INVESTIGATION

An undeveloped investigation of the relative importance of individual staples was started by Jakub Lála.

Initially, a single-removal (SR) type of simulation was run, where a single staple type is removed from the input parameters and thus is not inserted into the system. The pipeline for creating the input files was developed using Python scripts and is located on Jakub's GitHub forked repository in **scripts/investigation/single-removal/**:

- **create_sr-seq.py** – considers the input sequence file for the staples and iteratively creates sequence files, where a single type is removed
- **create_sr-jsons.py** – considers the input sequence staple and scaffold files and creates the JSON input file of the system for all simulation setups
- **create_sr-input-configs.py** – creates the .inp input files for all simulation setups according to the provided template
- **create_sr-slurms.py –** creates the .sh SLURM files necessary for cluster execution for all simulation setups
- **create_sr-folders.py** – prepares all the necessary files for all simulation setups into a sim_folders/ folder
- **create_sr-all.py** – runs all of the scripts above successively

Note that for proper execution one has to include the necessary input files in the **inps/** folder:

- **(bias_functions.json)**
- **input_template.inp**
- **movetypes_default.json**
- **num_walks.arch**
- **ops_default.json**
- **serial_template_slurm**
- **snoddin_scaffold.seq**
- **snoddin_staples.seq**

The files above are examples, thus if one wants to use the scripts with other DNA origami shapes other than the Snoddin tile, or wants to use different move type frequencies, one has to adjust the filenames accordingly in the script.

Taking the Snoddin example shown below, 13 different simulation setups were performed, where a different staple type was missing in each and one simulation setup had all of the staple types present. The figure below also colourfully groups staple types of similar characteristic, more specifically melting point. This comes from Alex's paper, where he studies the mean average occupancy for individual staple types in this Snoddin tile by performing REMC simulation. By analysing this order parameter for various temperatures, he was able to find the transition and identify it as the melting temperature. Various melting temperatures of staple types

then refer to various degrees of geometrical restrictions on the system.

More specifically, the staple type categories are:

- **same helix** - staples 1 and 12 => most stable, highest T melt
- **span-2** - staples 3, 6, 9
- **span-0, inside span-2** - staples 4, 7, 10
- **span-0, outside span-2** - staples 2, 5, 8, 11 => lowest T melt, thus most geometrically and physical scaffold restrictions

These are listed in the order of decreasing temperature, and thus increasing scaffold geometry restriction.

