



[Artificial Receptors](#)

[Flexible Sidechains](#)

[GOLD SAMPL](#)

[Optimum Docking & Rescoring Protocols](#)

[GOLD Validation Tests](#)

[GOLD References](#)

## Using GOLD for Structure-based Virtual Screening

GOLD is potentially a highly useful tool for Structure-based Virtual screening. However, in common with other docking programs, it is always important to ensure that the docking protocol and the set up of the protein model are optimum for the problem at hand. This may mean ensuring an appropriate active site definition is specified in the gold.conf file, an appropriate scoring function is chosen; and a protocol that allows sufficient time per ligand is used.

Given below are results of enrichment experiments with GOLD in which a number of different protocols have been applied. Hopefully these will be of some assistance to those wishing to use GOLD in the most optimum way for virtual screening. The protein models and the ligand structures of both decoy and active molecules used in this study come, with grateful thanks, courtesy of Dr H. Chen, AstraZeneca, and are reported in: Chen, H.; Lyne, P. D., Giordanetto, F., Lovell, T., Li, J., *J. Chem. Inf. Model.*, **46**, 401-405, 2006.

The enrichment experiments shown here have been carried out on four targets, Thrombin, Oestrogen Receptor (ER), Cox2 and secretory Phospholipase A2 (sPLA2). These proteins have binding sites which differ significantly in shape, size and polarity. Initial studies were carried out using in their entirety the ligand libraries supplied to us. The decoy set consists of 20,000 ligands whose size and make up is within the range expected for drug-like molecules. This set is seeded with 125 thrombin actives, 125 Cox2 actives, 53 ER actives and 17 sPLA2 actives. Where multiple ring conformers, tautomers or alternative protonation states occur, then appropriate additional structures are also present for both decoys and actives. The virtual library comprised 33294 structures in total. Later experiments used a set of decoys reduced in size by about 90%, leading to a library of 4141 structures in total.

### GOLD Protocols

All protocols used the binding site definitions defined by H. Chen *et al* in their work with GOLD v2.2. The version of GOLD used in this work is 3.0.1. No significant difference in performance was noticed between GOLD v.2.2 and v.3.0.1 in performance on these test sets (results not reported here). Enrichment rates are calculated at top cuts of 10% and 1% as determined by ranking to the appropriate fitness score. The enrichment rate is calculated by dividing the number of actives found in the top cut by the expected number to be found on random pick.

**Protocol GS\_1** - This protocol is identical to that used by H. Chen *et al*. This protocol uses "Library GA Parameter Settings" with a fixed 1500 operations per ligand (popsiz 50, sel. pressure 1.125, niche size 2). These settings are similar to the setting designated as 'Library Settings' in GOLD 2.2. However we currently would recommend setting the number of operations to at least 10,000 if using a fixed number of operations per ligand. Results are given in Table 1.

**Protocol CS\_1** - This protocol is identical to GS\_1 except that ChemScore is used as the scoring function instead of GoldScore. Results are given in Table 1.

**Protocol GS\_2** - This protocol uses "Automatic GA Parameter Settings" (first available in GOLD 3.0). This tailors the number of operations used by the size and number of rotatable bonds the ligand has. So large and flexible molecules are given more operations than small and rigid ones. 10% search efficiency has been applied, with a minimum of 1000 and a maximum of 12,500 operations per ligand. In terms of speed this is a fast protocol suitable for use in a Virtual Screening environment (~2x slower than GS\_1, though this factor depends on the protein target). Results are given in Table 1.

**Protocol CS\_2** - The ChemScore equivalent of GS\_2. Results are given in Table 1.

**Protocol GS\_3** - Also using automatic settings. However Search efficiency is set at 20% (min. 2000,max 25000 opps). Results are given in Table 1.

**Protocol CS\_3** - The ChemScore equivalent of GS\_3. Results are given in Table 1.

**Protocol GS\_2'** - Same protocol as GS\_2, except this run is carried out using a 90 % reduced set of active molecules. The effect of this on the enrichment rate is to reduce it slightly from that observed for the larger set of decoys. The run is repeated three times and the results averaged. Results are presented in Table 2.

**Protocol CS\_2'** - The ChemScore equivalent of GS\_2'. Results are given in Table 2

**Protocol GS\_4** - Same protocol as GS\_2', except "Internal Energy Offset" is switched on. GOLD estimates a lowest internal strain energy for each ligand during the course of the docking run and subtracts this from the final score. This compensates large and highly substituted ligands for strain energy that cannot be eliminated on conformational search. Results are given in Table 2.

