**Supplementary Information for:**

# ProPOSE: Direct exhaustive protein docking with side chain flexibility

Hervé Hogues, Francis Gaudreault, Christopher R. Corbeil, Christophe Deprez, Traian Sulea, and Enrico O. Purisima

*Human Health Therapeutics, National Research Council Canada, 6100 Royalmount Avenue, Montreal, Quebec, Canada, H4P 2R2*

**Antibody dataset selection**. The SAbDab Antibody database (http://opig.stats.ox.ac.uk/webapps/abdb/web_front/Welcome.php) (Sept. 28, 2014) was queried using the following criteria.

- Method=Xray
- Resolution<3.2A
- Complex=True
- Antigen=Protein

This resulted in a list of 517 Ab:Ag. From this list a hand-curated subset with the following selection criteria was extracted.

- Complete Fab region of the Ab (no breaks in the Ab)
- No small molecule at the interface
- Selection of Ab based on sequence/conformational diversity for a specific target. For example, if a single mutation results in the same loop conformation these cases are typically removed.
- Selection of Ab based on unique epitope for a specific target.

Only systems that went through the preparation steps and that executed to completion without error by all 6 docking programs in all 4 scenarios (BB, RB, BR and RR) are reported. The final Antibody dataset is comprised of 241 Ab:Ag complexes (Table S1).

**Table S1.** PDB codes of the selected Antibody complexes

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1ahw | 1bj1 | 1bql | 1bvk | 1dqj | 1e6j | 1egj | 1eo8 | 1fbi | 1fe8 |
| 1fj1 | 1fns | 1fsk | 1h0d | 1i9r | 1iqd | 1jhl | 1jps | 1jrh | 1k4c |
| 1kb5 | 1lk3 | 1mlc | 1nca | 1nfd | 1nsn | 1oaz | 1ors | 1osp | 1ots |
| 1p2c | 1pkq | 1r0a | 1rjl | 1sy6 | 1t03 | 1tpx | 1tqb | 1tzh | 1tzi |
| 1v7m | 1vfb | 1w72 | 1wej | 1xiw | 1yjd | 1ynt | 1yqv | 1yy9 | 1ztx |
| 2adf | 2aep | 2aeq | 2arj | 2b2x | 2bdn | 2cmr | 2dd8 | 2fd6 | 2fjg |
| 2fjh | 2h8p | 2h9g | 2hmi | 2i9l | 2j88 | 2jel | 2nr6 | 2ny1 | 2ny7 |
| 2oz4 | 2q8b | 2qqk | 2qqn | 2r0l | 2r56 | 2uzi | 2vxq | 2vxs | 2vxt |
| 2w9e | 2wuc | 2x7l | 2xqy | 2xra | 2xtj | 2xwt | 2yc1 | 2ypv | 2yss |
| 2zch | 3ab0 | 3b2u | 3b9k | 3bdy | 3bgf | 3bn9 | 3cvh | 3d85 | 3d9a |
| 3ehb | 3eoa | 3g04 | 3gbn | 3gi9 | 3gjf | 3grw | 3h42 | 3hfm | 3hi1 |
| 3hi6 | 3hmx | 3i50 | 3idx | 3iu3 | 3jwd | 3k2u | 3klh | 3kr3 | 3ks0 |
| 3l5x | 3l95 | 3ld8 | 3liz | 3lzf | 3ma9 | 3mj9 | 3mxw | 3nfp | 3nh7 |
| 3nps | 3o2d | 3p0y | 3pgf | 3pnw | 3q1s | 3q3g | 3qwo | 3r1g | 3raj |
| 3rkd | 3ru8 | 3rvv | 3s35 | 3s37 | 3se9 | 3skj | 3sm5 | 3so3 | 3sob |
| 3sqo | 3t3p | 3u30 | 3u9p | 3ubx | 3uc0 | 3v6o | 3vg9 | 3w2d | 3w9e |
| 3wd5 | 3zkm | 3ztn | 4aei | 4ag4 | 4al8 | 4am0 | 4bz1 | 4bz2 | 4cmh |
| 4cni | 4d9q | 4d9r | 4dkf | 4dn4 | 4dvr | 4ene | 4f2m | 4f3f | 4ffv |
| 4ffy | 4fp8 | 4fqi | 4fqj | 4g3y | 4g6j | 4g6m | 4g7v | 4gms | 4h88 |
| 4h8w | 4hc1 | 4hcr | 4hf5 | 4hfu | 4hj0 | 4hlz | 4ht1 | 4hwb | 4i2x |
| 4i3r | 4i77 | 4i9w | 4ij3 | 4irz | 4j6r | 4jan | 4jlr | 4jpk | 4jpw |
| 4k3j | 4k94 | 4k9e | 4kht | 4ki5 | 4kuc | 4kxz | 4l5f | 4leo | 4lf3 |
| 4liq | 4lmq | 4lst | 4lsu | 4lvh | 4lvn | 4m48 | 4m5z | 4mwf | 4np4 |
| 4o58 | 4ogy | 4oii | 4okv | 4olz | 4ot1 | 4py8 | 4u0r | 4u1g | 4u6h |
| 4u6v | | | | | | | | | |

