Supplemental Material for

Detecting Local Ligand-Binding Site Similarity in Non-Homologues Proteins by Surface Patch Comparison

Lee Sael & Daisuke Kihara
E-mail: dkihara@purdue.edu

Figure S1. Partial matching algorithm for comparing surface patches of pockets.

Surface patches of the two pockets are matched according to the distance (dissimilarity) so that the total distance of the matched pairs is minimized. This is similar to the weighted bipartite matching problem, which can be approximately solved by the auction algorithm [1]. Since we want to obtain pairs of patches that minimize total distance while the original auction algorithm maximizes the total weight values of pairs, we defined the weight for a pair of patches as *(Constant-value – the Euclidean distance of the 3DZD vectors)*. The pseudo code of the modified bipartite matching is shown below.

```
// Input: local surface patches of pocket A and pocket B, lpodA and lpodB.
//          lpodA = [ spd^A_0, spd^A_1, ... ,  spd^A_nA],   lpodB = [ spd^B_0, spd^B_1, ... ,  spd^B_nB]
//          The number of patches in pocket A is larger than pocket B (i.e. n_A ≥ n_B) else they are reversed.
// constant NUM is an arbitrary large value, larger than the maximum distance of patch pairs: NUM > d_ij

Initialization:
        SET δ ← 1/(n_A + 1)                              // δ is to control minimum "bid" in the auction
        Store all patches of lpodB i to queue Q ← i
        FOR j=1 to n_A DO                                //initializing values for patches in lpodA
              SET p_j ← 0 and SET pair_j ← –1
                                   //p_j stores the minimum bid for spd^A_j
                                   //pair_j stores the ID of the paired patch from lpodB for spd^A_j
        ENDFOR

Iteration:
        WHILE Q is not empty AND number of iteration is less than 10*n_A
              SET i ← value of front node in Q           //choose spd^B_i for a query and remove it from Q
              Delete the front node of Q
              Find j (spd^A_i) that maximizes w_ij – p_j  where w_ij is NUM  -  d_ij
                                                         // d_ij is the Euclidean distance of 3DZD
              IF w_ij – p_j >= 0 THEN
                     Push current pair of j, pair_j, into back of Q
                     SET pair_j ← i                      // spd^B_i  is assigned to spd^A_i
                     Update p_j ← p_j + δ //raise the minimum bid value for pairing with spd^A_i

              ENDIF
        ENDWHILE
Output:
        Output pairs of (pair_j, j) for all pair_j not equal to –1
```

The algorithm works as follows: First all patches in pocket B is stored in the queue *Q*. The queue *Q* becomes empty when each patch in pocket B either finds a satisfying patch in pocket A or does not find any similar patch. No more than one patch in B is assigned to a patch in A. For a query patch spd^B_i, when it finds a similar patch, spd^A_i, the previous patch in B that paired with spd^A_i is put back to the *Q* and new patch in B, spd^B_i is assigned to spd^A_i. The patch in B which is put back to *Q* has another round to be evaluated to find a patch in A. When patches are competed for a same spd^A_i, the value *p* for spd^A_i is increased, so that at the end a patch in B that is most similar to spd^A_i will be selected for its pair. This is the intention of raising the minimum bid value, p_j, at each iteration. In the end, the algorithm output the pairs of patches that minimize the overall distance between the patches in A and B.

Figure S2. Prediction success rates of different weight values. **A** shows the top-3 prediction rates for different combination of weight w1, in Equation 8, and weight w2, in Equation 9, are plotted. All four features, SHEV, were used. **B** shows the effect of removing n_A/N terms in Equation 6 and 7 with varying weights w1 and w2.

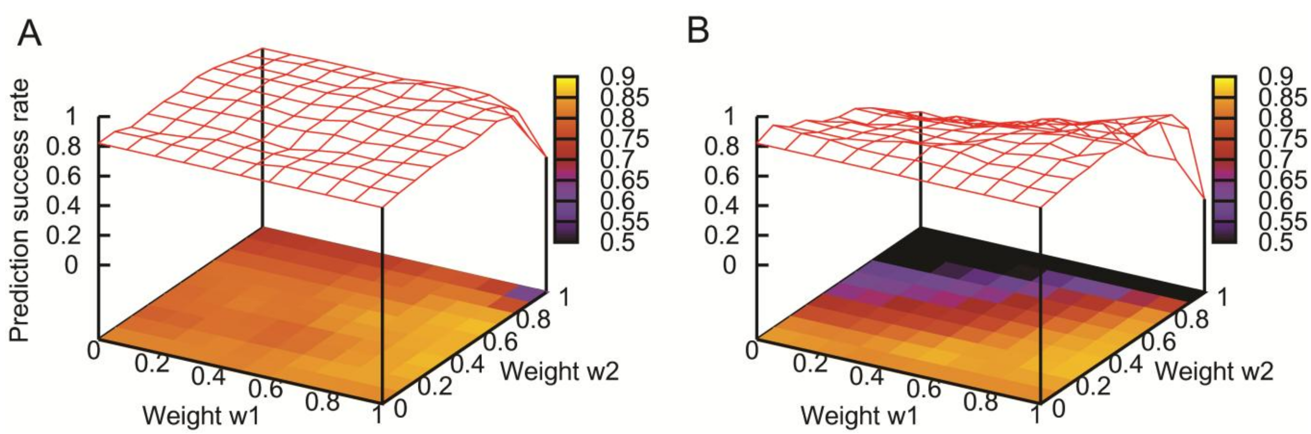
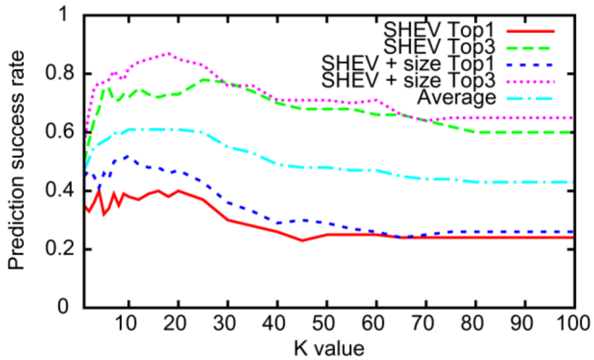


Figure S3. Prediction rate of different k-values on the Kahraman dataset. The prediction rates for different k value in Equation 11 are shown.



References

Demange, G.; Gale, D.; Sotomayor, M. Multi-Item Auctions. *The Journal of Political Economy* **1986a**, *94*, 863–872.