Supplementary Information for "InterComp: sequence-order independent structural matching applied to protein interface alignment"

Claudio Mirabello[1] and Björn Wallner[1,*]

S1 Parameter optimization

InterComp has three parameters that have a significant impact on the search for the optimal mapping between two molecules. The first parameter is d_0, which is a parameter of the Levitt-Gerstein score (see Eq. 1), the second parameter is the weight of the structural component of the scoring function in relation to the weight of the chemical component, W_{str} (see Eq. 3), and the third parameter is the percentage of null correspondences, i.e. the number of rows/columns that can be ignored during the annealing procedure in the distance matrix of the smallest of the two molecules. It should be noted that the weight on the sequence is $1 - W_{str}$, which means that increasing the structural score will in effect decrease the relative contribution from the sequence similarity. The default values for $d_0 = 0.5$ and $W_{str} = 0.5$ were found by a grid search trying all combinations of $d_0 \in [0.25, 1.75]$ in 0.25 steps and $W_{str} \in [0.50, 0.75, 1.0]$ maximizing the number of target monomers whose interfaces were correctly (PPV>0.5) found within the top 10 identified templates (Fig. S1). In addition, the combination $d_0 = 0.5$ and $W_{str} = 0.25$ was also tested (the single low-performing point in the "hard" category in Fig. S1), but considering the bad performance for the "hard" category and the high computational cost for obtaining additional points (35,000 core hours per point) no more combinations involving $W_{str} = 0.25$ were tried. The default values $d_0 = 0.5$ and $W_{str} = 0.5$ were selected to allow for better results in most cases (mainly on the "easy" and "medium" sets). Still, a higher weight on the structural component of the score ($W_{str} \geq 0.75$) or a higher d_0 could yield slightly better results in the "hard" category. For the optimized d_0 and W_{str} allowing for 10% null correspondences, did not improve the performance (data not shown). Thus, the percentage of null correspondences was set to 0.

[1]Division of Bioinformatics, Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden

[*]To whom correspondence should be addressed.

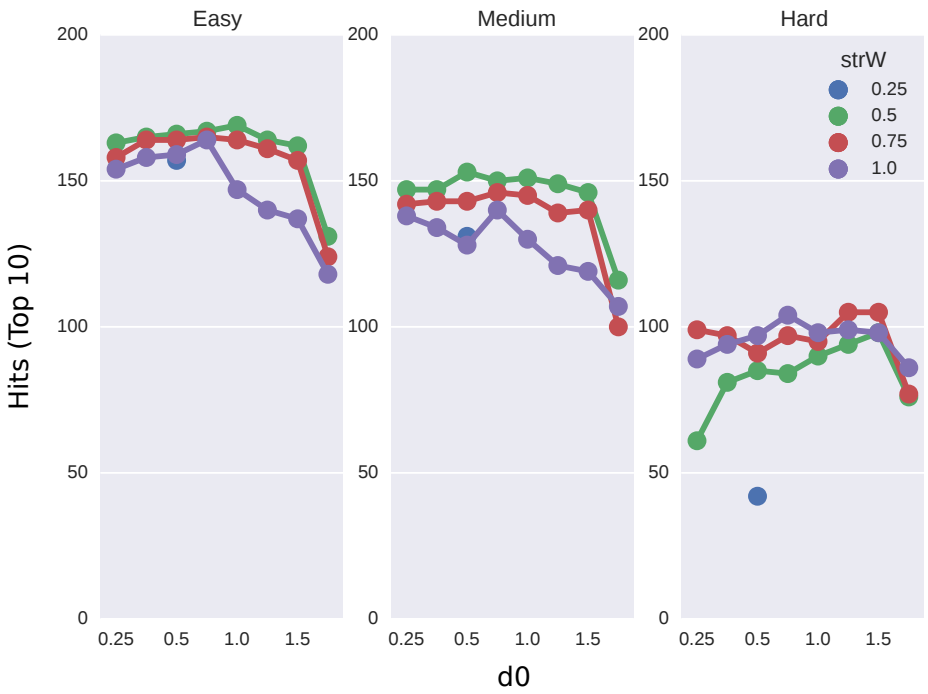

Figure S1: Number of targets for which the interface could be correctly detected at various parameter settings. This test was run on a subset of 200 targets randomly selected from set S568. The selected default parameters are $d_0 = 0.5$, and $W_{str} = 0.5$.