**meta-PPISP: a Meta Web Server for Protein-Protein Interaction Site Prediction**

*Supplementary Information*

Sanbo Qin and Huan-Xiang Zhou

## 1. Coefficients of linear regression

$c_0 = 0.0084$

$c_{ij}$

| $i$ = | $j$ = 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.0943 | 0.0215 | 0.0024 | 0.0190 | -0.0264 | -0.0359 | 0.0439 | 0.0633 | 0.0485 |
| 2 | 0.2860 | 0.0134 | -0.0271 | 0.0828 | 0.0625 | 0.0796 | 0.0496 | -0.0543 | 0.0381 |
| 3 | 0.3407 | 0.0010 | -0.1047 | -0.0845 | 0.0010 | -0.0223 | -0.0644 | 0.0403 | -0.0316 |

Note: $i$ = 1, 2, and 3 refer to cons-PPISP, PINUP, and Promate, respectively.

## 2. Real and Predicted Numbers of Interface Residues of Enz35 Proteins

| Unbound | Bound | Real | Pred | TP | FP | Cov | Acc |
|---|---|---|---|---|---|---|---|
| 1EGL_ | 1ACB_l | 15 | 28 | 15 | 13 | 1.000 | 0.536 |
| 1BA7_B | 1AVX_l | 13 | 19 | 12 | 7 | 0.923 | 0.632 |
| 1A19_B | 1AY7_l | 14 | 16 | 13 | 3 | 0.929 | 0.812 |
| 1RGH_B | 1AY7_r | 13 | 21 | 7 | 14 | 0.538 | 0.333 |
| 1HOE_ | 1BVN_l | 19 | 25 | 16 | 9 | 0.842 | 0.640 |
| 1PIG_ | 1BVN_r | 25 | 9 | 8 | 1 | 0.320 | 0.889 |
| 1HPT_ | 1CGI_l | 19 | 22 | 17 | 5 | 0.895 | 0.773 |
| 1K9B_A | 1D6R_l | 14 | 33 | 5 | 28 | 0.357 | 0.152 |
| 2BNH_ | 1DFJ_l | 34 | 12 | 11 | 1 | 0.324 | 0.917 |
| 9RSA_B | 1DFJ_r | 26 | 31 | 15 | 16 | 0.577 | 0.484 |
| 1CJE_D | 1E6E_l | 24 | 26 | 19 | 7 | 0.792 | 0.731 |
| 1E1N_A | 1E6E_r | 26 | 16 | 7 | 9 | 0.269 | 0.438 |
| 9PTI_ | 1EAW_l | 20 | 24 | 19 | 5 | 0.950 | 0.792 |
| 1CZP_A | 1EWY_l | 19 | 19 | 12 | 7 | 0.632 | 0.632 |
| 1GJR_A | 1EWY_r | 18 | 12 | 1 | 11 | 0.056 | 0.083 |
| 1ECZ_AB | 1EZU_l | 22 | 41 | 0 | 41 | 0.000 | 0.000 |
| 1F32_A | 1F34_l | 29 | 23 | 16 | 7 | 0.552 | 0.696 |
| 4PEP_ | 1F34_r | 33 | 19 | 10 | 9 | 0.303 | 0.526 |
| 1BX8_ | 1HIA_l | 15 | 25 | 15 | 10 | 1.000 | 0.600 |
| 2HPR_ | 1KKL_l | 20 | 18 | 13 | 5 | 0.650 | 0.722 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1JB1_ABC | 1KKL_r | 36 | 32 | 3 | 29 | 0.083 | 0.094 |
| 1FSC_ | 1MAH_l | 21 | 32 | 7 | 25 | 0.333 | 0.219 |
| 1J06_B | 1MAH_r | 25 | 11 | 7 | 4 | 0.280 | 0.636 |
| 1LU0_A | 1PPE_l | 14 | 25 | 14 | 11 | 1.000 | 0.560 |
| 1B1U_A | 1TMQ_l | 27 | 25 | 19 | 6 | 0.704 | 0.760 |
| 2UGI_B | 1UDI_l | 22 | 20 | 16 | 4 | 0.727 | 0.800 |
| 1UDH_ | 1UDI_r | 25 | 18 | 12 | 6 | 0.480 | 0.667 |
| 2RAC_A | 2MTA_l | 16 | 15 | 12 | 3 | 0.750 | 0.800 |
| 2BBK_JM | 2MTA_r | 30 | 25 | 1 | 24 | 0.033 | 0.040 |
| 1YCC_ | 2PCC_l | 16 | 18 | 6 | 12 | 0.375 | 0.333 |
| 1CCP_ | 2PCC_r | 15 | 12 | 4 | 8 | 0.267 | 0.333 |
| 3SSI_ | 2SIC_l | 13 | 18 | 13 | 5 | 1.000 | 0.722 |
| 1SUP_ | 2SIC_r | 22 | 30 | 18 | 12 | 0.818 | 0.600 |
| 1M08_B | 7CEI_l | 18 | 17 | 0 | 17 | 0.000 | 0.000 |
| 1UNK_D | 7CEI_r | 19 | 15 | 9 | 6 | 0.474 | 0.600 |
| Total | | 737 | 752 | 372 | 380 | 0.505 | 0.495 |

