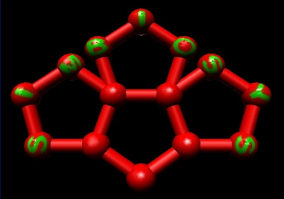
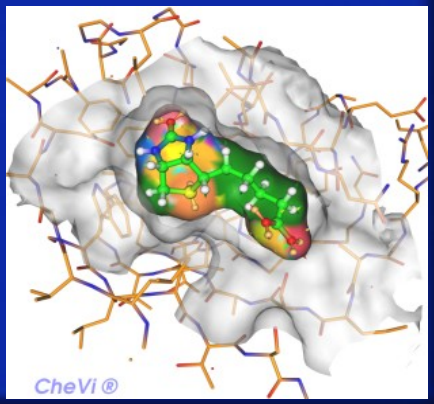
Fragment based docking and linking engine of eHiTS®

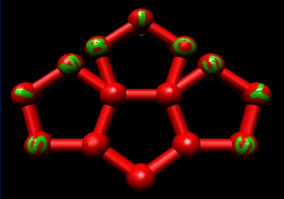


Zsolt Zsoldos

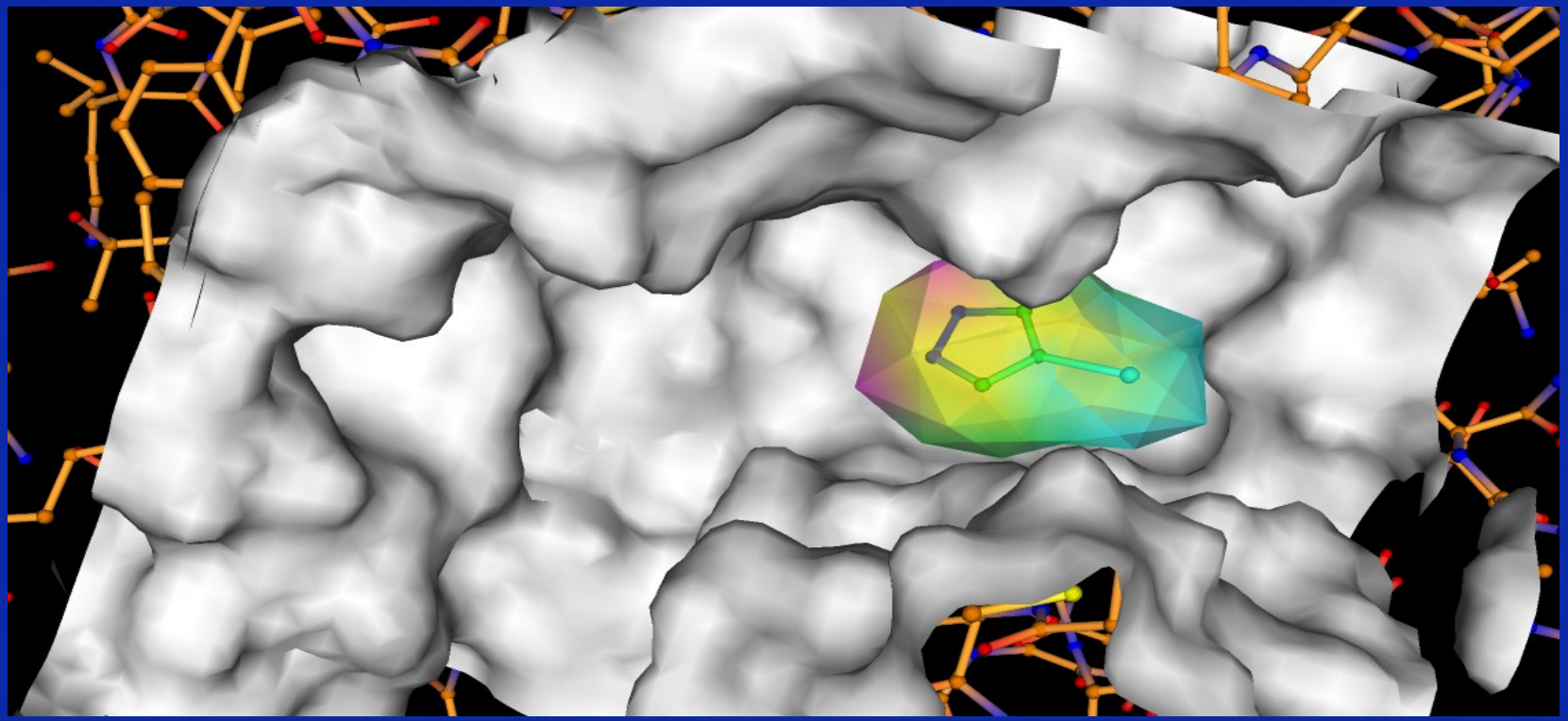
http://www.simbiosys.ca/



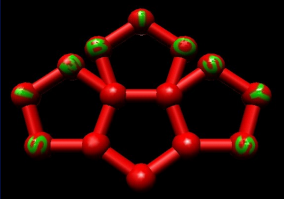
2. Introduction: the fragment docking problem for FBDD



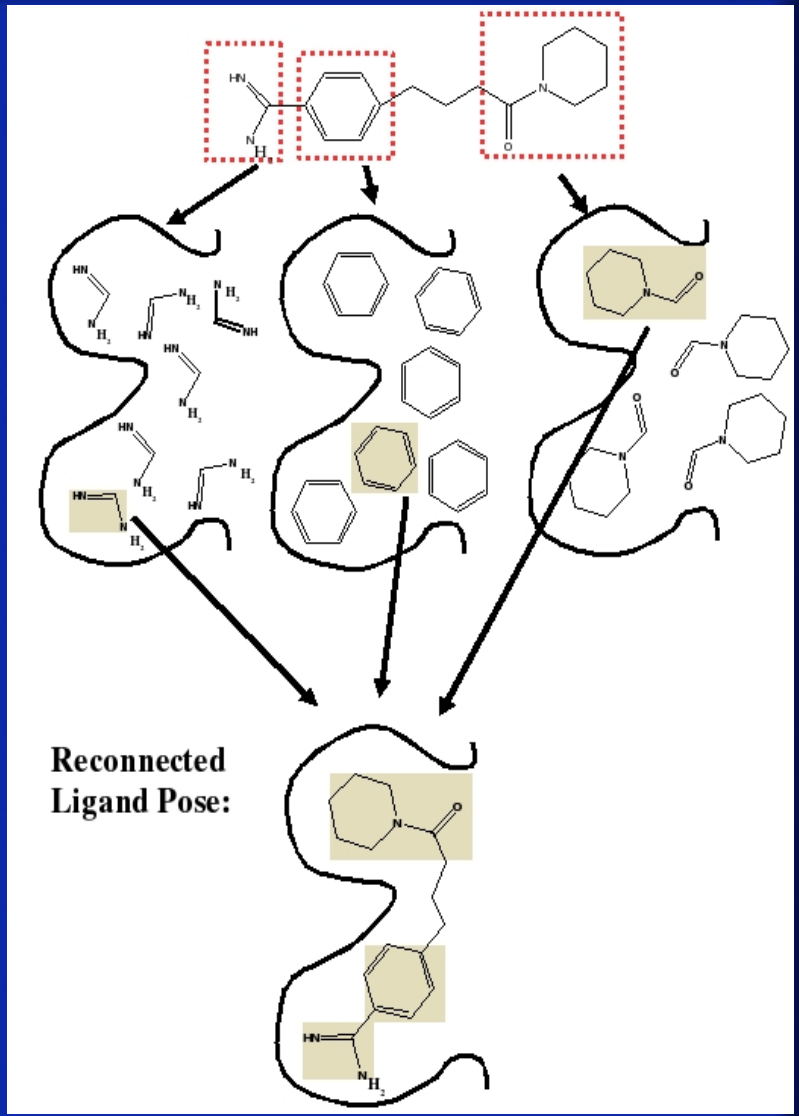
- For FBDD we need to dock small fragments into a large binding site
- Few or no rotatable bonds => small conformation space
- Fragment fits in many ways and places => large pose space
- Scoring problem: sub-optimal fragment poses in the best total score



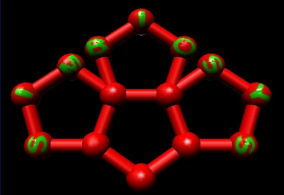
3. eHiTS algorithm overview: fragment based engine



