Minimising the Number of Kissing Loops in 3D RNA Origami Wireframes

Antti Elonen, Pekka Orponen
Aalto University
Computer Science

HALT Days, 29 Aug 2024

RNA basics: DNA vs RNA

DNA

- Nucleotide bases A, C, G, T
- Commonly double-stranded, comprising two W-C complementary helices
- B-form helix (base-planes about perpendicular to helical axis)
- Produced by replication from existing strands or (commercially) by ligation of short nucleotide sequences
- Kinetically inert in double-stranded form, reconfiguration requires either enzymes or denaturation (melting) and rehybridisation

RNA

- Nucleotide bases A, C, G, U
- Commonly single-stranded, folds upon itself following W-C pairing rules (and other effects)
- A-form helix (base-planes at ~19˚ angle from perpendicular to helical axis)
- Produced by polymerase transcription from a DNA template, template can be genetically engineered
- Kinetically unstable as open strand, folds at room temperature into stable conformation

RNA basics: The RNA helix

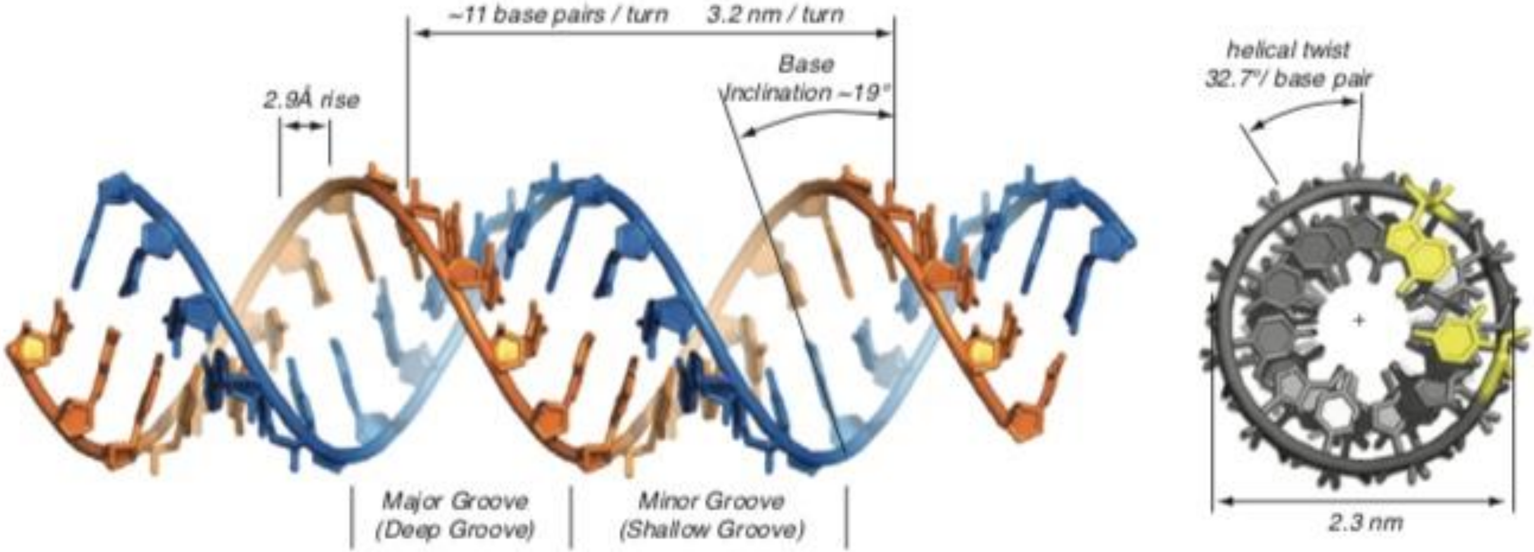
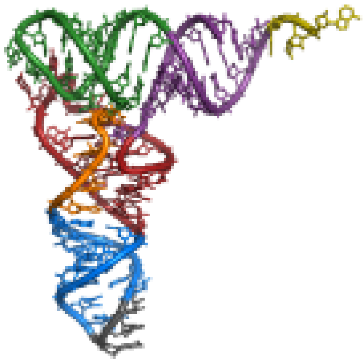


Figure: Geary et al., Science 2014

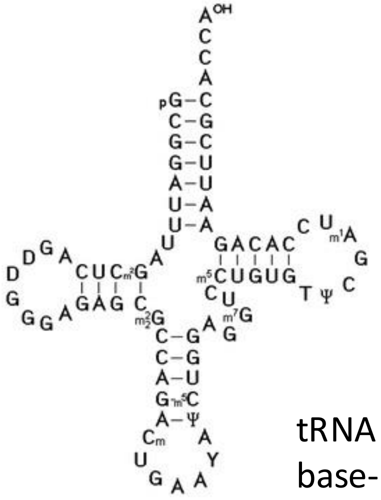
RNA basics: Levels of RNA structure

- *Primary structure*: linear sequence of bases, listed from the 5’ to 3’ end
  - Representation: linear string over alphabet {A, C, G, U}
- *Secondary structure:* pairing arrangement of bases in primary structure
  - Representations: base-pairing diagram, arc diagram
- *Tertiary structure:* actual [min-energy] 3D shape of a secondary structure
- (*Quaternary structure:* interactions across several molecules)

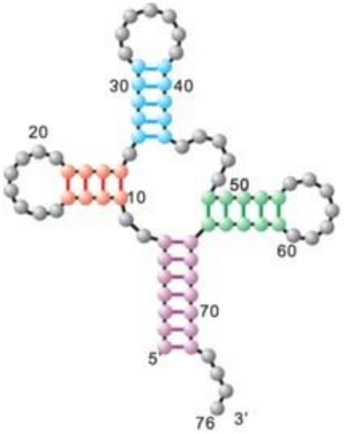
RNA basics: RNA secondary structure representations

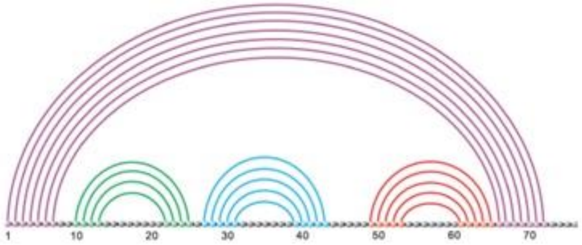


tRNA 3ary structure



tRNA 2ary structure base-pairing diagram





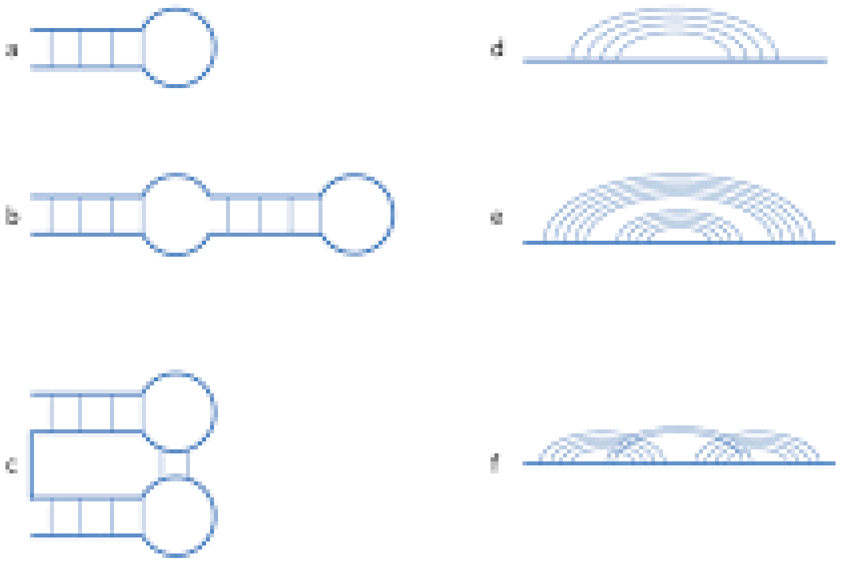
tRNA 2ary structure arc diagram

RNA basics: Structural motifs and representations

Base-pairing diagram

Arc diagram

Dot-bracket sequence



g ((((……))))

