The Mathematics of certain Polyhedral Self-Assemblies

Tomaž Pisanski (University of Primorska and University of Ljubljana), Roman Jerala (National Institute of Chemistry) and Nino Bašić (University of Ljubljana)

TBI Winterseminar 2015

Bled, February 18, 2015

Thanks

Thanks to organizers for inviting us!

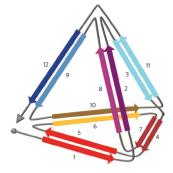
Thanks to all sponsors of my research!

Supported in part by the ARRS,





Motivation



TET12

In 2013 a group of Slovenian scientists and mathematicians under the leadership of Roman Jerala produced a polypeptide string that can self-assemble in the shape of a stable tetrahedron in such a way that each tetrahedral edge is composed of two intertwined peptide segments. They called the structure TET12. We give a mathematical interpretation of this remarkable bioengineering task. The model that best describes a self-assembly polyhedron comes from topological graph theory.

This model has been adequately described in a recent paper by G. Fijavž, T. Pisanski, J. Rus: Strong Traces Model of Self-Assembly Polypeptide Structures, MATCH Commun. Math. Comput. Chem. 71(2014) 199–212. However, there are several interesting question remaining. The model can be interpreted, on the one hand, as an Eulerian trail in a doubled skeleton graph of the corresponding polyhedron.

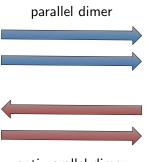
and on the other hand,

as a gluing process turning a fundamental polygon into a closed surface.

The design of such polypeptides requires a solution of several interesting combinatorial problems and problems of combinatorial optimization. Several questions concerning fundamental polygons are addressed.

高 ト イヨ ト イヨ ト

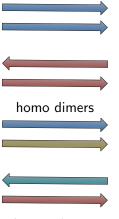
An Abstract Model



anti-parallel dimer

- The *word* alias *linear chain* alias *polypeptide* w has even length, say 2n.
- It contains *n* matching pairs or segments, called *dimers*.
- A dimer may be *parallel* or *anti-parallel*.
- A symbol (letter) appearing twice represents a parallel dimer.
- A letter appearing in lower and in upper case represents an anti-parallel dimer.^a

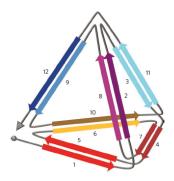
^aIn mathematics we may use a and a^{-1} instead of a and A. For human-computer interaction it was more convenient to use the former convention.



hetero dimers

In practice we have to distinguish homo-dimers from hetero-dimers. A homo-dimer is a pair of two identical segments while a hetero-dimer is a pair of different segments that attract each other. The problem of selecting an appropriate orthogonal set of segments resides outside of the current model. However, the model is sufficiently general that applies both to protein- and DNA- self-assemblies.

Tetrahedron TET12 as a single strand topofold.



TET12

In this model we may represent the original self-assembled tetrahedron TET12 as a word abcdAedbfeCf;

6 dimers: a,b,c,d,e,f,

2-antiparallel dimers: a A, c C,

4-parallel dimers; b b, d d, e e, f f.

The structure that we obtain in this way will be called a single strand *topofold*.

The protein is composed of 476 amino acids. [Note that there are 20 distinct amino acids. Each of them is encoded by a one-letter code.]

MYHHHHHHSR AGMKQLEKEL KQLEKELQAI EKQLAQLQWK AQARKKKLAQ LKKKLQASGP GSPEDEIQQL EEEIAQLEQK NAALKEKNQA LKYGSGPGDI EQELERAKAS IRRLEQEVNQ ERSRMAYLQT LLAKSGPGQL EDKVEELLSK NYHLENEVAR LKKLVGSGPG MKQLEKELKQ LEKELQAIEK QLAQLQWKAQ ARKKKLAQLK KKLQASGPGS PEDEIQALEE KNAQLKQEIA ALEEKNQALK YGSGPGQLED KVEELLSKNY HLENEVARLK KLVGSGPGSP EDKIAQLKQK IQALKQENQQ LEEENAALEY GSGPGSPEDE NAALEEKIAQ LKQKNAALKE EIQALEYGSG PGSPEDKIAQ LKEENQQLEQ KIQALKEENA ALEYGSGPGD IEQELERAKA SIRRLEQEVN QERSRMAYLQ TLLAKSGPGS PEDKNAALKE EIQALEEENQ ALEEKIAQLK YGSGTS

