Computational methods for polypeptide origami design

Daniel Silađi

Mentor: Andrej Brodnik
Co-mentor: Rok Požar

Introduction

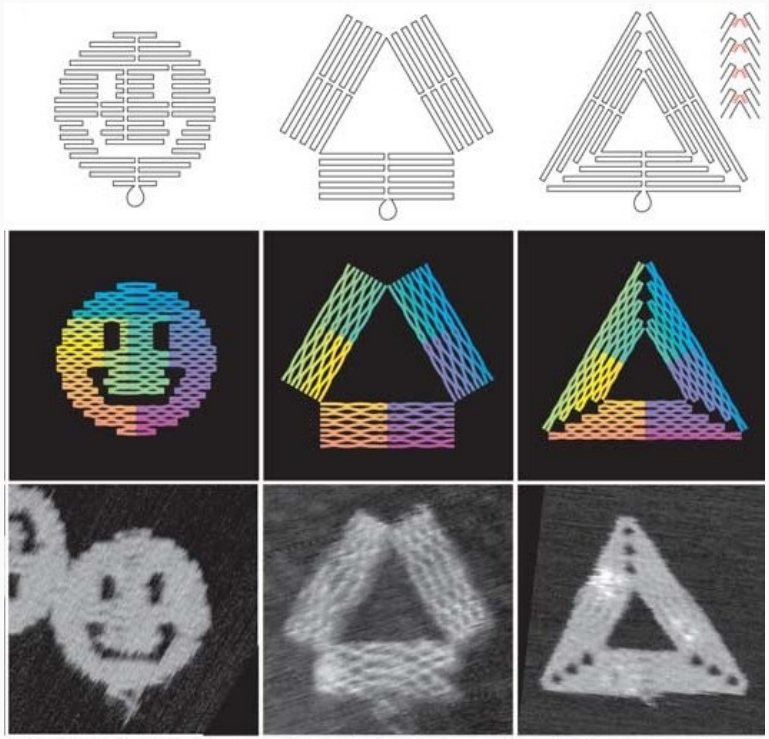
Talk structure

- Mathematical and biochemical background
- Interaction graphs and orthogonal sets
- An exact algorithm for the maximum orthogonal set problem
- Some heuristics for the maximum orthogonal set problem
- Results and conclusions

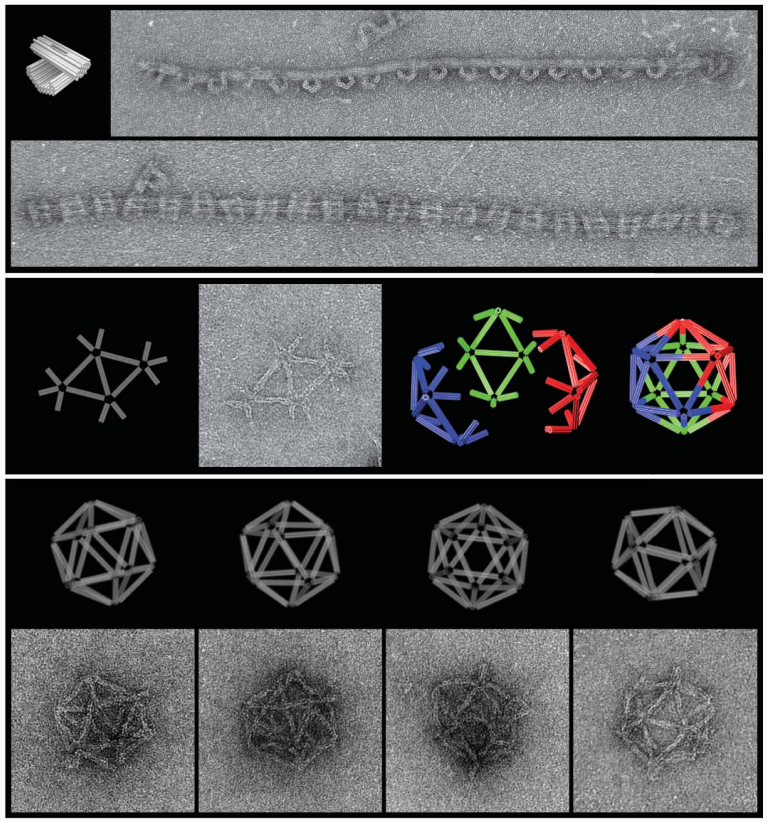
Biochemical background

Inspiration – DNA origami

By prescribing the sequence of bases in a single strand of DNA, we can build complex 2D and even 3D nanostructures



(a) 2D DNA origami, from the original paper, [10]



