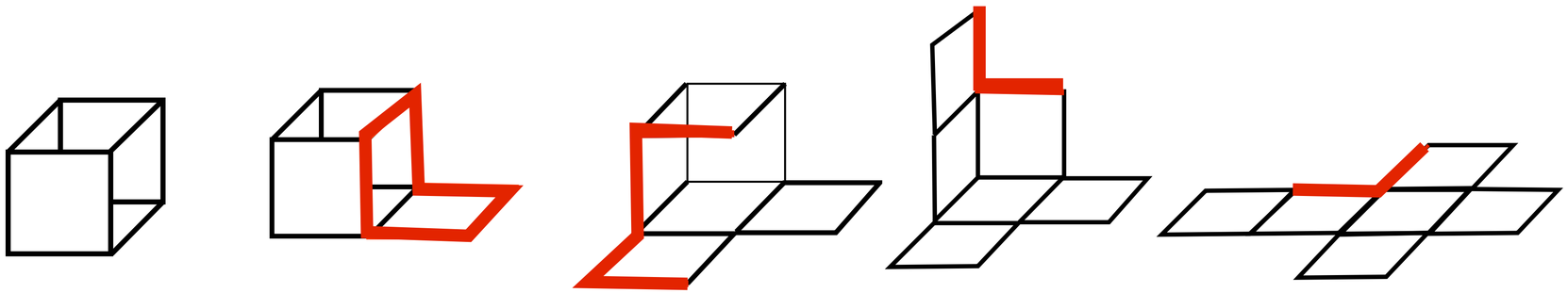


# A discrete folding algorithm

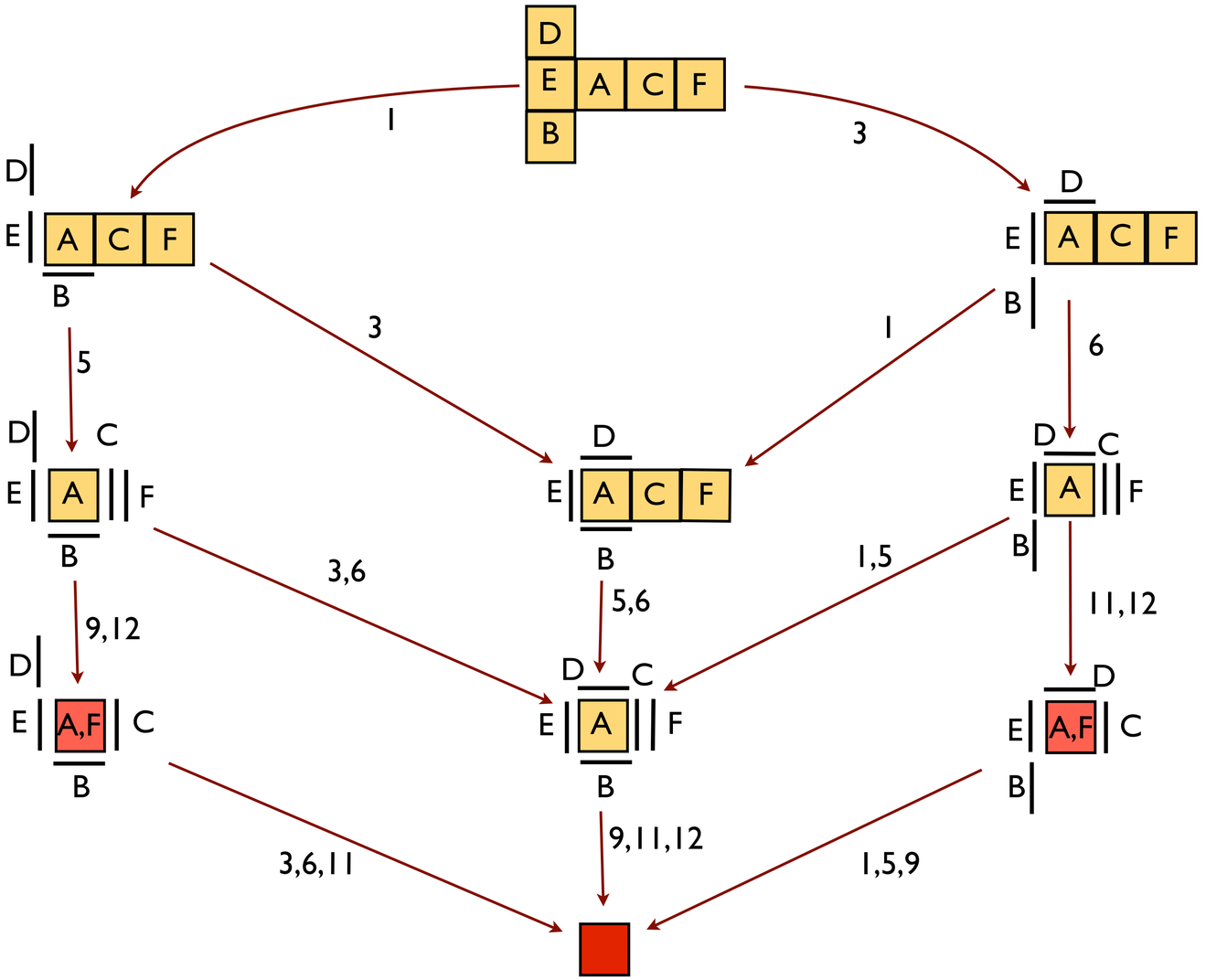
Fold a net  in discrete steps by gluing all edges at a vertex by rotating through the dihedral angle.

Glue only one edge at a time, unless additional faces collide. If so, glue all additional edges that are forced on you.

