**Algorithm: Patch based protein comparison**

*Idea* – Cover all c-alphas of a protein by overlapping patches which considers local properties of the protein. Comparing patches (or, local regions) is supposed to show higher local similarity in similar binding site regions.

*Input*: TSR triangles for two proteins.

*Output*: allxall patch similarity between two proteins.

*Method*:

Step 1. Generate TSR keys - Each TSR triangle will have one representative integer key, three c-alphas and their 3D co-ordinates.

Step 2. Form Patches - Find all c-alphas in each protein under consideration and form patch surrounding each c-alpha (patch center *c*).

2.1. Decide upon a patch radius *r* and maximum #of patch neighbors *n*,

2.2. For each patch center *c*,

2.2.1. calculate Euclidean distance between *c* and all other neighboring c-alphas,

2.2.2. select maximum of *n* neighbors with distance <= *r*,

2.2.3. calculate all keys formed with the neighboring c-alphas,

2.2.4. save patch center c and corresponding neighboring c-alphas as well as neighboring keys.

Step 3. Compare patches – Compare all patches across two proteins.

3.1. Retrieve all saved patches for two proteins in comparison,

3.2. Calculate Jaccard and generalized Jaccard Similarity between every pair of patches across two proteins.

3.2.1. The generalized Jaccard coefficient between two proteins X and Y is given by

J(X, Y)gen = ∑min(Xi,Yi)/∑max(Xi,Yi) , where i = 1 to n , n is the total #of keys in two proteins.

If a key is present in a protein it’s frequency is considered and if it not present it’s frequency is taken as 0, they are represented by Xi and Yi.

3.2.2. The Jaccard coefficient does not consider the frequency of keys, if a key is present, it’s frequency is given by 1, else, it is given by 0. So, the formula can be simplified as

J(X, Y) = #of common keys between X and Y / #of total keys between X and Y

We have also considered +/- 1 or 2 for Jaccard Similarity calculation. Here, to search if a key is common between two patches from proteins in comparison, we have searched for all keys in the +/- 1 or 2 range of Xi as a match in keys of Y. If there is a match found, we consider both keys as matched keys and include in list of common keys. Later, we calculate the protein of total keys which matched.

3.3. Save all pairs of patches wand their similarity in a file.

Step 4. Search for binding site residues with high patch similarity.