

Protein Structure Prediction Through Residue Co-Evolution

CSC412/2506 Course Project

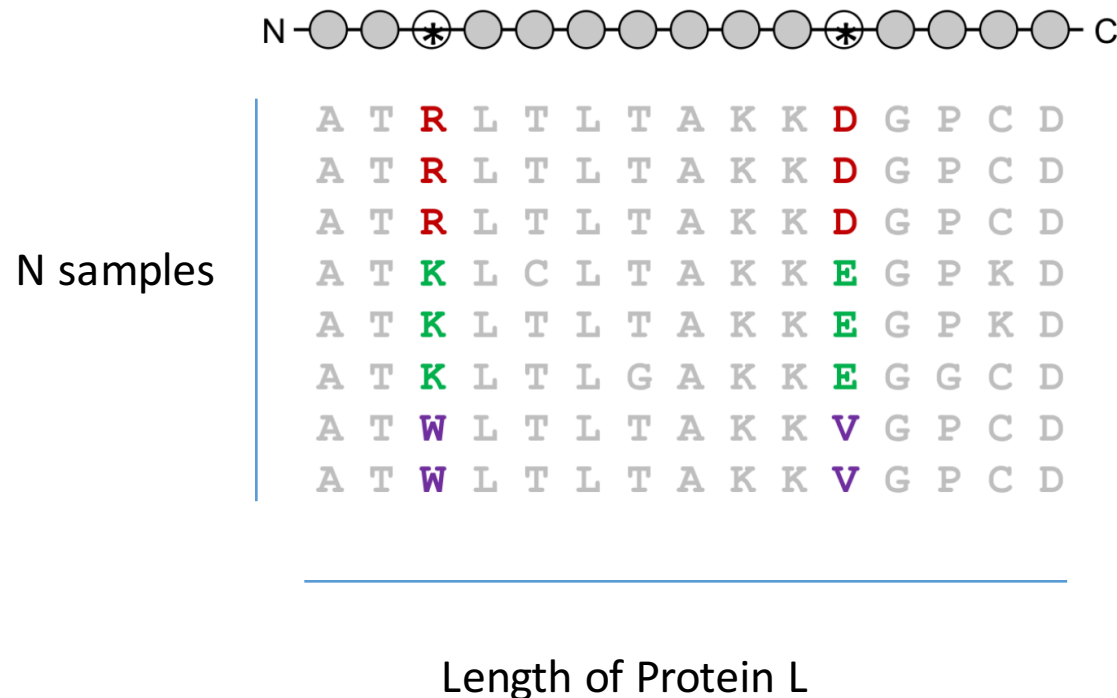
Chris Cremer

Goal

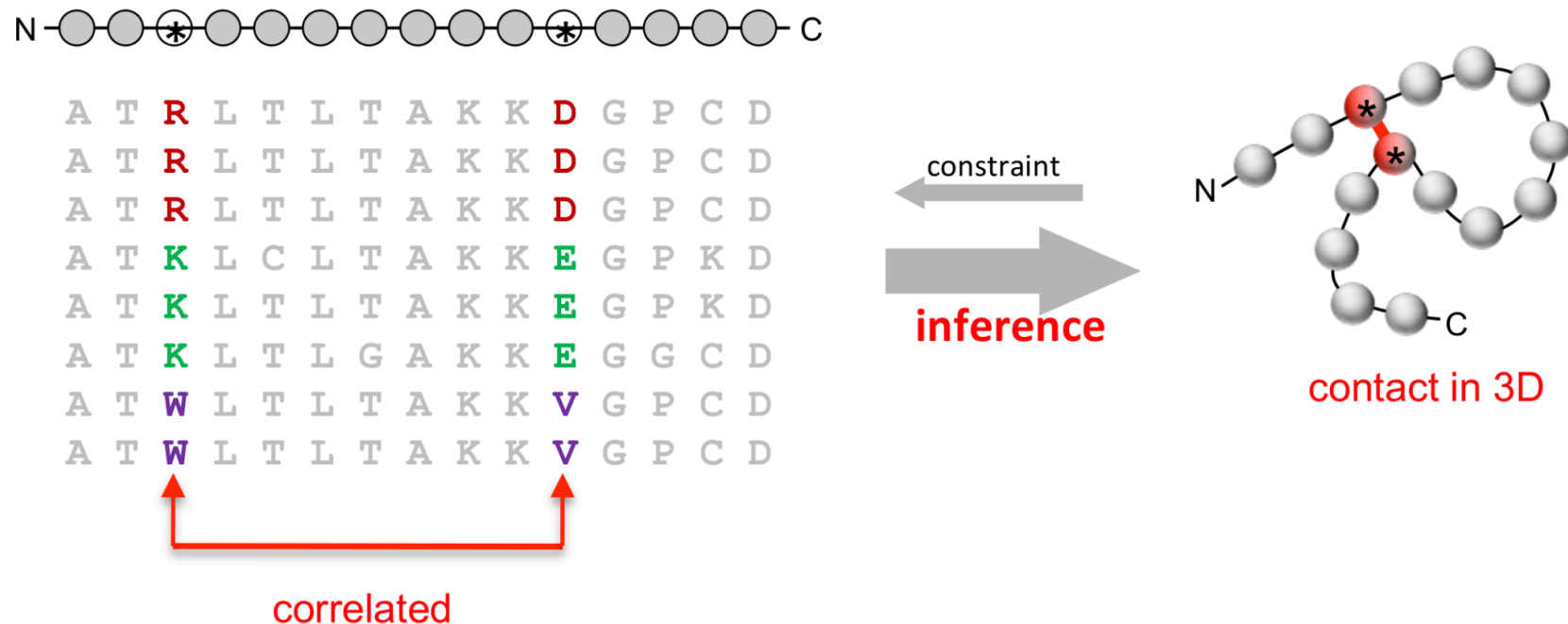
- Predict latent 3D positional vector for each amino acid of a protein
- Input: Protein Family Multiple Sequence Alignment