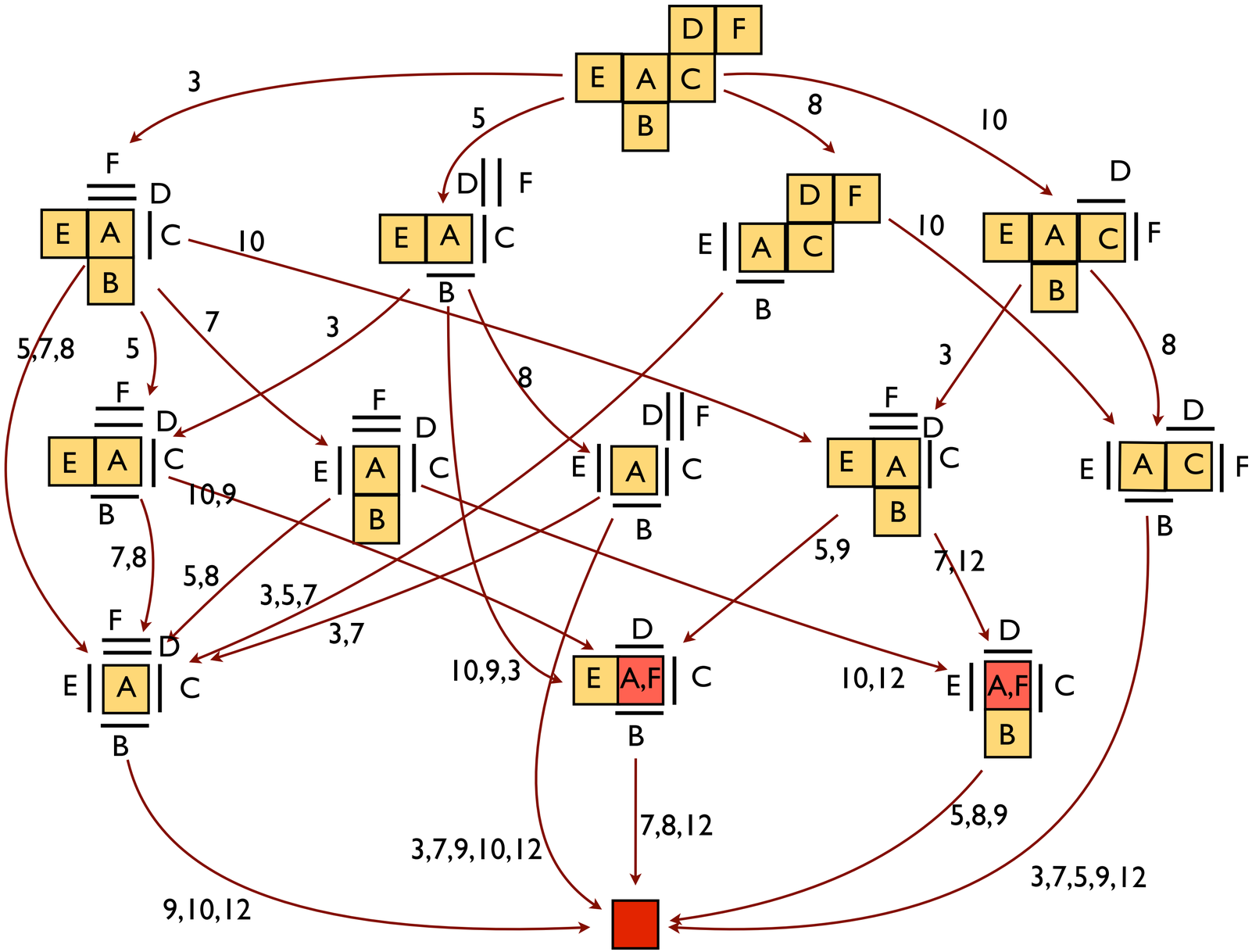


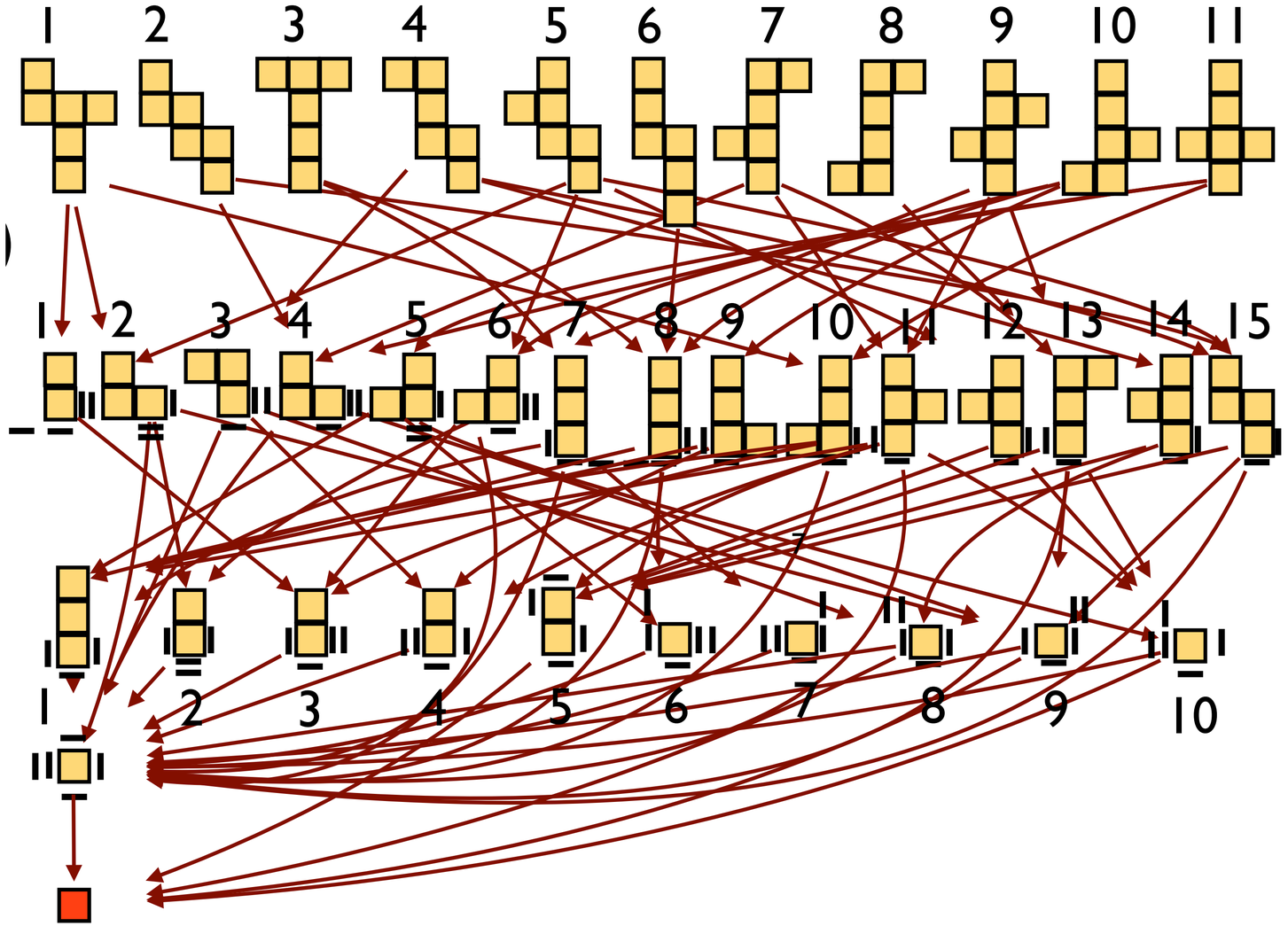
This has several  interpretations such as an  evolution of spanning trees on the vertex-edge graph by the removal of edges, or  the transformation of  an initial spanning tree on the face-edge graph by the addition of cycles.

