Understanding the origin of the 20 letter alphabet of proteins is a long-lasting biophysical problem. In particular, studies focused extensively on the effect of a reduced alphabet size on the folding properties [1, 2, 3]. However, the natural alphabet is a compromise between versatility and optimisation of the available resources.

Here, for the first time, we include the additional impact of the relative availability of the amino acids. We present a computational protein design scheme that involves the competition for resources between a protein and a potential interaction partner that, additionally, gives us the chance to investigate the effect of the reduced alphabet on protein-protein interactions. We identify the optimal reduced set of letters for the design of the protein, and we observe that even alphabets reduced down to 4 letters allow for single protein folding. However, it is only with 6 letters that we achieve optimal folding, thus recovering experimental observations. Additionally, we notice that the binding between the protein and a potential interaction partner could not be avoided with the investigated reduced alphabets. Therefore, we suggest that aggregation could have been a driving force for the evolution of the large protein alphabet.