h ((((…((((……))))…))))

i ((((….[[))))…((((]]….))))

(a)+(d)+(g) Hairpin loop
(b)+(e)+(h) Hairpin with internal loop
(c)+(f)+(i) Kissing-loop pair (a type of “pseudoknot”)

RNA basics: A 180˚ kissing-loop pseudoknot

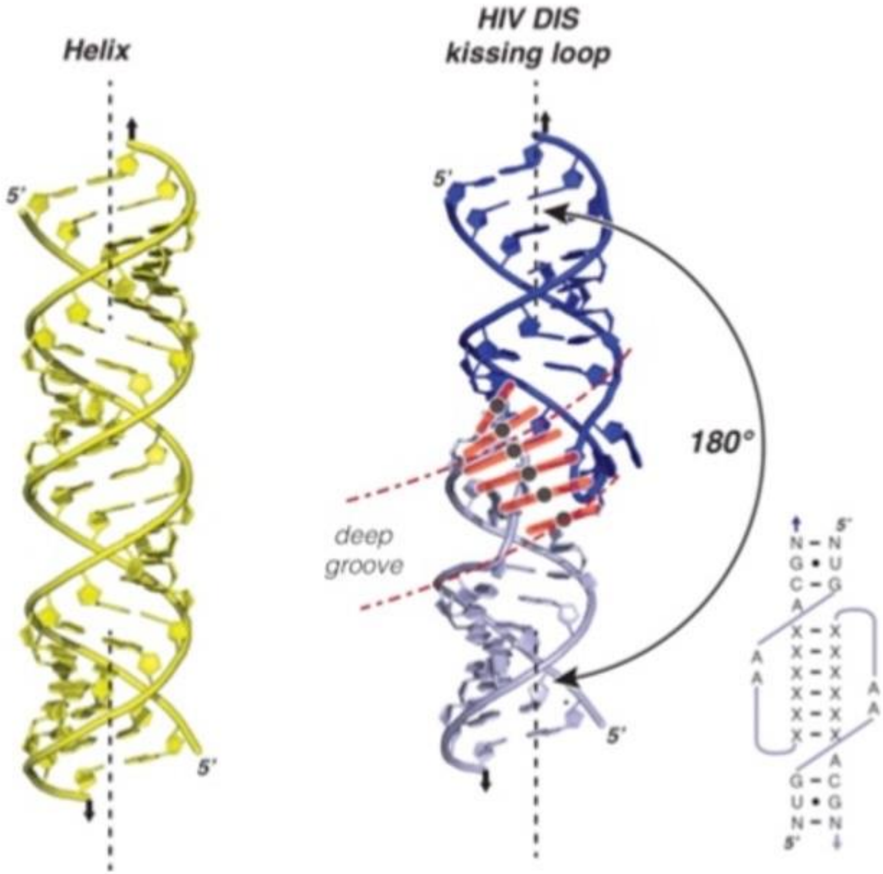
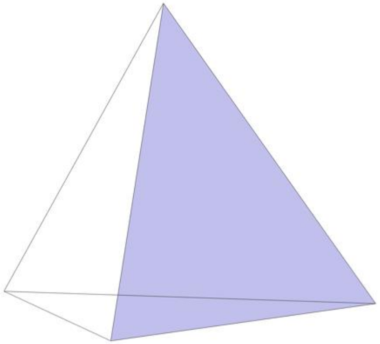
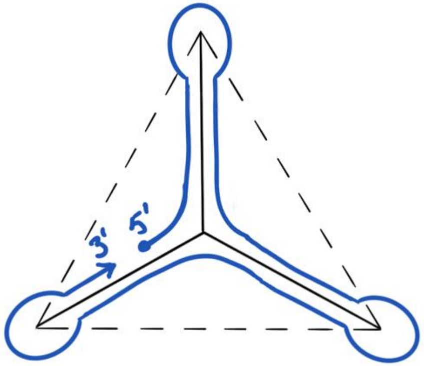


Figure: Geary et al., Science 2014

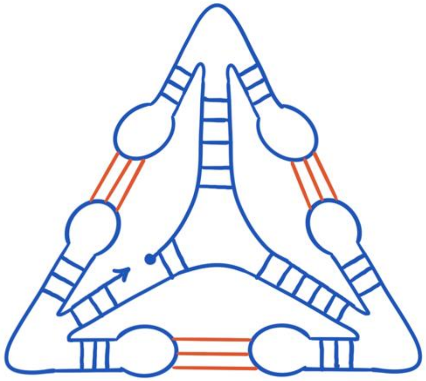
Design scheme for RNA origami wireframes



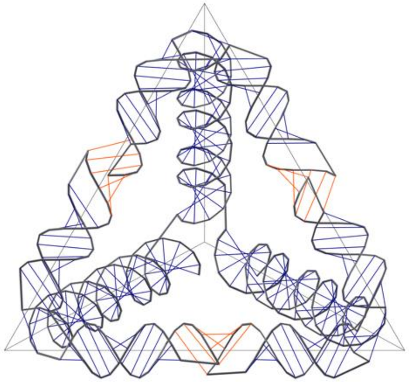
1. Target mesh model



2. Spanning tree & strand routing



3. Stem pairings, kissing loops



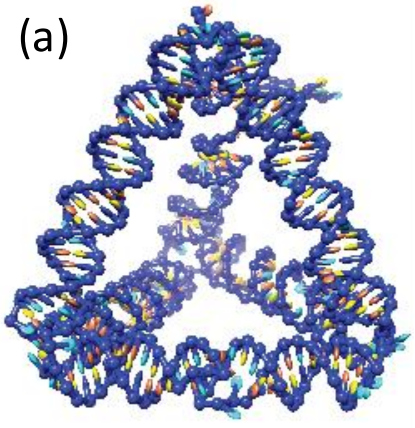
4. Helix diagram

Elonen et al.,
ACS Nano 2022

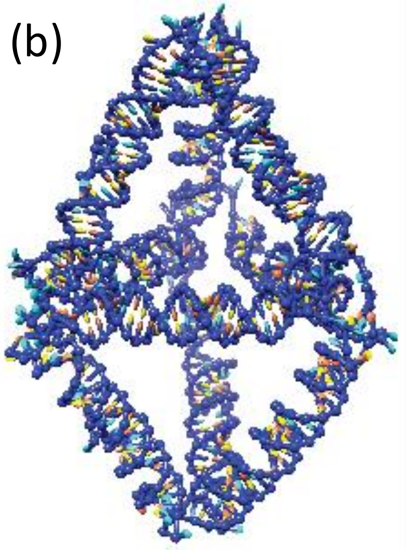
Experimental results

Three structures were designed by the ~~Sterna~~ DNAforge tool and synthesised in the laboratory:
(a) tetrahedron, (b) bipyramid, (c) triangulated prism