This list of 35 proteins is from the enzyme/inhibitor category of Docking Benchmark 2.0 of Mintseris et al. [Proteins 60, 214 (2005)]. Proteins with higher than 35% sequence identities were eliminated.

Predictions were done on unbound structures. The real interface residues (Real) were identified in the bound complex (defined with a 5 Å distance cutoff). These residues were then mapped to the unbound structure through sequence alignment. The two subunits of the bound complex are denoted as l (for ligand) and r (for receptor).

In this and the following two tables, predictions of meta-PPISP obtained at a threshold value of 0.34 for positive prediction are shown. The intersection of Real and predicted (Pred) interface residues is true positive (TP). The rest of Pred is false positive (FP). Coverage (Cov) is calculated as TP/Real; accuracy (Acc) is calculated as TP/Pred.

**3. Real and Predicted Numbers of Interface Residues of CAPRI Targets**

| Target | Real | Pred | TP | FP | Cov | Acc |
|---|---|---|---|---|---|---|
| T01_l | 12 | 16 | 10 | 6 | 0.833 | 0.625 |
| T01_r | 12 | 18 | 0 | 18 | 0.000 | 0.000 |
| T02_l | 18 | 30 | 13 | 17 | 0.722 | 0.433 |
| T02_r | 19 | 31 | 0 | 31 | 0.000 | 0.000 |
| T03_l | 25 | 20 | 15 | 5 | 0.600 | 0.750 |
| T03_r | 46 | 50 | 9 | 41 | 0.196 | 0.180 |
| T04_l | 24 | 36 | 14 | 22 | 0.583 | 0.389 |
| T04_r | 23 | 12 | 0 | 12 | 0.000 | 0.000 |
| T05_l | 19 | 30 | 10 | 20 | 0.526 | 0.333 |
| T05_r | 22 | 13 | 1 | 12 | 0.045 | 0.077 |