Multiple Sequence Alignment

- Proteins are made of amino acids (residues) (20)
- MSA is an alignment of many protein sequences from the same family

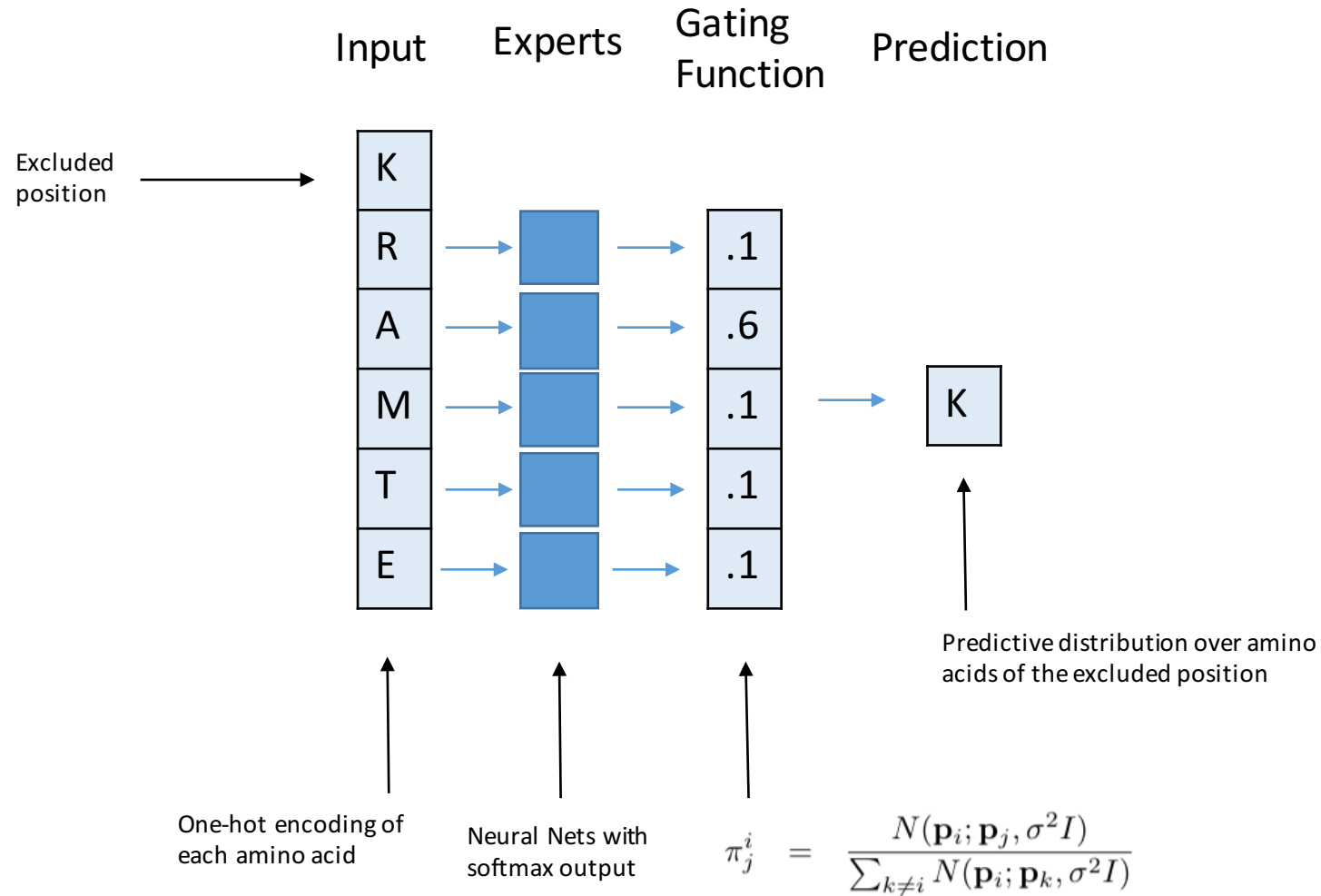


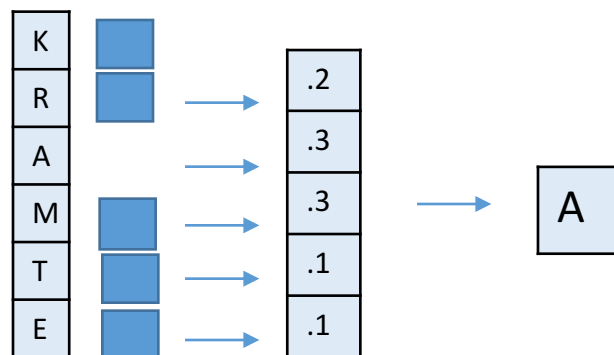
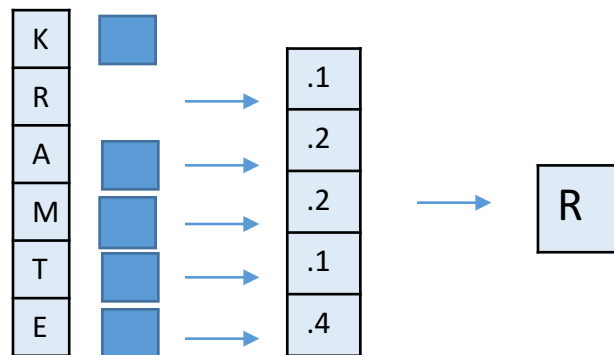
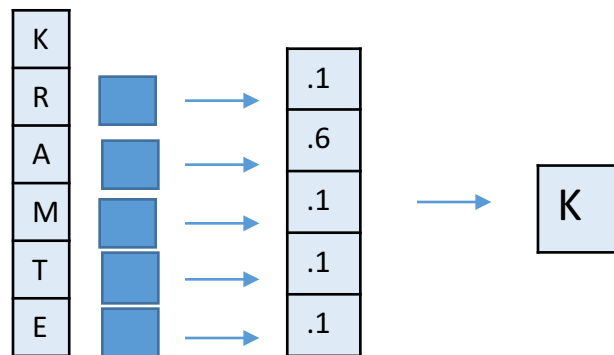
How can we use a protein family MSA to predict 3D structure?



Infer the latent 3D coordinates of each column based on which columns predict each other

Model Description





•
•
•

Objective Function

Cross-Entropy	Attract Adjacent	Repel All	Weight Decay
$E(\Theta, \Psi) = \sum_{m=1}^M \sum_{l=1}^L C(\mathbf{a}_l^m, h(m, l)) + \lambda \sum_{i=1}^{L-1} N(\mathbf{p}_i \mathbf{p}_{i+1}, \sigma^2 I)^{-1} + \nu \sum_{i=1}^L \sum_{j=1}^L N(\mathbf{p}_i \mathbf{p}_j, \sigma^2 I) + \rho \sum_{i \neq j}^L \ G_{l j}\ _F$			