(b) 3D DNA origami from a follow-up paper, [3]

From DNA origami to polypeptide origami

- Advantages of using polypeptides over DNA
  - Custom DNA synthesis becomes expensive very quickly
  - More diversity in the available number of functional groups (20 amino acids vs 4 nucleotides) [7]
  - Possibility of in-vivo production and folding
- We need a class of peptides that is both flexible and understood well-enough, so that we can predict whether and how will they bind

Coiled coils

- Building blocks are heptads (7 amino acids), positions within heptad denoted with *abcdefg*
- Empirical models developed for estimating the interaction free energy ("interaction score"), based solely on primary structure [5, 9]

$$\text{score} = \text{weights} \cdot \text{features}$$

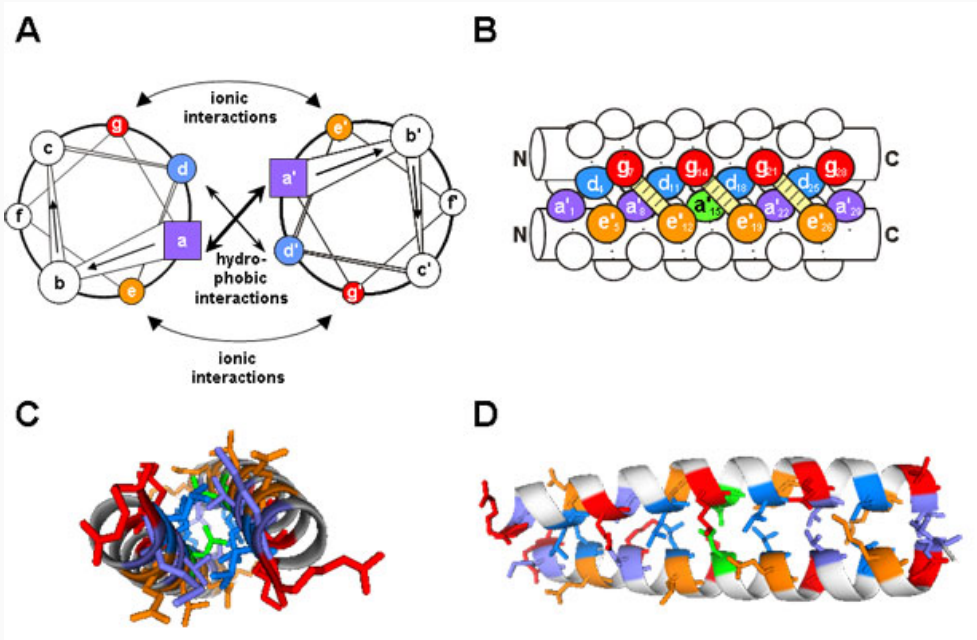


Figure 2: Different views of coiled-coil dimers

Single-chain polypeptide tetrahedron [6]