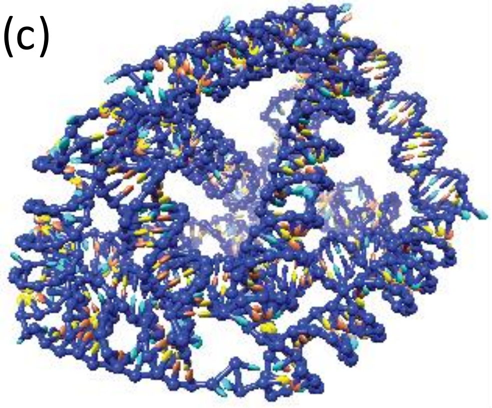
(a)



(b)



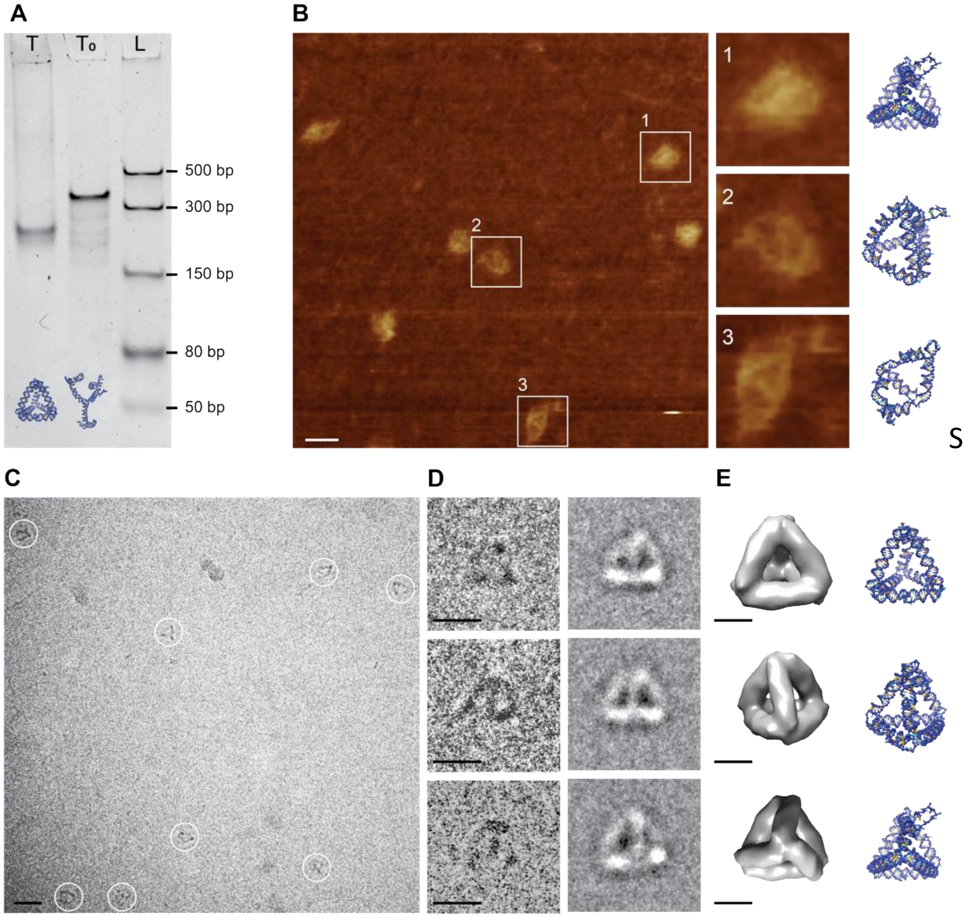
(c)



