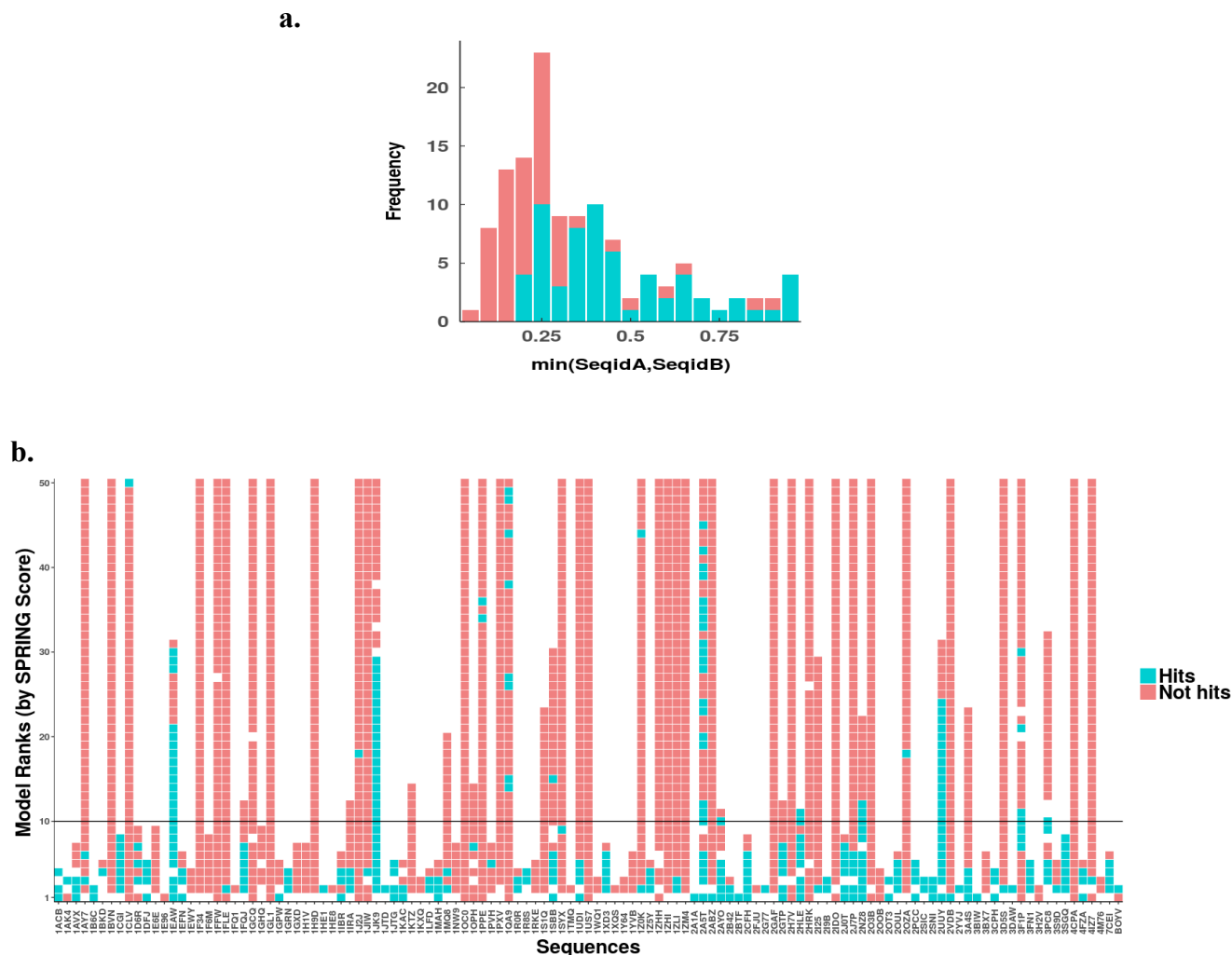


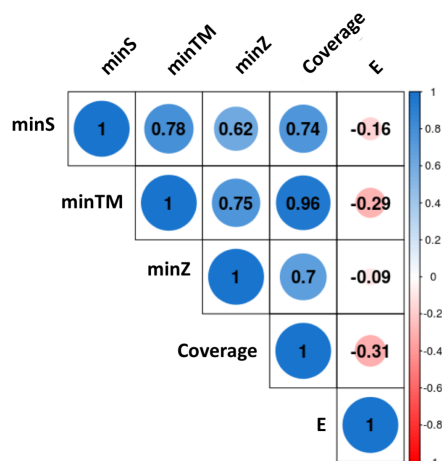
Supplementary Information for

Integrating *ab initio* and template-based algorithms for protein-protein complex structure prediction