伺 と く ヨ と く ヨ と …

3

It was artificially designed from smaller ingredients that we identify and name as follows:

```
TeT12 = "MYHHHHHHSRAG" +
```

"MKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQA" + "SGPG" + "SPEDEIQQLEEEIAQLEQKNAALKEKNQALKYG" + "SGPG" + "DIEQELERAKASIRRLEQEVNQERSRMAYQTLLAK" + "SGPG" + "QLEDKVEELLSKNYHLENEVARLKKLVG" + "SGPG" + "MKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQA" + "SGPG" + "SPEDEIQALEEKNAQLKQEIAALEEKNQALKYG" + "SGP" + "QLEDKVEELLSKNYHLENEVARLKKLVG" + "SGPG" + "SPEDKIAQLKQKIQALKQENQQLEEENAALEYG" + "SGPG" + "SPEDENAALEEKIAQLKQKNAALKEEIQALEYG" + "SGPG" + "SPEDKIAQLKEENQQLEQKIQLKEENAALEYG" + "SGPG" + "DIEQELERAKASIRRLEQEVNQERSRMAYLQTLLAK" + "SGPG" + "SPEDKNAALKEEIQALEEENQALEEKIAQLKYG" + "SGTS"

・ 同 ト ・ ヨ ト ・ ヨ ト ・ ヨ ・

The structure of the protein

This can be further abbreviated as:
TET12 = start + APH + linker + P3 + linker + BCR + linker
+ GCNsh + linker + APH + linker + P7 + linker + GCNsh +
linker + P4 + linker + P5 + linker + P8 + linker + BCR +
linker + P6 + stop
where
P3 = "SPEDEIQQLEEEIAQLEQKNAALKEKNQALKYG"
P4 = "SPEDKIAQLKQKIQALKQENQQLEEENAALEYG"
P5 = "SPEDENAALEEKIAQLKQKNAALKEEIQALEYG"
P6 = "SPEDKNAALKEEIQALEEENQALEEKIAQLKYG"
P7 = "SPEDEIQALEEKNAQLKQEIAALEEKNQALKYG"
P8 = "SPEDKIAQLKEENQQLEQKIQALKEENAALEYG"
APH = "MKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQA"
BCR = "DIEQELERAKASIRRLEQEVNQERSRMAYLQTLLAK"
GCNsh = "QLEDKVEELLSKNYHLENEVARLKKLVG"
start = "MYHHHHHHSRAG"
<pre>stop = "SGTS"</pre>
linker = "SGPG"

For our purposes the substrings start, stop and linker are irrelevant.

For each of the 9 segments we use one-letter abbreviation:

- b = 3 = P3 = "SPEDEIQQLEEEIAQLEQKNAALKEKNQALKYG"
- b = 4 = P4 = "SPEDKIAQLKQKIQALKQENQQLEEENAALEYG"
- f = 5 = P5 = "SPEDENAALEEKIAQLKQKNAALKEEIQALEYG"
- f = 6 = P6 = "SPEDKNAALKEEIQALEEENQALEEKIAQLKYG"
- e = 7 = P7 = "SPEDEIQALEEKNAQLKQEIAALEEKNQALKYG"
- e = 8 = P8 = "SPEDKIAQLKEENQQLEQKIQALKEENAALEYG"
- a = A = APH = "MKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQA"
- c = B = BCR = "DIEQELERAKASIRRLEQEVNQERSRMAYLQTLLAK"
- d = G = GCNsh = "QLEDKVEELLSKNYHLENEVARLKKLVG"

The abbreviated encoding is given here:

T12 = "A3BGA7G458B6"

or for our purposes

T12 = abcdAedbfeCf

伺 ト く ヨ ト く ヨ ト

Definition

A word (topofold) is called *parallel* if all of its dimers are parallel (aa or AA). It is called *anti-parallel* if all of its dimers are anti-parallel (aA).

