

## Technical Note on Blind Docking

### eHiTS 2009 as a Blind Docking Tool

by Orr Ravitz, SimBioSys, Inc.

As the molecular docking paradigm solidifies its status as a significant tool for drug discovery, chemists explore additional applications of the methods in ways that sometime stretch the existing algorithms to their limits. Most docking programs, including eHiTS, have not been designed or optimized to perform blind docking. In structure based drug discovery, the user is typically expected to define, at some level of accuracy, the binding pocket in the target of interest. The binding site is determined either based on known binding modes of ligands as found in crystal structures of complexes, or based on an educated hypothesis. There are cases, however, in which assumptions about the possible locations of binding hot spots are difficult or should be avoided altogether. This is the case, for example, when the existence of secondary binding sites is suspected, or when one would like to screen active ligands and other compounds on a range of targets to estimate the possibility for drug side-effects, toxicity, and other types of biological activities.

The standard eHiTS usage requires a rough definition of the binding pocket. This is done through the clip file. This file should contain at least two sets of coordinates (or two spatial points) that are located in the designated binding pocket. eHiTS then draws a box around those points, expands it to some extent in all directions and places the search grid inside that box. Then, the box is “flooded” with a virtual fluid to detect all the cavities which will define the binding surface. This is a highly automated process, but it still relies on that user-defined clipping. Commonly the native ligand, amino acids from the binding pocket, or a few atoms from either are chosen as a clip file. If eHiTS is run with the -complex option, the native ligand is inferred as the clipping coordinates. However, eHiTS could be used without any clipping. In this case, the entire receptor will be considered for docking. The whole protein will be flooded, and sufficiently deep clefts will be searched on its surface. The final space in which docking will be performed is defined by the interconnected pockets found on the target. The search grid in such scenarios is typically large, and extensive sampling is required. Nevertheless, the computational efficiency of the eHiTS algorithm allows good sampling in reasonable timescales.

Several eHiTS users expressed specific interest in blind docking in recent months, and therefore we decided to evaluate eHiTS' performance in this context. We used the set that was used in an earlier blind docking evaluation ([Hetenyi and van der Spoel, 2006 \[1\]](#)). We focused on the 43 complexes used in the paper and have not attempted to use the apo structures. 3 codes (1B70, 1FIW and 1QIZ) were left out because of uncertainty regarding the exact structure used in the paper for docking. The default accuracy (3) was used throughout the study. The average blind docking time was 9 minutes per receptor for this set.

**Results:**

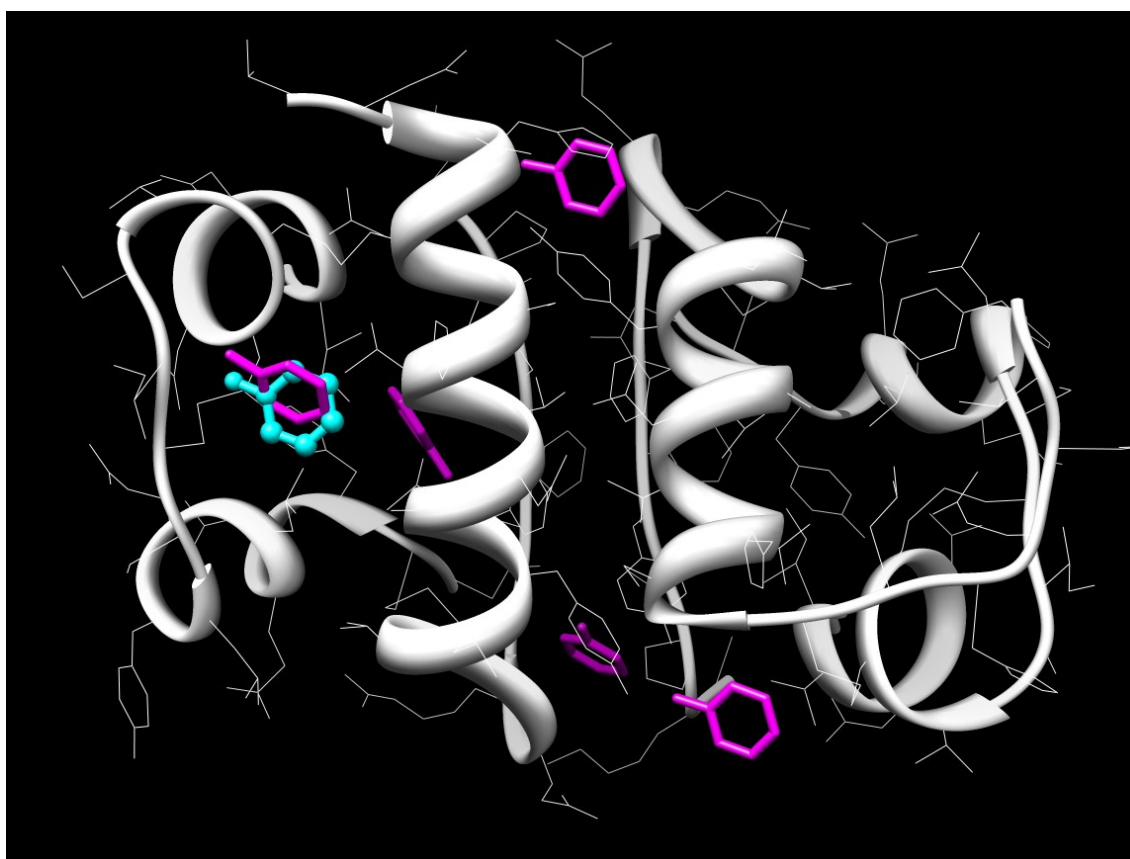
77.5% of the cases gave at least one conformation under 2 Å in the top 10 poses. In the other cases, one accumulative docking round using poses from the first round as clip files produced successful binding modes in the top 5 poses. The top rank pose is in most cases in the correct binding pocket, offering a good starting point for pose refinement.

