Tables

Table 1. Atom Formal Charge of GEMDOCK

Formal charge	Atom name
Receptor:	
0.5	N atom in His (ND1 & NE2) and Arg (NH1 & NH2)
-0.5	O atom in Asp (OD1 & OD2) and Glu (OE1 & OE2)
1.0	N atom in Lys (NZ)
2.0	Metal ions (MG, MN, CA, ZN, FE, and CU)
0	Other atoms
Ligand:	
0.5	N atom in $-C(NH_2)_2^+$
-0.5	O atom in $-COO^-$, $-PO_2^-$, $-PO_3^-$, $-SO_3^-$, and $-SO_4^-$
1.0	N atom in $-NH_3^+$ and $-N^+(CH_3)_3$
0	Other atoms

Table 2. Atom Types of GEMDCOK

Atom type	Heavy atom name
Donor	Primary and secondary amines, sulfur, and metal atoms
Acceptor	Oxygen and nitrogen with no bound hydrogen
Both	Structural water and hydroxyl groups
Nonpolar	Other atoms (suchas carbon and phosphorus)

Table 3. The 305 Test Complexes

1a07[a]_PPP(D)[b]	***1b59_OVA***	1ctt_DHZ	***1eoc_4NC***	***1hos_PHP***	***1lst_LYS***	***1ppc_PPP(I)***	***1tmn_PPP(I)***	***2cpp_CAM***	3nos_HAR
1a0q_HEP [c]	1b6n_PI3	***1cvu_ACD***	***1epb_REA***	***1hpv_478***	**1lyb_PPP(J)**	***1pph_CCC(I)***	***1tng_AMC***	***2ctc_LOF***	3pgh_FLP
1a1b_PPP(D)	***1b9v_RA2***	**1cx2_S58**	***1epo_PPP(I)***	**1hri_S57**	**1lyl_LYS**	***1ppi_GLC***	***1tnh_FBA***	**2dbl_S5H**	3ptb_BEN
1a1e_PPP(D)	**1baf_NPP**	***1d0l_BLG***	***1eta_T44***	***1hsb_PPI)***	***1mbi_IMD***	1ppl_PPP(I)	***1tni_PBN***	2er7_BOC	***3tpi_PPP***
1a28_STR	***1bbp_BLV***	***1d3h_A26***	***1etr_MQI***	***1hsl_HIS***	**1mcq_PPP(P)**	***1pso_PPP(I)***	***1tnl_TPA***	***2fox_FMN***	***4aah_PQQ***
1a42_BZO	***1bgo_I10***	***1d4p_BPP***	***1ets_NAS***	***1htf_G26****	1mcr_PPP(P)	***1ptv_PPP(B)***	1tph_PGH	***2qbp_GLC***	**4cox_IMN**
1a4g_ZMR	**1bl7_SB4**	**1dbb_STR**	**1ett_CCC(T)**[d]	1hti_PGA	***1mdr_SAA***	***1qbr_638***	***1tpp_APA***	***2h4n_AZM***	**4cts_OAA**
1a4k_FRA	***1blh_FOS***	**1dbj_AE2**	1etz_GAS	***1hvr_XK2***	1ml1_PGA	1qbt_146	***1trk_TPP***	***2ifb_PLM***	***4dfr_MTX***
1a4q_DPC	***1bma_PPP(B)***	1dbm_SIH	***1f0r_815***	***1hyt_BZS***	***1mld_CIT***	***1qbu_846***	***1tyl_TYL***	**2lgs_GLU**	4er2_PPP(I)
1a6w_NIP	**1bmq_MNO**	***1dd7_1PM***	***1f0s_PR2***	**1ibg_OBN**	1mmb_BAT	***1qcf_PP1***	***1ukz_AMP***	**2mcp__PC**	***4est_PPP(I)***
1a9u_SB2	***1byb_GLC***	***1dg5_TOP***	***1f3d_TPM***	1icn_OLA	***1mmq_RRS***	1qh7_XYP	**1ulb_GUN**	2mip_PPP(E)*	4fab_FDS
1aaq_PSI	***1byg_STU***	***1dhf_FOL***	**1fax_DX9**	***1ida_PPP(I)***	1mnc_PLH	1ql7_ZEN	**1uvs_I11**	***2pcp_1PC***	**4fbp_AMP**
1abe_ARA*	**1c12_TRZ**	**1did_DIG**	1fbl_HTA	1igj_DGX	***1mrg_ADN***	***1qpe_PP2***	**1uvt_I48**	**2phh_PHB**	***4lbd_961***
