Text S4. Statistical Analysis of the Highest Ranked Models Selected by DockRank and Other Methods on ZDock3-BM3 decoys

The non-parametric Friedman test (see discussion in [75]) shows that the null hypothesis that the highest ranked conformations produced by the different scoring functions have the same mean values for I-RMSD (where the mean is computed over the 67 cases) can be rejected with high confidence (p-value < 0.0001). We further applied the Nemenyi test (see discussions in [75]) to determine whether the differences in the means of I-RMSDs of the highest ranked conformations produced by any pair of scoring functions are statistically significant. Figure S4 shows that, at a significance level of 0.05, there is no statistically significant difference between the mean I-RMSD values of the highest ranked conformations returned by DockRank's scoring function using actual partner-specific interface residues and DockRank's scoring function using PS-HomPPI predicted interface residues. Moreover, when performance is measured in terms of I-RMSD of the highest ranked conformation, DockRank significantly outperforms other scoring schemes, whereas the differences between the two energy-based scoring functions, ZRank and IRAD, and the variants of DockRank's scoring function using three non partner-specific predictors, NPS-HomPPI, PrISE, meta-PPISP, are not statistically significant.