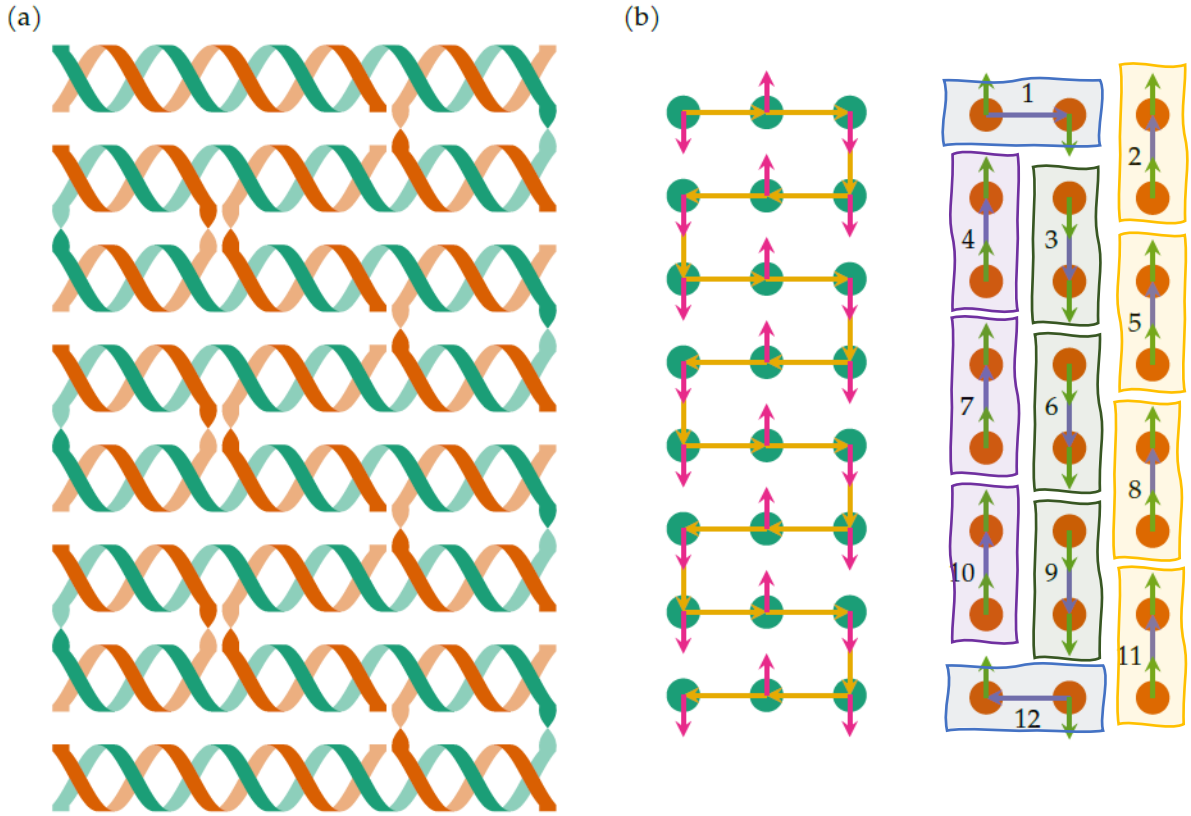
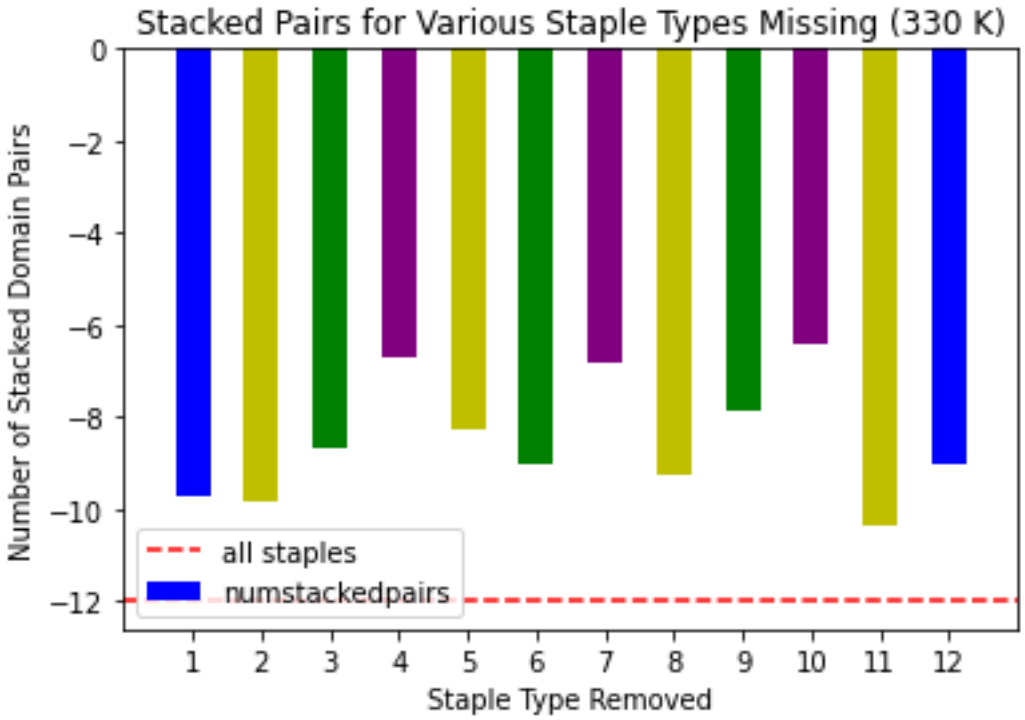


Figure 4.2: Schematic representations of the 24-binding-domain scaffold system. (a) Helical cartoon representation of the system in a fully stacked assembled configuration. (b) Representation of the system with the lattice model. The scaffold (left) and staples (right, numbered) are shown in the fully stacked assembled configuration, but for clarity have been drawn separately. A full legend for all diagram elements is provided in Figure 2.1.

It is unclear which order parameter to use for quantifying the extent of assembly. In this specific case, it was proposed by Alex to use the number of stacked domain pairs, as the fully assembled structure in a planar form should maximize the number of stacked pairs.