1abf_FCA*	***1c1e_ENH***	1die_DNJ	***1fen_AZE***	***1imb_LIP***	***1mrk_FMC***	***1qpq_NTM***	***1vgc_V36***	***2pk4_ACA***	***4phv_VAC***
1acj_THA	1c2t_NHS	***1dmp_450***	**1fgi_SU1**	***1ivb_ST1***	***1mts_BX3***	1rbp_RTL	1vrh_SD8	2plv_SPH	4tpi_PPP(S)
1acl_DME	***1c5c_TK4***	***1dog_NOJ***	1fig_TSA	1ivc_ST2	1mtw_DX9	***1rds_GPC***	***1wap_LTR***	***2qwk_G39***	***5abp_GLA****
1acm_PAL	***1c5x_ESI***	***1dr1_BIO***	***1fkg_SB3***	1ivd_ST1	***1mup_TZL***	***1rne_C60***	***1xid_ASC***	2r04_W71	***5cpp_ADN***
1aco_TRA	***1c83_OAI***	**1dwb_BEN**	***1fki_SB1***	1ive_ST3	***1nco_CHR***	***1rnt_2GP***	***1xie_ASO***	**2r07_W33**	***5er1_PPP(I)***
1aec_E64	***1cbs_REA***	**1dwc_MIT**	***1fl3_SPB***	**1ivq_PPP(I)**	***1ngp_NPA***	***1rob_C2P***	1xkb_4PP	2sim_DAN	5p2p_DHG
1aha_ANE	***1cbx_BZS***	**1dwd_MID**	***1flr_FLU***	1jao_PPP(I)	***1nis_NTC***	**1rt2_TNK**	***1ydr_IQP***	***2tmn_PPP(I)***	6abp_ARB
1ai5_MNP	***1cdg_MAL***	***1dy9_PPP(F)***	***1frp_FDP***	***1jap_PPP(I)***	1nsd_DAN	1sln_INH	1yds_IQS	***2tsc_CB3***	6cpa_ZAF
1aj7_NPE	1cf8_HAZ	**1eap_HEP**	1ghb_PPP(I)	***1kel_AAH***	***1okl_MNS***	***1slt_NAG***	***1ydt_IQB***	***2yhx_OTG***	***6rnt_2AM***
1ake_AP5	***1cil_ETS***	***1ebg_PAH***	***1glp_GTS***	1kno_PNP	***1okm_SAB***	***1snc_PTP***	***1yee_PNB***	**2ypi_PGA**	***6rsa_UVC***
1aoe_GW3	1cin_MTS	***1eed_PPP(I)***	***1glq_GTB***	***1lah_ORN***	***1pbd_PAB***	1srf_MTB	***25c8_GEP***	***3cla_CLM***	7cpa_FVF
1apt_PPP(I)	***1ckp_PVB***	***1ei1_ANP***	1gpy_G6P	***1lcp_PLU***	***1pdz_PGA***	1srg_MHB	***2aad_2GP***	***3cpa_PPP(S)***	***7tim_PGH***
1apu_PPP(I)	***1cle_CLL***	***1ejn_AGB***	**1hak_K21**	***1ldm_OXM***	1pgp_6PG	1srh_MOB	***2ack_EDR***	***3erd_DES***	8gch_PPP(C)
1aqw_PPP(H)	***1com_PRE***	1ela_PPP(B)	***1hdc_CBO***	***1lic_HDS***	1pha_PFZ	***1srj_NAB***	***2ada_HPR***	***3ert_OHT***	
1ase_NOP	***1coy_AND***	1elb_PPP(B)	1hdy_PYZ	1lkk_PPP(B)	***1phd_PIM***	1stp_BTN	***2ak3_AMP***	3gch_CIN	
1atl_SLE	***1cps_CPM***	1elc_PPP(B)	1hef_PPP(I)	1lmo_NAG	1phf_PIM	**1tdb_UFP**	2cgr_GAS	***3gpb_G1P***	
1azm_AZM	**1cqp_803**	1eld_PPP(I)	***1hfc_HAP***	***1lna_PPP(P)***	***1phg_MYT***	1tka_N3T	***2cht_BAR***	**3hvt_NEV**	
1b58_PPP(B)	1ctr_TFP	1ele_PPP(I)	***1hiv_PPP(I)***	***1lpm_MPA***	***1poc_GEL***	***1tlp_PPP(I)***	***2cmd_CIT***	3mth_MPB	

The whole set is all entries (305 complexes) and bold entries denote clean set (224 complexes). Italic entries denote the clean set with resolution below 2.5 Å (180 complexes) and Underlines entries denote clean set with resolution below 2.0 Å (92 complexes).