# Pathways for net 3

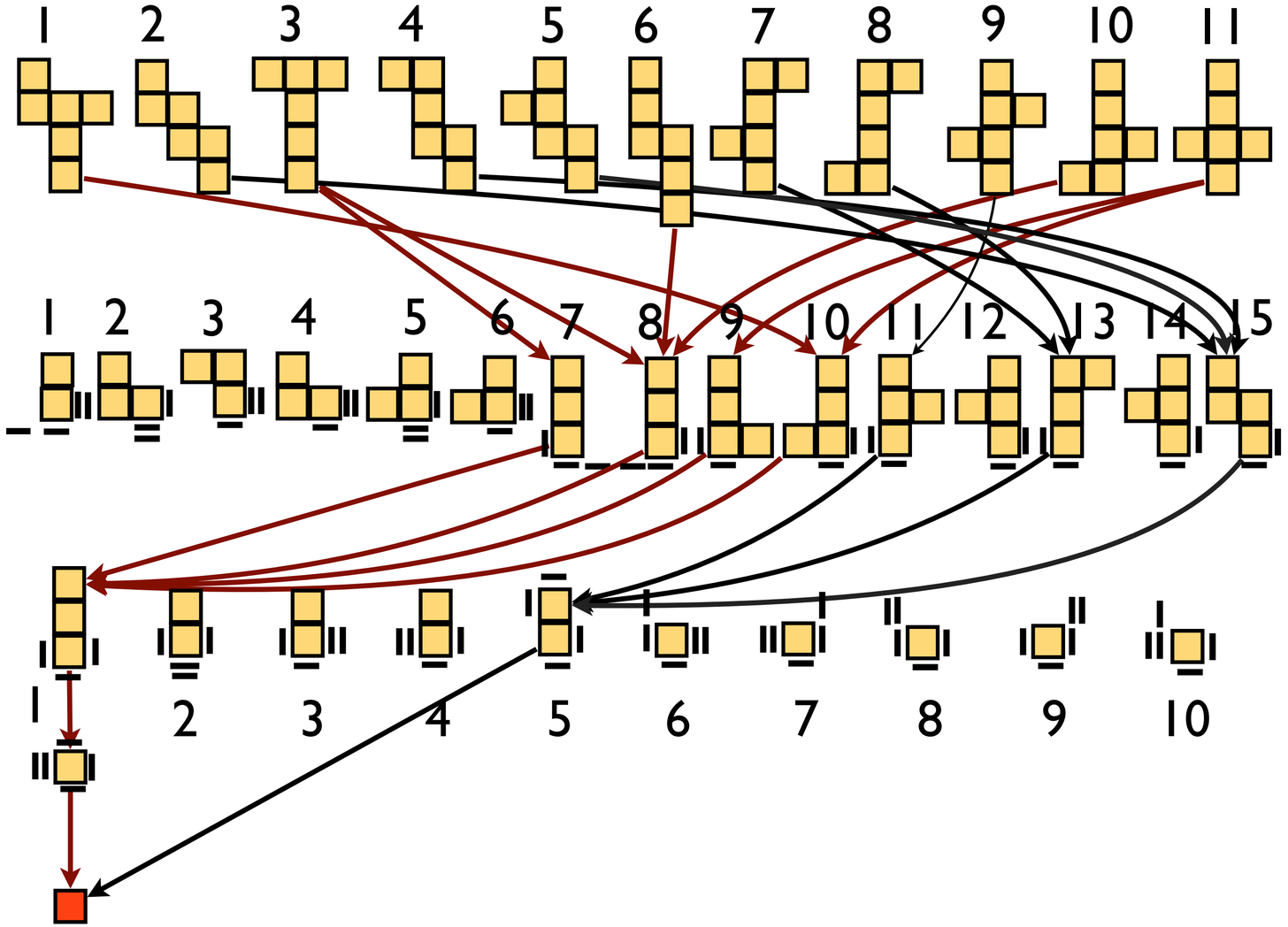


# Pathways for net 5

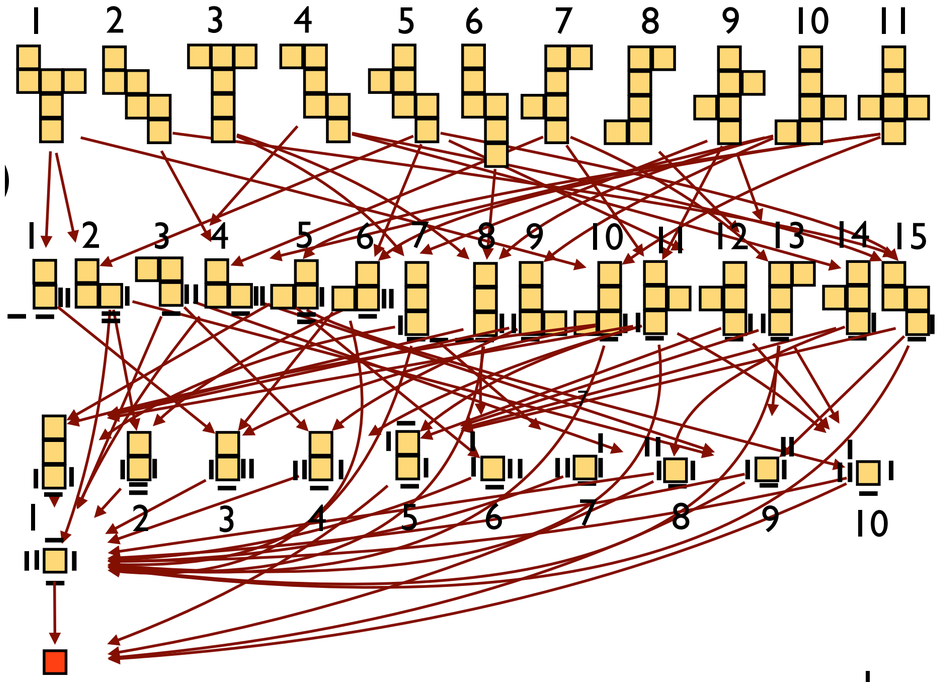




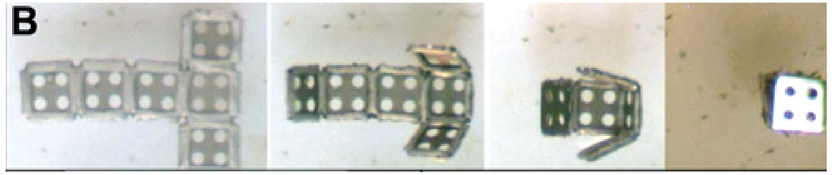

All folding pathways for the cube.



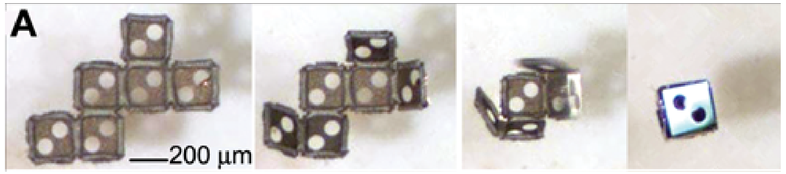

Geodesic folding pathways for the cube (for a hypothesized distance).