Results of an initial simulation are given below:

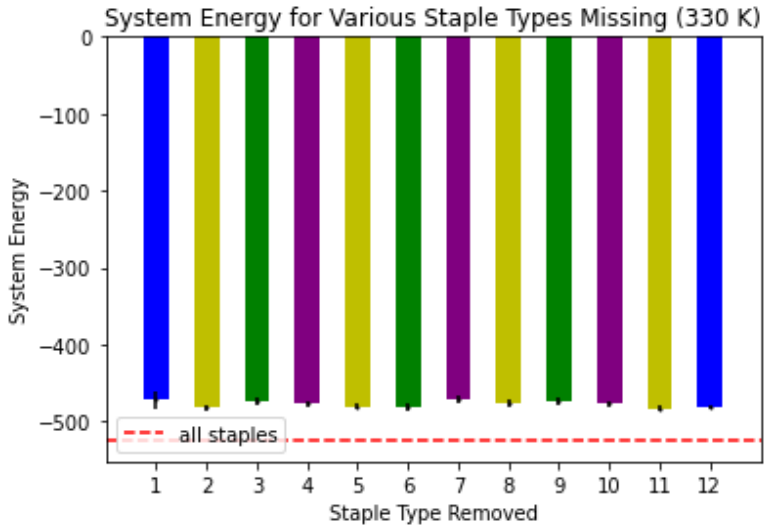


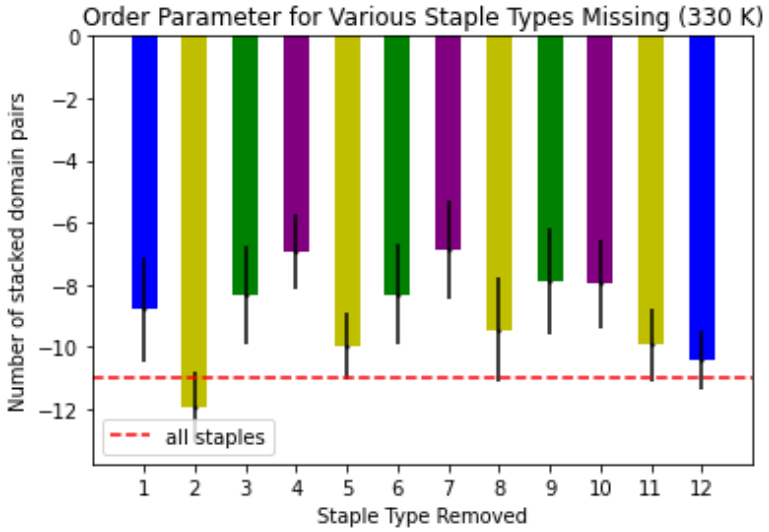
Simulation Details: 330 K, 10 000 000 MC steps, 10 000 MC step logging frequency, constant temperature, stacking energy = 1000

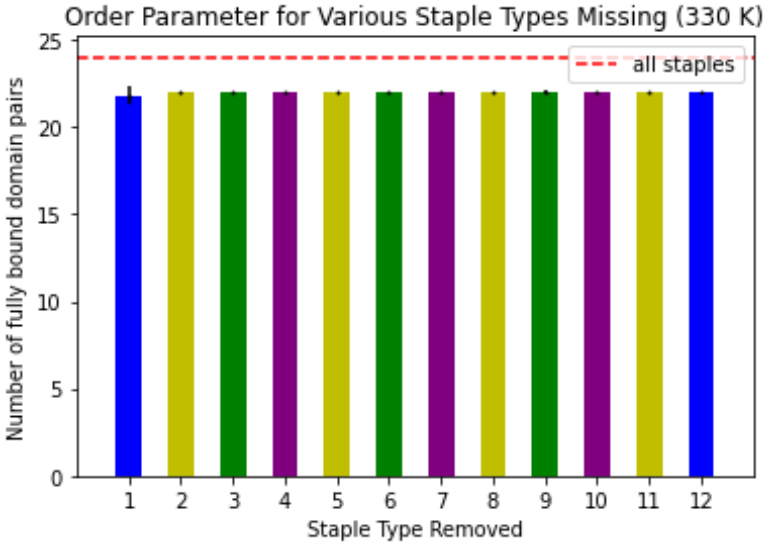
From the plot, staples 4, 7 and 10 seem to be crucial as their removal causes the greatest decrease in the number of stacked number pairs. In Alex's paper, these impose great physical restrictions, yet they are not the most restrictive, i.e. they do not have the lowest melting point. Notice that in this plot the number of stacked domain pairs for a system, where all staples are present was not actually simulated and was assumed to be 12. Moreover, the code for some reason prints out these values as negative rather than positive.