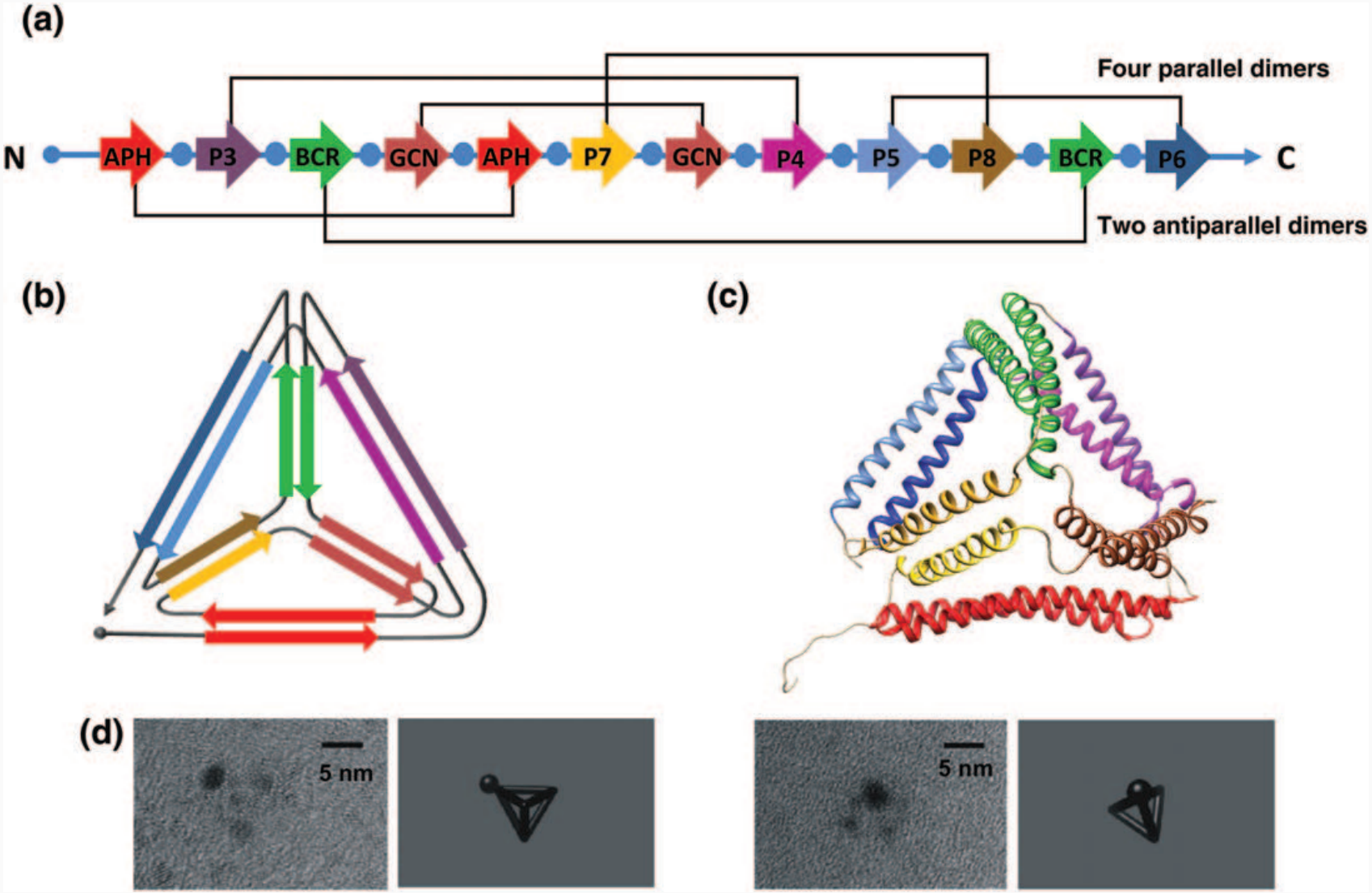


Figure 3: Construction steps for the tetrahedron

Workflow

1. Design the polyhedron
2. Find a suitable double trace of the polyhedron graph
3. Choose a set of peptides to be placed along the edges
4. Synthesize the peptide chain
5. Validate the design experimentally

Workflow

1. Design the polyhedron
2. Find a suitable double trace of the polyhedron graph
3. Choose a set of peptides to be placed along the edges
4. Synthesize the peptide chain
5. Validate the design experimentally

Workflow

1. Design the polyhedron
2. Find a suitable double trace of the polyhedron graph
3. Choose a set of peptides to be placed along the edges
4. Synthesize the peptide chain
5. Validate the design experimentally

Orthogonal sets

Preliminaries

- Let G be a polyhedron graph with n vertices and m edges, that we want to realize as a single chain – we need m pairs of peptides that interact only mutually
- We search for these pairs inside a (large) set of admissible peptides, A. The interactions between them are represented as a matrix M – the *interaction matrix*.
- Given $c_s, c_w \in \mathbb{R}$, construct the *interaction graph* $G_i = (V, E, E_s)$
  - i) $V = A$ (the set of peptides);
  - ii) $E = \{\{i, j\} | M_{ij} \leq c_w\}$ (the set of all interacting peptide-pairs)
  - iii) $E_s = \{\{i, j\} | M_{ij} \leq c_s\}$ (the set of all strongly interacting peptide-pairs/edges)