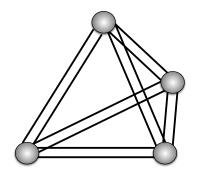
If a topofold is designed from proteins parallel dimers are preferred. On the other hand, if it is assembled from DNA only antiparallel dimers are possible.

Note that TET12 is neither parallel nor anti-parallel.

Question

Is there any other string that would self-assemble into an antiparallel (or parallel) tetrahedron?

The double trace model



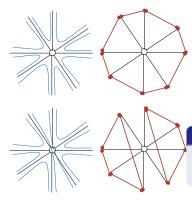
Take a polyhedron, double the edges of its graph and run a polypeptide chain as an Eulerian tour in the doubled skeleton. This is called a *double trace*. The original polyhedron graph is called the *skeleton* of the trace. Make sure the trace is **stable**.

What is a stable tetrahedron?



The double trace abcdeEfFDabc in the tetrahedron in the top figure is not stable. It breaks down at the bottom two vertices as shown in the middle figure. The actual skeleton in the bottom figure is not the original tetrahedron but a triangle with a pending tree.

Vertex-figure and stable double trace

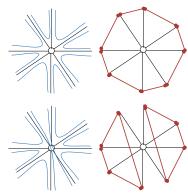


At each vertex v of the original polyhedron P we define a graph, called vertex-figure, on arcs (edges) incident with v. Two arcs a and bare adjacent if and only if the trace enters v using a and leaves it using b.

Definition

A double trace is stable if and only if each vertex-figure is connected.

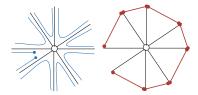
Vertex-figure and stable double trace



Vertex-figure of the top double trace is connected (and octagon): it is stable.

Vertex-figure of the bottom double trace is disconnected (a union of two four-cycles): unstable. The original vertex would split into two vertices.

Vertex-Figure: An Exception



There is exactly one vertex at which we begin and end our double trace. We may consider the vertex-figure at that vertex to be a path.

Proposition

In a stable double trace the vertex-figure at each vertex is either a cycle or a path. Finding a double trace is easy.

Theorem (L. Euler - 1735)

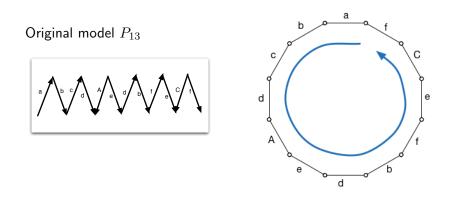
A graph has an Eulerian circuit if and only if it is connected and has each vertex of even valence.

Clearly every connected graph with doubled edges has even valence. Hence every connected graph has a double trace. This follows from the 18th century result of L. Euler but the actual algorithms for finding one were developed much later. However, finding a stable double trace is much more difficult.

A Different Model

Gašper Fijavž and I independently proposed a different model for the polypeptide TET12 = abcdAedbfeCf . Instead of a path consider a cycle.

New model: C_{12}



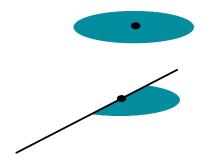
Observation

The model from the previous slide may be interpreted as a fundamental polygon of a closed surface.

What follows is an informal introduction to the theory of surfaces. We have to understand two concepts:

- closed surface
- fundamental polygon

What is a surface?