Also note that the colours of the different simulation setups are given accordingly to the colours as defined above by the categories of staple types. Although it may seem as though characteristically similar staple types have a similar effect on the order parameter studied, this is clearly not the case. The only confident distinction one can observe is for the aforementioned staples 4, 7 and 10, which clearly all show the least number of stacked domain pairs.

The simulation setup was revised. The system will all staples present was also included, as well as the stacking energy and the number of fully bound domain pairs was analysed. The results are given in the plots below:





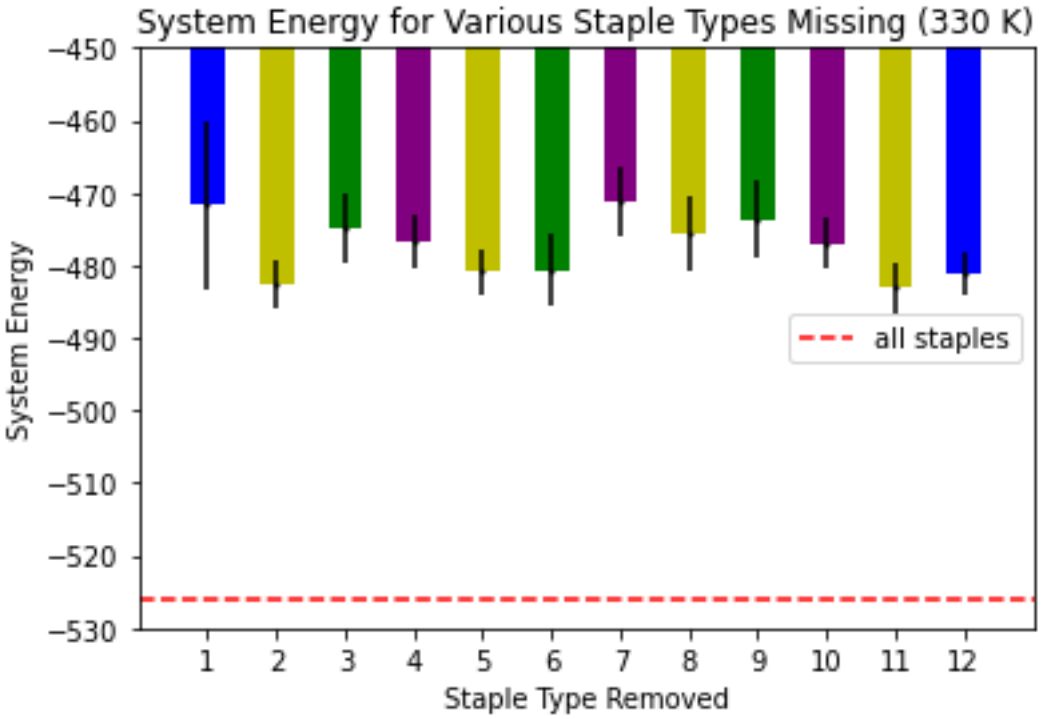


Simulation Details: 330 K, 30 000 000 MC steps, 1000 MC step logging frequency, constant temperature, stacking energy = 1000

Firstly, notice that the values for the system with all staples present is now actually represented by a simulation. Moreover, the error bars are also included. It is clear that the large uncertainty in the number of stacked domain pairs may explain why the simulation setup with staple type 2 missing shows a more stacked configuration than the system with all staples present, as this is not what we would expect to see.

The number of fully bound domain pairs does not seem to be a good indicator of the extent of the assembly to the target structure in this specific example, as this order parameter is nearly always 22 (the maximum we would expect) for all setups. Only for simulation setups with missing staple type 1, 8, 9 and 10 have a slightly lower values than 22.