Interaction graphs

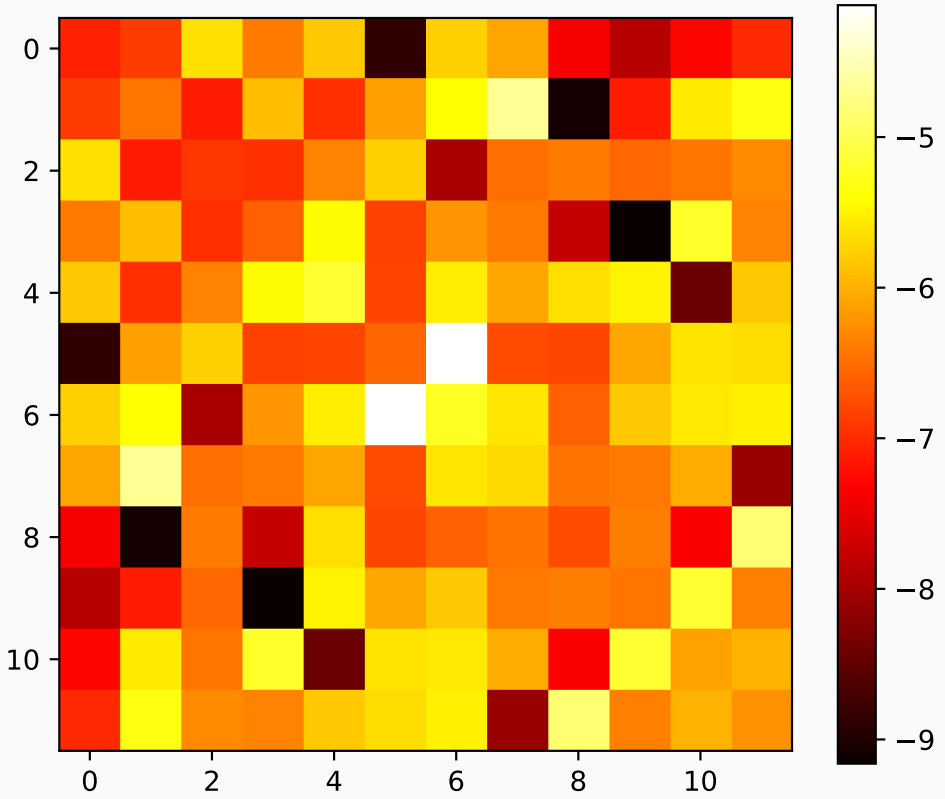


Figure 4: Full interaction matrix

Interaction graphs

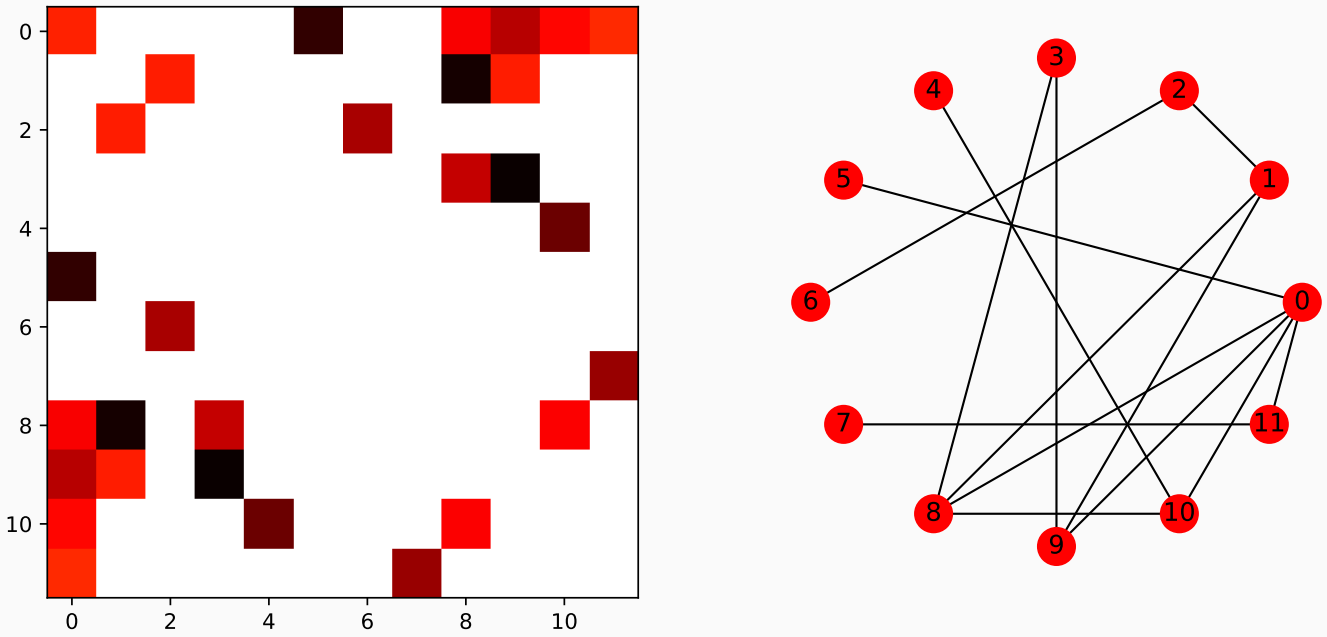


Figure 5: Edges of the interaction graph

Interaction graphs

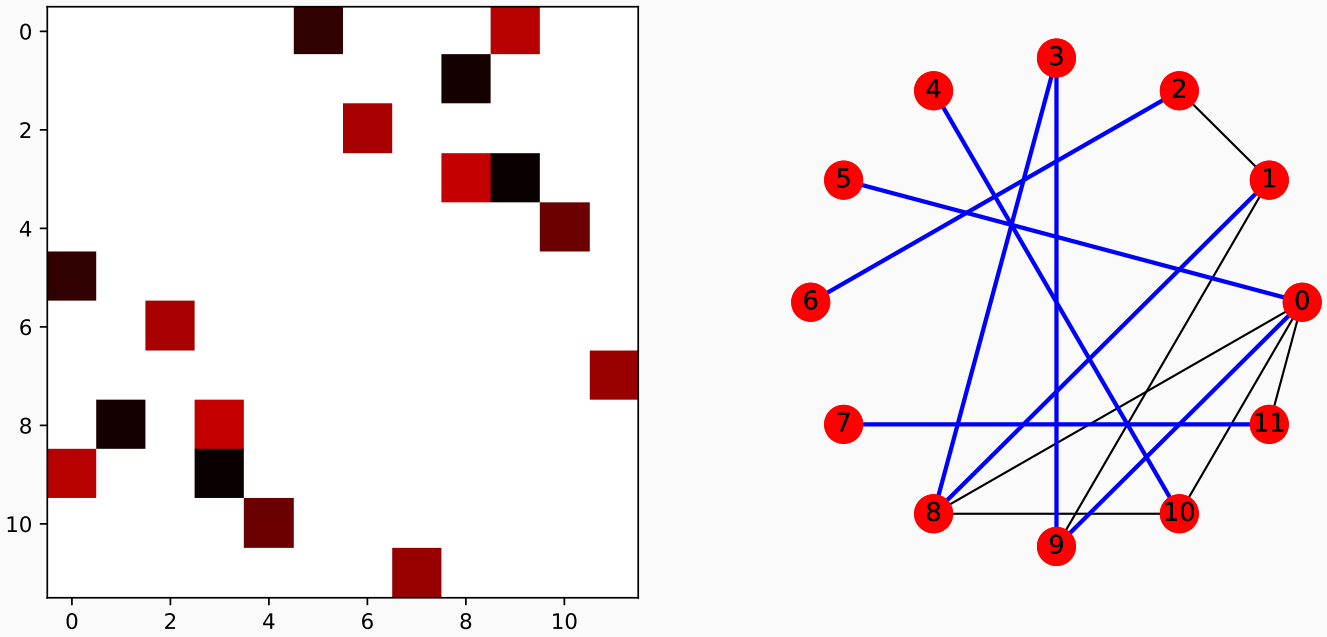


Figure 6: Strong edges of the interaction graph

Interaction graphs

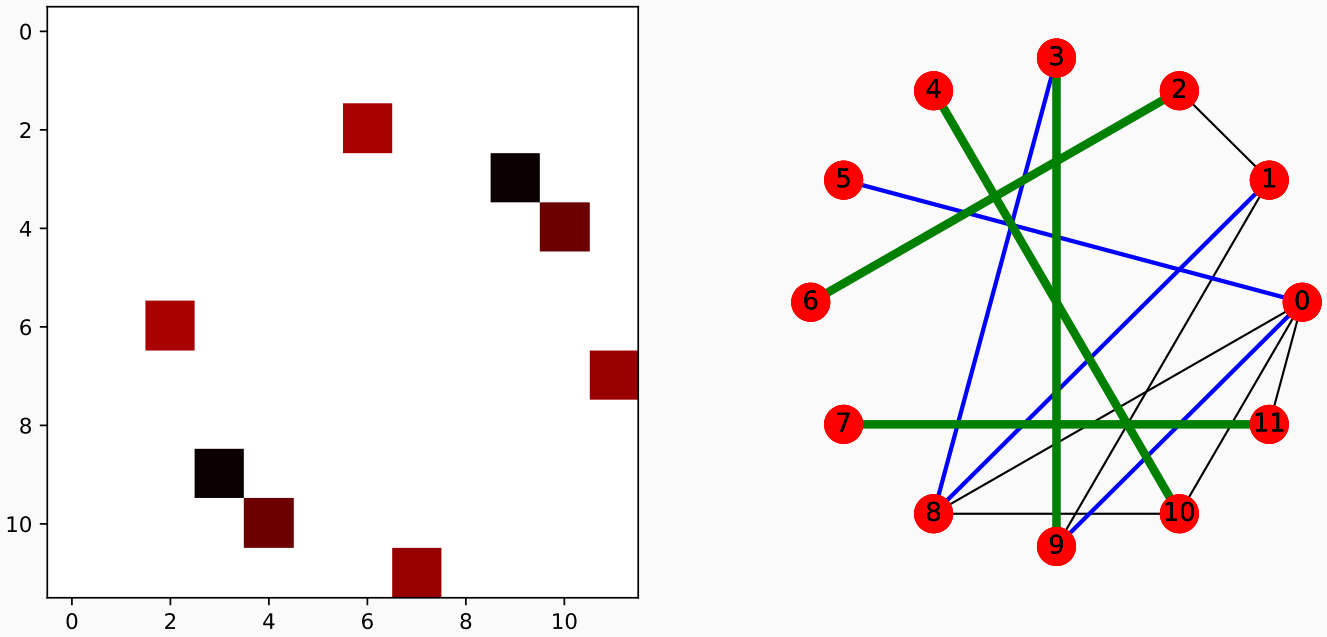


Figure 7: Independent edges of the interaction graph

Definition

We can define such a set of edges for any graph $G = (V, E)$ and a set $E_s \subseteq E$

Orthogonal set definition