| | | | | | | |
|---|---|---|---|---|---|---|
| T06_l | 23 | 33 | 9 | 24 | 0.391 | 0.273 |
| T06_r | 31 | 12 | 8 | 4 | 0.258 | 0.667 |
| T07_l | 12 | 13 | 4 | 9 | 0.333 | 0.308 |
| T07_r | 18 | 23 | 1 | 22 | 0.056 | 0.043 |
| T08_l | 21 | 24 | 2 | 22 | 0.095 | 0.083 |
| T08_r | 22 | 28 | 0 | 28 | 0.000 | 0.000 |
| *T09_l | 40 | 19 | 0 | 19 | 0.000 | 0.000 |
| *T10_l | 90 | 19 | 9 | 10 | 0.100 | 0.474 |
| **T12_l | 13 | 26 | 8 | 18 | 0.615 | 0.308 |
| T12_r | 25 | 20 | 9 | 11 | 0.360 | 0.450 |
| T13_l | 19 | 18 | 0 | 18 | 0.000 | 0.000 |
| T13_r | 25 | 21 | 18 | 3 | 0.720 | 0.857 |
| T14_l | 59 | 12 | 4 | 8 | 0.068 | 0.333 |
| T14_r | 37 | 26 | 5 | 21 | 0.135 | 0.192 |
| T15_l | 19 | 24 | 12 | 12 | 0.632 | 0.500 |
| T15_r | 23 | 18 | 13 | 5 | 0.565 | 0.722 |
| T16_l | 23 | 27 | 20 | 7 | 0.870 | 0.741 |
| T16_r | 28 | 13 | 7 | 6 | 0.250 | 0.538 |
| T17_l | 13 | 24 | 3 | 21 | 0.231 | 0.125 |
| T17_r | 19 | 27 | 0 | 27 | 0.000 | 0.000 |
| T18_l | 21 | 25 | 9 | 16 | 0.429 | 0.360 |
| T18_r | 23 | 33 | 12 | 21 | 0.522 | 0.364 |
| T19_l | 21 | 25 | 7 | 18 | 0.333 | 0.280 |
| T19_r | 26 | 29 | 18 | 11 | 0.692 | 0.621 |
| T20_l | 42 | 11 | 10 | 1 | 0.238 | 0.909 |
| T20_r | 35 | 22 | 12 | 10 | 0.343 | 0.545 |
| T21_l | 13 | 21 | 0 | 21 | 0.000 | 0.000 |
| T21_r | 15 | 24 | 5 | 19 | 0.333 | 0.208 |
| T22_l | 14 | 23 | 4 | 19 | 0.286 | 0.174 |
| T22_r | 16 | 17 | 1 | 16 | 0.062 | 0.059 |
| *T23_l | 44 | 15 | 8 | 7 | 0.182 | 0.533 |
| ***T25_l | 20 | 17 | 4 | 13 | 0.200 | 0.235 |
| T25_r | 23 | 22 | 17 | 5 | 0.739 | 0.773 |
| T26_l | 24 | 15 | 5 | 10 | 0.208 | 0.333 |
| T26_r | 33 | 10 | 6 | 4 | 0.182 | 0.600 |
| T27_l | 13 | 18 | 0 | 18 | 0.000 | 0.000 |
| T27_r | 10 | 12 | 0 | 12 | 0.000 | 0.000 |
| Total | 1173 | 1022 | 322 | 700 | 0.275 | 0.315 |

*Homo-dimeric or homo-trimeric targets

**T11 has the same bound complex as T12. For T12, one subunit is from the bound complex; for T11, that subunit is built by homology.

***T24 has the same bound complex as T25. For T25, one subunit is from the bound complex; for T24, that subunit is built by homology.