**Benchmark dataset selection.** Quality criteria similar to that of the Antibody dataset was applied to the Protein Benchmark 5.0 complexed and unbound structures in addition to the requirement that all interface residues be present in the bound and unbound forms of the proteins. All antibody systems were removed from the set.

Only systems that went through the preparation steps and that executed to completion without error by all 6 docking programs in all 7 scenarios (BB, RB, BR, RR, UB, BU, UU) are reported. The final Benchmark dataset contained 150 complexes (Table S2).

**Table S2.** PDB codes of the selected Protein benchmark dataset 5.0 complexes.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1acb | 1ak4 | 1akj | 1atn | 1avx | 1ay7 | 1b6c | 1bkd | 1buh | 1bvn |
| 1cgi | 1clv | 1d6r | 1e6e | 1eaw | 1eer | 1efn | 1ewy | 1ezu | 1f34 |
| 1f51 | 1fc2 | 1fcc | 1ffw | 1fle | 1fq1 | 1fqj | 1gcq | 1ghq | 1gl1 |
| 1gla | 1gpw | 1grn | 1gxd | 1h1v | 1h9d | 1he1 | 1hia | 1i2m | 1i4d |
| 1ibr | 1ijk | 1j2j | 1jiw | 1jtd | 1jtg | 1jwh | 1jzd | 1k74 | 1kac |
| 1kkl | 1ktz | 1kxp | 1lfd | 1m10 | 1m27 | 1mah | 1ml0 | 1mq8 | 1oc0 |
| 1ofu | 1oph | 1oyv | 1ppe | 1pxv | 1qa9 | 1r0r | 1r6q | 1r8s | 1rke |
| 1rlb | 1rv6 | 1s1q | 1sbb | 1syx | 1t6b | 1tmq | 1udi | 1us7 | 1xqs |
| 1xu1 | 1y64 | 1yvb | 1z5y | 1zhh | 1zhi | 2a1a | 2a5t | 2a9k | 2abz |
| 2ajf | 2ayo | 2b42 | 2b4j | 2btf | 2c0l | 2cfh | 2g77 | 2gaf | 2gtp |
| 2h7v | 2hle | 2hqs | 2hrk | 2i9b | 2ido | 2j0t | 2j7p | 2nz8 | 2o3b |
| 2oob | 2ot3 | 2oul | 2pcc | 2sic | 2sni | 2uuy | 2vdb | 2x9a | 2z0e |
| 3a4s | 3aaa | 3aad | 3biw | 3bp8 | 3bx7 | 3cph | 3d5s | 3daw | 3f1p |
| 3fn1 | 3h11 | 3h2v | 3k75 | 3l89 | 3p57 | 3pc8 | 3r9a | 3s9d | 3sgq |
| 3szk | 3vlb | 4cpa | 4fza | 4h03 | 4hx3 | 4iz7 | 4lw4 | 4m76 | 7cei |