Elonen et al.,
ACS Nano 2022

Tetrahedron

Characterisation of the tetrahedron structure



Scalebar 20 nm

Scalebar 20 nm

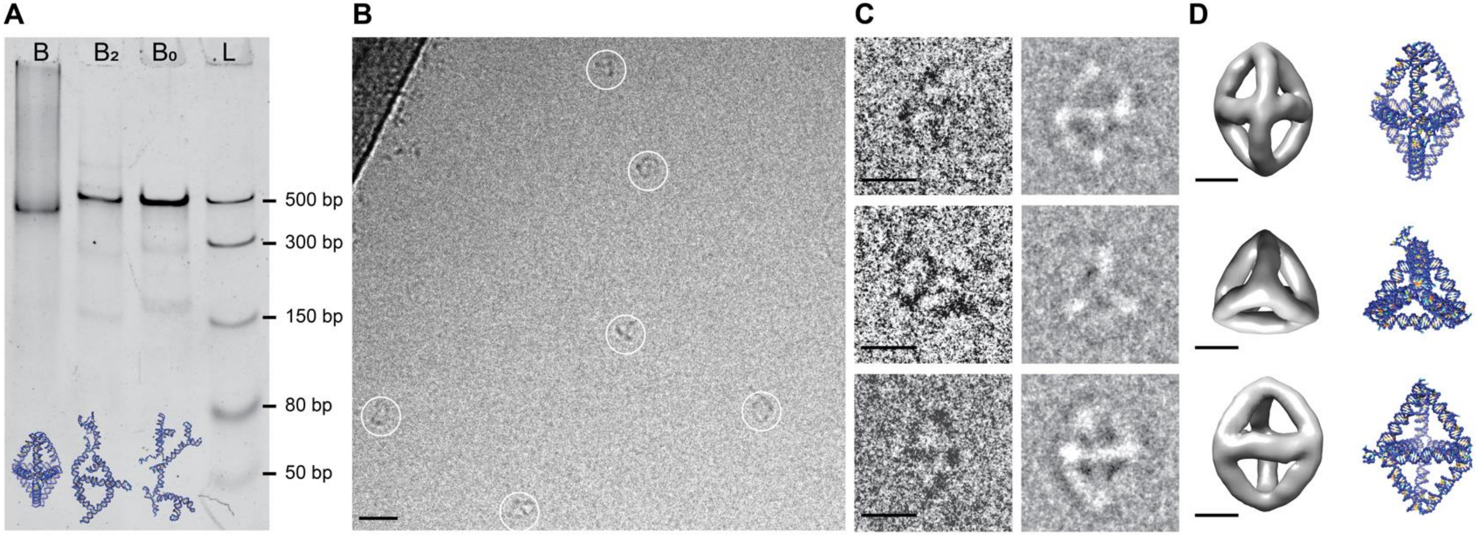
Scalebars 10 nm

Scalebars 5 nm

Elonen et al.,
ACS Nano 2022

Bipyramid

Characterisation of the bipyramid structure



Scalebar 20 nm

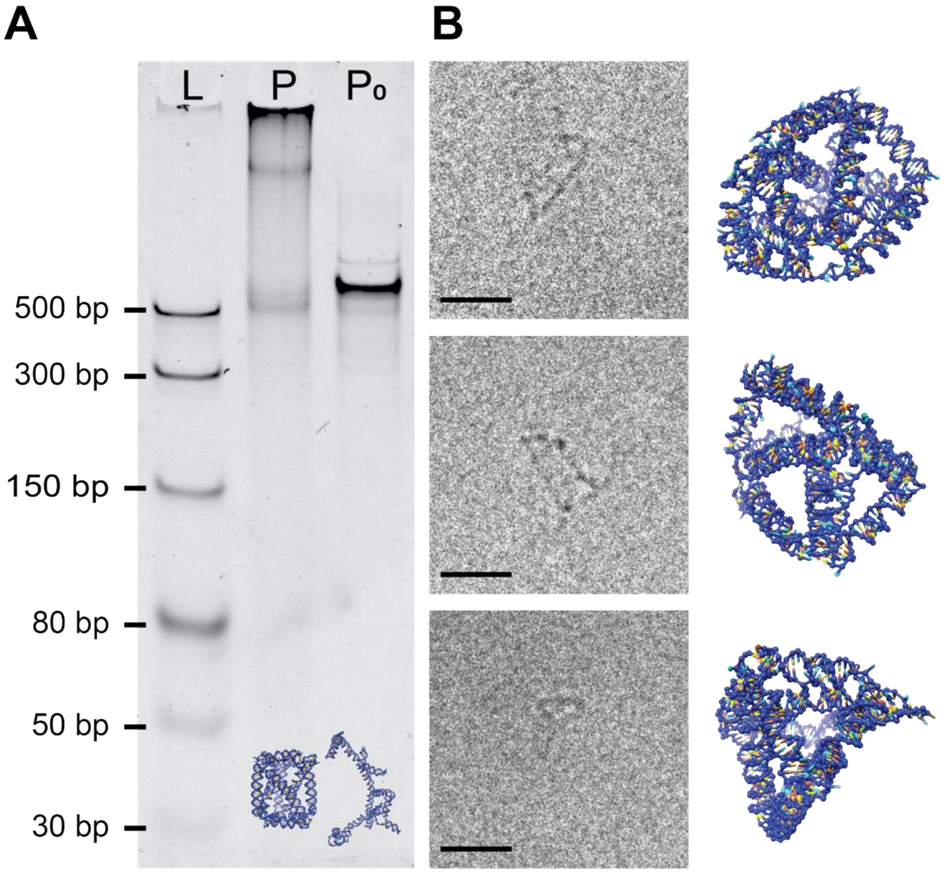
Scalebars 10 nm

Scalebars 5 nm

Elonen et al.,
ACS Nano 2022

Prism

Characterisation of the prism structure



Scalebars 20 nm

Elonen et al.,
ACS Nano 2022

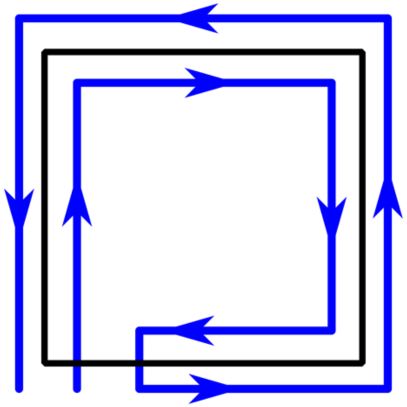
Problems with kissing loops

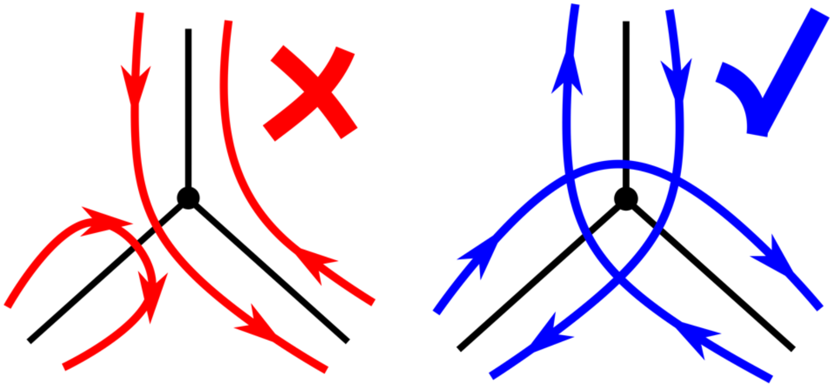
- Aggregation across particles
- Possible issues with kinetics and long-term stability
- Limited set of experimentally validated ones (~10)

Hence one might want to minimise the number of KL’s used in the designs.

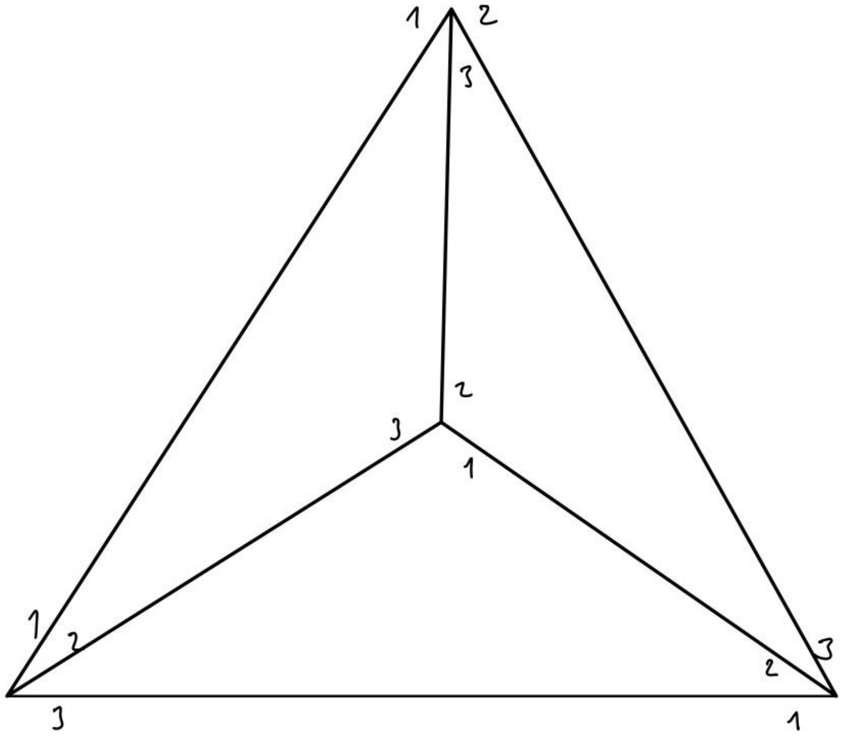
How about KL-free strand routing?