Looking at the system energy, there seem to be some deviations between the various setups, but a closer look is necessary to see any clear indications:



The system energy deviations do not seem to be giving a result that would agree with the number of stacked domain pairs, hence at least one of these is not the correct way to access the extent of assembly. Nevertheless, the error bars are fairly wide and thus presumably the simulation should be run for more MC steps.

The investigation was then expanded to analyse double-removal simulation setups, for which a similar set of scripts was developed. The results are displayed in the following pages.

Simulation Details: 330 K, 10 000 000 MC steps, 1000 MC step logging frequency, constant temperature, stacking energy = 1000

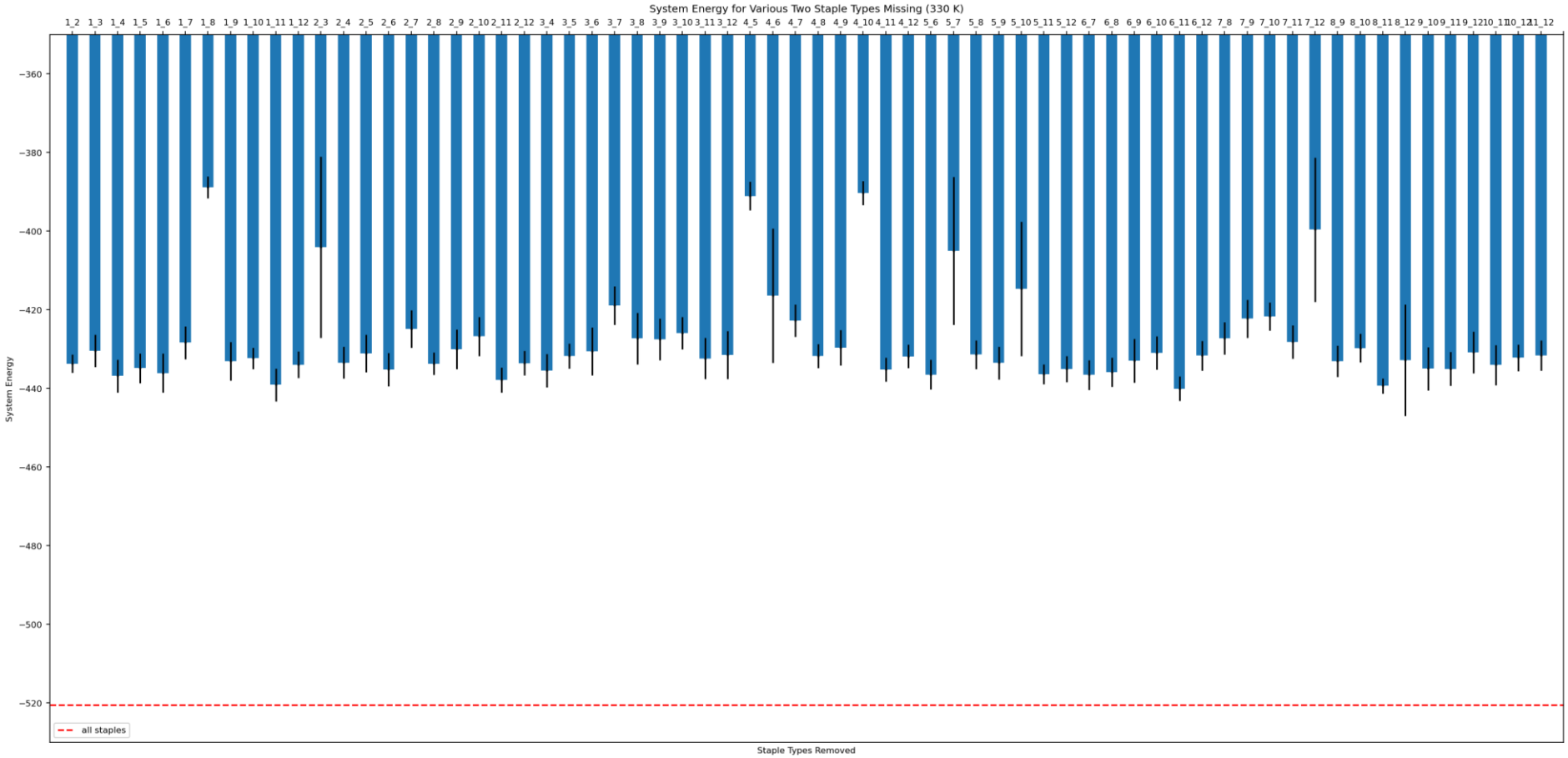
Notice that only 10 million steps were simulated for each setup, which is much less compared to the previous SR investigation. It is suggested that this should be re-run with more steps for statistically more meaningful results.