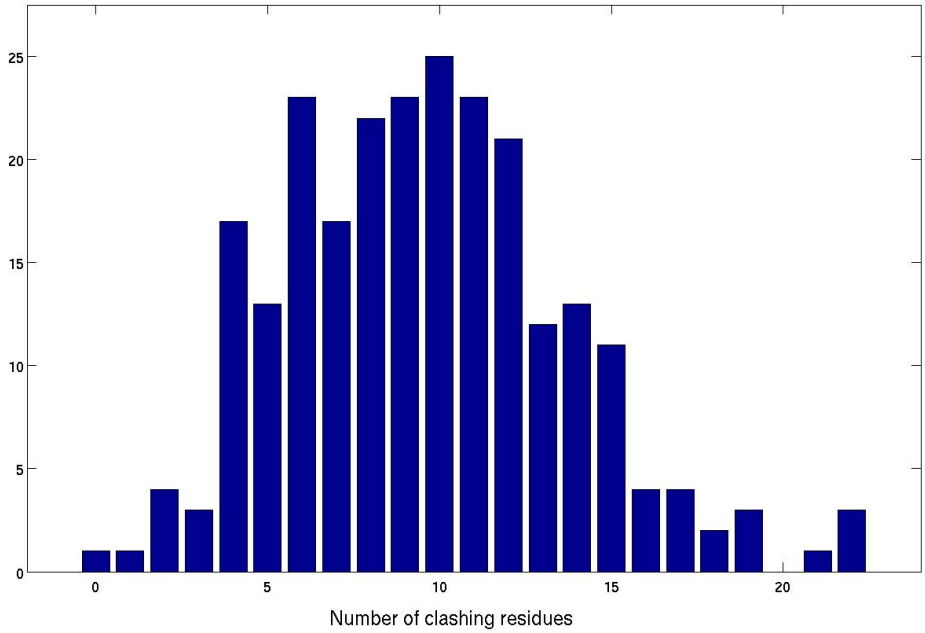


**Fig. S1**. Distribution of the number of interfering sidechains when both repacked proteins in the Antibody dataset are overlaid to the bound position. A sidechain is considered clashing when its AMBER van der Waals interaction energy > 5 kcal/mol.