- Criteria for good RNA strand routing (”antiparallel strong trace”; Fijavž, Pisanski, Rus, MATCH 2014):
  1. Every edge in the design needs to be covered twice, in antiparallel directions
  2. Vertices must be stable

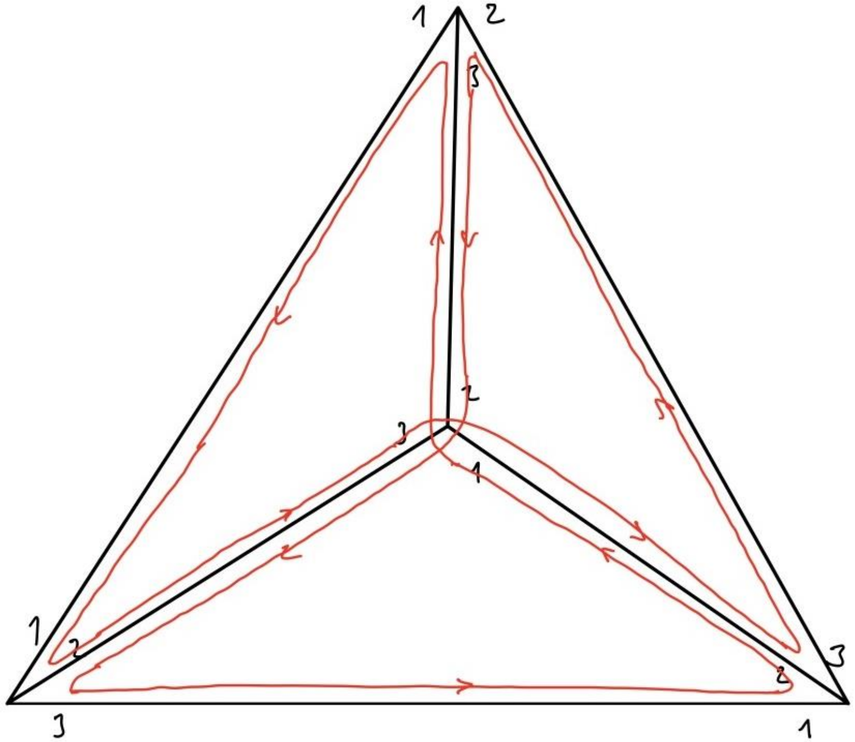




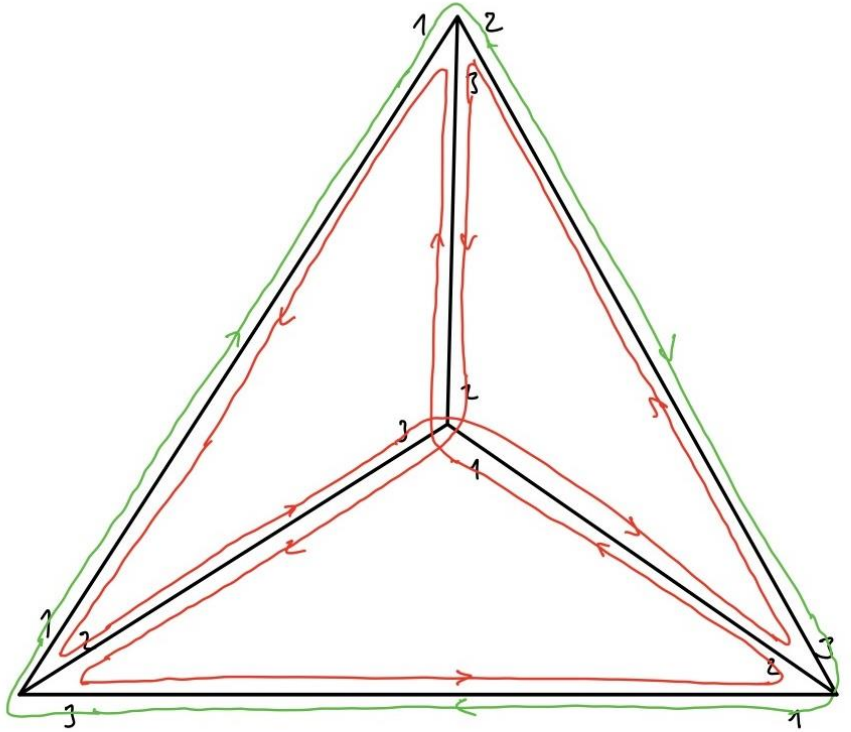
Let’s try the tetrahedron



Let’s try the tetrahedron



Let’s try the tetrahedron

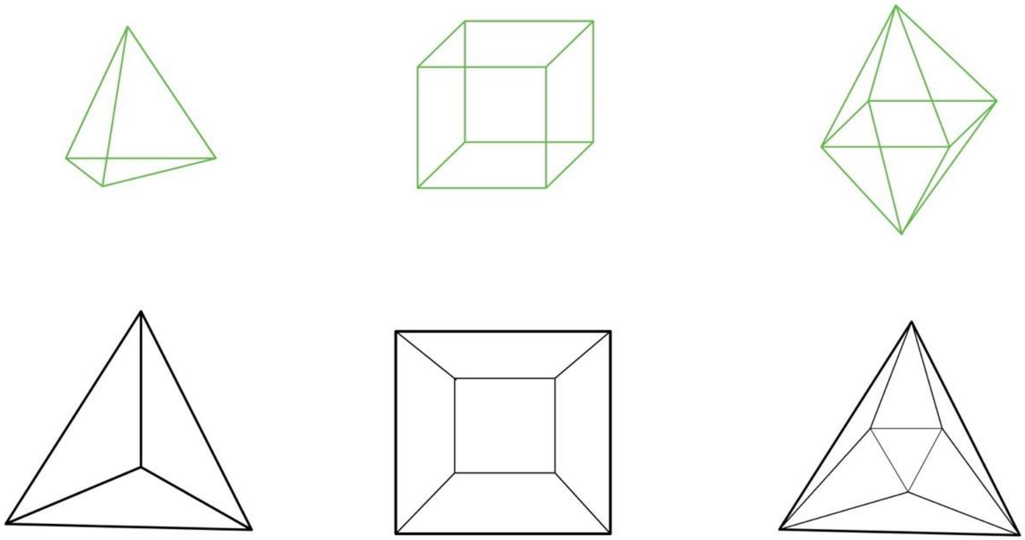


Need two strands!

- Klavžar & Rus (2013): A graph G admits an antiparallel “proper” trace if and only if $\delta(G) \geq 2$ and G has a spanning tree T such that each connected component of the co-tree $G \backslash T$ either (i) is of even size or (ii) contains a vertex of degree ≥ 4.
- Fijavž & al. (2014): A graph admits an antiparallel strong trace if and only if it has a *cellular 1-face embedding in some closed orientable surface*.

Graph embeddings

- In discussing 3D wireframe graph designs, it benefits mathematical accuracy to consider the model meshes as embedded in surfaces.
- E.g. polyhedral meshes can be viewed as embedded in genus-0 (sphere-equivalent) surfaces.
- Graphs of polyhedra are often presented as *Schlegel diagrams*, obtained by projecting the polyhedral mesh on a plane.
  - A Schlegel diagram of a convex polyhedral graph is always planar.