**4. Real and Predicted Numbers of Interface Residues of Dockground32 Proteins**

| Unbound | Bound | Real | Pred | TP | FP | Cov | Acc |
|---|---|---|---|---|---|---|---|
| 1g41_A | 1ofh_A | 13 | 13 | 1 | 12 | 0.077 | 0.077 |
| 1jjw_C | 1ofh_G | 14 | 23 | 0 | 23 | 0.000 | 0.000 |
| 1qlp_A | 1oph_A | 12 | 22 | 11 | 11 | 0.917 | 0.500 |
| 1hj9_A | 1oph_B | 15 | 16 | 5 | 11 | 0.333 | 0.312 |
| 1scn_E | 1oyv_B | 19 | 38 | 13 | 25 | 0.684 | 0.342 |
| 1oyv_I* | 1oyv_I | 10 | 30 | 5 | 25 | 0.500 | 0.167 |
| 1scn_E | 1r0r_E | 21 | 41 | 15 | 26 | 0.714 | 0.366 |
| 2gkr_I | 1r0r_I | 12 | 19 | 7 | 12 | 0.583 | 0.368 |
| 1t6e_X | 1t6g_A | 22 | 16 | 10 | 6 | 0.455 | 0.625 |
| 1bk1_A | 1t6g_C | 20 | 22 | 8 | 14 | 0.400 | 0.364 |
| 1ta3_B* | 1ta3_B | 28 | 12 | 7 | 5 | 0.250 | 0.583 |
| 1om0_A | 1ta3_A | 22 | 32 | 21 | 11 | 0.955 | 0.656 |
| 1om0_A | 1te1_A | 20 | 22 | 0 | 22 | 0.000 | 0.000 |
| 1te1_B* | 1te1_B | 19 | 29 | 6 | 23 | 0.316 | 0.207 |
| 2a31_A | 1tx6_A | 25 | 14 | 12 | 2 | 0.480 | 0.857 |
| 1c2a_A | 1tx6_I | 20 | 23 | 0 | 23 | 0.000 | 0.000 |
| 1sup_A | 1v5i_A | 26 | 26 | 12 | 14 | 0.462 | 0.462 |
| 1v5i_B* | 1v5i_B | 21 | 19 | 14 | 5 | 0.667 | 0.737 |
| 1cpy_A | 1wpx_A | 40 | 30 | 15 | 15 | 0.375 | 0.500 |
| 1wpx_B* | 1wpx_B | 32 | 14 | 13 | 1 | 0.406 | 0.929 |
| 1zhr_A | 1xx9_A | 36 | 17 | 16 | 1 | 0.444 | 0.941 |
| 1ecz_AB | 1xx9_CD | 33 | 37 | 15 | 22 | 0.455 | 0.405 |
| 1m4l_A | 1zlh_A | 27 | 15 | 12 | 3 | 0.444 | 0.800 |
| 1zlh_B* | 1zlh_B | 25 | 33 | 20 | 13 | 0.800 | 0.606 |
| 1p14_A | 2auh_A | 21 | 29 | 7 | 22 | 0.333 | 0.241 |
| 2auh_B* | 2auh_B | 24 | 33 | 23 | 10 | 0.958 | 0.697 |
| 1l8t_A | 2bkk_A | 12 | 13 | 0 | 13 | 0.000 | 0.000 |
| 2bkk_B* | 2bkk_B | 18 | 24 | 3 | 21 | 0.167 | 0.125 |
| 2bkq_A | 2bkr_A | 35 | 27 | 17 | 10 | 0.486 | 0.630 |
| 1ndd_A | 2bkr_B | 23 | 21 | 14 | 7 | 0.609 | 0.667 |
| 1hj9_A | 2fi4_E | 14 | 19 | 7 | 12 | 0.500 | 0.368 |
| 1bpi_A | 2fi4_I | 13 | 23 | 13 | 10 | 1.000 | 0.565 |
| Total | | 692 | 752 | 322 | 430 | 0.465 | 0.428 |

These proteins are collected from http://dockground.bioinformatics.ku.edu/. They consist of subunits of enzyme-inhibitor complexes deposited between 2003 and 2006 in the Protein Data Bank (http://www.rcsb.org/pdb). The sequence identities are less than 30%. Unbound proteins and their bound counterparts have at least 97% sequence identities. For eight proteins (indicated by *), no unbound structures are available and the bound structures are used for prediction.

The final predictions of cons-PPISP, Promate, and PINUP on these 32 proteins have coverages of 39.0%, 16.8%, and 34.7%, respectively; the corresponding accuracies are 38.2%, 47.0%,

41.2%. At the same coverages, the accuracy of meta-PPISP is 46.4%, 53.9%, and 47.0%, respectively. The increase in accuracy is 8.2, 6.9, and 5.8 percentage points, respectively.