A subset $S \subseteq E_s$ is an *orthogonal set* if for any two distinct edges $u_1v_1, u_2v_2 \in S$ the following holds:

i) The two edges are not incident to each other, i.e. $\{u_1, v_1\} \cap \{u_2, v_2\} = \emptyset$.

ii) The two edges are not incident to a common edge, that is,

$$\{u_1u_2,\ u_1v_2,\ v_1u_2,\ v_1v_2\} \cap E = \emptyset.$$

iii) Additionally, if $u_1 \neq v_1$ (i.e. the edge is not a loop), we require that u_1 and v_1 are not incident to any loops in E.

Orthogonal sets – original results [1]

Similarly to the maximum independent set (MIS) problem, we define the maximum orthogonal set (MOS) problem to be the tuple (V, E, E_s, k), and prove the following

Theorem

The Maximum Orthogonal Set Problem is NP-complete.

Proof idea.

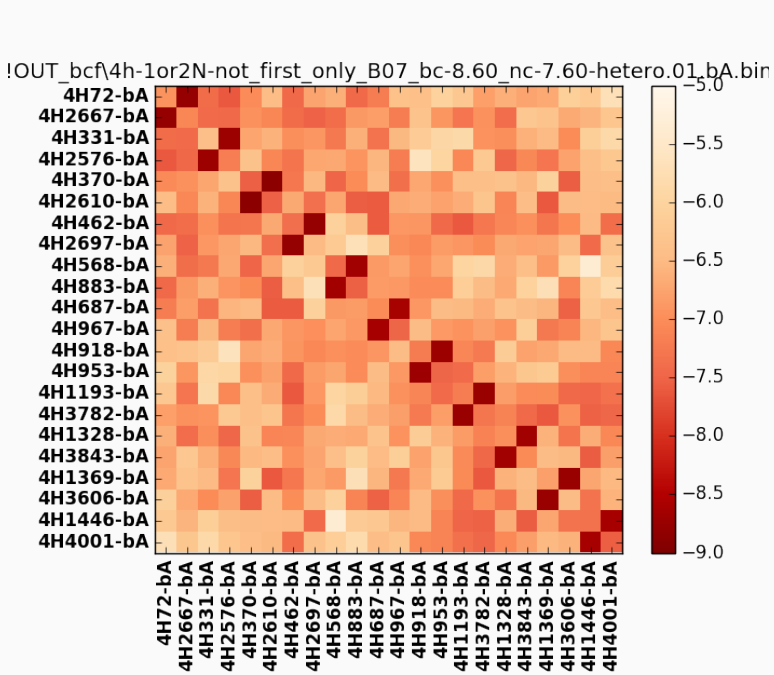
Starting from $G = (V, E)$, form $G' = (V', E')$ by adding v' for every $v \in V$, and connecting it only to the corresponding v. Prove that the MOS in G' consists only of edges of the form vv'. Then the v-endpoints of these edges form a MIS in G. Also, a MIS in G gives a MOS in G'. □