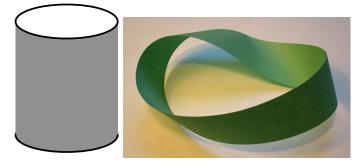
A surface is a topological space in which each point has a neighborhood homeomorphic to an open disk or a half of the disk. In the former case the point lies in the interior of the surface, in the latter it belongs to the boundary. A surface is *closed* if it compact and has no boundary. Surfaces can be classified according to several criteria.

- compact vs. non-compact;
- having boundary vs. having no boundary points;
- orientable vs. non-orientable.

Example

A plane is non-compact, orientable surface without boundary.

Classification of surfaces - Möbius strip



Example

A finite cylinder is a compact, orientable surface with boundary. A Möbius strip is a compact non-orientable surface with a single boundary component. In topology a *closed surface* is a compact topological space in which each point has a neighborhood homeomorphic to an open disk.

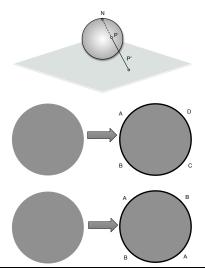
Closed surfaces are of two types:

- Orientable: sphere, torus, double torus, triple torus, ... S_g a sphere with g handles attached.
- *Non-orientable*: projective plane, Klein bottle, ... N_k a sphere with k Möbius strips attached.

Each other surface is a subspace of a closed surface (not in a unique way!).

Compactification

Each surface is a subspace of a closed surface (not in a unique way!).



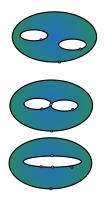
E.g. a *plane* can be compactified to the *sphere* (one point compactification) by the inverse of the stereographic projection +addition of the north pole N. A plane is homeomorphic to an open disk. It can be compactified to a *closed disk* or to the projective plane (add line at infinity and one point at infinity for each pencil of parallel lines).

Pseudosurfaces!

A subspace of a surface need not be a surface itself!

Definition

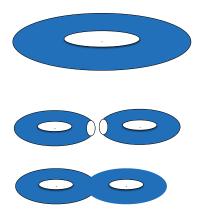
A pseudo surface is obtained from surfaces by a finite number of point identifications.



Top figure: Disk with two holes: a surface with three boundary components. *Middle figure*: A pseudo surface is a subspace of disk.

Bottom figure: By yellow point identification we obtain a pseudo surface in the middle figure.

Connected sum of surfaces



Attaching a handle is a special case of *connected sum of surfaces*. Double torus is a *connected sum* of two tori. Klein bottle is a *connected sum* of two projective planes. It is obtained from a sphere by attaching two Moebius strips. For maps on closed surfaces (i.e. cellular embeddings of graphs) having v vertices, e edges and f faces, the following equations hold, where χ is the Euler characteristic of the surface:

•
$$v - e + f = 2$$
 - sphere (convex polyhedron)

•
$$v - e + f = \chi = 2 - 2g$$
 - orientable genus g surface ($g \ge 0$).

•
$$v - e + f = \chi = 2 - k$$
 - non-orientable genus k surface $(k > 0)$.

Proposition

Euler characteristic of an orientable surface is even.

Symmetries of a surface are abstract, do not depend on the way the surface is embedded in space. Orientable surfaces come in two oriented forms. Two fundamentally different types of symmetries:

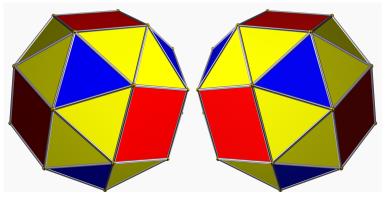
- orientation-preserving symmetry
- orientation-reversing symmetry

Any orientation-reversing symmetry of a surface transforms a map ${\cal M}$ to its mirror image ${\cal M}^*.$

Definition

A map M is chiral if no orientation-preserving symmetry of its surface transforms it into its mirror image M^* .

Chirality - Snub cube



Snube cube is vertex-transitive (uniform) chiral polyhedron. It comes in two oriented forms.

Unfortunately, any realization of a one chiral form is mathematically indistinguishable from its opposite orientation. In other words using our abstract model, we cannot distinguish between the two. It would be interesting to see if this is also the case in practice.

