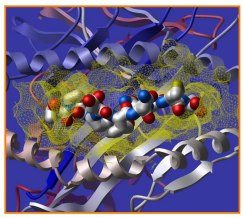
Protein-Ligand Docking Evaluations

Thomas Funkhouser
Princeton University
CS597A, Fall 2007

Introduction

Protein-ligand docking:

- Given a protein and a ligand, determine the pose(s) and conformation(s) minimizing the total energy of the protein-ligand complex

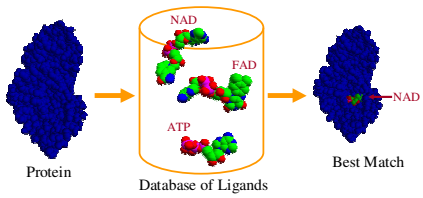


http://www.molsoft.com/

Introduction

Virtual screening:

- Given a protein and a database of ligands, use scores (produced by a docking tool) to determine which ligands are most likely to bind



How Good Are Docking Programs?

Questions:

- Docking accuracy?
- Screening accuracy?
- Binding affinity prediction accuracy?
- Computation speed?

How Good Are Docking Programs?

Questions:

ØDocking accuracy?

- Computation speed?
- Binding affinity prediction accuracy?
- Screening accuracy?

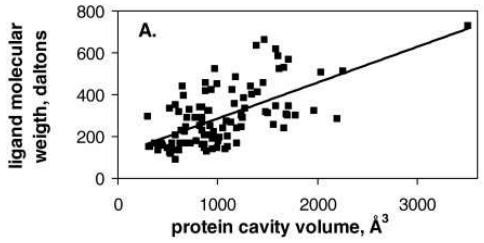
[Kellenberger04] Docking Study

8 Docking Programs:

- FRED (multiple conformers)
- DOCK (incremental construction)
- FLEXX (incremental construction)
- SLIDE (incremental construction)
- SURFLEX (incremental construction)
- GLIDE (Monte Carlo simulated annealing)
- QXP (Monte Carlo simulated annealing)
- GOLD (genetic algorithm)

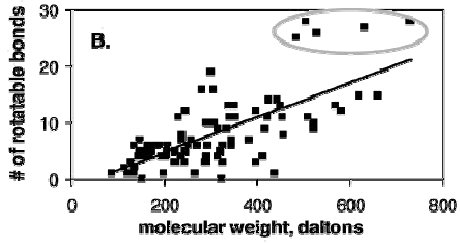
[Kellenberger04] Docking Study

100 Protein-Ligand Complexes:



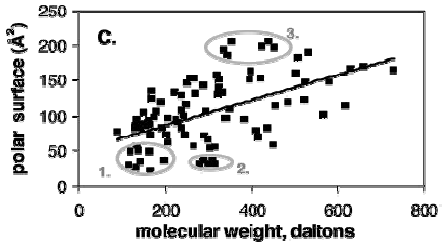
[Kellenberger04] Docking Study

100 Protein-Ligand Complexes:



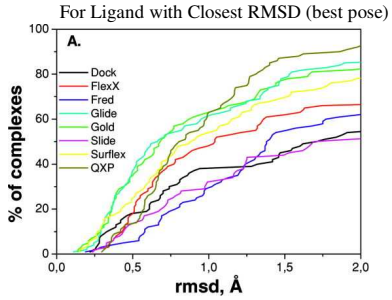
[Kellenberger04] Docking Study

100 Protein-Ligand Complexes:



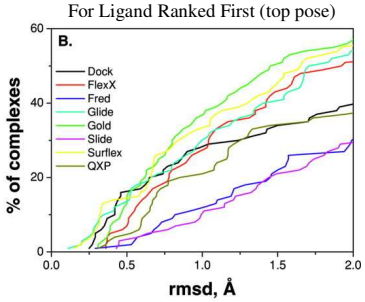
[Kellenberger04] Docking Study

Docking accuracy:



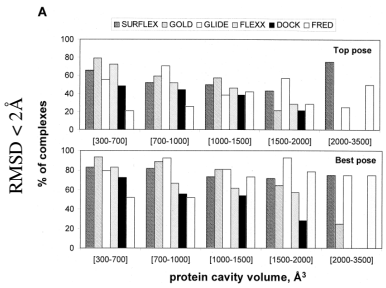
[Kellenberger04] Docking Study

Docking accuracy:



