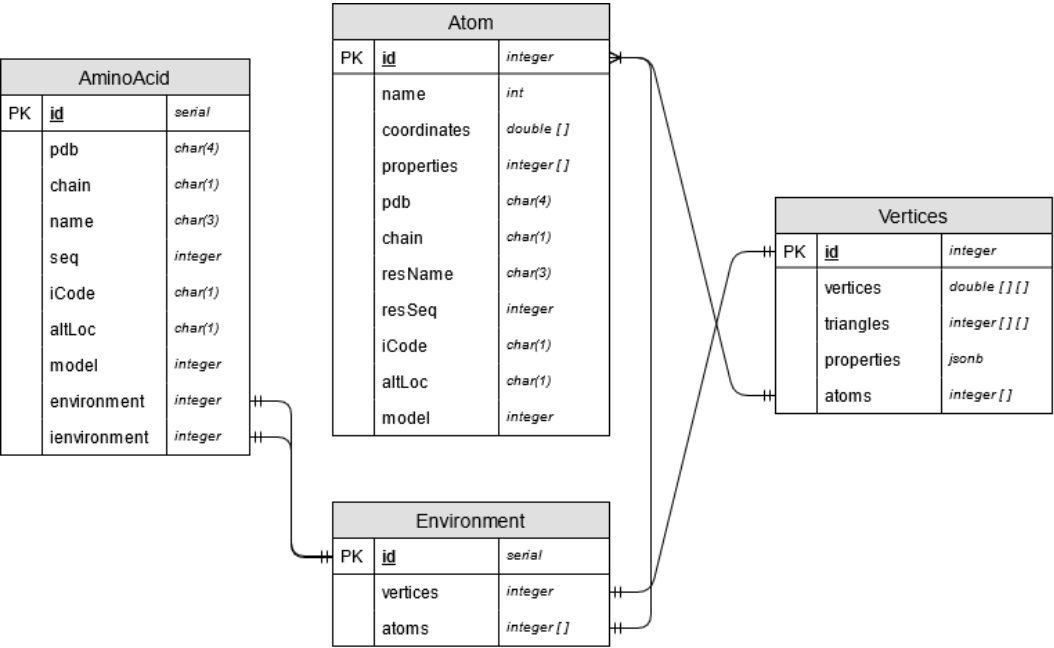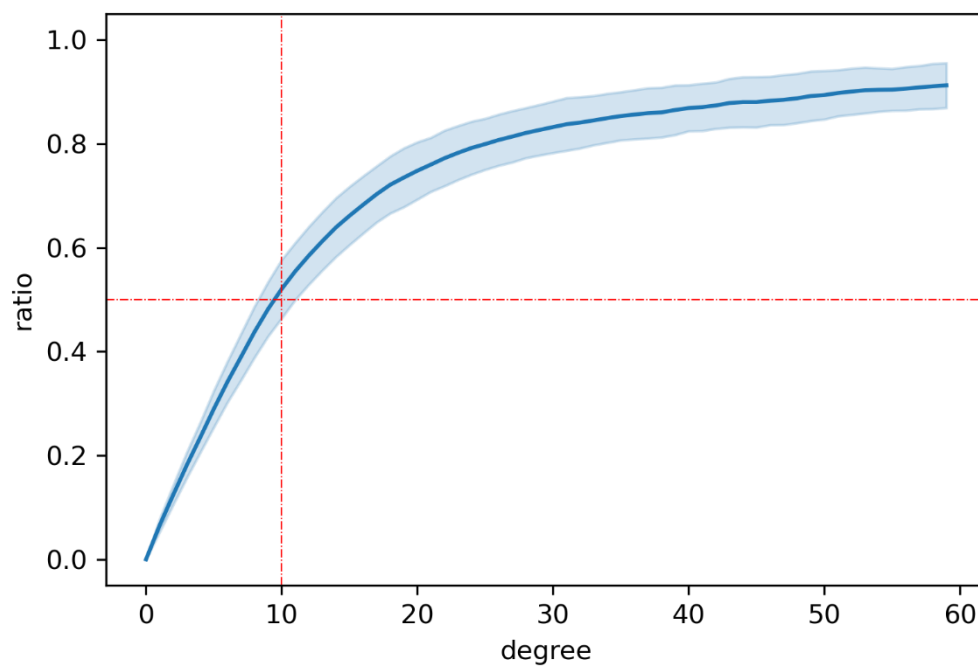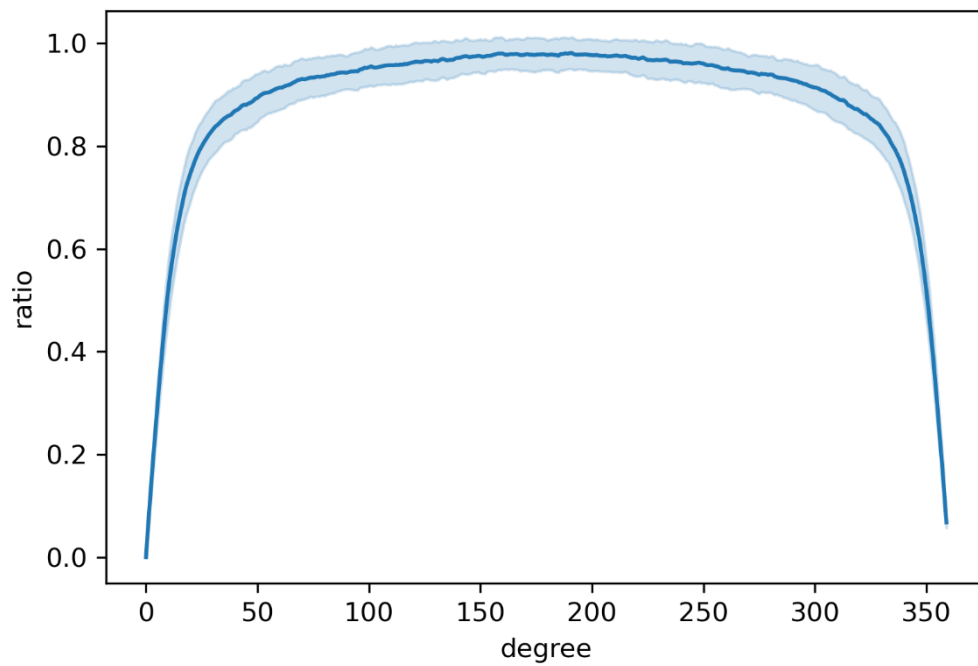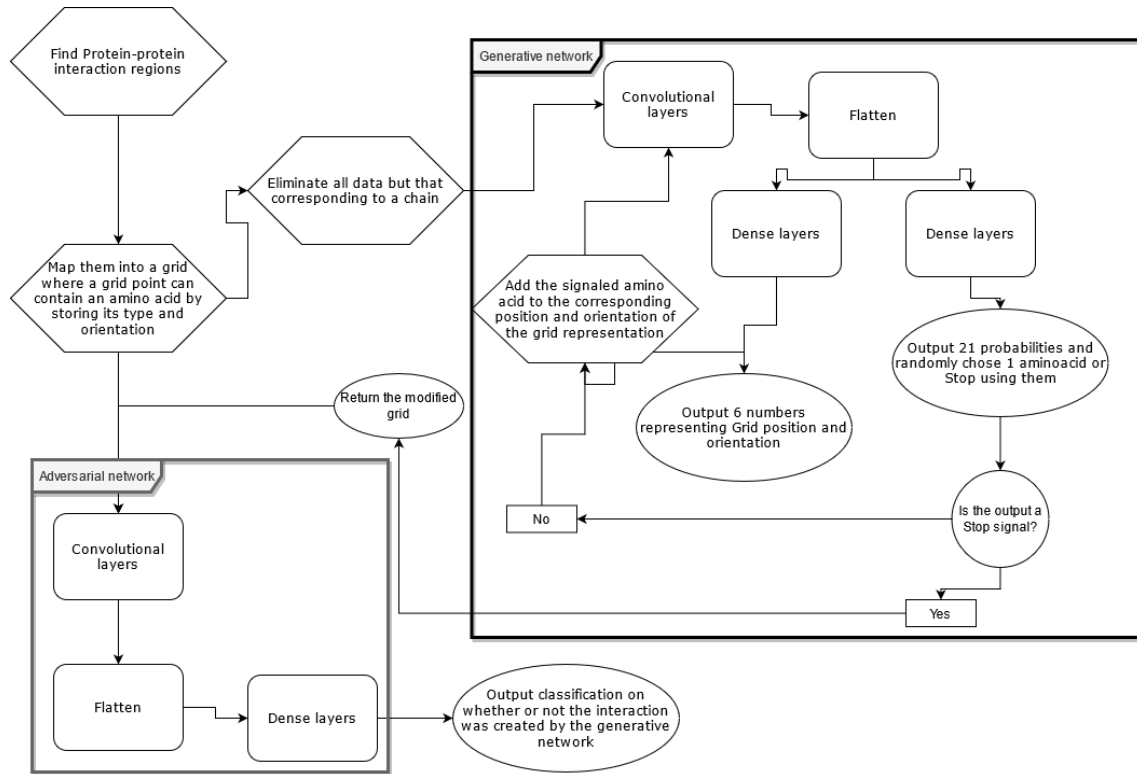# Supplementary material
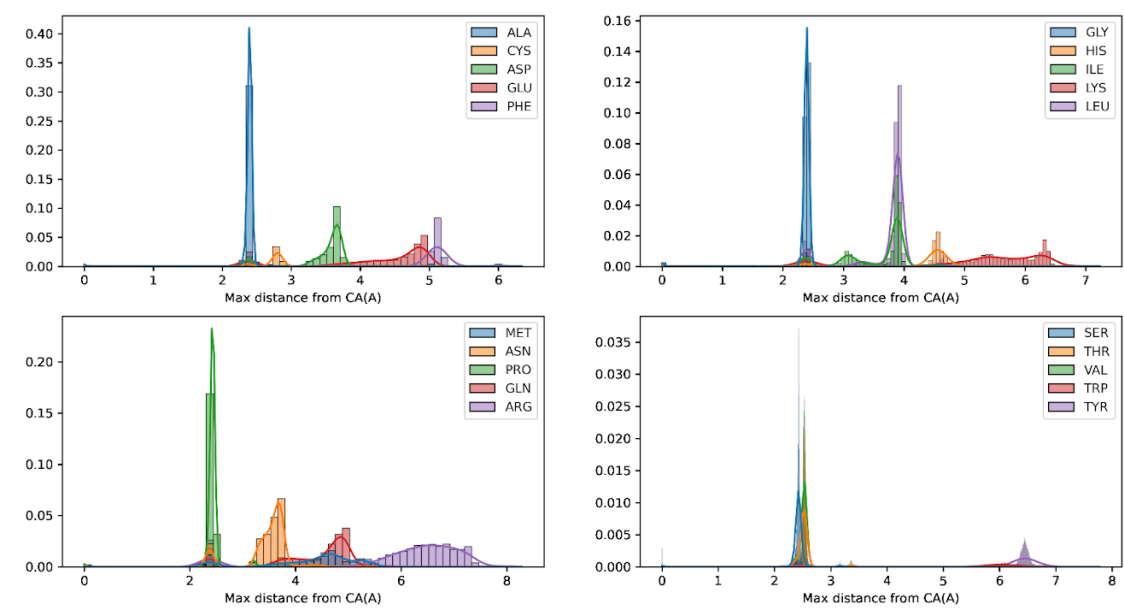
Supplementary Figure 1: Distribution of the distances from its CA to its most distant atom in all of the amino acids in the first 1000 PDB structures.

Supplementary Figure 2: The database schema for the environments and related information storage. [ ] Represents one dimensional arrays while [ ] [ ] represents two-dimensional arrays.

Supplementary Figure 3: Portion of initially true points of the grid that changed after a turn of x degrees. The shadow shows the standard deviation obtained after computing the same ratio for fifty environments.

Supplementary Figure 4: Alternative flowchart that was considered using a generative adversarial network approach. The idea was for the generative network to output the whole interaction when given one Protein-Protein interaction region with all chains erased but one. The grids used would not be the same as in the main flowchart and instead, contain only amino acids position and orientation because the generative network adds amino acids to the input grid. The needed code was written but it was not trained because of time constraints as the main workflow was already difficult to train and the generative network is computationally expensive. Moreover, this flowchart does not quite fit the project objectives as even if it perfectly worked it would output protein-protein-like interactions, which might not directly help with *de novo* peptide design. Additionally, it losses the surface representation to a more minimalistic one, most likely generating an important information loss.