Definition

Fundamental polygon is an even sided polygon in which edges are pairwise identified.

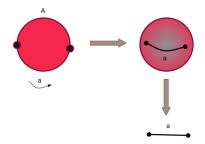
Recall the definition of the single strand *topofold* composed of lower-case and possibly upper-case letters!

Here are some standard fundamental polygons w:

aA	sphere
aa	projective plane
abAB	torus
aabb	Klein bottle
abABcdCD	double torus

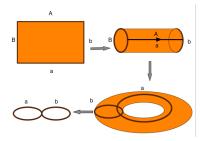
Note that the fundamental polygon does not change if we replace the word w by its cyclic shift or reversal.

Let the word be w = aA. The corresponding fundamental polygon is a digon. If we glue the sides in the antiparallel way the disk is zipped into a sphere. A scar is a graph with two vertices and one edge between them.



$w = abAB \dots$ The Torus and its Skeleton

Let the word be w = abAB. The corresponding fundamental polygon is a square. If we glue the opposing sides in the antiparallel way the disk first turns into a cylinder and then into a torus. A scar is a graph with a single vertex having two loops. This is the skeleton.



The surface is obtained from the fundamental polygon by gluing pairwise its sides.

Theorem

The surface determined by w is orientable if and only if each letter appears in w once in lower and once in upper case.

The Skeleton Determined by the Fundamental Polygon

We may monitor each step by considering its graph, called the *skeleton*. We start with a cycle (fundamental polygon) and obtain by successive vertex- and edge-identifications the skeleton at the end.

word	surface	v	е	f	χ	genus	Orientable
aA	sphere	2	1	1	2	0	yes
aa	projective plane	1	1	1	1	1	no
abcabc	projective plane	3	3	1	1	1	no
abAB	torus	1	1	1	0	1	yes
abcABC	torus	2	3	1	0	1	yes
aabb	Klein bottle	1	1	1	0	2	no
abABcdCD	double torus	1	2	1	-2	2	yes
abcdAedbfeCf	TET12	4	6	1	-1	3	no

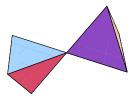
Advantage of the fundamental polygon model

Theorem

The double trace of the skeleton arising from a fundamental polygon is stable.

Follows from the definition of the surface. Unstable double traces may be interpreted in terms of pseudosurfaces.

Note: a pseudosurface is obtained from surfaces by a finite number of point identifications.



To each fundamental polygon w we associate a pair of numbers (p,q) where p+q=n and p denotes the number of parallel and q denotes the number of anti-parallel dimers. Clearly, the surface is orientable if and only if p=0.

Let P be a polyhedron with v vertices and e edges that is a skeleton of w. If |e-v| is even then the surface associated with w is non-orientable. This follows from the fact that $v-e+1=\chi$ and the Euler characteristic is odd.

Corollary

Tetrahedron does not admit an anti-parallel polypeptide.

Hence, it cannot be realized as a DNA single strand topofold!!!

Characterization of graphs admitting parallel and anti-parallel polypeptide realization

In a recent paper of Fijavž, Rus and TP (2014) the following has been shown:

Theorem

A graph admits an anti-parallel realization if and only if it admits a one-face orientable embedding.^a

^aKnown that it can be checked in polynomial time.

Characterization of graphs admitting parallel and anti-parallel polypeptide realization

It has been also shown:

Theorem

A graph admits a parallel realization if and only if it is Eulerian.

This was recently generalized by a Slovak mathematician Martin Škoviera:

Theorem (Škoviera)

For any realization of a graph, the subgraph on parallel edges is even. Furthermore, the directions determined by parallel traversal determine a balanced digraph. For any connected graph X a fundamental polygon F having X as a skeleton is therefore its *realization*. As noticed by Škoviera F also determines an even subgraph Y of X. This leads to an interesting problem.