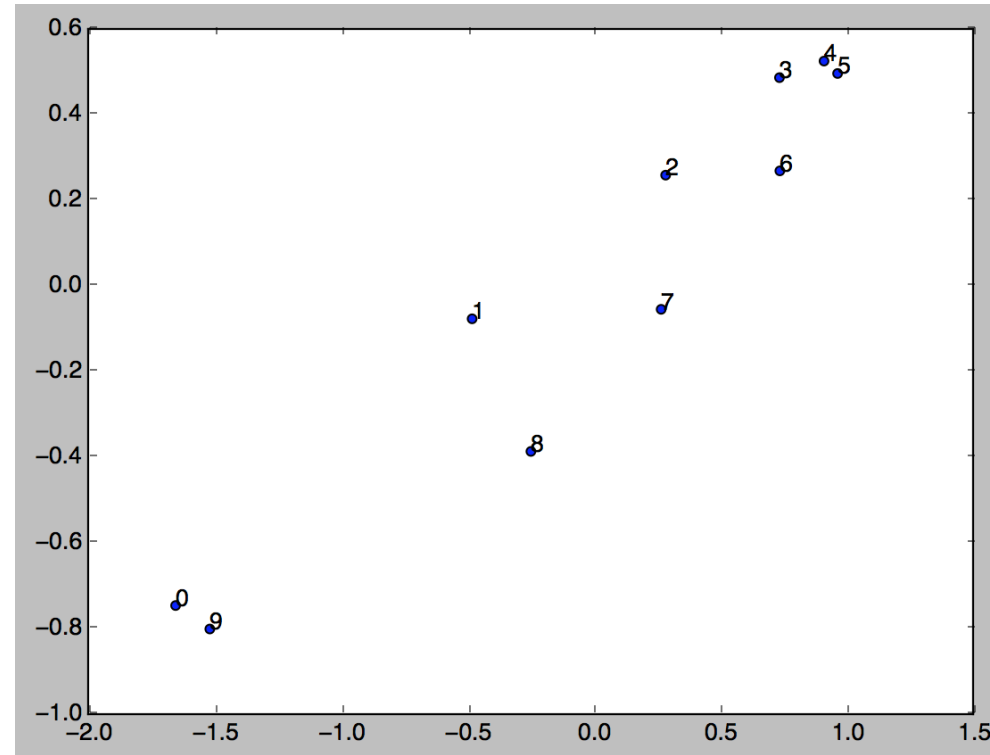
Parameters Θ :

- Weights and biases of the experts: G
- Coordinates of the columns: \mathbf{p}

Hyperparameters Ψ : λ, ν, ρ

Train via gradient descent

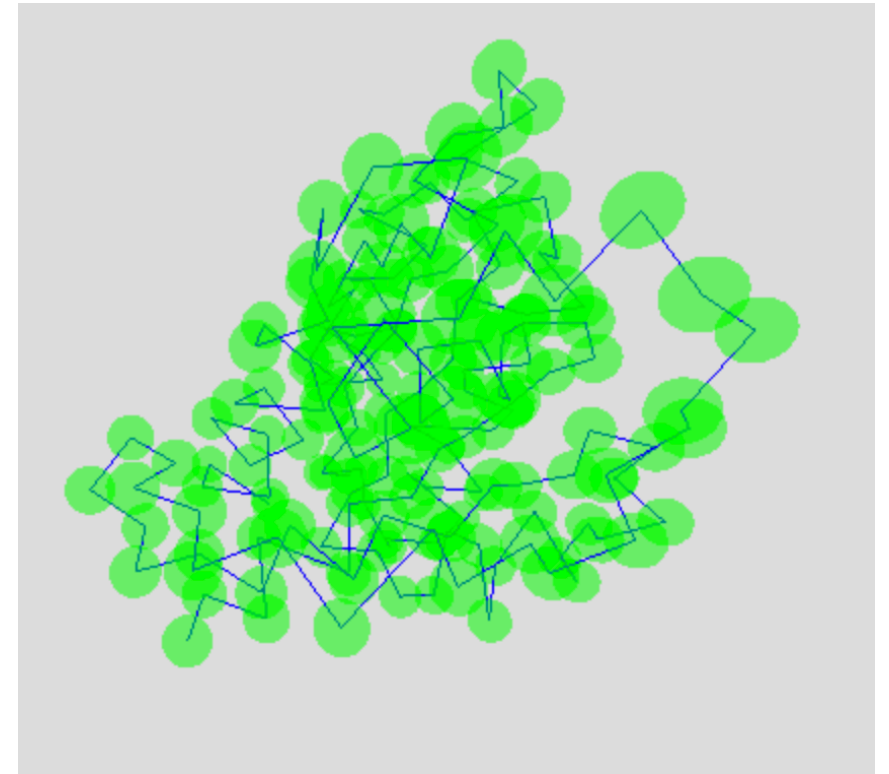
Preliminary Results



Only 0 and 9 predict each other

Upcoming Work

- Apply to real data
- Compare to other methods



RASH_HUMAN Crystallography