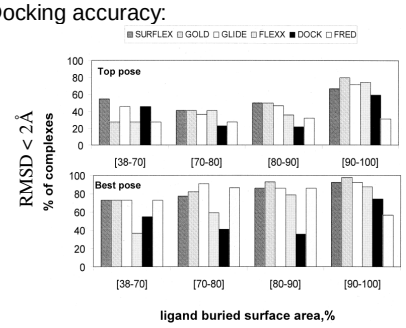
[Kellenberger04] Docking Study

Docking accuracy:



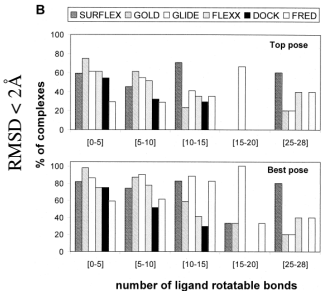
[Kellenberger04] Docking Study

Docking accuracy:



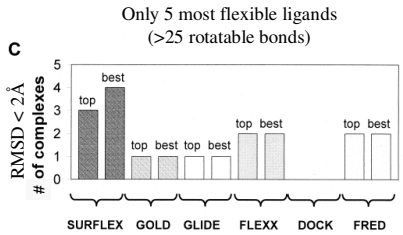
[Kellenberger04] Docking Study

Docking accuracy:



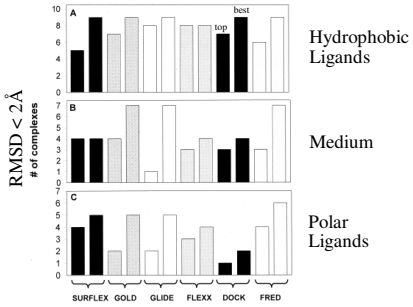
[Kellenberger04] Docking Study

Docking accuracy:



[Kellenberger04] Docking Study

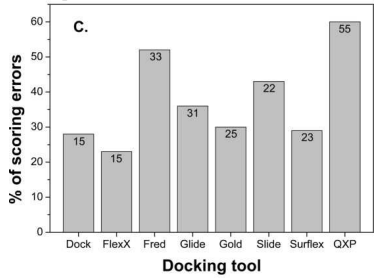
Docking accuracy:



[Kellenberger04] Docking Study

Docking accuracy:

% where pose <2Å RMSD considered, but not scored first



How Good Are Docking Programs?

Questions:

- Docking accuracy?

ØComputation speed?

- Binding affinity prediction accuracy?
- Screening accuracy?

[Kellenberger04] Docking Study

Computation time (seconds):

Program	Average[a]	Minimum[b]	Maximum[c]
FRED[d]	18	0.1	193
DOCK	46	1	667
FLEXX	67	2	595
QXP	108	37	378
SLIDE	118	1	1743
SURFLEX	135	9	1460
GOLD	137	55	479
GLIDE	234	9	2825

How Good Are Docking Programs?

Questions:

- Docking accuracy?
- Computation speed?

ØBinding affinity prediction accuracy?

- Screening accuracy?

[Marsden04] Study

Compare predicted and measured binding energies

Empirical methods:

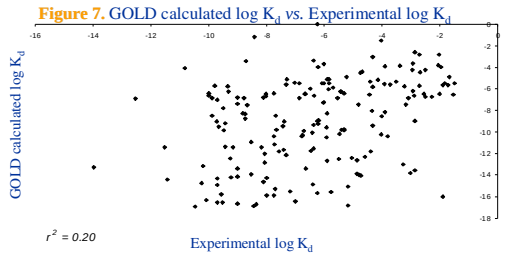
- Gold [Jones97]
- DOCK [Kuntz82]
- ChemScore [Eldridge97]

Knowledge-based methods:

- PMF [Muegge99]
- Bleep [Mitchell99]

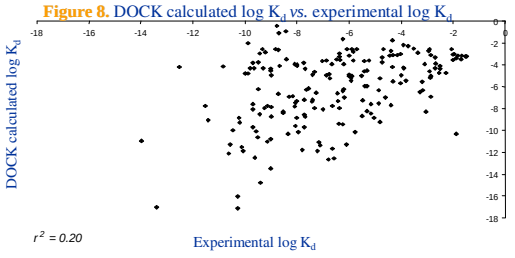
[Marsden04]

[Marsden04] Study



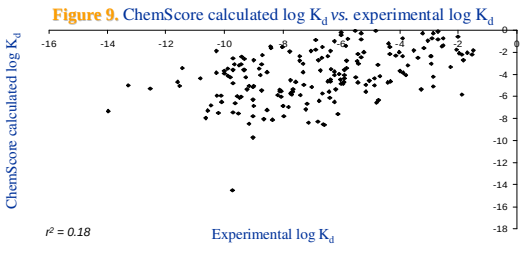
[Marsden04]