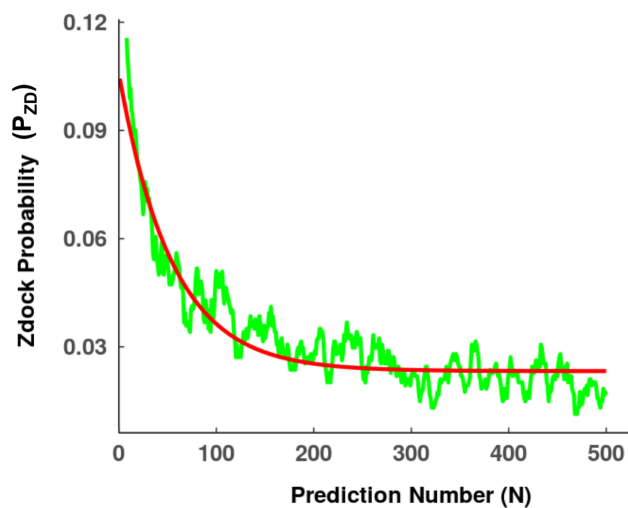
Sweta Vangaveti, Thom Vreven, Yang Zhang and Zhiping Weng*



Supplementary Figure 1. Performance of SPRING on 121 test cases. **a.** The distribution of the sequence identities between the query proteins and the top ranked SPRING prediction for the test cases. The smaller of the two sequence identities — seqidA, seqidB for query proteins A and B respectively, is plotted. **b.** Each column indicates one test case, represented by its PDB ID, and the grid squares in the column represent SPRING predictions for this test case, plotted according to the SPRING ranks (based on SPRING score, S_{SP}) on the y-axis. Cyan and red indicate predictions that are hits ($I\text{-RMSD} < 5 \text{ \AA}$) and non-hits respectively.



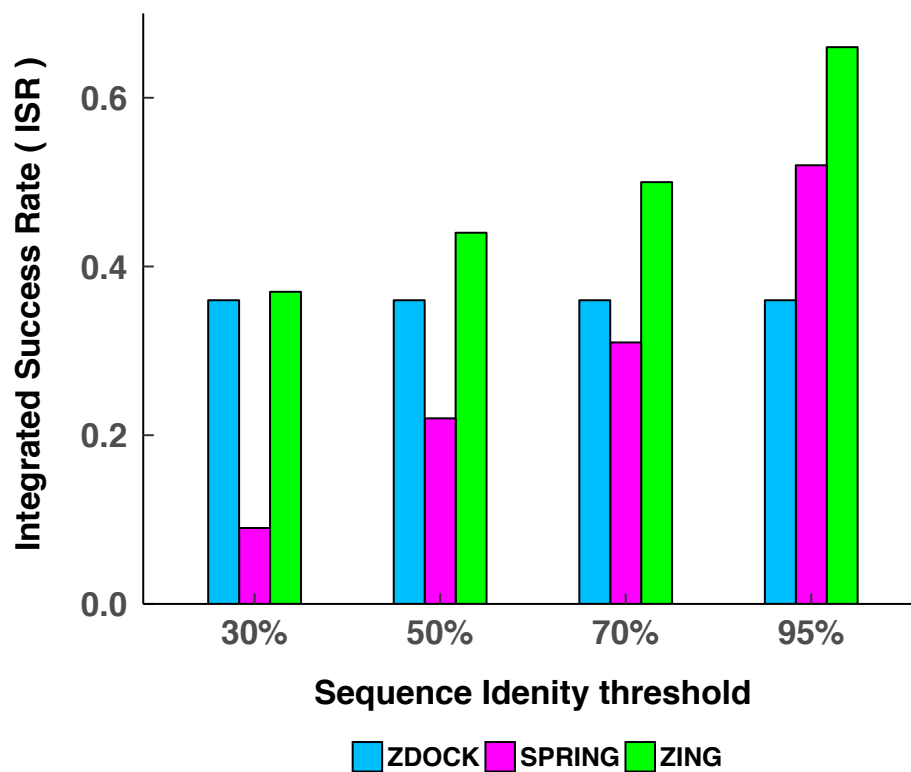
Supplementary Figure 2. Correlation between the five features considered for logistic regression modeling.



Supplementary Figure 3. Computing the probability that a ZDOCK prediction is a hit. The probability that a ZDOCK prediction is a hit is computed as a function of its ZDOCK rank. It is the fraction of the predictions that are hits among the predictions that have the same rank across the 142 single-chain test cases of version 5.0 of the docking benchmark. Because the ZDOCK frequency curve calculated directly from the benchmark (in green; smoothed by averaging with a window size of 8) was not monotonic, we fitted an exponential decay function to the frequencies (in red) and computed the probability accordingly. The fitted function is given by:

$$P_{ZD} = z_0 * e^{-z_1 * n} + z_2,$$

where $z_0 = 0.083$, $z_1 = 0.018$, $z_2 = 0.023$, and 'n' is the rank of the prediction.



Supplementary Figure 4. Integrated Success Rate (ISR) for ZDOCK, SPRING and ZING. The ISR for ZDOCK, SPRING and ZING is shown for different query-template sequence identity thresholds (30%, 50%, 70% and 95%) used to filter SPRING predictions. After filtering, the query-template sequence identity for each of the participating proteins in the template database is less than the threshold. At all thresholds, we used the same feature set and parameters as those trained and tested for the 95% threshold dataset. As expected, the ISR of SPRING decreases as the threshold decreases, nevertheless, the ISR of ZING is still higher than both ZDOCK and SPRING, albeit only slightly at 30% when the performance of SPRING drops well below that of ZDOCK.