Exact algorithm for orthogonal sets

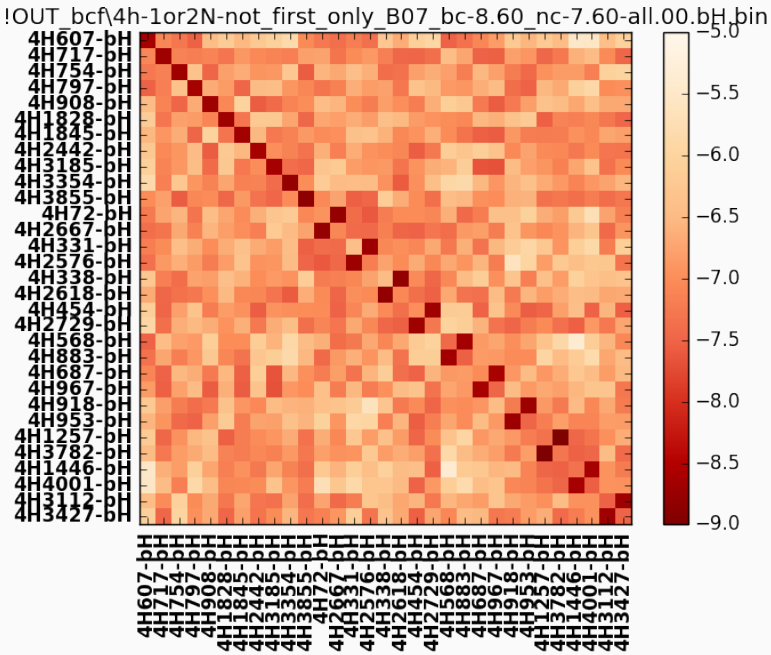
1. Start with a graph $G = (V, E)$ and a set $E_s \subseteq E$ (e.g. an interaction graph)
2. From E_s remove all pairs uv where $u \neq v$ and u or v is incident to a loop
3. Form a new graph $G' = (E_s, E')$, where we connect two vertices u_1v_1 and u_2v_2 if they can not be together in an orthogonal set
4. Find the maximum independent set in G' (= the maximum orthogonal set in G)

Some orthogonal sets

11-, and 21-pair orthogonal sets were constructed, currently undergoing experimental validation.



(a) Heterodimers only



(b) Both homodimers and heterodimers

Figure 8: Interaction scores for the constructed orthogonal sets

Heuristics

The need for heuristics

- Clique computation is still computationally expensive
- Instead of considering a large initial set of peptides, and finding the largest orthogonal subset, try to build the orthogonal set directly
- Two greedy approaches
  1. Start from a small library of heptads. Greedily build longer peptide pairs, by adding pairs of heptads that are known to bind to each-other.
  2. Iteratively extend a small orthogonal set, by alternatively taking its Cartesian product with another set, and determining the maximum orthogonal subset of the (still moderately sized) resulting product set.

The need for heuristics

- Clique computation is still computationally expensive
- Instead of considering a large initial set of peptides, and finding the largest orthogonal subset, try to build the orthogonal set directly
- Two greedy approaches
  1. Start from a small library of heptads. Greedily build longer peptide pairs, by adding pairs of heptads that are known to bind to each-other.
  2. Iteratively extend a small orthogonal set, by alternatively taking its Cartesian product with another set, and determining the maximum orthogonal subset of the (still moderately sized) resulting product set.

Iterative set building

- The algorithm fits into the intensification-diversification framework for combinatorial optimization metaheuristics

  diversification: Explore the search space, by taking the "Cartesian product"

  $$S_k \cdot H_{k+1} = \{a \cdot b | a \in S_k, b \in H_{k+1}\}$$

  intensification: Exploit the accumulated knowledge about the search space

  $$S_{k+1} = \mathrm{OrthogonalSubset}(S_k \cdot H_{k+1})$$

- S_k and H_k are the current orthogonal set, and the current *extension set*, respectively