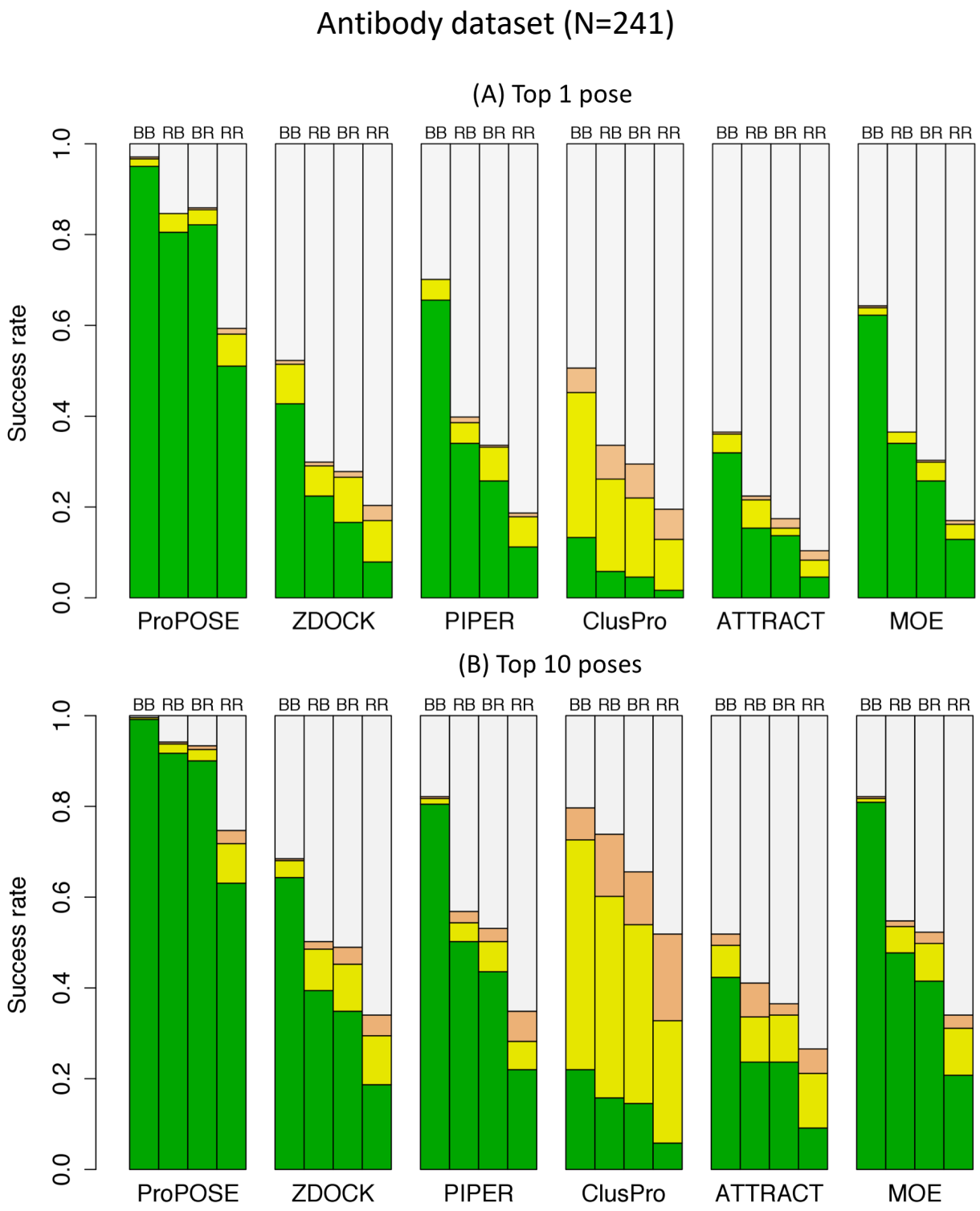


**Figure S2.** Absolute docking success rate using the Top 1 and Top 10 pose predictions under four scenarios: (BB) Bound ligand with bound receptor; (UB) Unbound ligand with bound receptor; (BU) Bound ligand with unbound receptor; (UB) Unbound ligand with bound receptor; (UU) Unbound ligand with unbound receptor; Colors indicate the quality of the predictions according to the CAPRI metric as High (green), Medium (yellow) or Acceptable (coral).