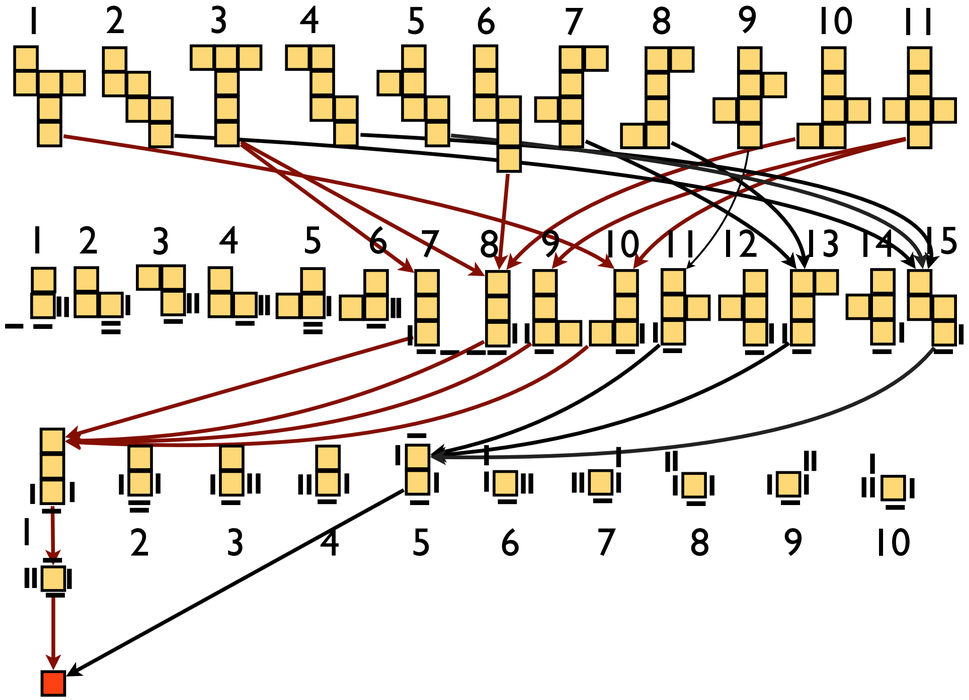



All folding pathways.



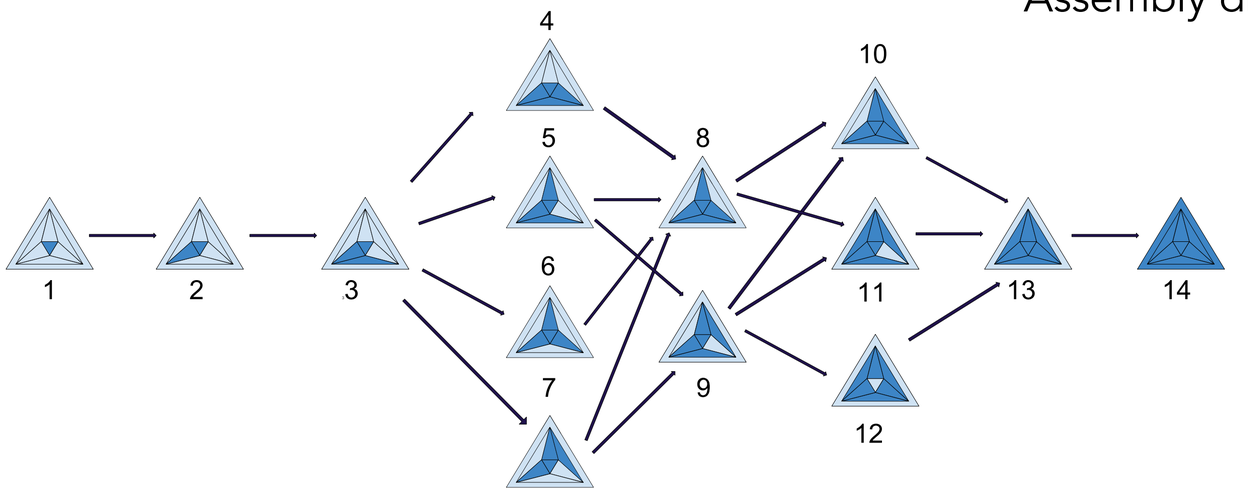

Optimal assembly pathways.



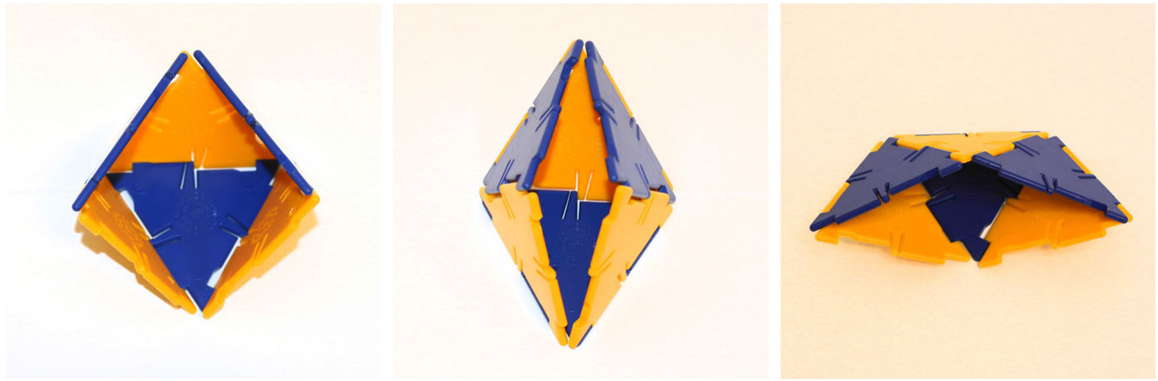
# Combinatorial structure and embeddings

Assembly diagram: bonds formed and partial order on states

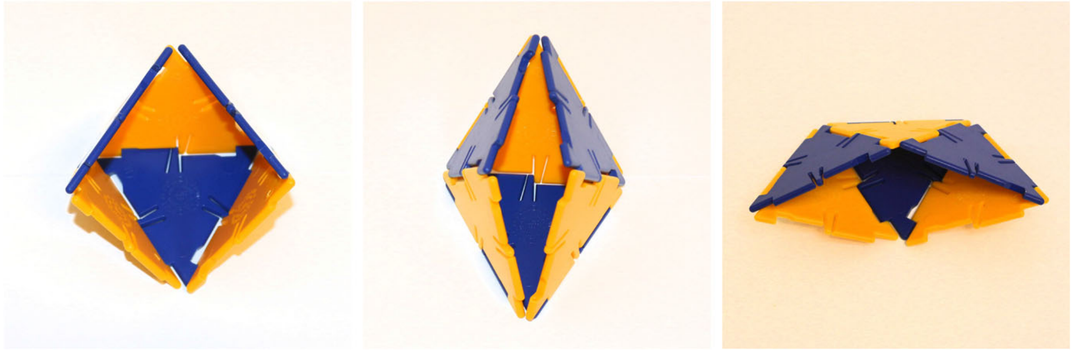



Each state sits in $\mathbb{R}^3$ in infinitely different ways.

Can we use this to obtain a minimalistic geometric understanding of the assembly process?

