Deutsch:

Die verschiedenen Dockings und Designs werden mit RosettaScripts durchgeführt. Mittels PyMol werden Wassermoleküle und weitere nicht benötigte Moleküle von der originalen Struktur (\_\_\_\_) entfernt. Danach erfolgt eine neue Nummerierung der PDB-Datei, damit jede Aminosäure ihre eigene Nummer erhält. Dies dient in späteren Schritten zur besseren Nachverfolgung von Mutationen, könnte aber weggelassen werden. Die Startstrukturen mit den neuen Peptiden des Spike-Proteins (6ZOZ) werden durch ein Python-Skript erstellt. Dabei wird die Sequenz des originalen Peptids (flg22) durch die gewünschte Sequenz ausgetauscht.

Die neue Struktur wird nun an den Hochleistungsrechner BinAC in Tübingen geschickt, auf dem dann für jede erstellte Struktur 10 Replikate des ersten Docking-Schritts durchgeführt werden. Es handelt sich hierbei um ein lokales Docking, bei dem weder das Peptid noch der Rezeptor großartig bewegt werden. Die Ergebnisse werden in einer Score-Datei zusammengefasst. Anhand dieser Datei werden die besten Dockings ausgewählt: Dafür werden die Ergebnisse nach ihrer Bindungsenergie sortiert.

Die Dateien werden neu benannt und für das erste Design an den BinAC geschickt. Nach dem ersten Design werden Histogramme mit der Verteilung der Bindeenergien aufgezeichnet. Hiervon werden 12 Designs ausgewählt. Dabei gilt, je negativer die Verteilung ist, desto besser ist die Designbarkeit. Außerdem gelten als Auswahlkriterien die Bindungsenergie (je negativer, desto besser) aber auch die Anzahl an eingebrachten Mutationen.

Diese Dateien werden erneut umbenannt und zum zweiten Docking an den BinAC geschickt. Die Auswahl erfolgt erneut nach der Bindeenergie: Von jedem der (nun) Unterordner wird wieder die beste Struktur ausgewählt.

Diese Dateien werden erneut umbenannt und zum zweiten Design geschickt. Die Finale Auswahl erfolgt anhand der Verteilungen der Bindeenergien. Aus jeweiligen Daten wird die beste Datei (Bindeenergie, Mutationen gesamt, wenig eingesetztes Tryptophan) ausgesucht. Am Ende sind dies die Finalen 3-8 Strukturen.

Wichtige Werte der nun finalen Strukturen sind der beiliegenden Excel-Datei, „important\_Datas“, zu entnehmen. Die eingefügten Mutationen sind in den jeweiligen Ordnern den jeweiligen Word-Dokumenten zu entnehmen. Außerdem sind die Fasta-Sequenzen unter Fastas und die Protein-Komplexe mit ihren angefärbten Bereichen unter PSE zu finden.

English:

The various dockings and designs are performed with RosettaScripts. Using PyMol, water molecules and other unneeded molecules are removed from the original structure (4MN8). Afterwards, a new numbering of the PDB file is done, so that each amino acid gets its own number. This is used in later steps to better track mutations but could be omitted. The start structures with the new peptides of the spike protein (6ZOZ) are created by a Python script. This involves replacing the sequence of the original peptide (flg22) with the desired sequence.

The new structure is now sent to the high-performance computer BinAC in Tübingen, where 10 replicates of the first docking step are then performed for each structure created. This is a local docking, where neither the peptide nor the receptor is moved much. The results are summarized in a score file. Based on this file, the best dockings are selected: For this, the results are sorted by their binding energy.

The files are renamed and sent to the BinAC for the first design. After the first design, histograms are plotted with the distribution of binding energies. Of these, 12 designs are selected. The more negative the distribution, the better the designability. In addition, the binding energy (the more negative, the better) but also the number of introduced mutations are considered as selection criteria.

These files are renamed again and sent to the BinAC for the second docking. The selection is again based on the binding energy: the best structure of each of the now subfolders is again selected.

These files are renamed again and sent to the second design. The final selection is made based on the distributions of the binding energies. From each data the best file (binding energy, total mutations, little used tryptophan) is selected. At the end these are the final 3-8 structures.

Important values of the now final structures can be found in the enclosed Excel file, "important\_Datas". The inserted mutations are to be taken documented in the respective folders to the respective Word documents. In addition, the Fasta sequences can be found under Fastas and the protein complexes with their stained regions under PSE.