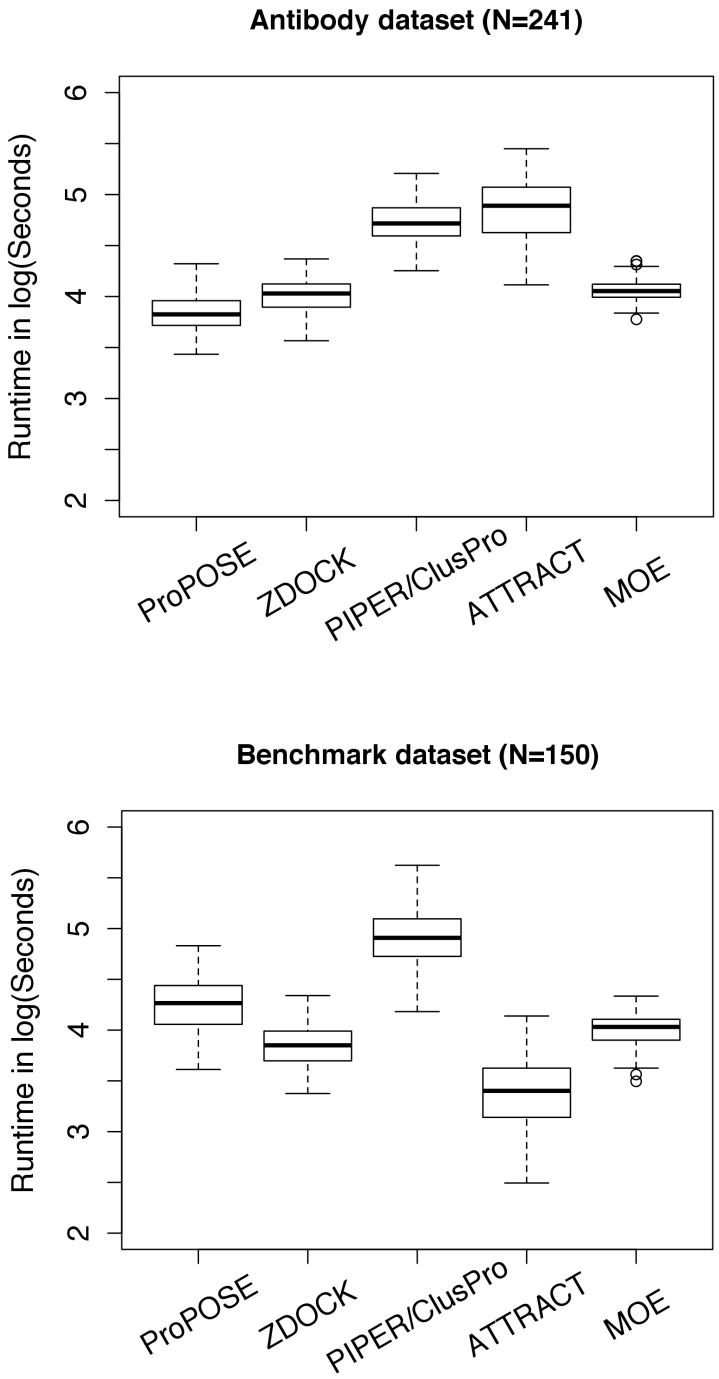


**Figure S3.** Runtime of the 6 programs tested on the Antibody and Benchmark dataset.

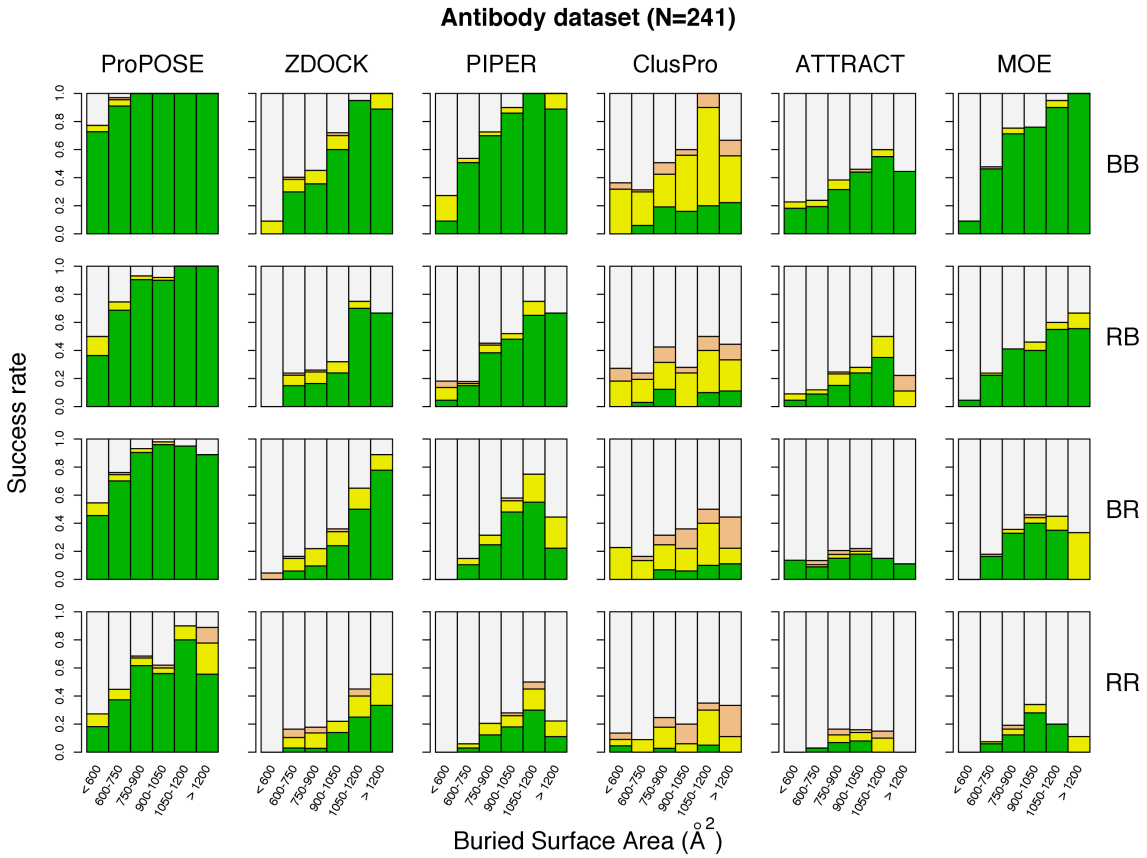


**Figure S4.** Absolute docking success rate for the top-pose prediction dependence on the buried surface area on the Antibody dataset under four sidechain perturbation scenarios: (RR) Repacked ligand with repacked receptor; (BR) Bound ligand with repacked receptor; (RB) Repacked ligand with bound receptor; (BB) Bound ligand