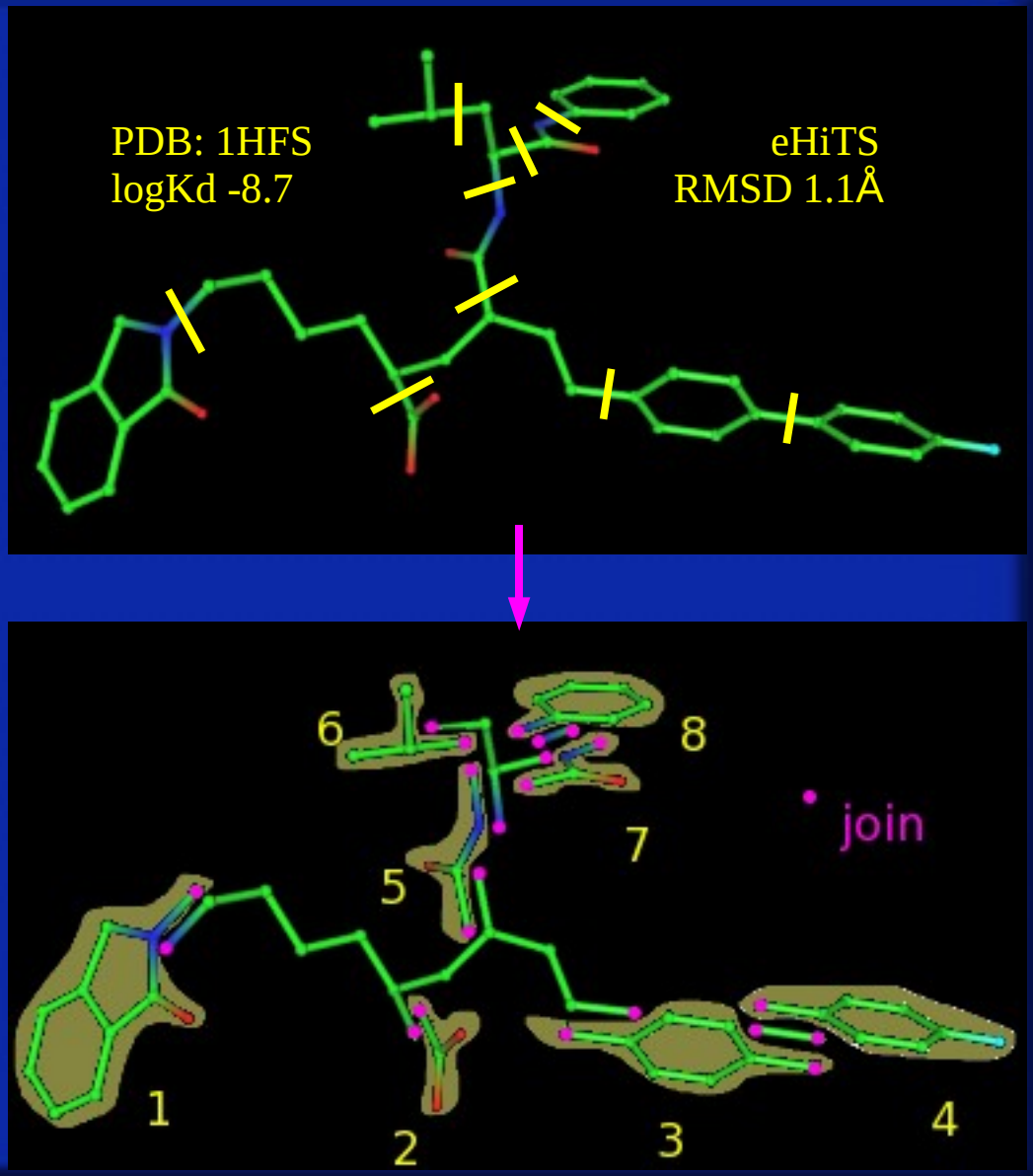
- Ligand is divided into rigid fragments and flexible chains
- All rigid fragments are docked independently
  - many poses,
  - steric filter only
- Pose matching
  - clique detection
  - global score
- Flexible chain fitting (continuous)
- Local energy minimization



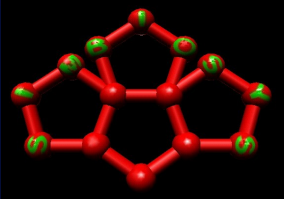
4. Fragmentation of the ligand



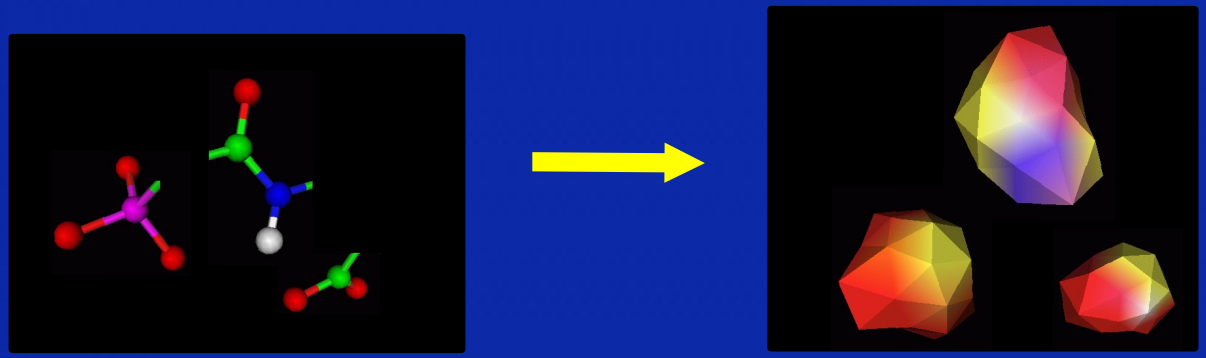
- Rigid fragments
  - Ring systems
  - Double & resonance bonds
  - Terminal fragments
  - No rotatable bonds inside
- Flexible chains
  - Single bonds, sp^3 atoms
  - May contain junctions
- Join bonds duplicated

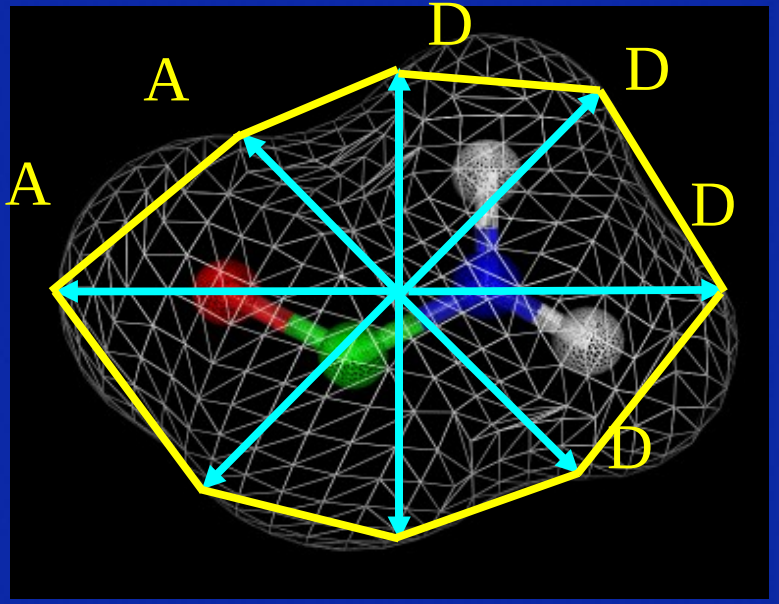


5. Rigid fragment docking based on Chemical feature mapped polyhedra

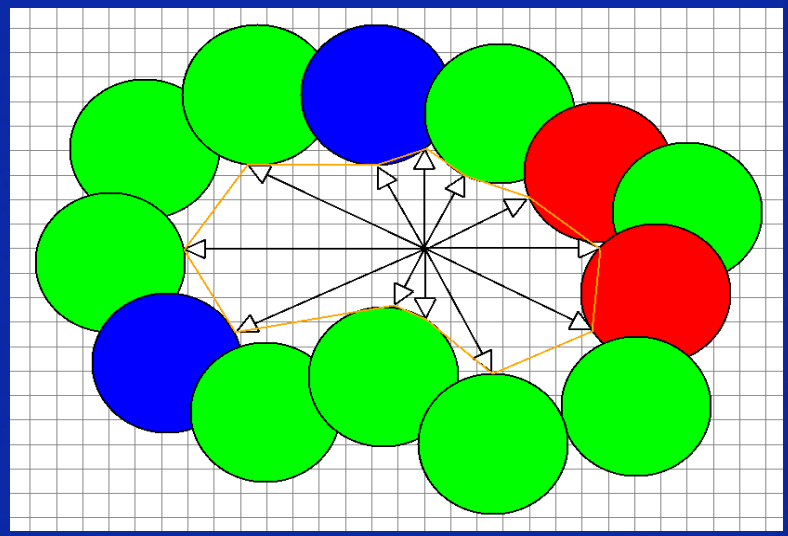


- Polyhedron shrink-wrapped onto molecular surface (Connolly)

