In the course of evolution, proteins undergo substantial changes in their amino-acid sequences, while conserving their three-dimensional fold and their biological functionality. Modern sequencing techniques provide us with increasingly large families of evolutionary related proteins. Such data can be used to infer statistical models of sequence variability.

I will discuss the surprising efficiency of models including pairwise epistasis (Potts models / Markov random fields), which are able to reproduce non-fitted statistical features of protein families, help in predicting 3d structure and protein-protein interactions from sequence, and open novel ways toward evolution-guided protein design.