* The ligand with multi-crystal geometries in protein-ligand complex.

[a] A 4-character PDB code used in the Protein Data Bank.

[b] A 3-character ligand code used in the Protein Data Bank.

[c] "PPP" denotes a peptide ligand and the uppercase character in ( ) denotes the chain code of the peptide ligand.

[d] "CCC" denotes a composition ligand and the uppercase character in ( ) denotes the chain code of this composition.

Table 4. Parameters of GEMDOCK for Molecular Docking

Parameter	Value of parameters
Initial step sizes	s = 0.8, v = ? = 0.2 (in radius)
Family competition length	L = 2
Population size	N = 300
Recombination rate	p_c = 0.3
No. of the maximum generation	70

Table 5. GEMDOCK Validation Results for Different Lists [a]

	RMSD <0.5Å	RMSD <1.0Å	RMSD <1.5Å	RMSD <2.0Å	RMSD <2.5Å	RMSD <3.0Å
All entries [b]	14 %	51 %	71 %	78 %	84 %	86 %
Clean list [c]	16 %	55 %	73 %	82 %	86 %	89 %
Clean list, R<2.5Å [d]	17 %	55 %	73 %	83 %	86 %	89 %
Clean list, R<2.0Å [e]	22 %	57 %	75 %	83 %	87 %	89 %

[a] The success rate of GEMDOCK is based on solutions having the first rank.

[b] All entries include full test set (305 complexes).

[c, d, e] Clean lists defined in Table 3 are divided into three sets by different resolution thresholds: all (224 complexes), < 2.5 Å (180 complexes) and < 2.0 Å (92 complexes).

Table 6. Comparing GEMDOCK with Gold on Different Lists

RMSD	GEMDOCK[a]				GOLD[b]			
	All entries	Clean list	Clean list R<2.5Å	Clean list R<2.0 Å	All entries	Clean list	Clean list R<2.5 Å	Clean list R<2.0 Å
<0.5 Å	14 %	16 %	17 %	22 %	14 %	17 %	19 %	19 %
<1.0 Å	51 %	55 %	55 %	57 %	44 %	50 %	51 %	56 %
<1.5 Å	71 %	73 %	73 %	75 %	59 %	65 %	66 %	72 %
<2.0 Å	78 %	82 %	83 %	83 %	68 %	72 %	73 %	78 %
<2.5 Å	84 %	86 %	86 %	87 %	75 %	78 %	80 %	85 %
<3.0 Å	86 %	89 %	89 %	89 %	80 %	82 %	83 %	88 %

[a]The success rate of GEMDOCK was based on the docked solution with the first rank.

[b]The success rate of GOLD, a steady-state genetic algorithm, was based on the docked solution with the first rank.

Table 7. Parameters of GEMDOCK for Virtual Screening

Parameter	Value of parameters
Initial step sizes	s = 0.8, v = ? = 0.2 (in radius)
Family competition length	L = 2
Population size	N = 300
Recombination rate	p_c = 0.3
No. of the maximum generation	60

Table 8. Ligand preferences evolved from known active ligands are used to screen lead compounds for the human α-thrombin inhibitor complexes

Ligand name	Electrostatic preferences (Equation 9)			Hydrophilic preferences (Equation 10)			Molecular weight (Equation 11)	
	q_{elec}	s_{elec}	UB_{elec}	q_{hb}	s_{hb}	Ur_{hb}	m_{mw}	K
Thrombin inhibitor	4.00	0.97	4.97	0.32	0.04	0.36	33.2	0.17

Table 9. Comparing GEMDOCK with GOLD with Respect to Docking Ten Ligands back into Reference Protein (1dwd)

Ligand Id[a]	GEMDCOK		GOLD
	E_{tot} [c]	E_{bind} [c]	
7kme_BNN	1.05	1.56	7.80
2hpp_PPP	1.73	1.78	7.71
1k22_MEL	1.15	1.30	2.56
1k21_IGN	1.42	1.18	1.28
1dwd_MID	1.56	1.13	1.78
1dwc_MIT	1.03	1.84	7.82
1d4p_BPP	0.80	1.06	0.83
1bmn_BM9	0.76	1.19	4.83
1b7x_PPP	0.93	1.40	8.62
1a61_MOL	0.73	1.25	2.46

[a] Four characters followed by three characters (separated by a underline) denote the PDB code and the ligand name in the PDB, respectively.