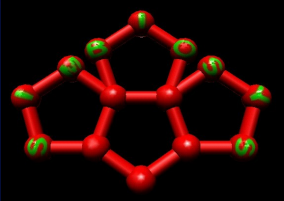




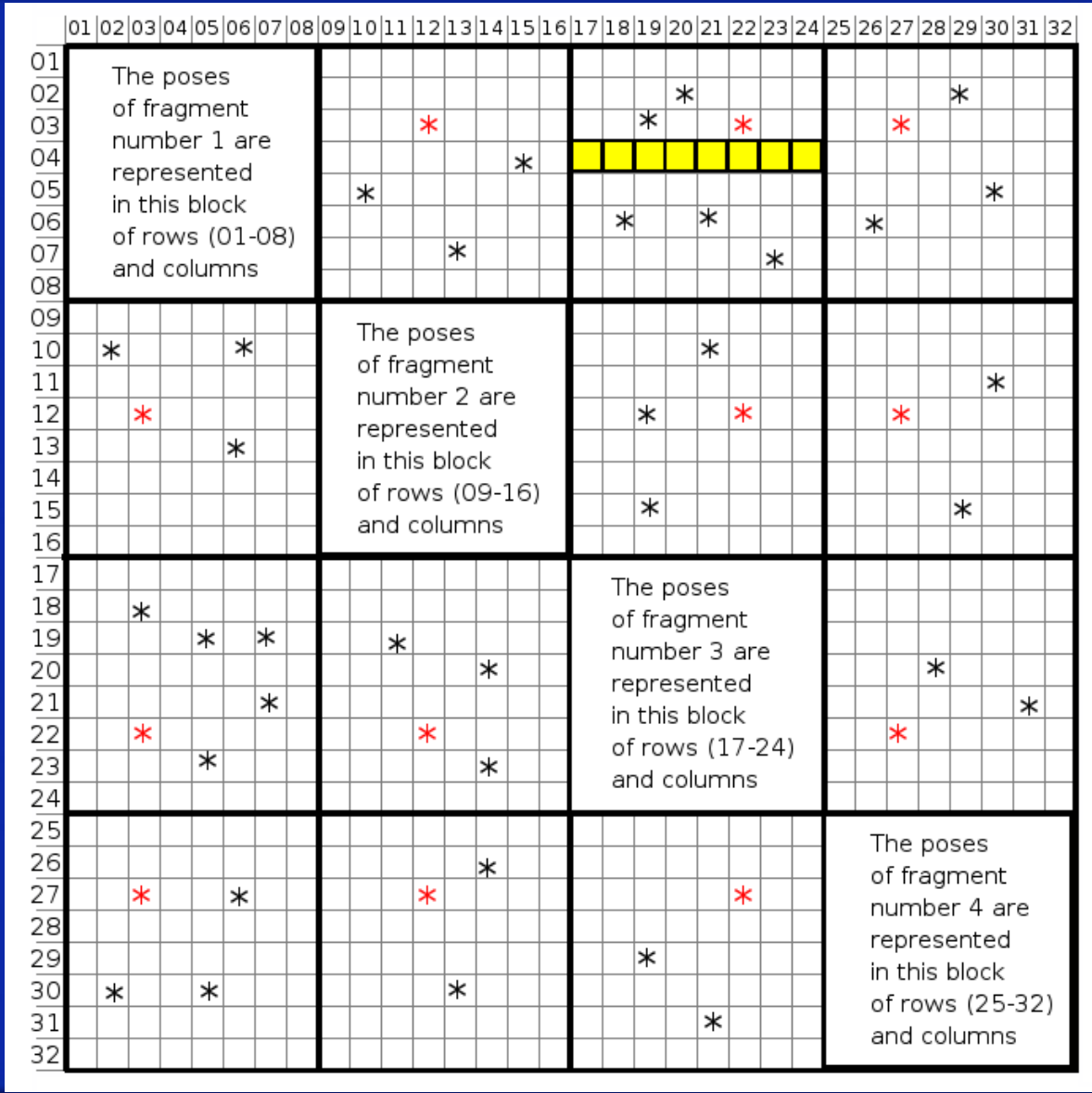
- Chemical feature flags on vertices
- Analogue cavity representation
- Rapid mapping of ligand and cavity polyhedra



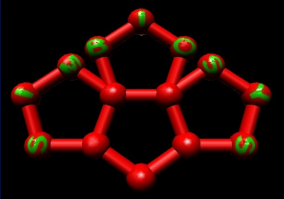
6. Pose matching by clique detection



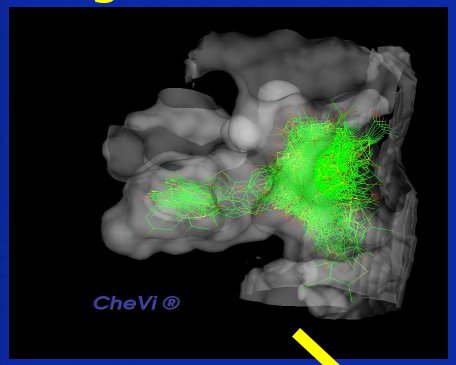
- Nodes:
  - Rigid fragment poses
- Edges:
  - Different fragments
  - No bump
  - Compatible distance
- Special attributes:
  - Fragment blocks
  - 1 pose per block