Background: definitions

- *Surface S* = topological space of dimension two = every point has a neighbourhood homeomorphic to a closed unit disk.
  - *Homeomorphism* = topological isomorphism = continuous bijection between topological spaces with a continuous inverse.
- *Genus* of *S* = number of nonintersecting cycles that can be drawn on *S* without separating it = number of "handles" or "3D holes" in *S.*
- *Graph embedding G → S* = continuous 1-1 mapping of *G* to *S* as a system of 0-D points and 1-D arcs connecting them.
- An embedding *G* → *S* divides *S* into *regions* or *faces*. If the interior of each region is homeomorphic to an open disk, the regions are called *cells* and the embedding a *cellular embedding.*

- A surface *S* is *closed*, if it is topologically compact and without boundary, and *orientable*, if there is a consistent sense of clockwise/ counterclockwise at each point, i.e. there is no embedding of the Möbius strip in *S*.
- *We shall only be considering closed orientable surfaces.*

Background: facts

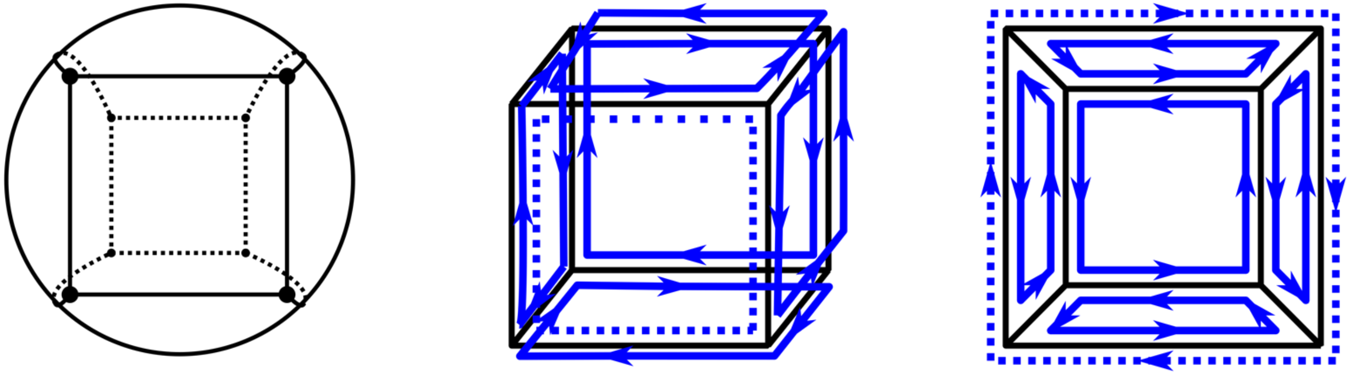
- *Classification theorem of closed surfaces*: Any two closed surfaces that have the same genus are homeomorphic, and vice versa.
- *Corollary*: Any closed orientable surface is homeomorphic either to the sphere (genus 0) or k tori sewn together (a ”k-hole donut” of genus k).

- Any embedding of a graph $G = (V, E)$ in a closed orientable surface S of genus $\gamma(S)$, with faces F, satisfies the *Euler polyhedral equation:*

$$|V| - |E| + |F| = 2 - 2\gamma(S)$$

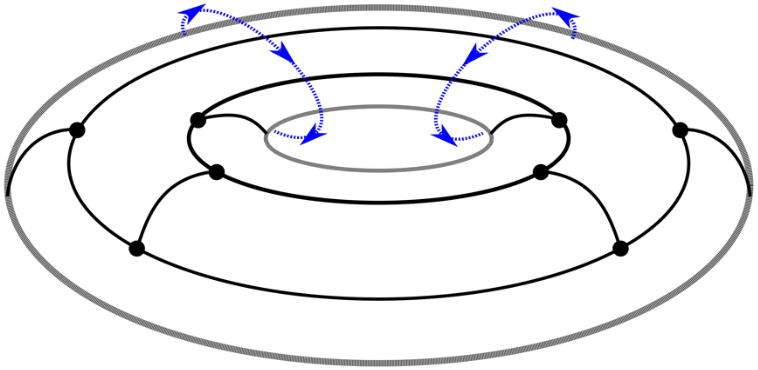
(Generalisation of Euler’s classic polyhedral formula $|V| - |E| + |F| = 2$.)

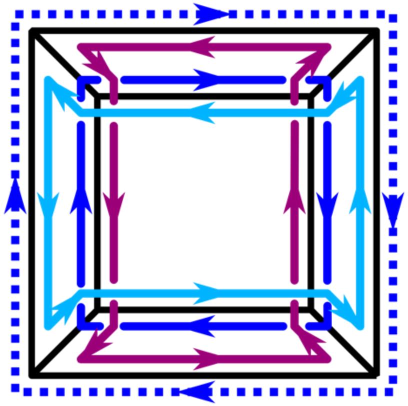
Embedding the cube in a sphere



- This is a 6-cell embedding and can be constituted, with stable vertices, of 6 RNA/DNA strands as presented.
- Criterion for stable vertices: at each vertex, the incident edges are arranged in some cycle (permutation), and this cycle is followed in the strand routing.
- Different permutation systems across the vertices induce different strand cycles, and accordingly different cellular partitions across the embedding surface. (Here the all-clockwise permutations are used, generating 6 cells corresponding to the cube's 6 faces on the sphere.)

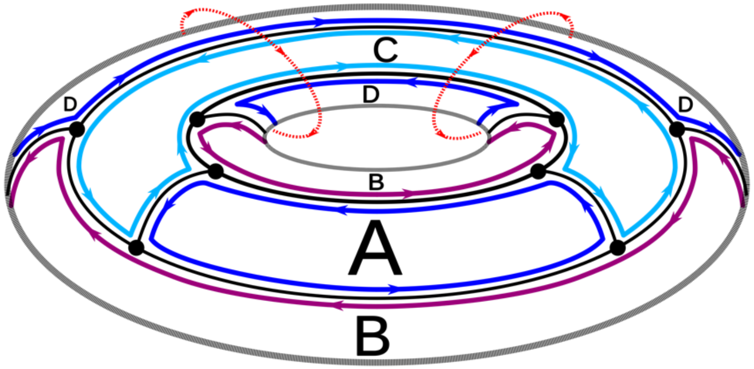
Embedding the cube in a torus

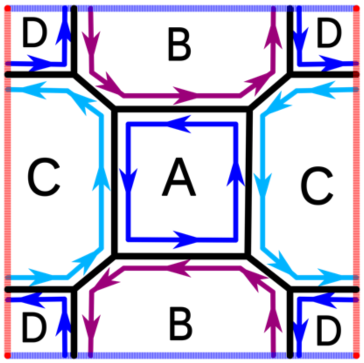




- Now there are only 4 cells and correspondingly 4 strand cycles!
- The cube is still perfectly constituted, but now from only 4 strands.

Embedding the cube in a torus





Torus diagram

- Now there are only 4 cells and correspondingly 4 strand cycles!
- The cube is still perfectly constituted, but now from only 4 strands.
- Different strand cycles are induced by different local edge-order permutations at the vertices.
- There is a 1-1 correspondence between the different permutation systems and antiparallel double-cycle covers of the graph’s edges.