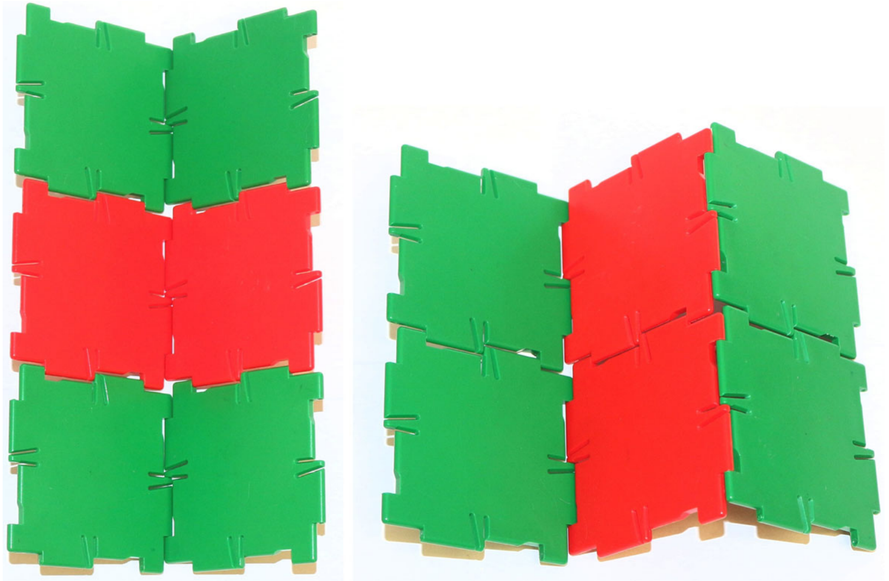


# Polyhedral linkages = real algebraic varieties= interesting mathematics



How do we describe the fluctuations  (conformational diffusion) in a strictly geometric manner?

Does this linkage have two or three degrees of freedom?



# 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 on the building game from: Johnson-Chyzhykov, Menon (2016).

# Part 4: How do we build shapes?

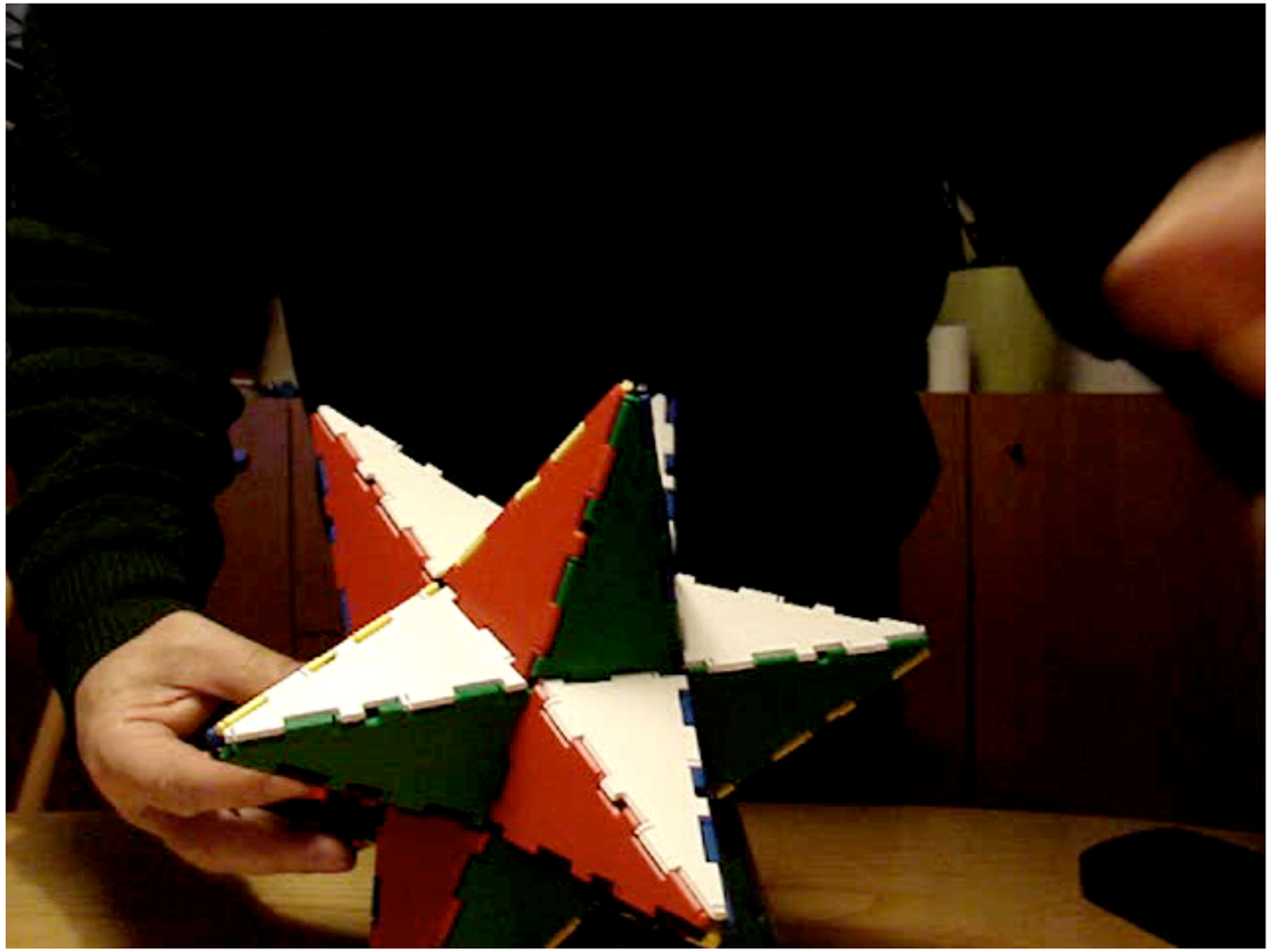
# Is there harmony between mathematical and biological structure?

(1) Well-defined mathematics : Assembly diagrams, Brownian motion of linkages, sequence-specific folding pathways, tiling problems.

(2) What does biology build? What can humans design? There is a theory for simple viruses, but that's a very small part of the story.

(3) How is the construction of shape in molecular biology related to the way in which mathematicians conceive of geometric constructions? Is this process computation?