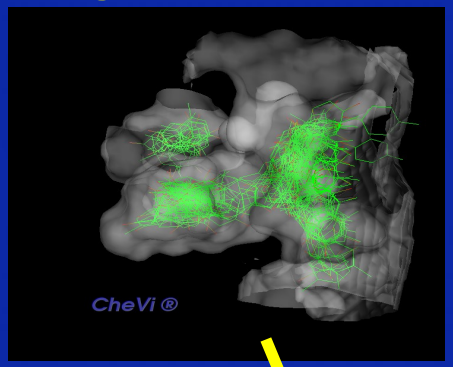
7. Fragment pose matching example



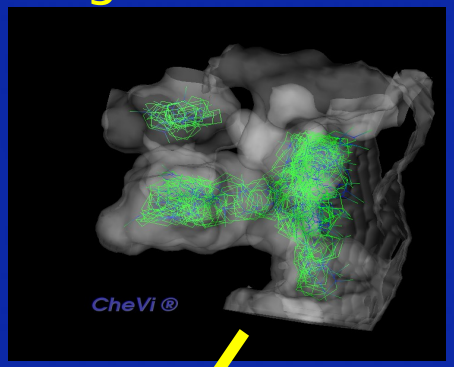
Frag 1



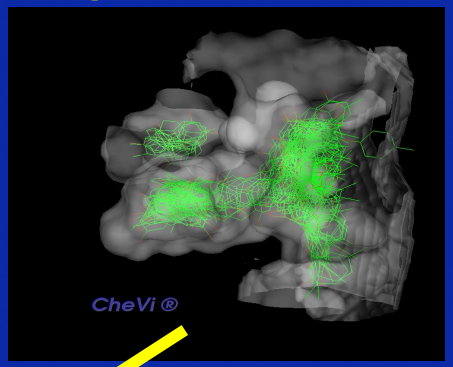
Frag 2

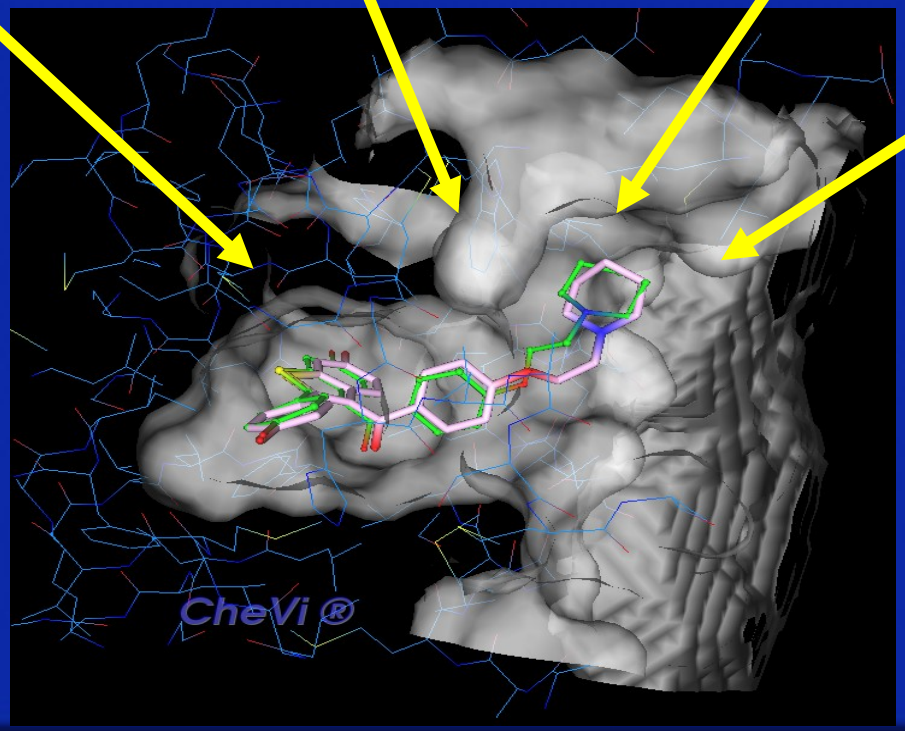


Frag 3



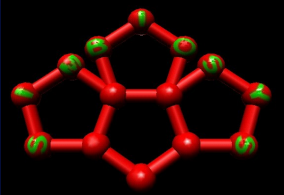
Frag 4



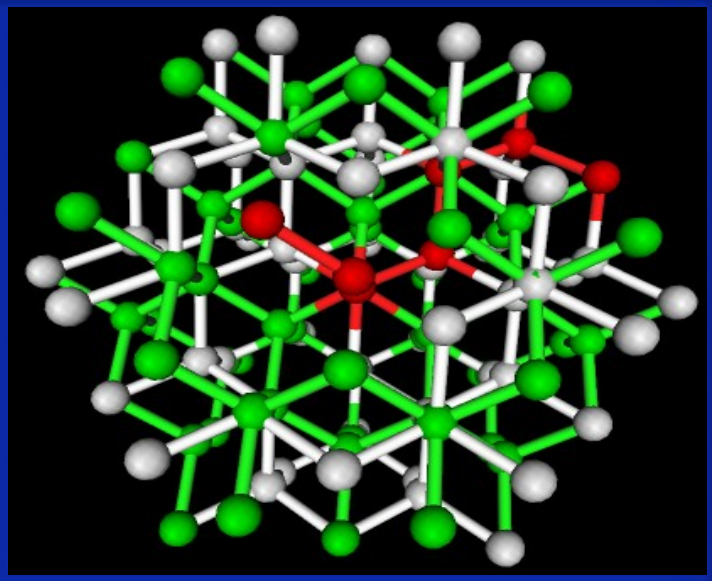


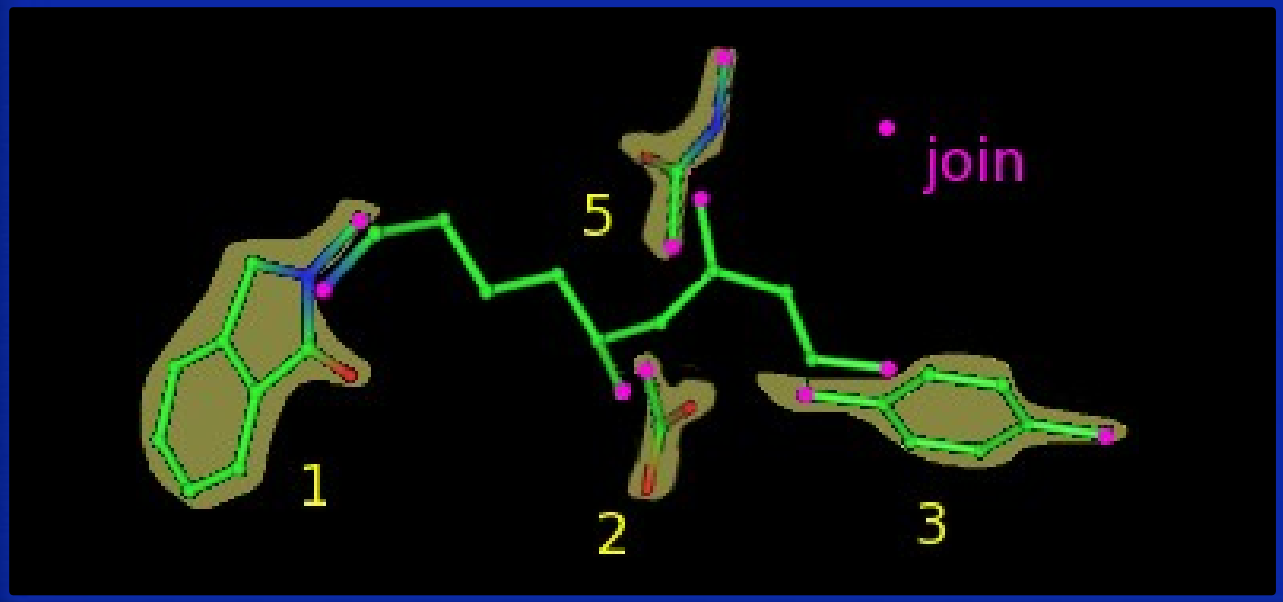
Top Ranked Pose
Xray-RMSD: 0.7Å
eHiTS Score: -8.5

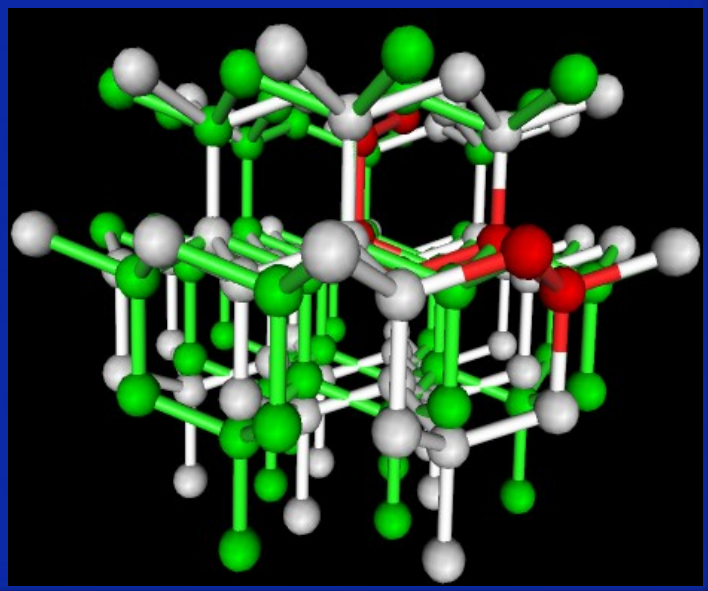
8. Flexible chain fitting



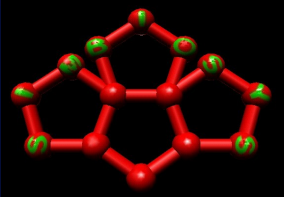
- Find chain conformation to link given rigid fragment poses:
  - Pick starting conformation from double diamond latice lookup table
  - Tweak the chain to fit end point pairs and avoid boundary (PLSF)



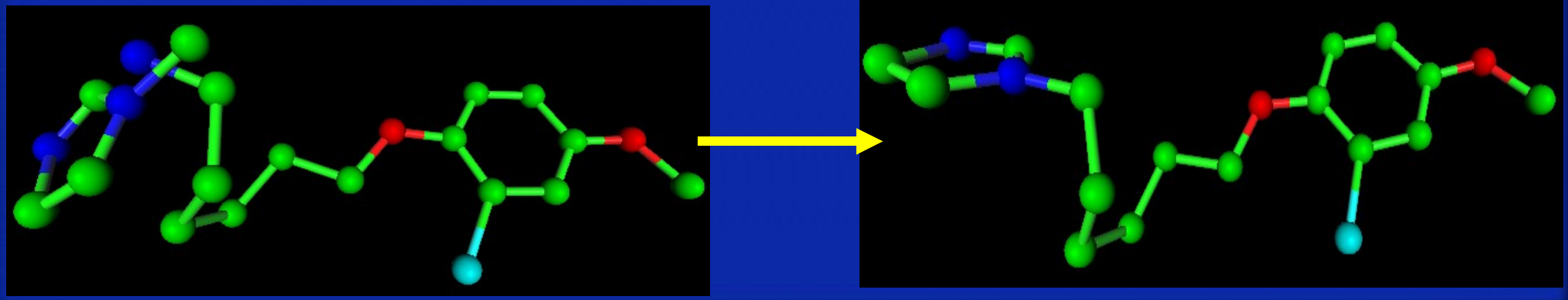




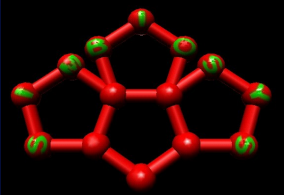
9. Reconstruction of the ligand by joining the fragments