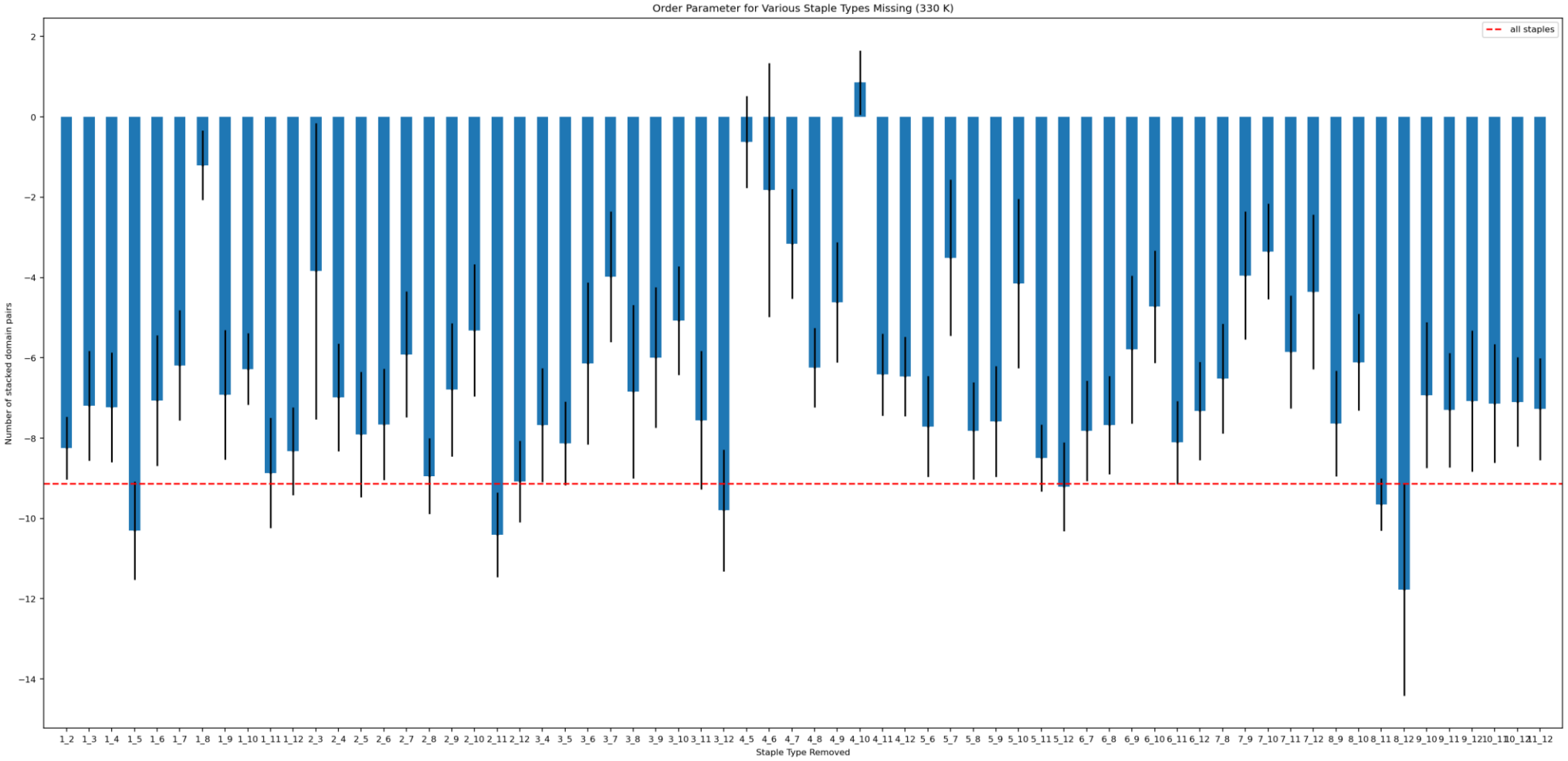
Firstly, looking at the system energy results, one may see that there are certain staple removal combinations that show a less negative energy. With some careful examination, some of these combinations such as 4 with 5 or 4 with 10, could be argued for as they form the key turning points (or sides) of the origami shape. Nevertheless, the removal of 1 with 8 or 7 with 12, should not be expected to have such a different energy, as staples 1 and 12 have both domains on the same chain, and thus should not be that important for the geometrical restrictions of assembling.