# Polyhedra as machines?



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

# A fundamental biomolecular machine: $Ca^{2+}$-ATPase

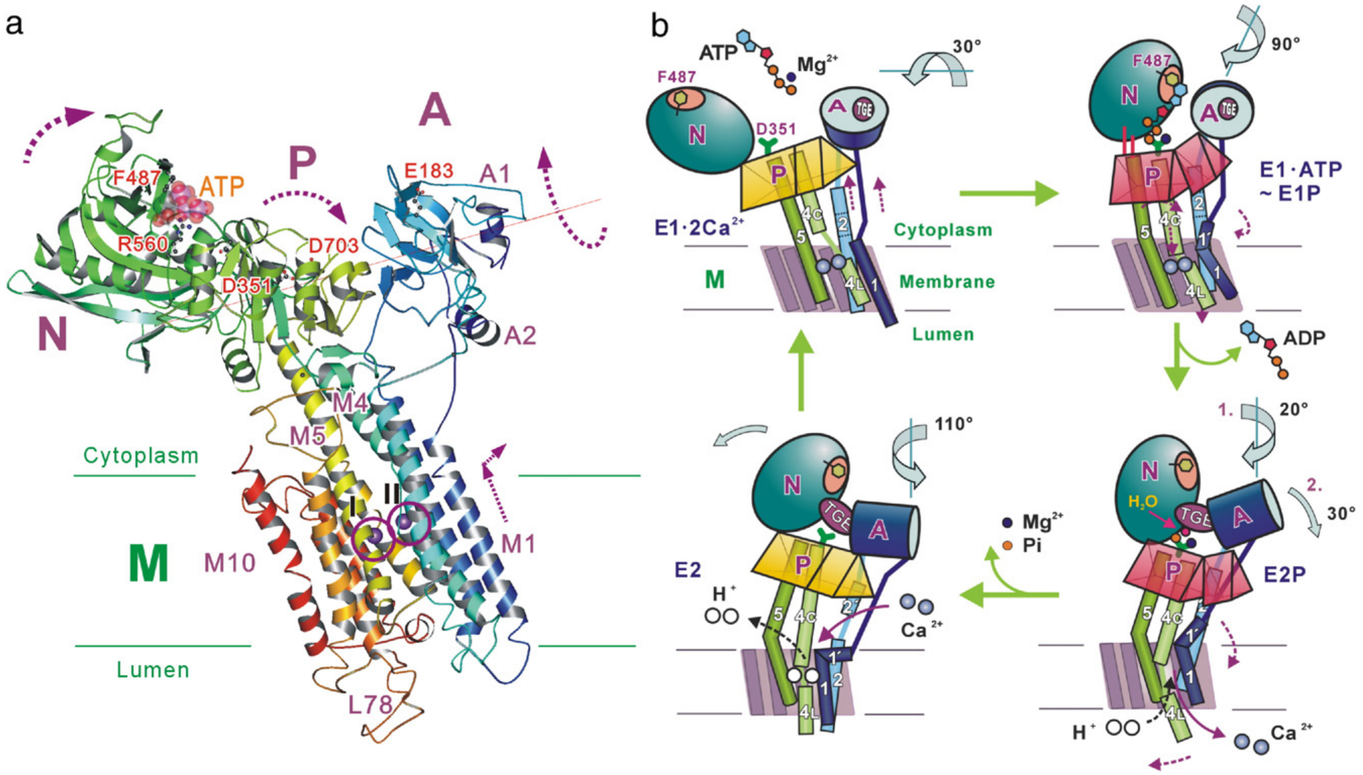



**Fig. 1.** Architecture of $Ca^{2+}$-ATPase and its ion pumping mechanism. *a*, A ribbon representation of $Ca^{2+}$-ATPase in the E1·2$Ca^{2+}$ state, viewed parallel to the membrane plane. Colours change gradually from the amino terminus (blue) to the carboxy terminus (red). Purple spheres (numbered and circled) represent bound $Ca^{2+}$. Three cytoplasmic domains (A, N and P), the α -helices in the A-domain (A1–A3) and those in the transmembrane domain (M1–M10) are indicated. M1′ is an amphipathic part of the M1 helix lying on the bilayer surface. Docked ATP is shown in transparent space fill. Several key residues—E183 (A), F487 and R560 (N, ATP binding), D351 (phosphorylation site), D627 and D703 (P) are shown in ball-and-stick. Axis of rotation (or tilt) of the A-domain is indicated with thin orange line. PDB accession code is 1SU4 (E1·2$Ca^{2+}$). *b*, A cartoon illustrating the structural changes of the $Ca^{2+}$-ATPase during the reaction cycle, based on the crystal structures in 7 different states. Modified from [7].

C. Toyoshima: Biochimica et Biophysica Acta (2009)

(F0-F1 ATP-synthase has an "almost human" rotary engine construction)

# Sequence evolution preserves the essential geometry

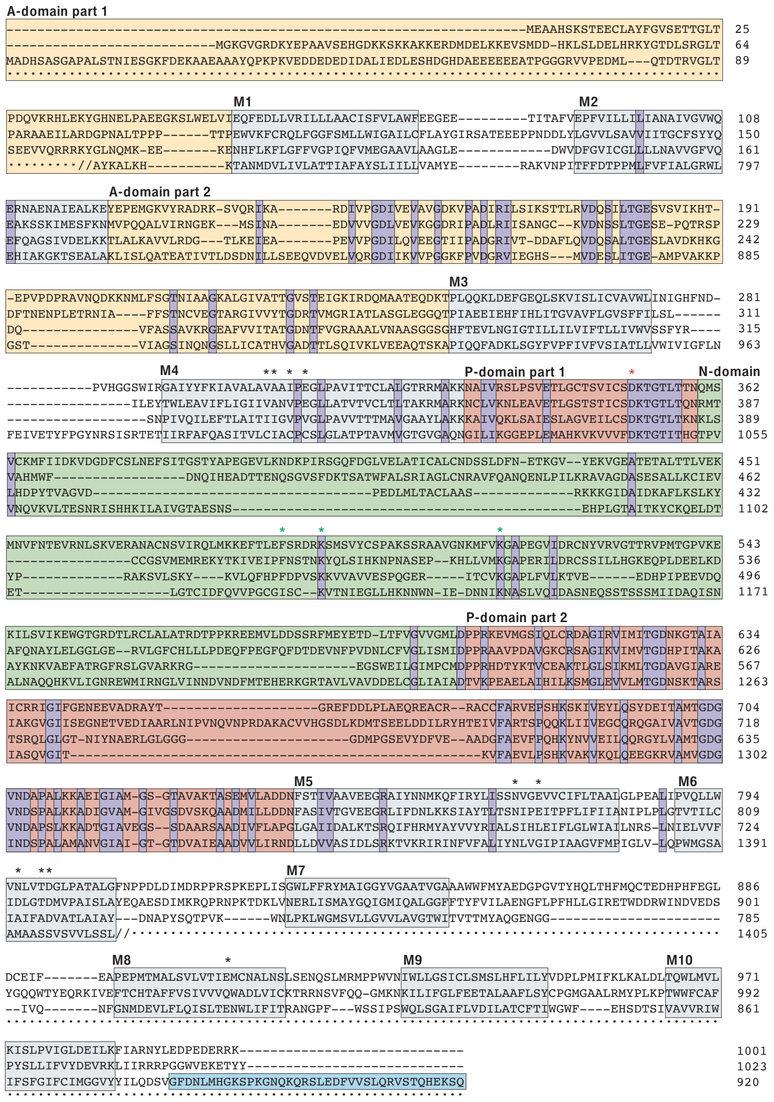



Figure 1 | **Conserved residues in P-type ATPases.** Sequence alignment of four representative P-type ATPases. From the top, the sequences of rabbit sarcoplasmic-reticulum $Ca^{2+}$-ATPase, rat $Na^+/K^+$-ATPase, *Neurospora crassa* plasma-membrane $H^+$-ATPase and the human $Cu^+$-ATPase that is affected in Menkes disease are shown. The actuator (A)-domain is shaded in yellow, the phosphorylation (P)-domain in red, the nucleotide-binding (N)-domain in green, and the carboxy-terminal regulatory domain of the $H^+$-ATPase in blue. The membrane-spanning helices M1–M10 are shaded grey. Identical residues and conservative substitutions are shaded purple. Coloured asterisks mark the phosphorylated aspartate (red), residues in the ion-binding site (black) and the nucleotide-binding site (green). For clarity, for the $Cu^+$-ATPase, the large metal-binding, amino-terminal extension and the carboxy-terminal part beyond M6 are not included. Sequences were aligned using CLUSTAL[103].

