Grant Proposal Mark

Mark Rozanov

October 2019

1 Multithreading Cascade Algorithm for Protein Threading

1.1 Inputs:

- cryo-EM map of resolution 6 Åand better
- Sequences of all chains

1.2 Output:

Full atomic model - location and rotamer angles of all amino acids.

1.3 Preprocessing and Neural Networks

- *AAnchor* - finds position, type and confidence of an amino acid in a cryo EM map
- $NextPos(Type_0, Type_1)$. Finds position (and confidence) of an amino acids of $Type_1$, given that the previous amino acid in chain is of $Type_0$. Number of pretrained deep networks for each pair $(Type_0, Type_1)$.
- $DeepRot(Type)$ Finds rotamer angles for a given residue type, cryo-EM map , and C_α location.

1.4 Procedure (assume one big chain with amino acids $a_1 \dots a_M$)

Denote by :

- $B^k_{n,m}$ candidate location and folding of amino acids $a_n, a_{n+1}, \dots, a_m$. k is a candidate number.
- $B^k_{n,m} = \{\overline{x}^k_n, \overline{x}^k_{n+1}, \overline{x}^k_m, \}$, where x^i_j is the i-st candidate for the position of the C_α of an amino acid number j.
- $P(B^k_{n,m})$ probability of the path $B^k_{n,m}$

Initialize Path Calculate $B^k_{i,i}$ and $P(B^k_{i,i})$ by running AAnchor.

Step For each k, candidate $B^k_{n,m}$ and $P(B^k_{n,m})$, calculate $B^k_{n,m+1}$ and $P(B^k_{n,m+1})$

1. Find 3D coordinates and probability of next amino acid $B^k_{m+1,m+1} = \{\overline{x}^k_{m+1}\}$, $P(B^k_{m+1,m+1})$ using deep NN $NextPos(a_m, a_{m+1})$
2. Calculate folding energy of the new path $G(B^k_{n,m+1})$ use Rosette Energy function
3. update probility: $P(B^k_{n,m+1}) = f\left(P(B^k_{n,m}), P(B^k_{m+1,m+1}), G(B^k_{n,m+1})\right)$

Branch If in the step there is more then one candidate from $NextPos(a_m, a_{m+1})$ initiate new path $B^{k+1}_{n,m+1}$

End Given two thresholds : P_{low} and P_{high}. For each path $B^k_{i,j}$ in the memory

- **Delete** $B^k_{i,j}$ if $P(B^k_{i,j} < P_{low}$
- **Save local path** $B^k_{i,j}$ if $P(B^k_{i,j} > P_{high}$

Refine

- Run $DeepRot(Type)$ and obtain rotamer angles

1.5 Highlights

- We can achieve full modelling for resolutions up to 5 Å.
- While the results of $NextPos(a_m, a_{m+1})$ and $AAnchor$ could be of low confidence, using multyhypothesis approach will enable high precision threading
- In calculation of propabilities $P(B^k_{n,m})$ evolutionary, SSE and other data can be considered
- The precision of NNs : $AAnchor$, $NextPos$, $DeepRot$ can be improved be retraining on homolog molecules with simulated cryo-EM maps