

Computational methods for polypeptide origami design

Thesis summary

Daniel Silađi

mentor: dr. Andrej Brodnik

co-mentor: dr. Rok Požar

May 26, 2017

During the last few decades significant advances have been made in our understanding of the (bio)chemical processes in living organisms. Consequently, there has been an ever-increasing demand placed on laboratories to synthesize increasingly complex organic molecules. Such molecules can then be used as scaffolds for attaching other molecules, or as highly targeted particle delivery systems.

In 2006, [4] demonstrated a method of folding a single strand of DNA into a 3D shape, just by specifying its nucleotide sequence. This is done by designing the sequence such that some of its regions bind to each other and form the “wireframe” of the object after folding. Although for biological reasons it would be useful to replicate this construction using sequences of amino acids (peptides, polypeptides or proteins, depending on their length), there is no known computationally efficient way of determining whether and how two peptides will interact in the general case. Luckily, in the specific case of *coiled coil* peptides, [3] presented an algorithm for determining the interaction strength of two such peptides. On the other hand, [2] presented both experimental and graph-theoretical results about the classes of graphs which can be constructed from a single peptide chain.

Most notably, their algorithm requires the chain to go along the *double (Eulerian) trace* of the graph. This further creates the need for a large set of pairs of short peptides (that will be placed along the graph edges) that interact only mutually. Such a set is called an *orthogonal set*. Given an undirected interaction graph $G = (V, E)$, whose vertices are peptides, and whose edges indicate the presence of interaction between them, an orthogonal set $S \subseteq E$ is any set with the following property:

$$\{u_1, v_1\} \cap \{u_2, v_2\} = \emptyset,$$

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

for all distinct edges $u_1v_1, u_2v_2 \in S$. In other words, S is an independent set in the line graph $L(G)$ of G with the additional restriction that no two vertices share a neighbor. By

reducing the maximum independent set problem to the problem of finding the maximum orthogonal set in a specific graph, we have proved that finding the maximum orthogonal subset of a graph is NP-hard. Similarly, by reducing it to the maximum independent set problem, we have proved that our problem is also in NP, and thus that it is NP-complete. We also describe an end-to-end exact algorithm that can be used for determining the maximum orthogonal subset of a peptide set, where the peptides are given by their amino acid sequences. The independent set algorithm is a modified version of the maximum clique algorithm presented in [1].

We also present some heuristics that exploit the problem structure (i.e. that the graphs in question are not general graphs, but are constructed from weighted peptide interaction graphs) in order to build larger orthogonal sets from scratch. More specifically, two greedy approaches are presented:

1. Directly building pairs of hopefully strongly interacting chains and determining their maximum orthogonal subset, and
2. Iteratively extending a smaller 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.

It should be noted that the latter approach gives us orthogonal sets that are 3-10 times larger than the ones obtained by exact methods as subsets of smaller sets. The obtained results are as of now among the best ones in the literature.

Finally, we discuss some possible applications of the newly developed bottom-up heuristic for solving the maximum orthogonal set on general graphs.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Paul WK Rothmund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.