[b] The RMSD value (Å) for docking each ligand back into its reference complex, 1dwd.

[c] E_{tot} and E_{bind} are defined in Equation 1.

Table 10. Comparison of GEMDOCK with GOLD on screening 1010 compounds with false positive rates (%)

	GEMDCOK		GOLD	
True Positive (%)	None [a]	Both [b]	Ori [c]	New [d]
80	8.50 %	3.00 %	77.90 %	55.70 %
90	9.50 %	3.10 %	91.30 %	76.30 %
100	10.70 %	3.40 %	99.30 %	99.80 %

[a] GEMDOCK without pharmacological interaction and ligand preferences (e.g., E_{bind} for the scoring function).

[b] GMDDCOK with pharmacological interaction and ligand preferences (e.g., E_{tot} for the scoring function).

[c] GOLD with external *H-bond*, external *vdw*, internal *vdw* and internal *torsion* of GoldScore.

[d] GOLD with external *vdw* and external *H-bond* of GoldScore.

Table 11. GEMDOCK and GOLD Screening Accuracies Using Different Combinations of Scoring Functions

	Thrombin inhibitor (reference protein: 1dwd)			
Measure factor	None [a]	Both [b]	Ori [c]	New [d]
Average hit rate (%)	22.94	34.88	4.23	4.49
Average enrichment	23.16	35.23	4.28	4.53
Average false positive rate (%)	4.18	1.38	53.66	33.67
Average GH score	0.30	0.39	0.07	0.10

[a,b] Using E_{bind} and E_{tot}, respectively, for the scoring function. These energy terms are defined in Equation 1.

[c] GOLD with external *H-bond*, external *vdw*, internal *vdw* and internal *torsion* of GoldScore.

[d] GOLD with external *vdw* and external *H-bond* of GoldScore.

Table 12. The Ranks of Ten Known Thrombin Inhibitors Using GEMDOCK and GOLD with Different Combinations of Scoring Functions

Thrombin inhibitors (reference protein: 1dwd)				
Ligand Id [a]	None [b]	Both [c]	Ori [d]	New [e]
7kme_BNN	6	2	322	14
2hpp_PPP	93	5	653	772
1k22_MEL	41	9	346	565
1k21_IGN	104	11	205	92
1dwd_MID	5	12	787	217
1dwc_MIT	16	15	922	1008
1d4p_BPP	12	17	3	12
1bmn_BM9	4	38	700	36
1b7x_PPP	117	40	480	398
1a61_MOL	75	44	1003	308

[a] Defined in Figures 8.

[b,c] Using E_{bind} and E_{tot}, respectively, for the scoring function. These energy terms are defined in Equation 1.

[d] GOLD with external *H-bond*, external *vdw*, internal *vdw* and internal *torsion* of GoldScore.

[e] GOLD with external *vdw* and external *H-bond* of GoldScore.

Table 13. Statistics of Overall Screening Accuracy for Fusion Screening Ranking of HSV-1 Thymidine Kinase (TK)

Measurement factors	Non-combination [a]	Two combination [b]	Three combination [c]	Four combination [d]
Enrichment factor				
Average	12.95	27.20	35.36	-[e]
Standard deviation	10.30	20.78	17.07	-[e]
Maximum value	28.16	54.44	52.14	36.70
Minimum value	6.20	9.92	11.81	36.70
False positive rate				
Average	5.65%	2.97%	1.94%	-[e]
Standard deviation	3.17	2.00	1.51	-[e]
Maximum value	8.52%	5.21%	4.18%	1.36%
Minimum value	1.43%	0.61%	1.05%	1.36%

[a] Four individual methods

[b] Combination of two selected methods, six compositions

[c] Combination of three selected methods, four compositions

[d] Combination of four selected methods and only one composition

[e] Average and standard deviation could not calculate when one value exist

Table 14. Statistics of Overall Screening Accuracy for Fusion Screening Ranking of Human Dihydrofolate Reductase (DHFR)

Measurement factors	Non-combination [a]	Two combination [b]	Three combination [c]	Four combination [d]
Enrichment factor				
Average	69.27	79.54	86.78	-[e]
Standard deviation	13.05	8.40	2.73	-[e]
Maximum value	88.13	91.65	89.47	87.19
Minimum value	58.89	68.33	83.10	87.19
False positive rate				
Average	1.59%	0.55%	0.26%	-[e]
Standard deviation	1.44	0.35	0.16	-[e]
Maximum value	3.57%	0.94%	0.43%	0.27
Minimum value	0.13%	0.09%	0.12%	0.27