[Marsden04] Study



[Marsden04]

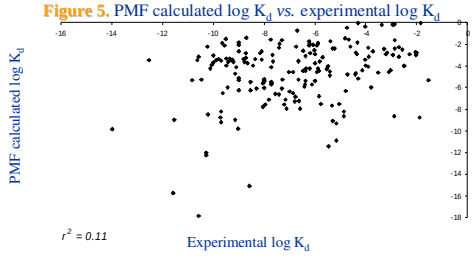
[Marsden04] Study



[Marsden04]

[Marsden04] Study

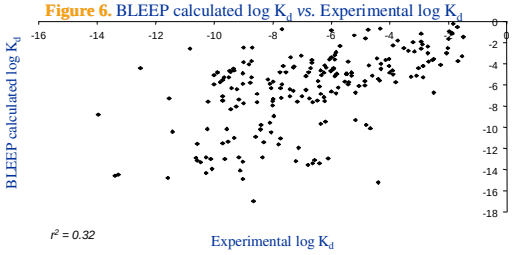
Figure 5. PMF calculated log K_d *vs.* experimental log K_d



[Marsden04]

[Marsden04] Study

Figure 6. BLEEP calculated log K_d *vs.* Experimental log K_d



[Marsden04]

[Marsden04] Study

Table 1. Correlations Between Experimental and Calculated log K_d Values Given by Five Scoring Functions.

	Dataset	No. of complexes	BLEEP R_p	BLEEP r^2	PMF R_p	PMF r^2	GOLD R_p	GOLD r^2	DOCK R_p	DOCK r^2	ChemScore R_p	ChemScore r^2
	All	205	0.59	0.32	0.31	0.11	0.50	0.20	0.43	0.20	0.45	0.18
A	Serine proteinases	35	0.82	0.74	0.75	0.51	0.61	0.47	0.83	0.69	0.13	0.00
B	Metalloproteinases	25	0.72	0.44	0.45	0.33	0.44	0.14	0.42	0.20	0.43	0.17
C	Carbonic anhydrase ii	18	0.53	0.47	0.46	0.32	0.42	0.34	0.54	0.01	0.50	0.28
D	Sugar binding proteins	30	0.76	0.58	0.45	0.09	0.02	0.00	0.05	0.00	0.29	0.09
E	Aspartic proteinases	38	0.08	0.01	-0.52	0.13	0.02	0.00	0.08	0.01	-0.08	0.00

[Marsden04]

[Wang04] Study

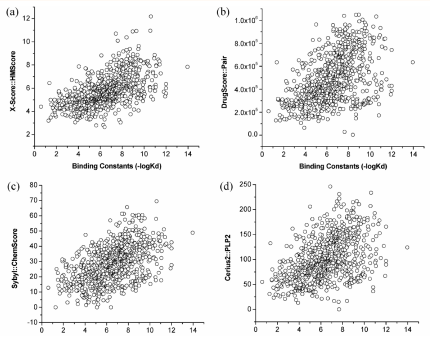
Data set:

- 800 protein-ligand complexes from PDB (PDBBind) with measured binding affinities

Scoring functions:

- Force-field methods:
  - § D-Score, GoldScore
- Empirical methods:
  - § X-Score, F-Score, ChemScore, LigScore, PLP, LUDI, HINT
- Knowledge-based methods:
  - § DrugScore, PMF