with bound receptor. Colors indicate the quality of the predictions according to the CAPRI metric as High (green), Medium (yellow) or Acceptable (coral). From left to right the BSA bins contain 22, 67, 73, 50, 20 and 9 systems, respectively.

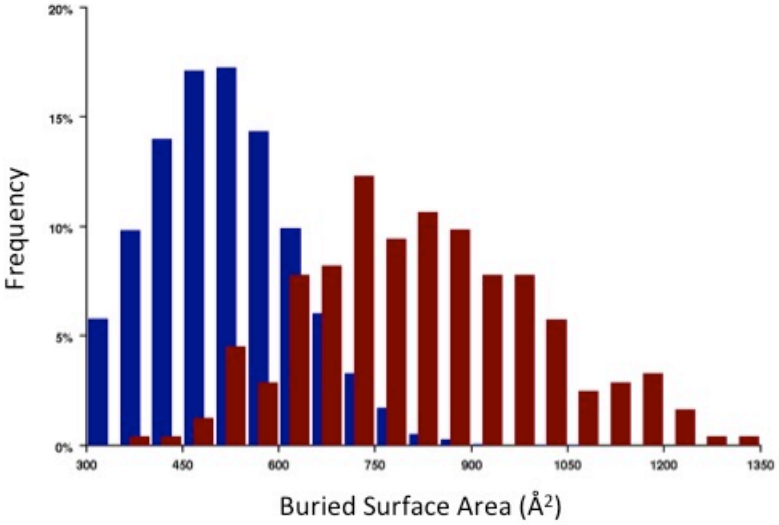


**Figure S5.** BSA distribution for the Antibody complexes (red) compared to the best decoys (blue). Decoys are defined as best 5000 poses of each Antibody that are more than 10 Å RMSD away from their native state