The table below details the results for the specific codes. The Job# column describes whether the results were obtained with a single blind docking run, or with 2 cycles. The Rank# and RMSD columns indicate the rank of the first pose under 2 Å and its RMSD from the crystallographic conformation. The last two columns indicate the top-rank and closest poses RMSDs.

PDB Code	Job#	Rank#	RMSD	RMSDtop	RMSDclose
1A0Q	1	1	1.52	1.52	1.52
1A53	1	4	1.91	2.30	1.21
1A8U	1	1	1.38	1.38	0.68
1ALW	1	3	1.75	2.72	1.33
1AZ8	2	2	1.43	7.09	0.86
1BZJ	1	8	1.00	7.11	1.00
1C83	1	1	0.79	0.79	0.38
1C84	1	1	0.82	0.82	0.64
1C85	1	1	0.85	0.85	0.79
1CA7	1	1	1.73	1.73	1.25
1D1Q	1	1	1.36	1.36	0.59
1DY4	1	9	1.24	5.96	1.00
1E7A	1	8	1.91	48.67	1.91
1ECV	1	1	1.19	1.19	0.77
1EQG	2	1	0.70	0.70	0.70
1EV3	1	1	0.43	0.43	0.33
1F5K	2	1	1.12	1.12	0.40
1GAF	2	1	1.53	1.53	1.01
1GUH	2	1	1.95	1.95	1.35
1HD2	1	1	0.64	0.64	0.38
1HDU	1	3	0.90	2.42	0.86
1HZ4	1	2	0.60	3.46	0.39
1IVB	2	3	1.56	3.26	1.02
1JU4	1	1	0.47	0.47	0.44
1KEL	2	1	1.34	1.34	1.03
1MPJ	2	1	0.57	0.57	0.57
1NGP	1	8	1.27	28.59	1.11
1PTH	2	4	1.30	2.54	0.48