**5. Comparison of meta-PPISP to Three Individual Methods and to Alanine Mutation Results on Ribonuclease Inhibitor**

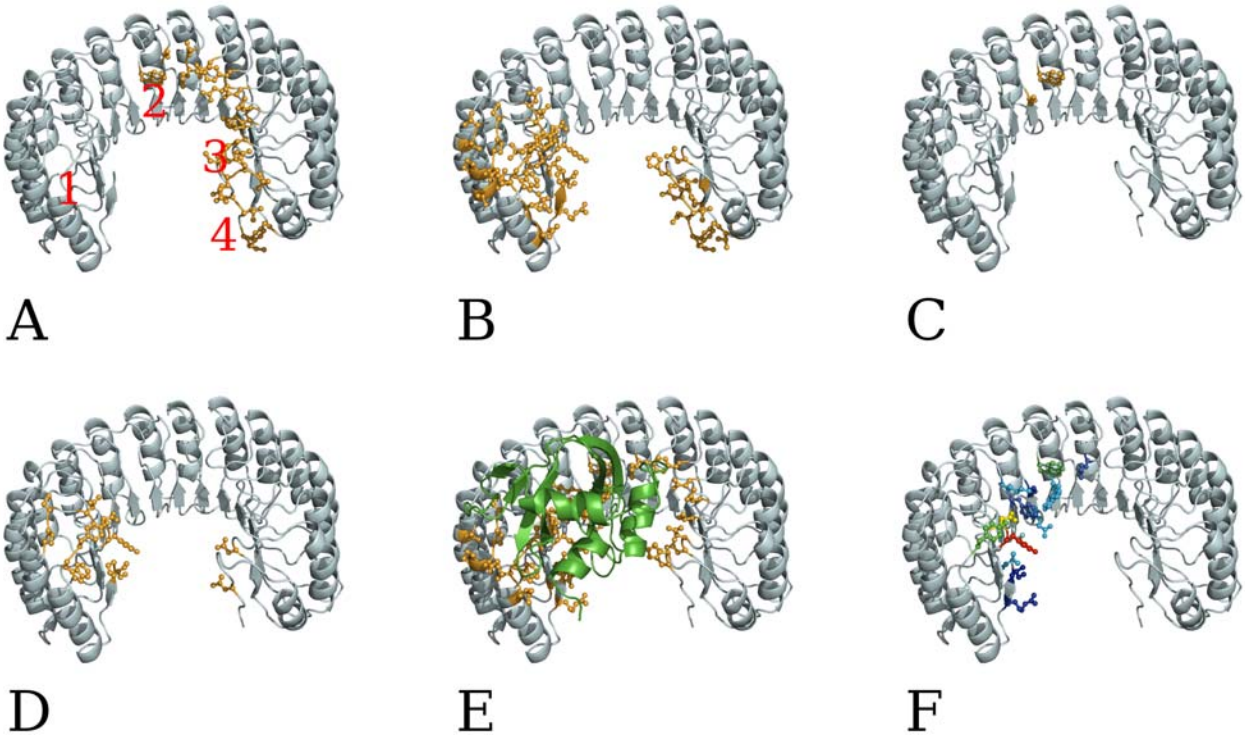


**Fig. S5.1** Detailed Comparison of interface predictions on ribonuclease inhibitor by (A) cons-PPISP; (B) PINUP; (C) Promate; and (D) meta-PPISP. To simplify the comparison, the predicted interface residues, shown by gold side chains, are divided into four sub-sites [labeled in (A) as 1 to 4]. In (E) the residues of ribonuclease inhibitor found in the interface with ribonuclease A (PDB entry 1dfj) are shown by side chains (ribonuclease A shown as green ribbon). It can be seen that sub-sites 1, 2, and 3 are correct predictions, but sub-site 4 is false positive. Sub-site 1 has the most interface residues.

cons-PPISP predicts sub-sites 2, 3, and 4. PINUP predicts sub-sites 1, 3, and 4; in addition to sub-site 4, there are also a number of false-positive predictions around sub-site 1. Promate predicts only sub-site 2. In comparison, meta-PPISP predicts subs-site 1 and 2 and misses the few interface residues in subs-site 3; it makes a single false-positive prediction. Relative to cons-PPISP and PINUP, the number of false positives is substantially reduced; relative to Promate, the number of true positives is substantially increased.

In Fig. S5.1(F), changes in the binding energy ($\Delta\Delta G$) between ribonuclease inhibitor and ribonuclease A by alanine mutations are mapped by color. The $\Delta\Delta G$ values of 14 mutations,

obtained from ASEdb (http://nic.ucsf.edu/asedb/) and ranging from 0.3 to 5.9 kcal/mol, are display by the color spectrum from blue to red. The mutations with the highest $\Delta\Delta G$ values are located in sub-site 1. This is where the predicted interface residues of meta-PPISP are mostly located; the raw scores of sub-site 1 residues are also the highest.