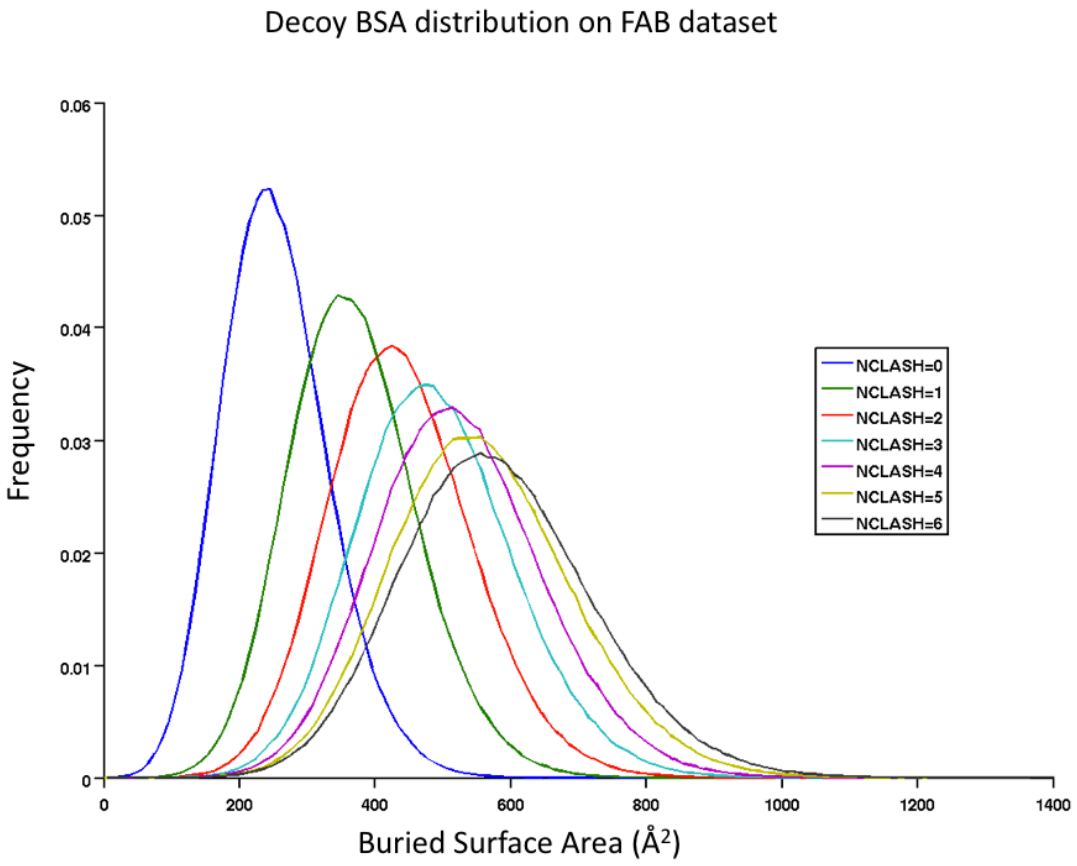


**Figure S6**. BSA distribution of best decoy poses with different number of tolerated sidechain conflicts on the Fab dataset. NCLASH=0 means that no sidechain clash was tolerated. NCLASH=N means that at most N clashing sidechains are tolerated on the ligand and N are tolerated on the receptor.

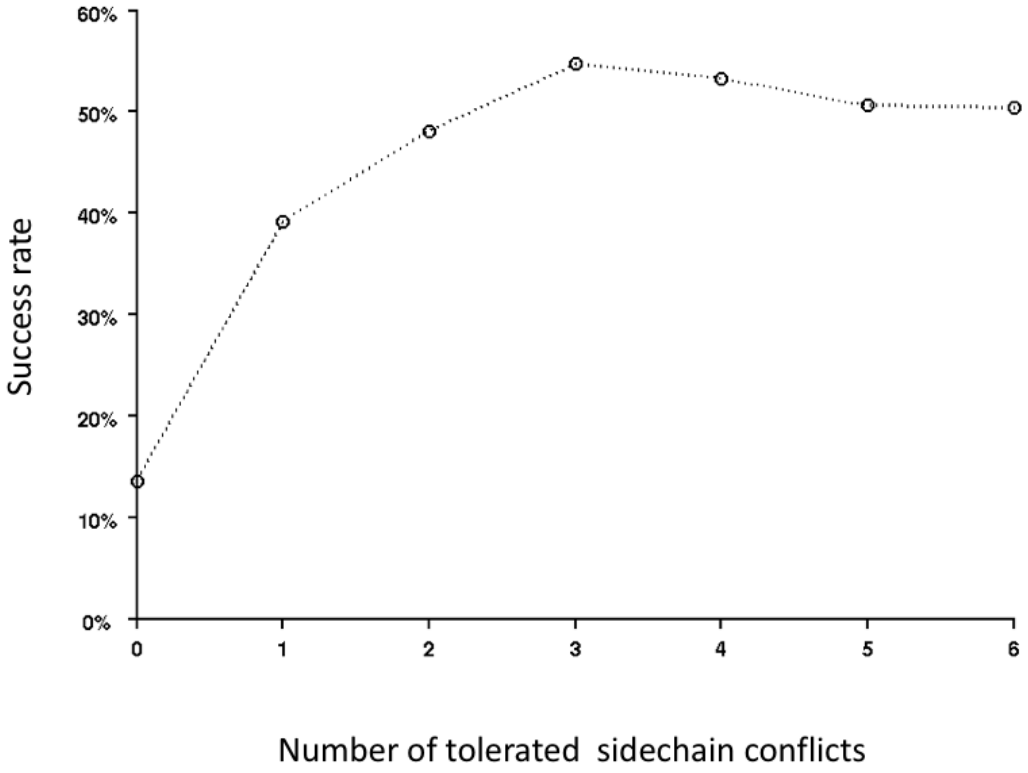


**Figure S7.** Success rate on the repacked-repacked (RR) Antibody dataset for various number of sidechain conflicts tolerated on each protein in the initial docking stage (Stage1).