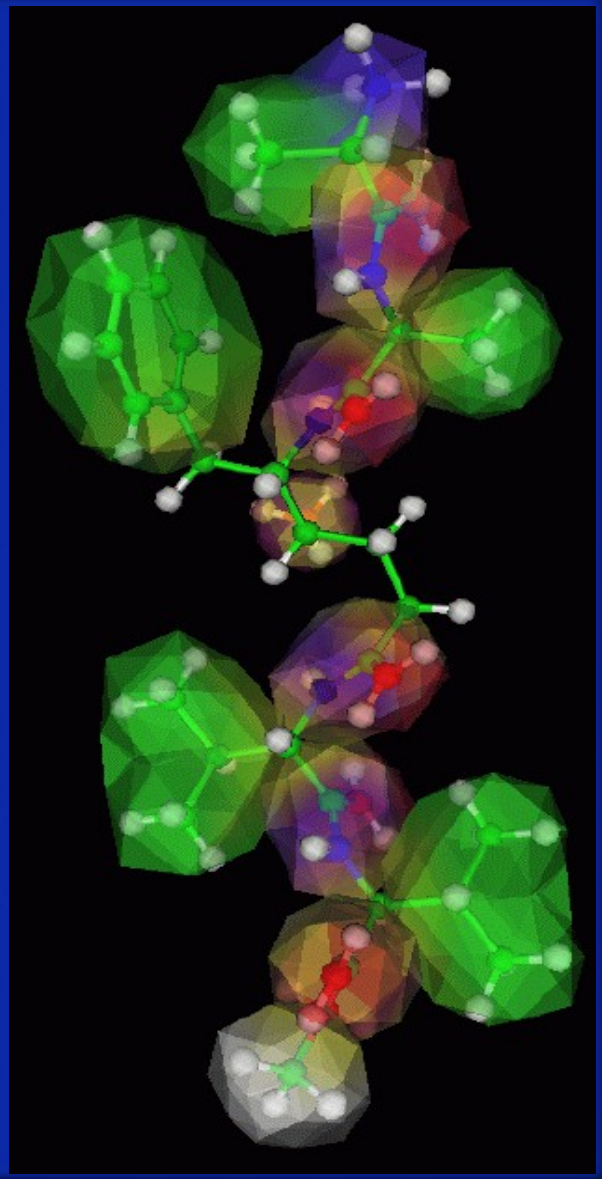
- Fragments are joined pairwise iterations
- Bond lengths (including join) and angles are maintained
- 3D transformations are applied to the fragments to align the join bonds without distorting fragment conformations
- Both fragments are transformed, the amount of movement is inversely proportional to the size, i.e. the smaller one moves more



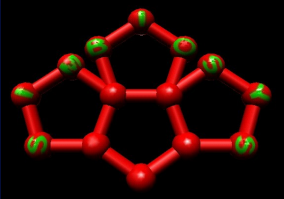
10. Local energy minimization



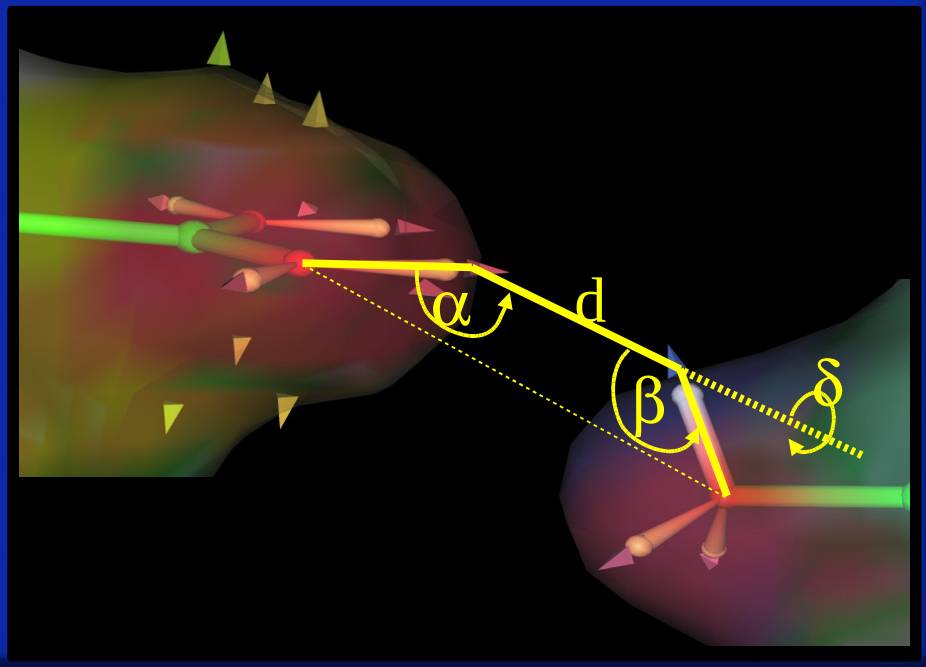
- Modified Powell algorithm
- Rigid body rotations and translations about 3 axes
- Dihedral angles of rotatable single bonds
- Bond lengths and angles are preserved
- Goal function includes:
  - receptor-ligand interaction score
  - conformation strain
  - internal ligand-ligand interaction score
  - de-solvation score (both ligand and receptor)



11. Surface point based statistical scoring function

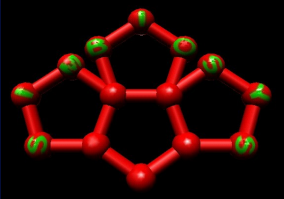


- Based on Interacting Surface Points (ISP) instead of atom centers
- De-solvation: continuous model, ISP type dependent
- Steric clash penalty: distance-square from Connolly surface
- Pocket depth: signed distance of atoms from convex hull
- Protein family data based coverage (ISP type pairs)
- Ligand strain energy
  (torsional probability, vdw LJ 6-12)
- Ligand intra-molecular interaction score (ISP pair ~ receptor)
- Ligand entropy loss
  (frozen rotatable bonds)