One more orthogonal set

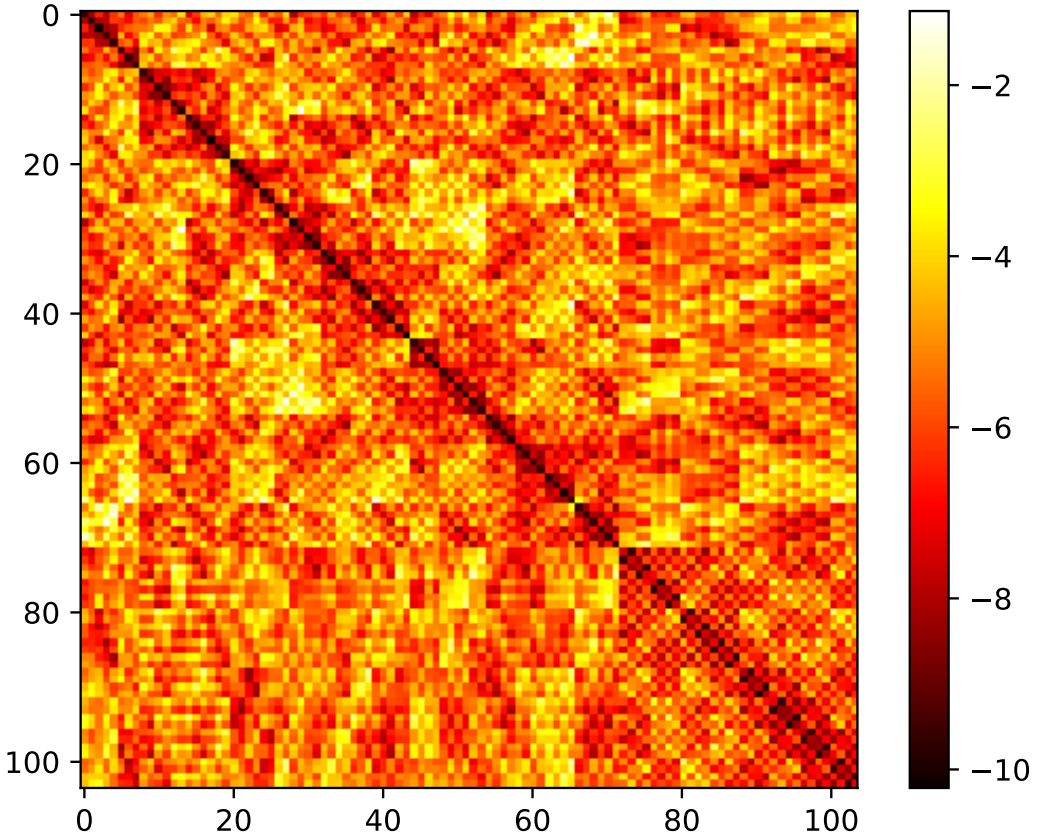


Figure 9: A 104-peptide orthogonal set, obtained using the previously described algorithm

Results and conclusions

Results and conclusions

- Significance of molecular self-assembly techniques
- Methods for doing polypeptide origami with coiled coils
- Algorithmic way of determining an orthogonal subset
- Heuristics for building orthogonal sets directly
- Possible applications to independent sets of product graphs?

Questions?

Thank you for your attention!

Bibliography I

Andrej Brodnik, Vladan Jovičić, Marko Palangetić, and Daniel Silađi.

Construction of orthogonal CC-sets.

In *Proceedings of the Middle-European Conference on Applied Theoretical Computer Science (MATCOS 2016)*, pages 36–39, 2016.

Matjaz Depolli, Janez Konc, Kati Rozman, Roman Trobec, and Dusanka Janezic.

Exact parallel maximum clique algorithm for general and protein graphs.

Journal of chemical information and modeling, 53(9):2217–2228, 2013.

Bibliography II


Shawn M Douglas, Hendrik Dietz, Tim Liedl, Björn Högberg, Franziska Graf, and William M Shih.
Self-assembly of DNA into nanoscale three-dimensional shapes.
Nature, 459(7245):414–418, 2009.


Gašper Fijavž, Tomaž Pisanski, and Jernej Rus.
Strong traces model of self-assembly polypeptide structures.
MATCH Communications in Mathematical and in Computer Chemistry, (71):199–212, 2014.

Bibliography III


Jessica H Fong, Amy E Keating, and Mona Singh.
Predicting specificity in bZIP coiled-coil protein interactions.
Genome biology, 5(2):R11, 2004.


Helena Gradišar, Sabina Božič, Tibor Doles, Damjan Vengust, Iva Hafner-Bratkovič, Alenka Mertelj, Ben Webb, Andrej Šali, Sandi Klavžar, and Roman Jerala.
Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments.
Nature chemical biology, 9(6):362–366, 2013.

Bibliography IV

Vid Kočar, Sabina Božič Abram, Tibor Doles, Nino Bašić, Helena Gradišar, Tomaž Pisanski, and Roman Jerala.
TOPOFOLD, the designed modular biomolecular folds: polypeptide-based molecular origami nanostructures following the footsteps of DNA.
Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology, 7(2):218–237, 2015.

Janez Konc and Dušanka Janezic.
An improved branch and bound algorithm for the maximum clique problem.
MATCH Communications in Mathematical and in Computer Chemistry, 4(5), 2007.

Bibliography V


Vladimir Potapov, Jenifer B Kaplan, and Amy E Keating.
Data-driven prediction and design of bZIP coiled-coil interactions.
PLoS Comput Biol, 11(2):e1004046, 2015.


Paul WK Rothemund.
Folding DNA to create nanoscale shapes and patterns.
Nature, 440(7082):297–302, 2006.