Minimising the number of strands

- Note that in embedding the cube graph, both the sphere and torus embeddings satisfy Euler's equation:
  - Sphere: $|V| - |E| + |F| = 8 – 12 + 6 = 2 = 2 - 2\gamma(\text{sphere})$
  - Torus: $|V| - |\mathrm{E}| + |F| = 8 – 12 + 4 = 0 = 2 - 2\gamma(\text{torus})$
- Key lesson*: In an embedded graph, the number of faces is not a function of only the graph, but also the (genus of the) embedding surface.*
- By Euler's equation

  $$|V| - |E| + |F| = 2 - 2\gamma(S),$$

  to minimise the number of faces in an embedding of a graph *G*, and hence the number of the RNA strands needed to render the structure, one needs to find a *surface S of maximum genus* in which *G* can be embedded.
- (NB: Considering this problem will eventually help also in the task of minimising the number of kissing loops in RNA designs.)

Maximum-genus embeddings

- [Xuong (J. Comb. Theory 1979) and Furst et al. (J. ACM 1988)]
- Let $G = (V, E)$ be a connected graph and T a spanning tree of G. The edge-complement of T, $\mathrm{co}(T) = G \setminus T$, is a *co-tree* of G.
- Since any spanning tree of G has $|V| - 1$ edges, any co-tree of G has

  $$\beta(G) = |E| - |V| + 1$$

  edges. This is called the *Betti number* (or *cycle rank*) of G.
- The *deficiency* $\xi(G, T)$ of a spanning tree T of G is the number of odd-sized connected components in $\mathrm{co}(T)$. The deficiency $\xi(G)$ of a graph G is the minimum of this over all spanning trees T of G.

Theorem (Xuong 1979). The maximum embedding genus of a graph G is

$$\gamma_{\mathrm{M}}(\mathrm{G}) = (\beta(\mathrm{G}) - \xi(\mathrm{G})) / 2.$$

Corollary. A graph G has a 1-face embedding if and only if $\xi(G) = 0$.

Proof. Consider an embedding of G in a surface of genus $\gamma = \gamma_{\mathrm{M}}(G)$. In this embedding,

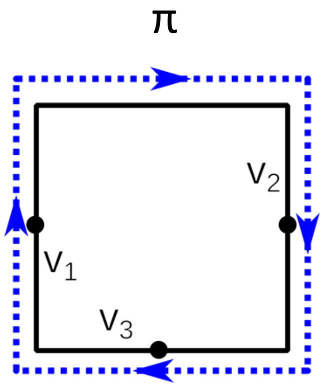
$$\boldsymbol{|F|} = |E| - |V| + 2 - 2\gamma = |E| - |V| + 2 - 2\cdot(E| - |V| + 1 - \xi(G))/2 = \mathbf{1 + \xi(G)}.$$

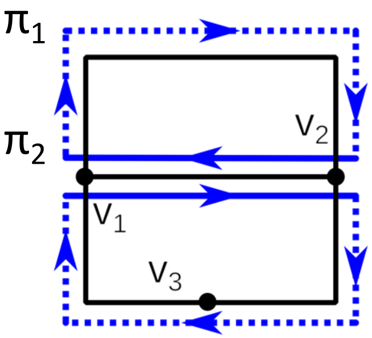
From maximising genus to minimising KL’s

- A *Xuong tree* for a graph *G* is a spanning tree *T** of *G* for which *co*(*T**) has a minimal number = $\xi(G)$ of odd-sized components.
- Xuong trees can be found efficiently by a reduction to the Matroid Parity problem for graphic matroids (Furst et al. 1988), for which a polynomial time algorithm exists (Stallman & Gabow, IEEE FOCS 1984).
- A minimal kissing-loop RNA design for a graph G can now be found as follows:
  1. Determine a Xuong tree *T** for graph *G*, together with the components of co(*T**)
  2. For each odd-sized component in *T**, remove one edge to be reconstituted later by a kissing-loop pairing, for a total of $\xi(G)$ kissing-loop pairs.
  3. Now all the components of *T** are of even size. Find an adjacency pairing for the edges in each component. (That is, group the edges in disjoint pairs so that the edges in each pair have a common vertex. This can be done by a simple depth-first search method.)
  4. Determine a 1-face (“twice-around-the-tree”) boundary walk for the tree *T** and merge all the adjacency pairs into this one by one by the process presented on the next slide. Complement with the $\xi(G)$ kissing-loop pairs.

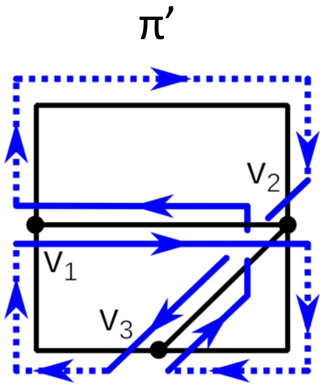
Adjacency pair augmentation

- [Xuong 1979, Furst et al. 1988]
- Let π be a 1-face boundary walk for a graph $G = (V, E)$ (= every edge in E is traversed twice by π, in antiparallel directions). Let $\{\{v_1,v_2\}, \{v_2,v_3\}\}$ be an adjacency pair of edges in $G \setminus E$.
- The pair can be merged to the circuit π in two steps (first $\{v_1,v_2\}$, then $\{v_2,v_3\}$), so that the result is a 1-face boundary walk π' for graph $G + \{\{v_1,v_2\}, \{v_2,v_3\}\}$.