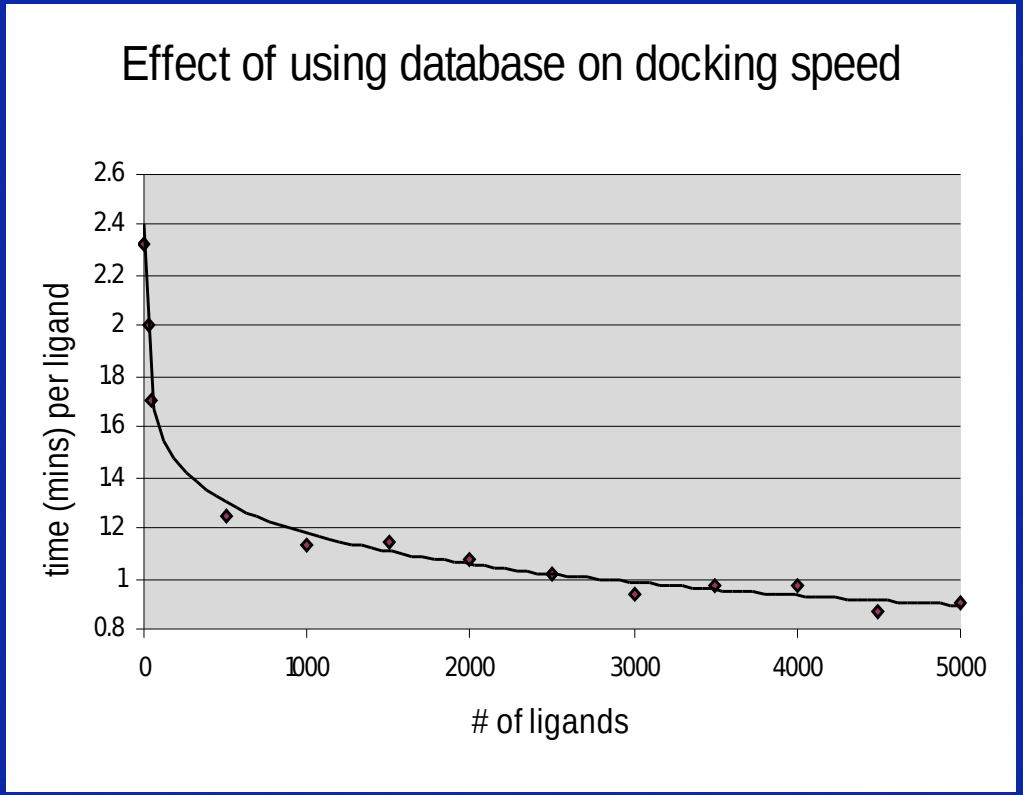


Wed. Mar.25th 10:00, Room 254A

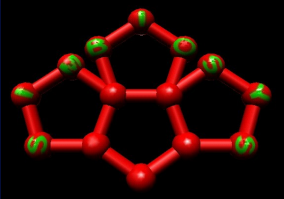
12. Docked fragment pose database

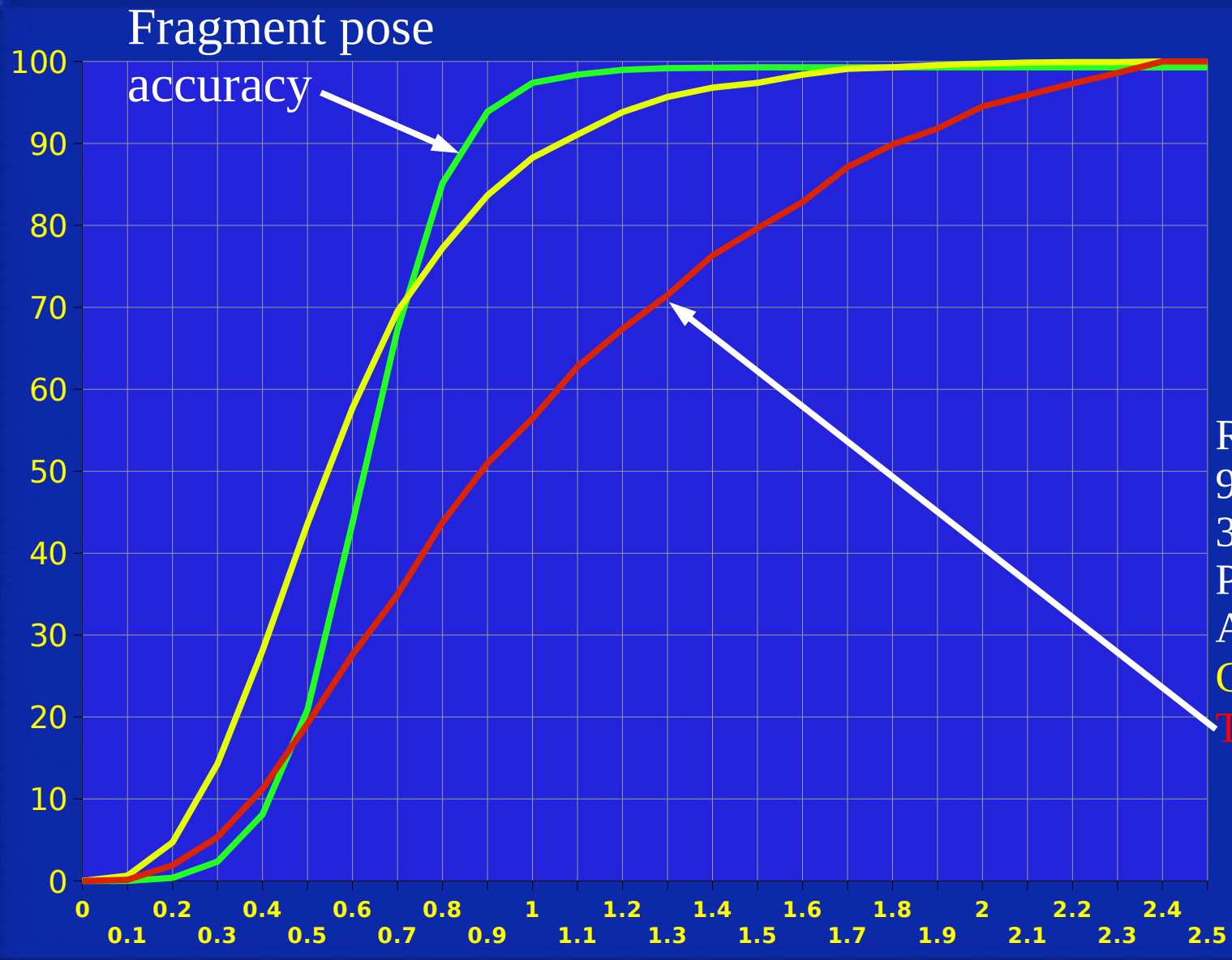


- Significant portion of CPU time spent in RigiDock
- Rigid fragment poses & scores are independent
- Stored in SQL DB and re-used for other ligands
- Knowledge could be used to construct novel ligands by linking fragment poses
- Disk space required:
  - Separate table per receptor
  - ~1MB per rigid fragment
  - Limited to 10K => 10GB



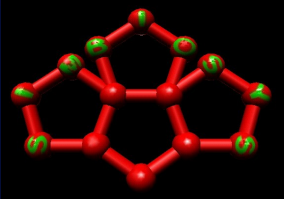
13. RMSD results on 1568 PDB complex with drug-like ligands

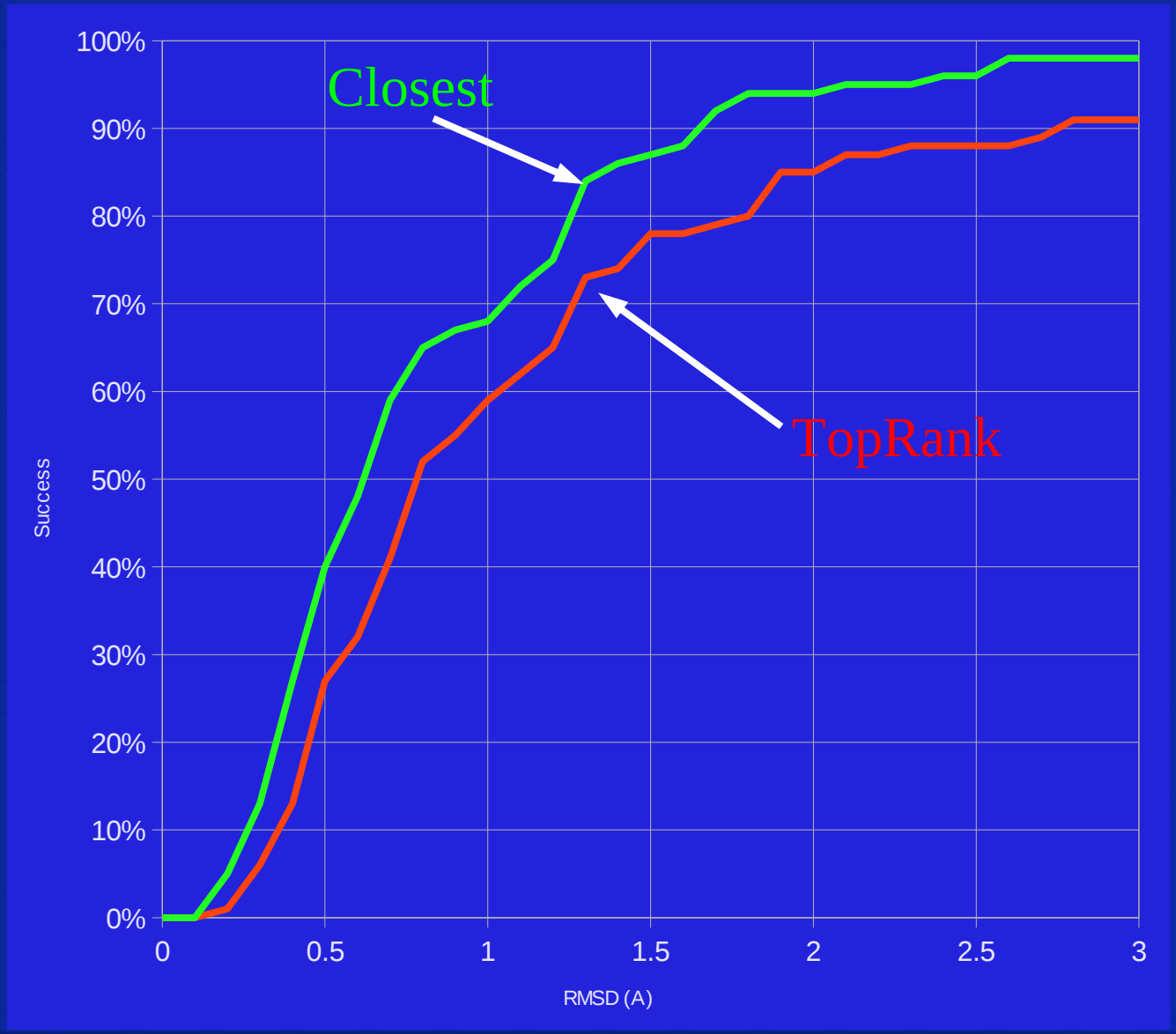




Resolution <= 2.5Å
97 protein families (5+)
349 singletons
PDB-bind 2004
Astex-GOLD validation
Closest average: **0.73Å**
TopRank aver.: **1.10Å**

14. Validation results on the Astex diverse benchmark set (85 cases)





	TopRank	Closest
Average	1.41Å	0.91Å
Median	0.84Å	0.67Å
<1Å	59%	68%
<2Å	85%	94%
<3A	91%	98%

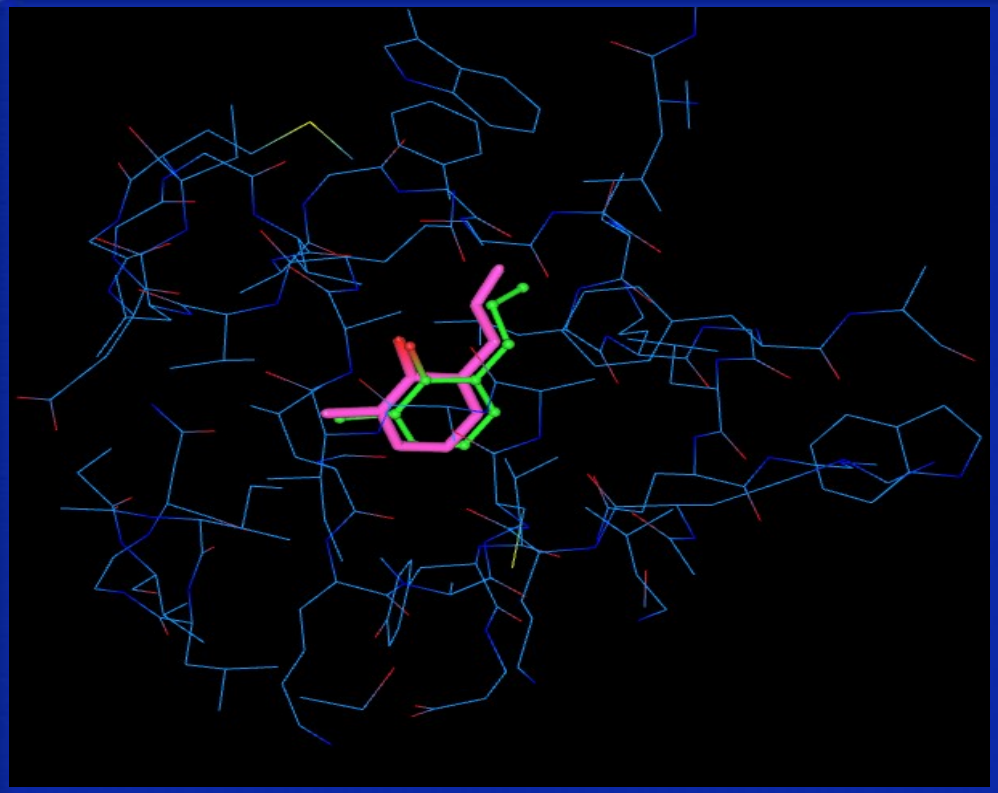
Diverse, High-Quality Test Set for the Validation of Protein-Ligand Docking Performance.