**Protocol CS\_4** - The ChemScore equivalent of GS\_4. Results are given in Table 2.

	<b>Thrombin</b>	<b>Cox2</b>	<b>ER</b>	<b>sPLA2</b>
<b>H.Chen et al</b> Enrich. (10%(1%))	<b>1.44</b>	<b>3.55</b>	<b>0.75</b>	<b>0</b>
<b>GS_1</b> Enrich. (10%(1%))	<b>2.8 (5.4)</b>	<b>4.1(14)</b>	<b>2.2(7.5)</b>	<b>3.5</b>
<b>CS_1</b> Enrich. (10%(1%))	<b>1.31 (2.7)</b>	<b>2.6(0)</b>	<b>8.8(57)</b>	<b>0.59</b>
<b>GS_2</b> Enrich. (10%(1%))	<b>6.5 (15)</b>	<b>4.7(8.8)</b>	<b>3.3(17)</b>	<b>5.9(41)</b>
<b>CS_2</b> Enrich. (10%(1%))	<b>3.8 (11)</b>	<b>3.1(1.6)</b>	<b>9.1(60)</b>	<b>4.7(23)</b>
<b>GS_3</b> Enrich. (10%(1%))	<b>5.8(12)</b>	<b>4.6(6.4)</b>	<b>3.4(21)</b>	<b>7.1(47)</b>
<b>CS_3</b> Enrich. (10%(1%))	<b>4.7 (18)</b>	<b>3.1(1.6)</b>	<b>8.9(64)</b>	<b>4.1(24)</b>

*Table 1.* GOLD enrichment rates for virtual screens carried out over the full virtual library. Notes: a) Best enrichment rates are highlighted in red. b) There is a significant difference between the results obtained by H. Chen *et al* and ourselves using ostensibly the same protocol (GS\_1). The reason for this is unclear. c) A protocol using variable settings at 5% search efficiency showed significantly poorer retrieval rates than that using 10% (results not shown) d) The Oestrogen receptor structure used is an antagonist form. The ER actives are a mixture of small agonists and larger antagonists. GoldScore appears to perform poorly for this protein. However if only antagonist structures are classed as true actives, the

enrichment rate for GoldScore under protocol GS\_2, at a 10% top cut, is 7.9.

	<b>Thrombin</b>	<b>Cox2</b>	<b>ER</b>	<b>sPLA2</b>
<b>GS_2'</b> Enrich. (10%(1%))	<b>5.4 (11)</b>	<b>4.3 (4.4)</b>	<b>2.9(16)</b>	<b>6.2 (38)</b>
<b>CS_2'</b> Enrich. (10%(1%))	<b>3.0 (7.2)</b>	<b>2.8 (0.8)</b>	<b>8.7(39)</b>	<b>4.4 (23)</b>
<b>GS_4</b> Enrich. (10%(1%))	<b>8.6 (17.9)</b>	<b>3.9 (3.0)</b>	<b>4.7 (27)</b>	<b>6.5 (35)</b>
<b>CS_4</b> Enrich. (10%(1%))	<b>2.7 (6.1)</b>	<b>3.9(4.8)</b>	<b>8.7(41)</b>	<b>4.9(16)</b>

Table 2.. GOLD enrichment rates for virtual screens carried out using only 10% of the decoy structures. Notes: a) Significantly improved enrichment rates over that achieved using GS\_2' or CS\_2' are highlighted in red. b) Maximum possible enrichment rate at the 1% cut is substantially lower for thrombin, Cox2 and ER than that for the experiments reported in table 1 due to the smaller number of decoys (18.4, 18.4 and 43.4 respectively).

### Summary of Results

1. Fast GA settings with a fixed number of operations per ligand do not give optimum enrichments. A good compromise between speed and accuracy suitable for Virtual Screening applications is achieved using "Automatic GA Parameter Settings" at 10% search efficiency. Some targets (e.g. sPLA2) can benefit from increased search efficiency so some care is needed to ensure a long enough protocol is used.
2. Enrichment rate depends on the scoring function used. Good enrichments were found using GoldScore for three targets whereas ChemScore was more effective against the fourth target. We would generally recommend GoldScore for binding sites which allow for significant numbers of hydrogen bonding ligand-protein interactions. ChemScore is recommended for binding sites with substantial hydrophobic character.
3. It is recommended that "Internal Energy Offset" be switched on, when using GoldScore. This leads to clear cut improvements in enrichment rates for some targets.