Template based protein structure prediction

List of functions

function	input	output	description
Get data phase			
Seq blast	Target seq Threshold (90%)	Alignments with scores Pdbs	Submit the sequence to blast and get the results
Get alignment	List of alignment (from Seq_blast)	X, Y, Z coordinates for each position in each sequence	Get the coordinates from each sequence
Model generation phase			
Map pdb	Matching XYZ coordinates of top alignment	XYZ of model	Get the xyz of the backbone atoms in the template that matches the target
Gap detection	Target sequence Top template	List of gaps	Each gap will contain the missing residues as well as one residue before gap and one residue after gap
Gap filling	gap	Filled gap	For each gap, take the residues in the gap in addition to one residue before and one residue after the gap. Search the alignments and fill it with the highest score possible
Generate model	All XYZ (from model ad from gaps)	Model generated	Put all coordinates from top alignment and the ones obtained from gap filling together
Optimization			
Get pdf	Get XYZ coordinates from each position in each alignment	Pdf for X, Y, Z for each position for each alignment	Assume each pdf is a gaussian pdf and compute the “mean” and the “standard deviation”
Pdb optimization	XYZ from generate model Pdf functions	Modify every X, Y, Z coordinates to maximize corresponding pdf	Use the gradient descent to obtain the maximum probability

			Record the corresponding (X, Y, Z) value
Generate model optimized	Updated X, Y, Z coordinates	PDB file	Update the pdb file for the target using the updated coordinates
SWRL4	pdb	Pdb with residue	Add the side chain to each residue