[Wang04] Study



[Wang04] Study

Pearson's correlation coefficient

scoring function	N^b	R_p	SD	ME	a	b
X-Score::HPScore	*800*	*0.514*	*1.89*	*1.47*	*0.71*	*2.03*
X-Score::HMScore	*800*	*0.566*	*1.82*	*1.42*	*0.92*	*1.18*
X-Score::HSScore	*800*	*0.506*	*1.90*	*1.48*	*0.93*	*1.24*
~~*DrugScore::Pair*~~	~~*800*~~	~~*0.473*~~	~~*1.94*~~	~~*1.51*~~	~~*4.9 × 10⁻⁶*~~	~~*4.19*~~
DrugScore::Surf	*800*	*0.463*	*1.95*	*1.53*	*7.2 × 10⁻²*	*4.48*
DrugScore::Pair/Surf	*800*	*0.476*	*1.94*	*1.50*	*4.7 × 10⁻⁶*	*4.09*
Sybyl::D-Score	800	0.522	2.09	1.67	9.7 × 10⁻³	5.00
Sybyl::PMF-Score	785	0.147	2.16	1.74	6.4 × 10⁻³	5.92
Cerius2::LUDI1	790	0.334	2.08	1.66	2.6 × 10⁻³	4.88
Cerius2::LUDI2	799	0.379	2.04	1.62	4.2 × 10⁻³	4.28
Cerius2::LUDI3	800	0.331	2.08	1.67	3.2 × 10⁻³	4.68
GOLD::GoldScore	694	0.285	2.16	1.72	2.4 × 10⁻²	5.33
GOLD::GoldScore_opt	772	0.365	2.06	1.63	3.0 × 10⁻²	4.70
GOLD::ChemScore	741	0.423	2.00	1.56	8.5 × 10⁻²	4.65
GOLD::ChemScore_opt	762	0.449	1.96	1.52	8.6 × 10⁻²	4.41
HINT	800	0.330	2.08	1.65	0.20	6.36

HPScore = $C_{0,1} + C_{VDW,1}$*(VDW) + $C_{HB,1}$*(H-Bond) + C_{HP}*(Hydrophobic Pair) + $C_{RT,1}$*(Rotor)

HMScore = $C_{0,2} + C_{VDW,2}$*(VDW) + $C_{HB,2}$*(H-Bond) + C_{HM}*(Hydrophobic Match) + $C_{RT,2}$*(Rotor)

HSScore = $C_{0,3} + C_{VDW,3}$*(VDW) + $C_{HB,3}$*(H-Bond) + C_{HS}*(Hydrophobic surface) + $C_{RT,3}$*(Rotor)

[Wang04] Study

scoring function	success rate in classification[a] low-affinity group ($-\log K_d < 5.0$)	medium-affinity group ($5.0 \le -\log K_d \le 8.0$)	high-affinity group ($-\log K_d > 8.0$)
X-Score::HPScore	33/205 = 16%	358/402 = 89%	48/193 = 25%
X-Score::HMScore	41/205 = 20%	348/402 = 87%	65/193 = 34%
X-Score::HSScore	29/205 = 14%	359/402 = 89%	53/193 = 27%
DrugScore::Pair	24/205 = 12%	359/402 = 89%	45/193 = 23%
DrugScore::Surf	11/205 = 5%	362/402 = 90%	45/193 = 23%
DrugScore::Pair/Surf	24/205 = 12%	358/402 = 89%	47/193 = 24%
Sybyl::D-Score	0/205 = 0%	384/402 = 96%	2/193 = 1%
Sybyl::PMF-Score	0/196 = 0%	395/396 = 99%	0/193 = 0%
Sybyl::G-Score	12/205 = 6%	359/402 = 89%	30/193 = 16%
Sybyl::ChemScore	38/204 = 19%	349/400 = 87%	40/193 = 21%
Sybyl::F-Score	0/182 = 0%	362/362 = 100%	0/188 = 0%
Cerius2::LigScore	11/186 = 6%	340/366 = 93%	16/165 = 10%
Cerius2::PLP1	24/205 = 12%	364/401 = 91%	35/193 = 18%
Cerius2::PLP2	30/205 = 15%	363/402 = 90%	32/193 = 17%
Cerius2::PMF	0/202 = 0%	390/400 = 97%	3/193 = 2%
Cerius2::LUDI1	1/203 = 0%	379/394 = 96%	9/193 = 5%
Cerius2::LUDI2	6/205 = 3%	378/401 = 94%	15/193 = 8%
Cerius2::LUDI3	1/205 = 0%	387/402 = 96%	9/193 = 5%
GOLD::GoldScore	0/178 = 0%	331/339 = 98%	4/177 = 2%
GOLD::GoldScore_opt	3/200 = 1%	366/385 = 95%	11/187 = 6%
GOLD::ChemScore	8/177 = 5%	345/376 = 92%	37/188 = 20%
GOLD::ChemScore_opt	20/187 = 11%	346/386 = 90%	38/189 = 20%
HINT	2/205 = 1%	388/402 = 97%	11/193 = 6%

[a] The denominator is the total number of samples that an individual scoring function could compute in a certain group; the numerator is the total number of correctly classified samples in this group by the given scoring function.

How Good Are Docking Programs?

Questions:

- Docking accuracy?
- Computation speed?
- Binding affinity prediction accuracy?