Problem

Let X be a connected graph and Y some of its even subgraphs. Does there exist a fundamental polygon F with skeleton X and parallel edges defining Y?

The answer is yes if Y = X and X is Eulerian. However, the answer is no if Y has no edges and X admits no anti-parallel realizations. There are no non-trivial counter-examples for the tetrahedron, but there exists some for some other polyhedra.

It turns out that the tetrahedron works. It has two non-trivial even subgraphs:

- triangle C₃ three anti-parallel dimers
- four-cycle C_4 two anti-parallel dimers (TET12)

Both are realizable!



Which self-assembled polyhedra exist?

The answer follows from an old result of J. Edmonds (1960):

Theorem

Each connected graph is a skeleton of some fundamental polygon.

Corollary

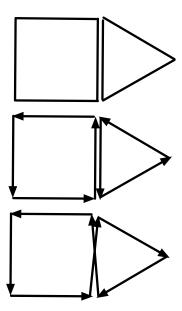
Each connected graph is a self-assembling realization of some polypeptide.

Corollary

Each convex polyhedron is a self-assembling realization of some polypeptide.

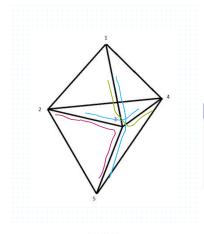
Take a connected graph G and any of its 2-cell embeddings determined by some rotation projection. If there is only one face, we are done and have a 1-face embedding. Otherwise select an edge belonging to two faces and make a *cross along an edge*; i.e. glue the two faces together by tracing first one face and then the other one in the reverse direction. This decreases the number of faces by one. After a finite number of steps we are done.

Glue Two Faces Along a Common Edge



On the left we have two faces with a common edge: a square abcd and a triangle AEF. We may glue them together into a single face abcdafe

No crossing realization



BAMBOO PA

We were trying to answer the following question airing from practice:

Question

Which convex polyhedra admit polypeptide realization in such a way that the chain does not cross itself at a vertex?

Image: A = A

Here we give a positive answer to this question:

Theorem

Given any 2-cell embedding of connected graph G, we may turn it into a 1-face embedding without changing any of the local rotations.

Proof follows directly from the construction on one of the previous slides.

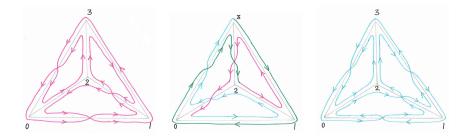
Corollary

The self-assembling of a polypeptide in a shape of a convex polyhedron can be achieved on the sphere in such a way that the polypeptide does not cross itself at any of its vertices.

伺 ト イヨト イヨト

How many polypeptides determine a polyhedron?

The answer to this question is a bit tricky. One may show that there are exactly three essentially different fundamental polygons determining tetrahedron: 'beDbcfEcadFa', 'eDbeFCBAfDAC', and 'eDbcfEBAfDAC'. All others are obtained by simple letter substitution.



However, when choosing linear chains more choices are available. In principle, we could choose different origins and different orientations. Hence each fundamental polygon gives rise to $2 \times 12 = 24$ polypeptide templates¹, 72 in total. For the tetrahedron each cyclic string is isomorphic to its reverse and only the first string has 3 automorphisms, the total number of non-equivalent polypeptide templates is (12/3 + 12 + 12) = 28.

¹In a template a symbol is just a place-holder for a peptide segment. = → = → へ Tomaž Pisanski (University of Primorska and University of Ljublic The Mathematics of certain Polyhedral Self-Assemblies For small polyhedra (or graphs) the following algorithm applies.

- Take the set of faces of a convex polyhedron.
- E.g. doubletetra = (adF, beD, cfE, agI, bhG, ciH)
- Recursively try out all glueings of all pairs of edges in different faces until you get a single face.
 - Record the solution.
- For each solution take any starting position.
 - Discard solutions that are not in canonical form with respect to the automorphisms.
 - Compute the canonical form of the reversed string and assign pointers from one to another.

