E.A.5.10 (Protein Folding)

1.1 Question 1 *(the protein must be valid)*

Define the schema of a problem instance I suitable for CSP modelling.

Let $S = \{S_1, S_2, ..., S_N\}$ be a sequence of length N with $S_n \in \{H, P\}$, $\mathcal{N} = \{1, ..., N\}$ and (X, D, C) s.t.

$$X = \{X_n \mid n \in \mathcal{N}\} \cup \{Y_n \mid n \in \mathcal{N}\} \tag{1}$$

– X_n is the x coordinate of protein n
– Y_n is the y coordinate of protein n

All the variables have domains $D = \{-N, -N+1, ..., 0, ...N-1, N\}$

$$C = C_{\text{adj}} \cup C_{\text{alldiff}} \tag{2}$$

All the amino-acids have a different position

1.2 Question 2 *(it must at least some contacts)*

Define the problem of finding a 2D-folding of the protein given as instance I with at least a given number $k \in \mathbb{N}$ of contacts as a CSP specification.

1.3 Question 3 *(now optimize it)*

Define the problem of finding a 2D-folding of the protein given as instance I with the maximum number $k \in \mathbb{N}$ of contacts as a Constraint Optimisation Problem specification.