To support my PhD comprehensive exam and allow initial investigation of robustness in biological sequences, I developed this dynamic programming algorithm to count the number of compatible k-mutants with the structure of a given sequence. This algorithm is a simpler and faster way to compute Z_k^* in the uniform energy model, while the algorithm proposed in aim 1 of the proposal is suitable to be extended to the Turner energy model.

For a given sequence $\mathbf{a}$, folding into structure s_0, define $Z^*(i,j,\mu)$ to be the number of sequences $\mathbf{a}[i,j]$, with $\mu \leq k$ number of mutations compatible with $s_0[i,j]$. For a nucleotide pair (x,y), let $m1(x,y)$ and $m2(x,y)$ respectively be the number of single-point and two-point mutations in (x,y), that keep base pair compatibility. For instance $m1(G,U) = 2$, because the only possible single-point mutations without destroying the base pair are $G \to A$ and $U \to C$. Therefore

$$m1(x,y) = \begin{cases} 2 & if \quad (x,y) = (G,U) or (U,G) \\ 1 & otherwise \end{cases}$$

$$m2(x,y) = \begin{cases} 3 & if \quad (x,y) = (G,U) or (U,G) \\ 4 & otherwise \end{cases}$$

The algorithm is as follows:

Base case: For for $0 \leq i \leq n$, define $Z^*(i,i,0) = 1$ and

$$Z^*(i,i,1) = \begin{cases} 3 & if \quad i \quad is \quad unpaired \\ 0 & otherwise \end{cases}$$

Case 1: j is unpaired:

$$Z^*(i,j,\mu) += 3 \cdot Z^*(i,j-1,\mu-1) + Z^*(i,j-1,\mu) \tag{1}$$

Case 2: If j is paired with position r and $r < j$:

$$Z^*(i,j,\mu) += Z^*(i,j-1,\mu) + Z^*(i,j-1,\mu-1) \cdot m1(\mathbf{a}_r, \mathbf{a}_j) + Z^*(i,j-1,\mu-2) \cdot m2(\mathbf{a}_r, \mathbf{a}_j) \tag{2}$$

Case 3: If j is paired with position r and $r > j$:

$$Z^*(i,j,\mu) += Z^*(i,j-1,\mu) \tag{3}$$

For a valid secondary structure $s_0[i,j]$ with a balanced number of base pairs, $Z^*(i,j,\mu)$ is the number of μ-mutants of $\mathbf{a}[i,j]$, compatible with $s_0[i,j]$. The time complexity of the above algorithm is $O(kn^2)$.