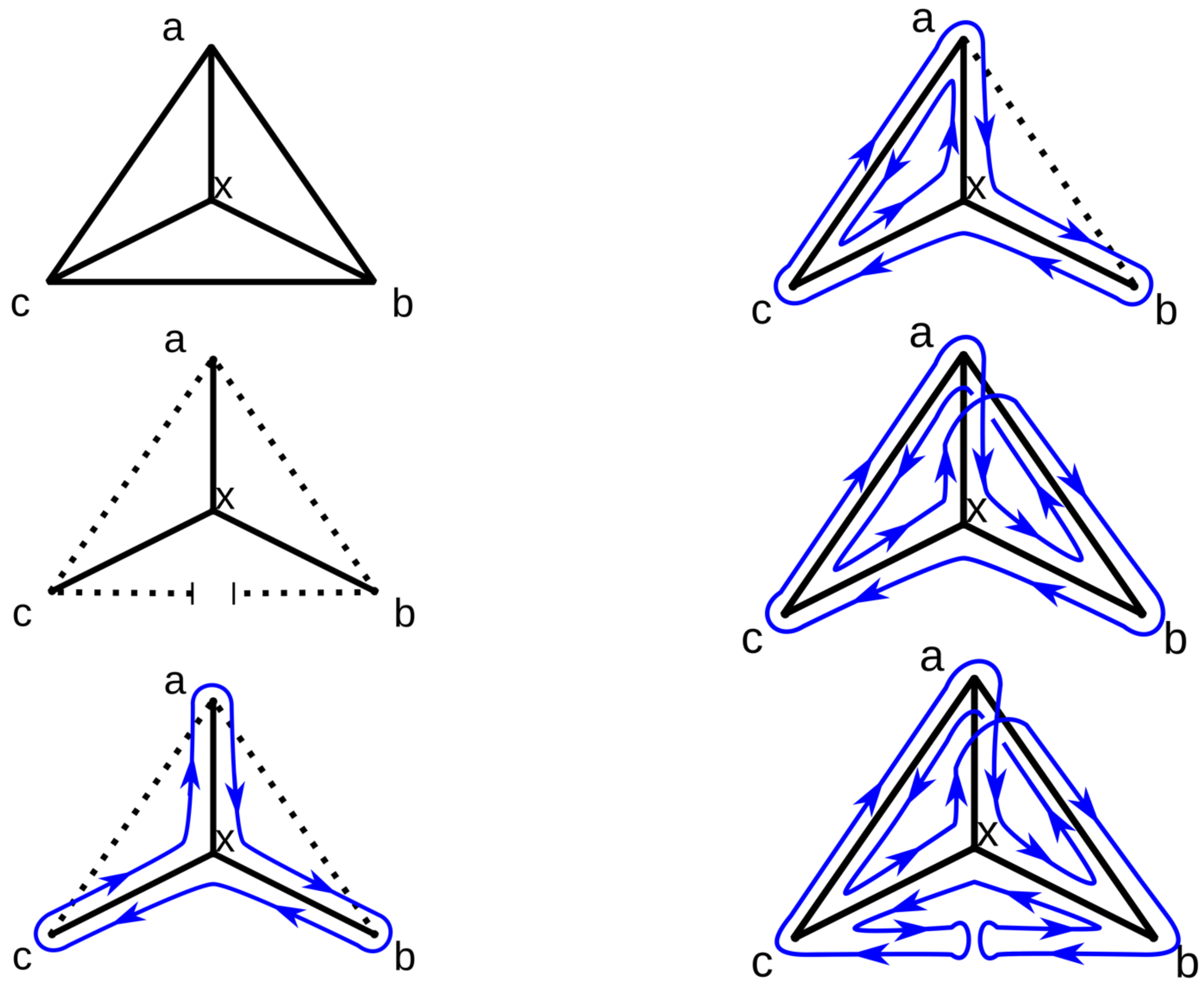


1. Adding $\{v_1,v_2\}$ splits π in two circuits π_1 and π_2



2. Adding $\{v_2,v_3\}$ merges the two circuits π_1 and π_2 into a new circuit π'

A 1-KL tetrahedron



Upper-embeddable graphs

- A graph G is *upper-embeddable* if $\xi(G) = 0$ or $\xi(G) = 1$, that is if G can be rendered in RNA with 0 or 1 kissing-loops.
- For instance, the following classes of graphs are known to be upper-embeddable:
  a. Locally connected graphs
     Locally connected = the (punctuated, induced) neighbourhood of every vertex is connected
  b. k-regular vertex-transitive graphs of girth $k \geq 4$ or $g \geq 4$
     Vertex-transitive = all local neighbourhoods of vertices are isomorphic
  c. Cyclically edge-4-connected graphs
     Graph contains two cycles that cannot be separated by removing fewer than 4 edges
  d. Simple graphs of diameter 2
     Simple = no loops, no multiedges

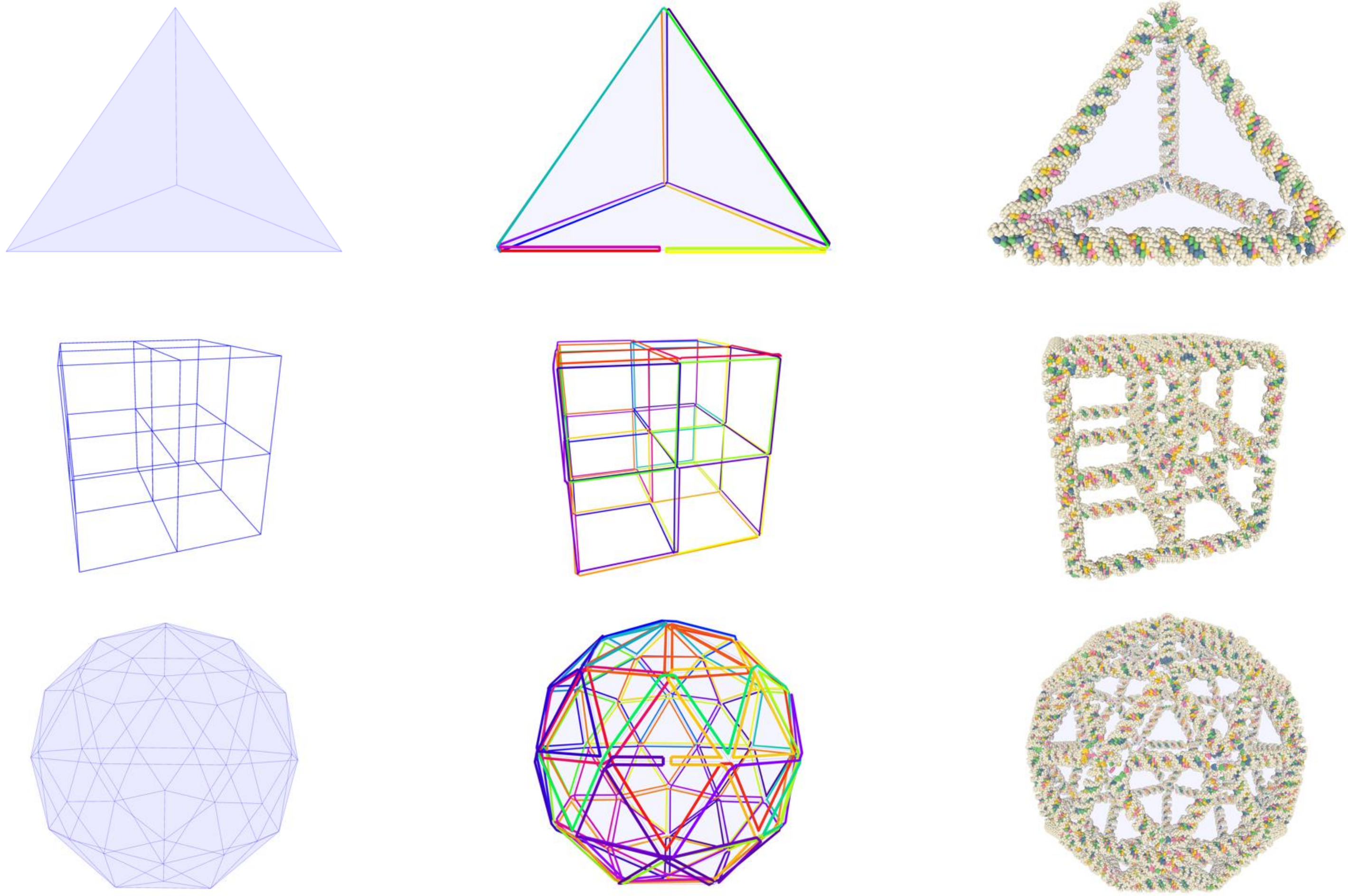
Corollary 1. All fully triangulated graphs are upper-embeddable.

Proof. By condition (a) above.

Corollary 2. All wireframes of Platonic solids are upper-embeddable.

Proof. Condition (a): Tetrahedron, Octahedron, Icosahedron. Condition (b): Cube, Dodecahedron.

The DNAforge design tool



https://dnaforge.org

Thank you for your attention!