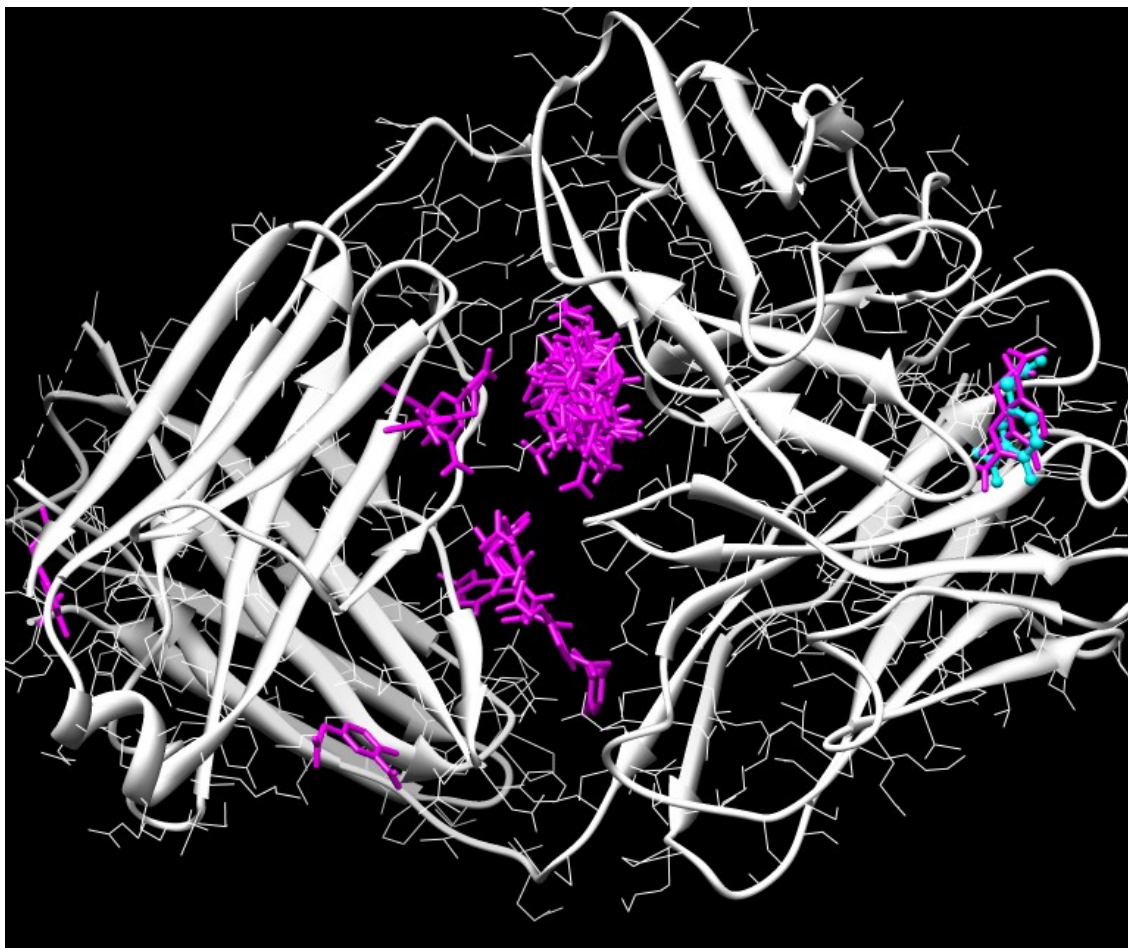
1RFN	1	3	0.58	3.99	0.58
1SRI	2	5	1.05	2.50	1.00
1TNJ	1	1	1.03	1.03	0.97
1TYM	1	1	1.71	1.71	1.55
2AY5	1	2	1.23	4.32	1.23
3CPA	1	1	0.85	0.85	0.85
3ERT	1	1	1.30	1.30	0.68
3PAX	1	3	1.17	3.81	0.90
3PCN	1	4	1.77	2.02	0.96
43CA	1	5	1.16	3.98	0.57
4DFR	1	2	1.25	2.25	1.22
4TS1	1	1	1.74	1.74	1.14

The blind docking of phenol into insulin (1MPJ) is shown in the picture below. The crystallographic pose is shown in cyan, and sample poses are shown in “hot spots” detected during docking. Those poses can be used to clip the receptor in accumulative docking runs in which the sampling is finer, and the binding pockets are better modelled. It should be noted that this code generates an unusually big number (5) of hot spots. In most cases in the set we observed three, two and often one hot spot, manifesting the detection of the correct binding pocket.



*Phenol binding to Insulin. Several potential binding pockets are detected for this small ligand.*

1NGP (N1G9 FAB fragment) is a case where the majority of poses are generated far from the native ligand. The illustration below shows that most of the poses are located in the big cavity between chains L and H of the crystal structure. Several poses, however, reproduce the x-ray binding mode (in cyan) with close to 1 Å RMSD.



*2-(4-hydroxy-3-nitrophenyl)acetic acid docked into N1G9 FAB fragment. The majority of poses are located in the big cavity between chains L and H.*

### **Conclusions:**

The above results clearly demonstrate the viability of eHiTS as a blind docking tool. In all cases the correct binding pocket has been identified in the top 32 solutions, and in most cases good poses under 2 Å and even 1 Å were found at the top of the generated poses. The conformations may be further refined by clipping the receptor for subsequent runs, and by working at higher accuracies. As always in eHiTS, the jobs are extremely easy to setup with a simple command line, and with no required preparation for the receptor or the ligand. This, and the speed of the calculations make eHiTS a high throughput blind docking solution.

### **Reference:**

1. Hetenyi, C. Van der Spoel, D.: "Blind docking of drug-sized compounds to proteins with up to a thousand residues." 2006 Feb 20;580(5):1447-50. Epub 2006 Jan 31.