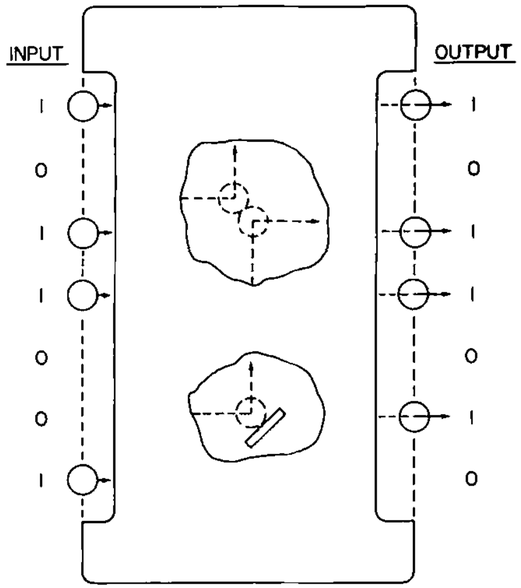
Sequences for P-type ATP-ases in four species (rabbit, rat, fungus, human)

W. Kuhlbrandt: Nature Reviews (2004)

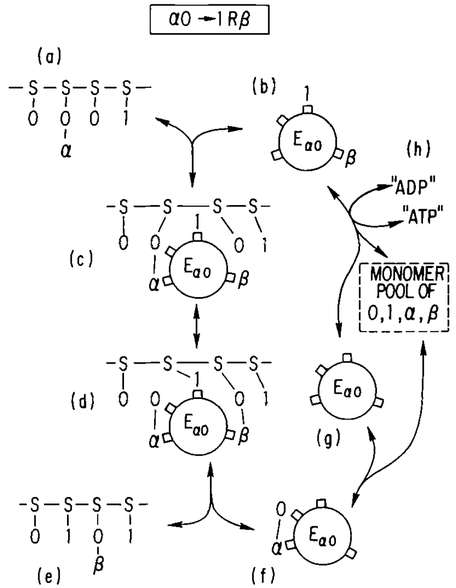
# The thermodynamics of computation (Bennett)

"A computer may be thought of  an engine that transforms free energy into waste heat and mathematical work"

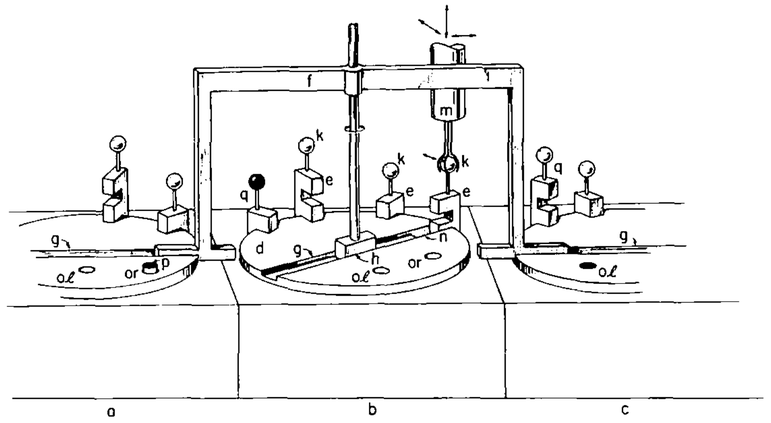
If "mathematical work = evaluation of a Boolean function" then we may seek different physical realizations of computers by seeking different ways of constructing logic gates.




Fredkin-Toffoli  ballistic computer

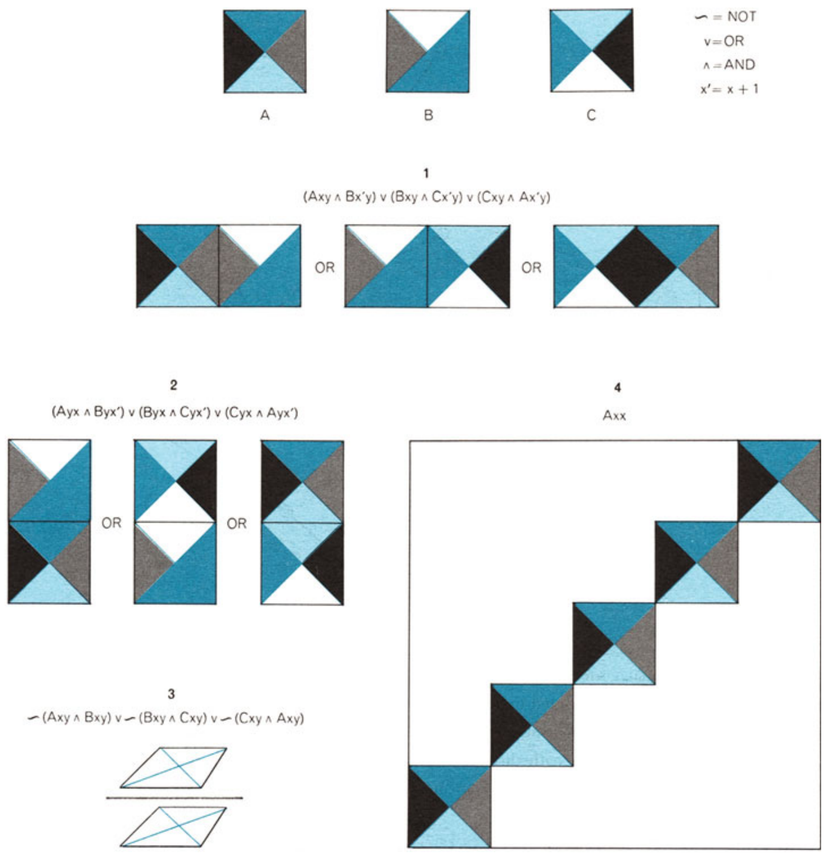



Hypothetical enzymatic machine




Brownian Turing machine

# Computation as a tiling problem (Hao Wang)




RULES FOR DOMINO PROBLEMS are set forth in the formal shorthand used by students of mathematical logic (*glossary is at top right*). At top center is a set of dominoes: *A*, *B* and *C*. The first expression states that colors must match on left and right edges, second that colors must match on top and bottom edges. The third rule is that dominoes must not be placed one atop another. The fourth expression, a constraint typical of those used to complicate games in approximating difficult problems of computation, states that only *A* can lie on the main diagonal of the plane. The positions on the plane are described by Cartesian coordinates. In designation such as "*Ayx*" domino's position on horizontal axis is given by the first variable, *y*, and vertical position by the second.