[a] Four individual methods

[b] Combination of two selected methods, six compositions

[c] Combination of three selected methods, four compositions

[d] Combination of four selected methods and only one composition

[e] Average and standard deviation could not calculate when one value exist

Table 15. Statistics of Overall Screening Accuracy for Fusion Screening Ranking of Estrogen Receptor α (ER)

Measurement factors	Non-combination [a]	Two combination [b]	Three combination [c]	Four combination [d]
Enrichment factor				
Average	59.07	71.76	70.62	-[e]
Standard deviation	29.24	19.79	17.44	-[e]
Maximum value	92.19	96.52	96.69	68.78
Minimum value	34.07	51.57	59.97	68.78
False positive rate				
Average	5.69%	4.13%	2.66%	-[e]
Standard deviation	9.85	4.41	1.81	-[e]
Maximum value	20.44%	9.22%	4.16%	1.86%
Minimum value	0.13%	0.04%	0.04%	1.86%

[a] Four individual methods

[b] Combination of two selected methods, six compositions

[c] Combination of three selected methods, four compositions

[d] Combination of four selected methods and only one composition

[e] Average and standard deviation could not calculate when one value exist

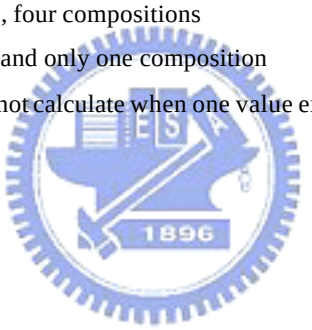

Table 16. Statistics of Overall Screening Accuracy for Fusion Screening Scoring of HSV-1 Thymidine Kinase (TK)

Measurement factors	Non-combination [a]	Two combination [b]	Three combination [c]	Four combination [d]
Enrichment factor				
Average	12.95	18.86	19.80	-[e]
Standard deviation	10.30	12.61	14.64	-[e]
Maximum value	28.16	35.16	37.96	25.96
Minimum value	6.20	6.54	7.93	25.96
False positive rate				
Average	5.65%	4.04%	3.88%	-[e]
Standard deviation	3.17	2.97	2.91	-[e]
Maximum value	8.52%	7.91%	6.39%	1.70%
Minimum value	1.43%	1.07%	1.01%	1.70%

[a] Four individual methods

[b] Combination of two selected methods, six compositions

[c] Combination of three selected methods, four compositions

[d] Combination of four selected methods and only one composition

[e] Average and standard deviation could not calculate when one value exist

Table 17. Statistics of Overall Screening Accuracy for Fusion Screening Scoring of Human Dihydrofolate Reductase (DHFR)

Measurement factors	Non-combination [a]	Two combination [b]	Three combination [c]	Four combination [d]
Enrichment factor				
Average	69.27	67.44	69.41	-[e]
Standard deviation	13.05	3.04	1.85	-[e]
Maximum value	88.13	72.13	71.89	70.38
Minimum value	58.89	64.51	67.46	70.38
False positive rate				
Average	1.59%	0.95%	0.70%	-[e]
Standard deviation	1.44	0.18	0.05	-[e]
Maximum value	3.57%	1.29%	0.77%	0.61%
Minimum value	0.13%	0.80%	0.67%	0.61%

[a] Four individual methods

[b] Combination of two selected methods, six compositions

[c] Combination of three selected methods, four compositions

[d] Combination of four selected methods and only one composition

[e] Average and standard deviation could not calculate when one value exist

Table 18. Statistics of Overall Screening Accuracy for Fusion Screening Scoring of Estrogen Receptor α (ER)

Measurement factors	Non-combination [a]	Two combination [b]	Three combination [c]	Four combination [d]
Enrichment factor				
Average	59.07	71.24	82.20	-[e]
Standard deviation	29.24	22.18	14.79	-[e]
Maximum value	92.19	94.85	91.18	88.77
Minimum value	34.07	48.25	60.08	88.77
False positive rate				
Average	5.69%	2.68%	0.33%	-[e]
Standard deviation	9.85	5.43	0.37	-[e]
Maximum value	20.44%	13.74%	0.88%	0.15%
Minimum value	0.13%	0.07%	0.13%	0.15%

[a] Four individual methods

[b] Combination of two selected methods, six compositions

[c] Combination of three selected methods, four compositions

[d] Combination of four selected methods and only one composition

[e] Average and standard deviation could not calculate when one value exist