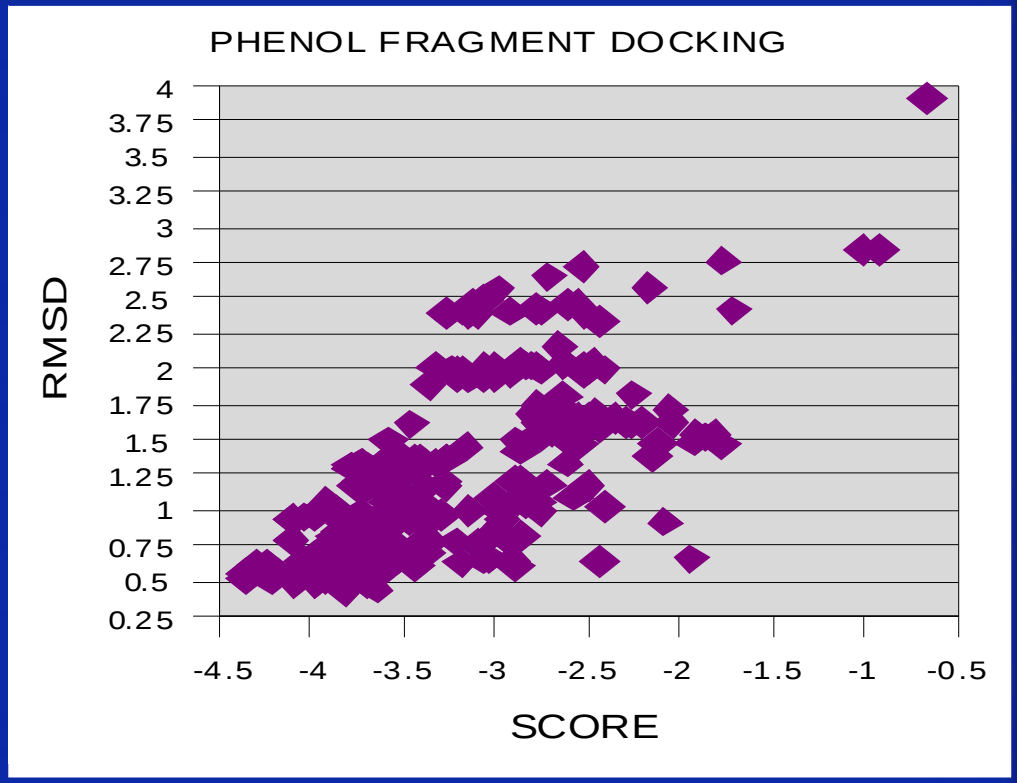
M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson, C. W. Murray

J. Med. Chem., 50, 726-741, 2007.

[DOI:10.1021/jm061277y]

15. Fragment docking case study on β-secretase





Congreve, C.; Aharony, D; Albert, J; Callaghan, O; Campbell, J; Carr, R.A.E.; Chessari, G; Cowan, S; Edwards, P.D.; Frederickson, M; McMenamin, R; Murray, C.W. ; Patel, S; and Wallis, N.

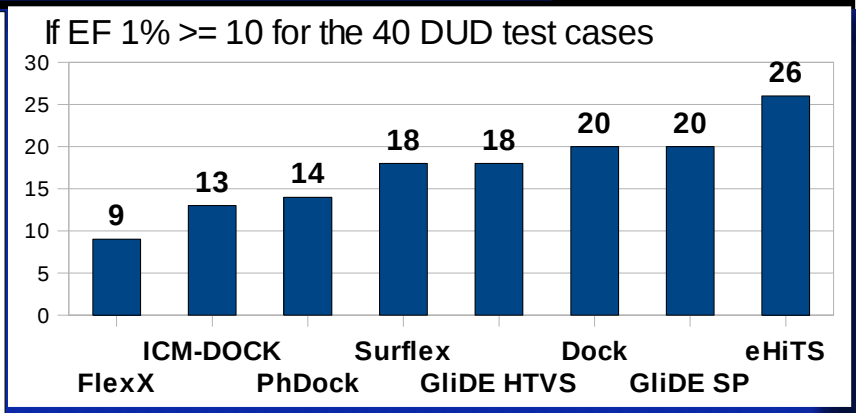
Application of Fragment Screening by X-ray Crystallography to the Discovery of Aminopyridines as Inhibitors of β-Secretase.

J. Med. Chem. **2007**, 50, 1124-1132.

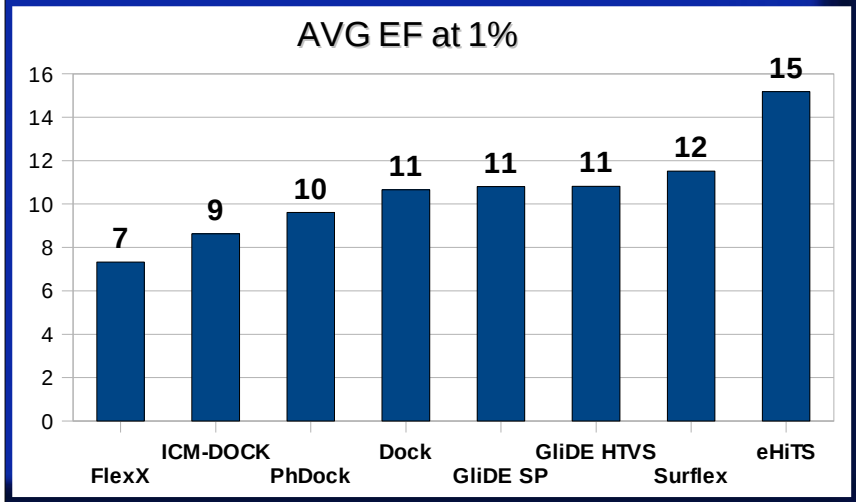
More details: eHiTS application talk by Danni Harris, Thursday Mar.26th 8:30 am, room 258