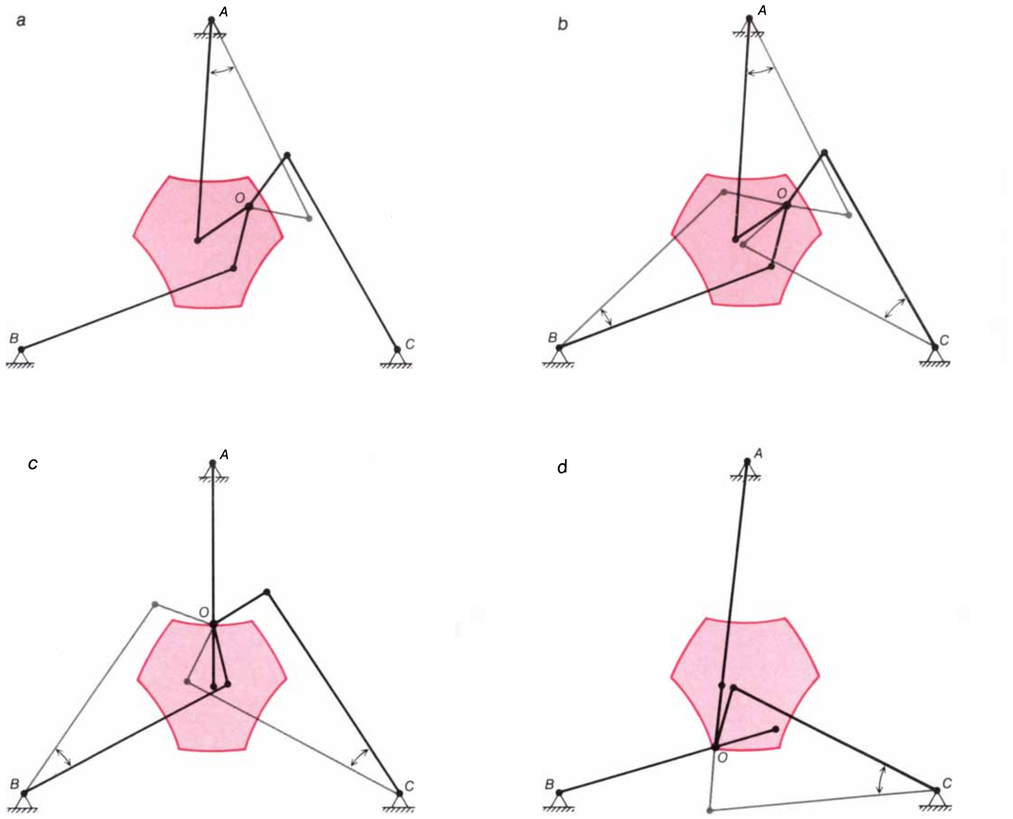
a) Logic gates can be realized using tiles with compatible edges

b) Decision problems can be made equivalent to a tiling problem

Experimentally realized using "sticky" DNA by Seeman, Rothemund, Winfree creating the field of DNA origami.

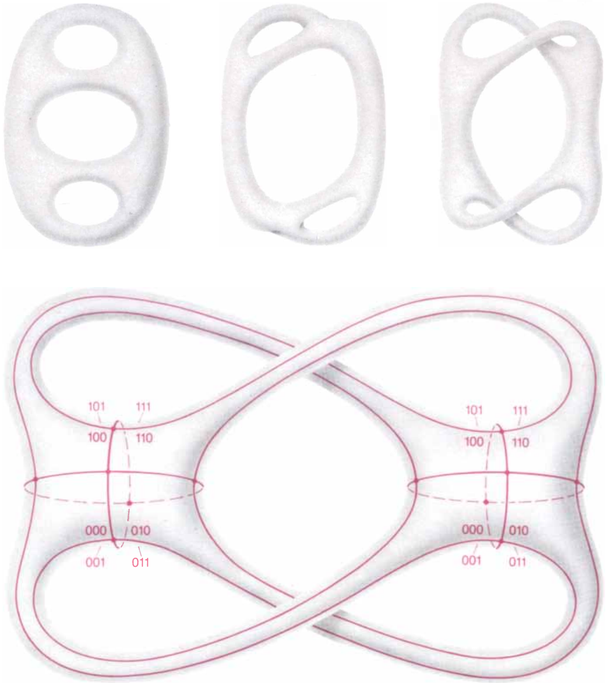
H. Wang, *Games, logic and computers*. Scientific American (1965)

# "Human" mathematical work (construction of manifolds)




**CONFIGURATION SPACE** of the three double cranks is a two-manifold on which distinct points represent distinct configurations, or possible arrangements, of the linkage. Every point inside the curvilinear hexagon traced by the central pin of the linkage can be reached when any one of the double cranks is bent in either of two ways (*a*). The bending of each double crank is independent of the bending of the other cranks, and so every point inside the hexagon gives rise to $2^3$, or eight, configurations of the linkage (*b*). Whenever the central pin reaches an edge of the hexagon, only two double cranks can be bent; the linkage can assume only four configurations (*c*). When the central pin reaches a vertex of the hexagon, only one of the double cranks can bend; the linkage can assume only two configurations (*d*).

Linkages define building blocks




**EIGHT ABSTRACT HEXAGONS** make up the configuration space of the three double cranks; each hexagon corresponds to one of the eight ways the three double cranks can bend to reach one of the points in the interior of the curvilinear hexagon. If the elbow of a double crank is bent clockwise, its configuration is labeled 0; if it is bent counterclockwise, its configuration is labeled 1. The abstract hexagons are labeled with three binary digits. In order to visualize the eight hexagons and the relations imposed on them in the configuration space the hexagons can be placed on the surface of a three-hole doughnut and distorted as if they were made of rubber. The three-hole doughnut has been deformed to the topologically equivalent manifold at the bottom of the illustration in order to show the eight hexagons more symmetrically. The binary digits assigned to the abstract hexagons reflect the pattern of gluings. If the digits for two hexagons match at two of the three positions, the hexagons are glued along two opposite edges; the two edges correspond to the edges of the curvilinear hexagon where the double crank associated with the nonmatching binary digit is straight. Any four hexagons with one matching binary digit meet at a vertex in the configuration space. The diametrically opposite vertexes of the four hexagons meet at a second point in the space. The two points represent straight configurations of the two double cranks associated with the two nonmatching digit positions.

Glue building blocks

Thurston and Weeks: Scientific American (1984)

# What is the mathematical structure of self-assembly?

AlphaFold provides an astonishing resolution of a fundamental bottleneck in molecular biology. It certainly computes in theory and practice. But what would it mean to understand how it works?

Our approach to a mechanistic understanding of self-assembly is based on minimalistic geometry (distance information, combinatorial structure). It is well-grounded and surprisingly rich in "human" mathematics, but requires new ideas (theorems, computation, evolution) in order to bridge the gap between proof of concept and practical application.

The main argument in favor of an approach that is based on minimal rules ("linkage rules" and "gluing rules") is its internal mathematical consistency and independence of scale. We are optimistic that there can be a harmonious way of efficiently combining these ideas with biology.