ØScreening accuracy?

[Kellenberger04] Screening Study

Dock 1000 ligands into HIV-1 TK

- 10 known TK inhibitors
- 990 randomly chosen “drug-like” molecules

Measure how often TK inhibitors are highly ranked

[Kellenberger04] Screening Study

Screening accuracy:

TABLE I Description of Hit Lists generated by 8 Docking Tools on the Thymidine Kinase Example
A hit list is generated from the top-scoring compounds selected at a given threshold.

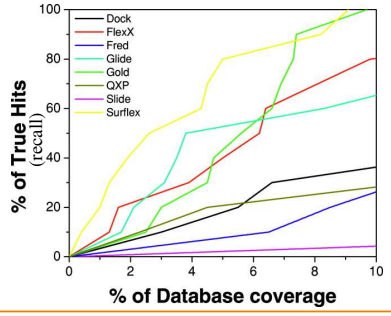
	Top 2.5%		Top 5%		Top 10%	
	Hit Rate[a]	Yield[b]	Hit Rate	Yield	Hit Rate	Yield
DOCK	0	0	2	10	3	30
FLEXX	8	20	8	40	8	80
FRED	0	0	0	0	2	20
GLIDE	8	20	10	50	6	60
GOLD	4	10	8	40	10	100
SLIDE	0	0	0	0	0	0
SURFLEX[c]	16	40	16	80	10	100
QXP	0	0	4	20	2	20

[a] Hit rate: (AH/TH) × 100
[b] Yield: (AH/A) × 100, where TH is the total number of compounds in the hit list, AH the number of true hits in the hit list, and A the total number of true hits in the library.
[c] Figures reported for SURFLEX were obtained by using a protein penetration threshold value of −6.

Precision Recall

[Kellenberger04] Screening Study

Screening accuracy:



Conclusions

Docking accuracy?

- Correct pose can be predicted (within 2Å RMSD) for majority of cases (60-80%) – depends on properties

Computation speed?

- Minutes per complex

Binding affinity prediction accuracy?

- Scoring functions generally have modest correlation with measured K_d values

Screening accuracy?

- Determining which ligand binds best to protein is very difficult – correct ligand not always amongst top 10%

Discussion

?

References

[Eldridge97] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, "Empirical scoring functions. I: The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes," J. Comput.-Aided Mol. Des., 11, 1997, pp. 425-445.

[Friesner04] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, "Glide: A New Approach for Rapid, Accurate Docking and Scoring.," J. Med. Chem, 47, 2004, pp. 1739-1749.

[Gohlke00] H. Gohlke, M. Hendlich, G. Klebe, "Knowledge-based scoring function to predict protein-ligand interactions," J. Mol. Biol., 295, 2000, pp. 337-356.

[Jones97] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, "Development and Validation of a Genetic Algorithm for Flexible Docking," J. Mol. Biol., 267, 1997, pp. 727-748.

[Kellenberger04] E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, "Comparative evaluation of eight docking tools for docking and virtual screening accuracy," Proteins, 57, 2, 2004, pp. 225-242.

[Kuntz82] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, "A geometric approach to macromolecule-ligand interactions," J. Mol. Biol, 161, 1982, pp. 269-288.

[Marsden04] Marsden PM, Puvanendrampillai D, Mitchell JBO and Glen RC., "Predicting protein ligand binding affinities: a low scoring game?", Organic Biomolecular Chemistry, 2, 2004, p. 3267-3273.

[Mitchell99] J.B.O. Mitchell, R. Laskowski, A. Alex, and J.M. Thornton, "BLEEP - potential of mean force describing protein-ligand interactions: II. Calculation of binding energies and comparison with experimental data", J. Comput. Chem., 20, 11, 1999, pp. 1177-1185.

[Morris98] Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. "Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function", J. Computational Chemistry, 19, 1998, p. 1639-1662.

[Muegge99] I. Muegge, Y.C. Martin, "A general and fast scoring function for protein-ligand interactions: A simplified potential approach," J. Med. Chem., 42, 1999, pp. 791-804.

[Rarey96] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, "A Fast Flexible Docking Method using an Incremental Construction Algorithm," Journal of Molecular Biology, 261, 3, 1996, pp. 470-489.

[Wang04] R. Wang, Y. Lu, X. Fang, and S. Wang, "An Extensive Test of 14 Scoring Functions Using PDBbind Refined Set of 800 Protein-Ligand Complexes," J. Chem Inf Comput Sci, 44, 2004, 2114-2125.