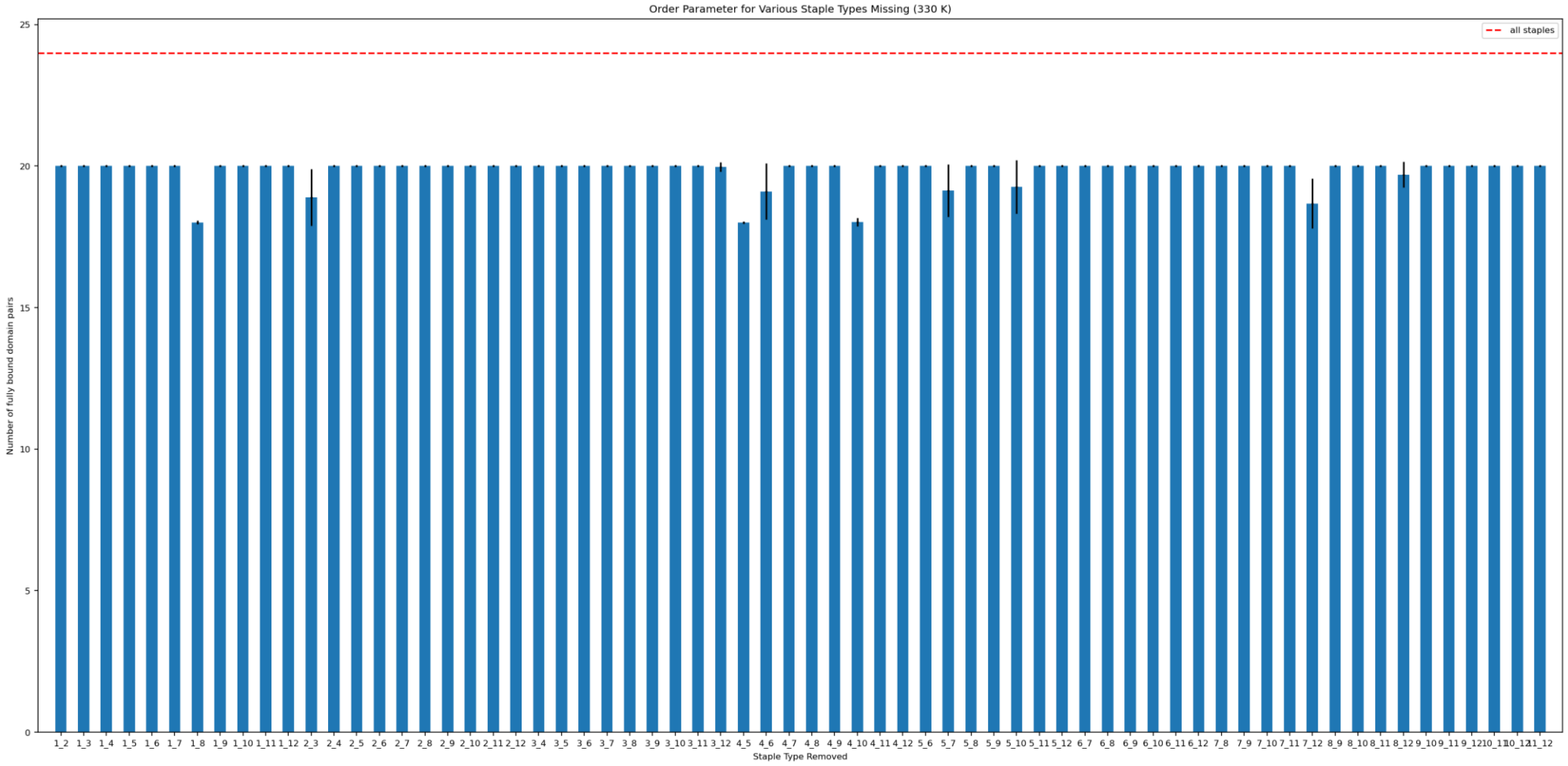
Secondly, looking at the number of stacked domain pairs, there are huge fluctuations, hence it is probably wise to not comment on these. The thing that only needs to be stressed again is therefore, that advanced sampling methods should be employed to obtain meaningful results.