高 と く ヨ と く ヨ と

For a graph with m edges its running time is $O(m2^m)$.

Observation

Let w be a polypeptide template consisting of n dimer templates with p parallel and q anti-parallel dimer templates, h homo-dimer and e hetero-dimer templates. Then n = p + q = h + e and w gives rise to

$$s(w) = p!q!2^e$$

polypeptides. In case e = 0 or if we do not care about the order of hetero-dimers we get the value:

$$r(w) = p!q!$$

Number of solutions for tetrahedron

Note that

$$S(w) = T(w)s(w)$$
$$R(w) = T(w)r(w)$$

id	w	r-id	par	anti-par	homo	hetero	autos	T(w)	S(w)	R(w)
0	beDbcfEcadFa	0	3	3	?	?	3	4	?	144
1	eDbeFCBAfDAC	1	4	2	3	3	1	12	4176	522
2	eDbcfEBAfDAC	2	3	3	?	?	1	12	?	432
	TOTAL							28	?	1098

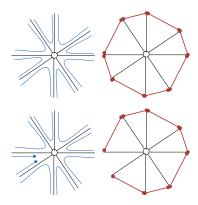
Table : Enumeration of tetrahedral polypeptides and their templates. There are 28 essentially different strings (templates). By permuting separately parallel and antiparallel symbols 1098 solutions are obtained. If one would distinguish the order in which the three hetero-dimers appear in the case of two antiparallel pairs, the number of solutions increases to 4176. Suppose we are building a stable double trace from several linear chains of polypeptides (multi-strand construction).

Question (Jerala)

Given an arbitrary convex polyhedron. What is the maximum number of linear chains that may self-assemble into a stable double trace?

Theorem

Given an arbitrary convex polyhedron P. The maximum number of linear chains that may self-assemble into a stable double trace is bounded by above by the number of vertices of P. For each P the maximum may be attained.

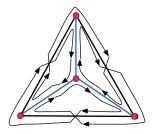


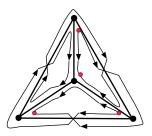
The vertex-figure of a stable trace must be connected at each vertex. It is therefore either a cycle or a path. The maximum is achieved if and only if all vertices have paths as vertex-figures. We have shown that for each

P a stable solution with cyclic vertex-figures exists. Now cut such a trace exactly once at every vertex. There will be vchains obtained in this way.

Parallel Stable Tetrahedron Exists!

It turns out that a parallel stable tetrahedron can be achieved if more than one chain is used.





Single chain stable double trace yielding tetrahedron with three antiparallel edges.

If the chain is cut at four corners and partially reversed a stable double trace with four chains yielding tetrahedron with all edges parallel is obtained. This observation leads to the following problems:

Problem

Classify graphs or convex polyhedra that admit stable parallel realization with multiple polypeptide chains.

Problem

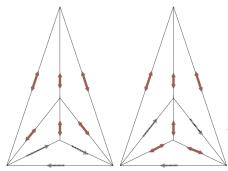
Classify graphs or convex polyhedra that admit stable anti-parallel realization with multiple polypeptide chains.

Theorem (The Brijuni Theorem)

Every connected graph admits a stable parallel realization with multiple polypeptide chains.

A B > A B >

Any finite connected graph G on n vertices has an even number, say 2k, odd vertices. By paiwise connecting these vertices with knew edges, an Eulerian graph X is obtained. By a theorem from our slides X admits a single-strand parallel stable trace. By removing the k edges we obtain a discontinuous stable parallel double trace in the original graph G with 2k beginnings and ends, hence the number of strands is k and $2k \leq n$.



Find a stable double trace with parallel triangle and the rest anti-parallel in a trigonal bipyramid. There are two cases: a facial or non-facial triangle.

Thanks for your attention.

Hope to see at least some of you in Kranjska Gora! 8th Slovenian International Conference on Graph Theory June 21 – 26, 2015 Kranjska Gora, Slovenia