Since meta-PPISP is built on the raw scores of the individual methods, it is thus also interesting to compare the raw scores of meta-PPISP with those of the individual methods (Fig. S5.2). Relative to cons-PPISP, the highest raw scores in meta-PPISP move from sub-sites 2 and 3 to sub-site 1. While relatively high raw scores in Promate are spread over a large portion of the protein surface, the high scores in meta-PPISP are concentrated. In sub-site 2, relatively high raw scores are missing in PINUP but are present in meta-PPISP (although these scores are not high enough to pass the threshold $S_{th}$).

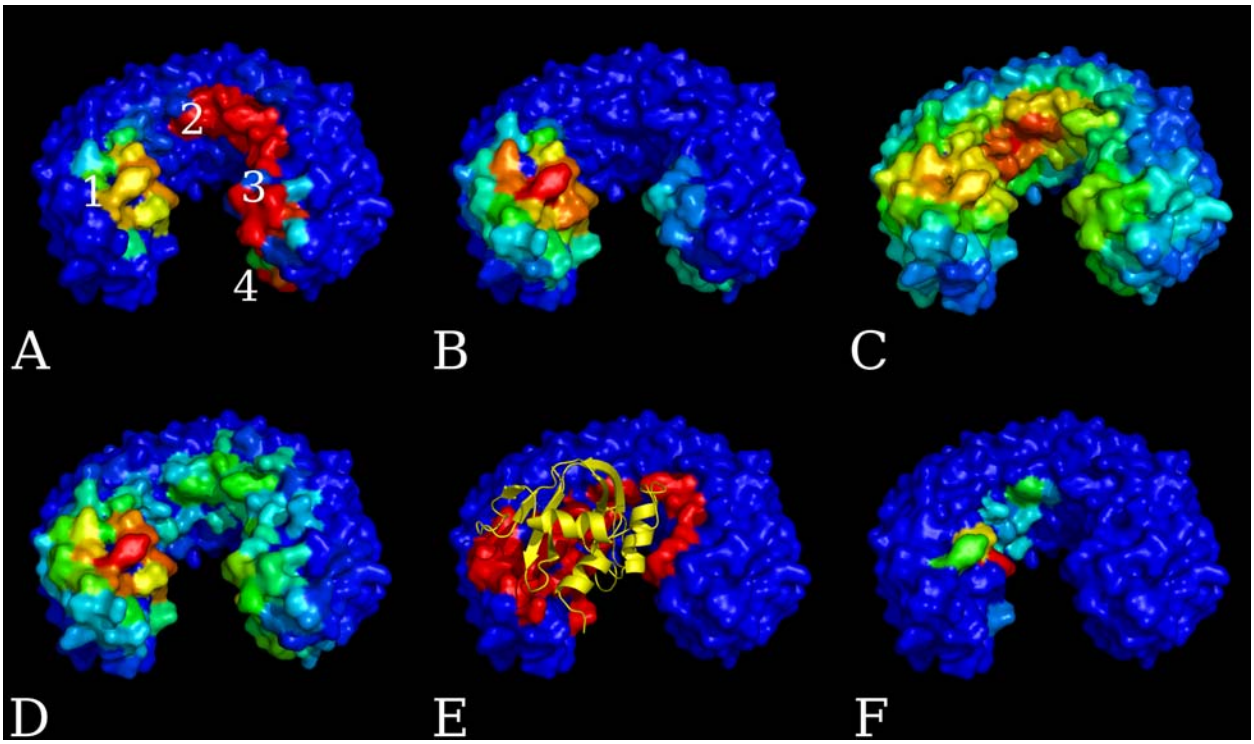


**Fig. S5.2** Comparison of raw scores for interface prediction. The labels A-F and 1-4 have the same meanings as in Fig. S5.1; ribonulease inhibitor is now represented by surface. In (A)-(D), raw scores from high to low values are represented by the spectrum from red to blue. In (E), ribonuclease inhibitor residues in contact with ribonuclease A in PDB 1dfj are shown in red, and ribonuclease A is shown as yellow ribbon. In (F), changes in binding energy by alanine mutations are color mapped as in Fig. S5.1(F).