Lastly, looking at the number of fully bound domain pairs, we once again see that this order parameter correctly mirrors the trends seen from the stacking energy, as explained above. Therefore, notice that removing staple 4 with staple 10 or staple 5 causes some other staple to not be able to bind, as on average the number of fully bound domain pairs is 18 rather than 20. Two investigation approaches could follow from this. Initially, it should be checked with better sampling and more MC steps, that this result is accurate, and we have not been stuck in a local energy minimum during the simulation. Afterwards, it should be checked what staple type was missing, which can be retrieved from the simulation output files.

Looking at both 4/10 and 4/5 combination, in both the staple 6 is completely missing from the final configuration (at least on average). Then staple 7 is the one that appears twice in the system for both simulation setups. Now again some important things must be stressed. Firstly, we are not that confident that these are meaningful results. Secondly, although it may come to one’s mind that staple 7 may be quite similar in the gene sequence in this specific simulation setup, and thus take over the spot of staple 6, this should not be the case. Energetically, it would still be favourable to exchange staple 7 for staple 6, hence something else must be in play. Furthermore, it is suggested that the it should be analysed where the staple 7 is actually bound to the system – whether both of the domains are (mis)bound or not, and if so, whether they still do not allow for some geometrical restriction and thus contribute at least slightly to some assembly.


System Energy for Various Two Staple Types Missing (330 K)
1_2 1_3 1_4 1_5 1_6 1_7 1_8 1_9 1_10 1_11 1_12 2_3 2_4 2_5 2_6 2_7 2_8 2_9 2_10 2_11 2_12 3_4 3_5 3_6 3_7 3_8 3_9 3_10 3_11 3_12 4_5 4_6 4_7 4_8 4_9 4_10 4_11 4_12 5_6 5_7 5_8 5_9 5_10 5_11 5_12 6_7 6_8 6_9 6_10 6_11 6_12 7_8 7_9 7_10 7_11 7_12 8_9 8_10 8_11 8_12 9_10 9_11 9_12 10_11 10_12 11_12
−360
−380
−400
−420
−440
−460
−480
−500
−520
System Energy
all staples
Staple Types Removed


Order Parameter for Various Staple Types Missing (330 K)
all staples
Number of stacked domain pairs
Staple Type Removed
2
0
−2
−4
−6
−8
−10
−12
−14
1_2 1_3 1_4 1_5 1_6 1_7 1_8 1_9 1_10 1_11 1_12 2_3 2_4 2_5 2_6 2_7 2_8 2_9 2_10 2_11 2_12 3_4 3_5 3_6 3_7 3_8 3_9 3_10 3_11 3_12 4_5 4_6 4_7 4_8 4_9 4_10 4_11 4_12 5_6 5_7 5_8 5_9 5_10 5_11 5_12 6_7 6_8 6_9 6_10 6_11 6_12 7_8 7_9 7_10 7_11 7_12 8_9 8_10 8_11 8_12 9_10 9_11 9_12 10_11 10_12 11_12


Order Parameter for Various Staple Types Missing (330 K)
all staples
Number of fully bound domain pairs
Staple Type Removed
0
5
10
15
20
25
1_2 1_3 1_4 1_5 1_6 1_7 1_8 1_9 1_10 1_11 1_12 2_3 2_4 2_5 2_6 2_7 2_8 2_9 2_10 2_11 2_12 3_4 3_5 3_6 3_7 3_8 3_9 3_10 3_11 3_12 4_5 4_6 4_7 4_8 4_9 4_10 4_11 4_12 5_6 5_7 5_8 5_9 5_10 5_11 5_12 6_7 6_8 6_9 6_10 6_11 6_12 7_8 7_9 7_10 7_11 7_12 8_9 8_10 8_11 8_12 9_10 9_11 9_12 10_11 10_12 11_12

To continue, the method of the mean staple occupancy used by Alex would be ideal to use in this investigation as well for all the different simulation setups. Moreover, some development of the bias functions or the use of umbrella sampling may be useful to improve convergence.