16. Enrichment results on DUD set (EF at 1% of database)

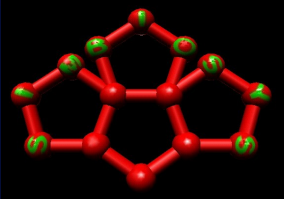
	DOCK	FlexX	GliDE HTVS	GliDE SP	ICM-DOCK	PhDOCK	Surflex	eHiTS
AR	3.9	6.4	18.0	12.8	15.4	24.4	7.7	7.7
Erag	7.6	18.2	19.7	7.6	25.7	22.7	9.1	7.6
Erant	13.6	16.3	1.6	16.3	19.1	10.9	16.3	23.7
GR	3.9	2.6	2.6	3.9	14.2	16.8	16.8	5.19
MR	0.0	7.2	7.2	21.7	36.2	43.5	36.2	14.29
PARP	30.7	30.5	6.1	15.4	5.9	0.0	21.3	29.41
PR	4.0	0.0	0.0	0.0	27.7	7.9	4.0	23.08
RXR	33.0	5.5	33.0	5.5	33.0	33.0	11.0	26.32
CDK2	11.4	8.5	14.2	8.5	0.0	9.9	11.4	14.1
EGFr	22.2	12.5	19.7	16.5	0.6	3.4	4.9	6.8
FGFr1	15.2	0.9	0.9	3.4	11.9	3.4	4.2	2.5
HSH90	5.5	0.0	0.0	0.0	0.0	2.8	2.8	8.3
P38	12.0	2.0	1.3	2.9	4.2	2.0	5.6	18.32
PDGFrb	10.7	5.9	0.0	5.9	3.6	3.0	1.2	14.79
SRC	24.8	7.0	14.6	23.6	13.4	5.7	6.4	27.9
TK	0.0	0.0	4.7	0.0	0.0	23.4	9.3	0.0
VEGFr2	17.6	9.4	16.4	9.4	8.2	3.5	10.6	4.6
FXA	13.2	24.4	15.3	20.2	5.6	10.4	17.4	6.9
Thrombin	9.8	8.4	19.7	14.1	1.4	14.1	2.8	32.39
Trypsin	12.4	0.0	18.5	20.9	6.2	0.0	8.2	10.42
ACE	12.6	8.4	4.2	12.6	2.1	0.0	6.3	12.5
ADA	0.0	0.0	8.3	16.6	2.8	0.0	11.0	18.4
COMT	21.9	0.0	21.9	11.0	0.0	11.0	0.0	10.0
PDE5	15.3	1.2	7.1	3.5	8.2	0.0	11.8	22.99
DHFR	8.4	16.8	8.1	14.5	17.8	2.2	18.2	17.6
GART	2.6	5.2	15.5	20.6	0.0	0.0	10.3	0.0
Ache	3.7	1.9	3.7	0.9	0.0	2.8	2.8	9.4
ALR2	27.5	0.0	3.9	7.9	7.9	3.9	19.6	28.0
AmpC	14.4	0.0	0.0	4.8	0.0	0.0	0.0	5.0
COX-1	0.0	12.5	12.5	8.3	16.6	25.0	4.2	0.0
CoX-2	0.7	1.7	29.5	29.4	3.3	1.2	16.5	24.7
GPB	0.0	0.0	21.2	3.9	1.9	5.8	9.7	29.4
HIVPR	6.5	0.0	4.8	14.5	6.5	0.0	8.1	9.84
HIVRT	2.4	0.0	14.5	12.1	14.5	4.8	14.5	14.29
HMGA	17.4	14.5	23.2	23.2	0.0	37.7	40.6	26.47
InhA	24.8	27.2	4.7	13.0	1.2	1.2	4.7	12.9
NA	4.1	4.1	14.5	0.0	26.9	22.8	26.9	10.4
PARP	12.2	30.5	6.1	6.1	3.1	0.0	21.3	27.38
PNP	0.0	0.0	8.8	4.4	0.0	8.8	17.5	12.24
SAHH	0.0	3.3	6.5	16.2	0.0	16.2	9.7	31.25
AVG	**10.65**	**7.33**	**10.81**	**10.8**	**8.63**	**9.61**	**11.52**	**15.18**



235th ACS New Orleans, COMP 144: "Comparison of pose generation and virtual screening accuracy for several molecular docking programs", Jason B. Cross, David C. Thompson, Brajesh K. Rai, J. Christian Baber, Kristi Yi Fan, Yongbo Hu, and Christine Humblet. Wyeth



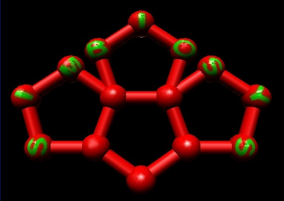
17. Summary of Features



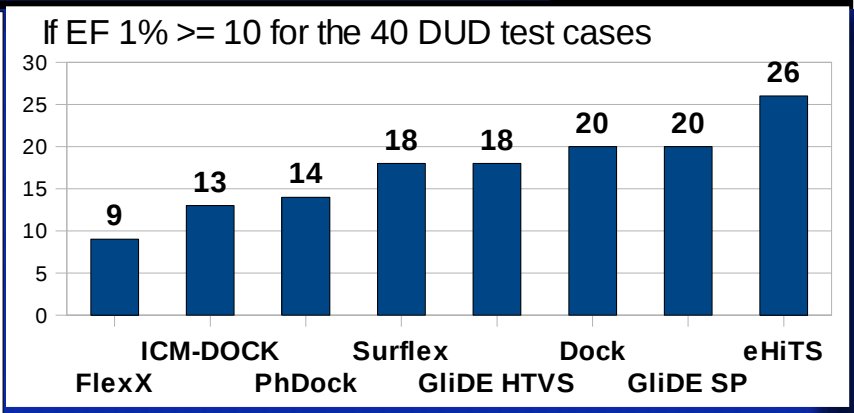
- Exhaustive, highest accuracy - don't miss a potential drug!
- Deterministic – no need to wait for a lucky day
- Fully automated - no manual setup, protonation, charges
- Customized scoring function with automatic training tool
- Integrated VHTS filter (thousand ligands / minute, great enrichment)
- Very fast – electronic High Throughput Screening with parallel/cluster support: SMP, linux clusters, PBS, LSF, SGE
  eHiTS Lightning on the Cell/BE processor 30X faster!

- Built-in option to pre-filter ligand libraries for specified fragments
- Other talks about eHiTS at this ACS meeting:
  - Scoring (Zsolt Zsoldos): Wednesday Mar.25th 10:00 am, room 254A
  - Application (Danni Harris): Thursday Mar.26th 8:30 am, room 258
- Visit us at Booth #316 at the exhibition
  Request a free evaluation online: http://www.simbiosys.ca/

18. Enrichment results on DUD set (EF at 1% of database)



	DOCK	FlexX	GliDE HTVS	GliDE SP	ICM-DOCK	PhDOCK	Surflex	eHiTS
AR	3.9	6.4	18.0	12.8	15.4	24.4	7.7	7.7
Erag	7.6	18.2	19.7	7.6	25.7	22.7	9.1	7.6
Erant	13.6	16.3	1.6	16.3	19.1	10.9	16.3	23.7
GR	3.9	2.6	2.6	3.9	14.2	16.8	16.8	5.19
MR	0.0	7.2	7.2	21.7	36.2	43.5	36.2	14.29
PARP	30.7	30.5	6.1	15.4	5.9	0.0	21.3	29.41
PR	4.0	0.0	0.0	0.0	27.7	7.9	4.0	23.08
RXR	33.0	5.5	33.0	5.5	33.0	33.0	11.0	26.32
CDK2	11.4	8.5	14.2	8.5	0.0	9.9	11.4	14.1
EGFr	22.2	12.5	19.7	16.5	0.6	3.4	4.9	6.8
FGFr1	15.2	0.9	0.9	3.4	11.9	3.4	4.2	2.5
HSH90	5.5	0.0	0.0	0.0	0.0	2.8	2.8	8.3
P38	12.0	2.0	1.3	2.9	4.2	2.0	5.6	18.32
PDGFrb	10.7	5.9	0.0	5.9	3.6	3.0	1.2	14.79
SRC	24.8	7.0	14.6	23.6	13.4	5.7	6.4	27.9
TK	0.0	0.0	4.7	0.0	0.0	23.4	9.3	0.0
VEGFr2	17.6	9.4	16.4	9.4	8.2	3.5	10.6	4.6
FXA	13.2	24.4	15.3	20.2	5.6	10.4	17.4	6.9
Thrombin	9.8	8.4	19.7	14.1	1.4	14.1	2.8	32.39
Trypsin	12.4	0.0	18.5	20.9	6.2	0.0	8.2	10.42
ACE	12.6	8.4	4.2	12.6	2.1	0.0	6.3	12.5
ADA	0.0	0.0	8.3	16.6	2.8	0.0	11.0	18.4
COMT	21.9	0.0	21.9	11.0	0.0	11.0	0.0	10.0
PDE5	15.3	1.2	7.1	3.5	8.2	0.0	11.8	22.99
DHFR	8.4	16.8	8.1	14.5	17.8	2.2	18.2	17.6
GART	2.6	5.2	15.5	20.6	0.0	0.0	10.3	0.0
Ache	3.7	1.9	3.7	0.9	0.0	2.8	2.8	9.4
ALR2	27.5	0.0	3.9	7.9	7.9	3.9	19.6	28.0
AmpC	14.4	0.0	0.0	4.8	0.0	0.0	0.0	5.0
COX-1	0.0	12.5	12.5	8.3	16.6	25.0	4.2	0.0
CoX-2	0.7	1.7	29.5	29.4	3.3	1.2	16.5	24.7
GPB	0.0	0.0	21.2	3.9	1.9	5.8	9.7	29.4
HIVPR	6.5	0.0	4.8	14.5	6.5	0.0	8.1	9.84
HIVRT	2.4	0.0	14.5	12.1	14.5	4.8	14.5	14.29
HMGA	17.4	14.5	23.2	23.2	0.0	37.7	40.6	26.47
InhA	24.8	27.2	4.7	13.0	1.2	1.2	4.7	12.9
NA	4.1	4.1	14.5	0.0	26.9	22.8	26.9	10.4
PARP	12.2	30.5	6.1	6.1	3.1	0.0	21.3	27.38
PNP	0.0	0.0	8.8	4.4	0.0	8.8	17.5	12.24
SAHH	0.0	3.3	6.5	16.2	0.0	16.2	9.7	31.25
AVG	**10.65**	**7.33**	**10.81**	**10.8**	**8.63**	**9.61**	**11.52**	**15.18**



235th ACS New Orleans, COMP 144: "Comparison of pose generation and virtual screening accuracy for several molecular docking programs", Jason B. Cross, David C. Thompson, Brajesh K. Rai, J. Christian Baber, Kristi Yi Fan, Yongbo Hu, and Christine Humblet. Wyeth