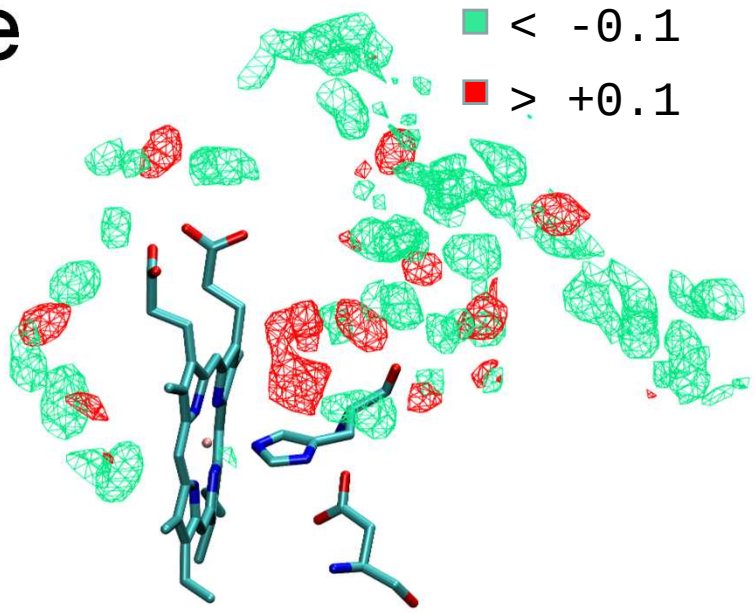
DOCK 3.7: How It Works – and How to Tweak your Setup

Trent E. Balius

Shoichet Lab Group Meeting

2014/10/31

Outline



- DOCK
  - History
  - Differences between 3.6 and 3.7
  - Sampling method
  - Scoring function
  - New Input file
  - New dir structure
  - My work – Vitamin D Receptor; Receptor desolvation
  - Methods of evaluation docking performance

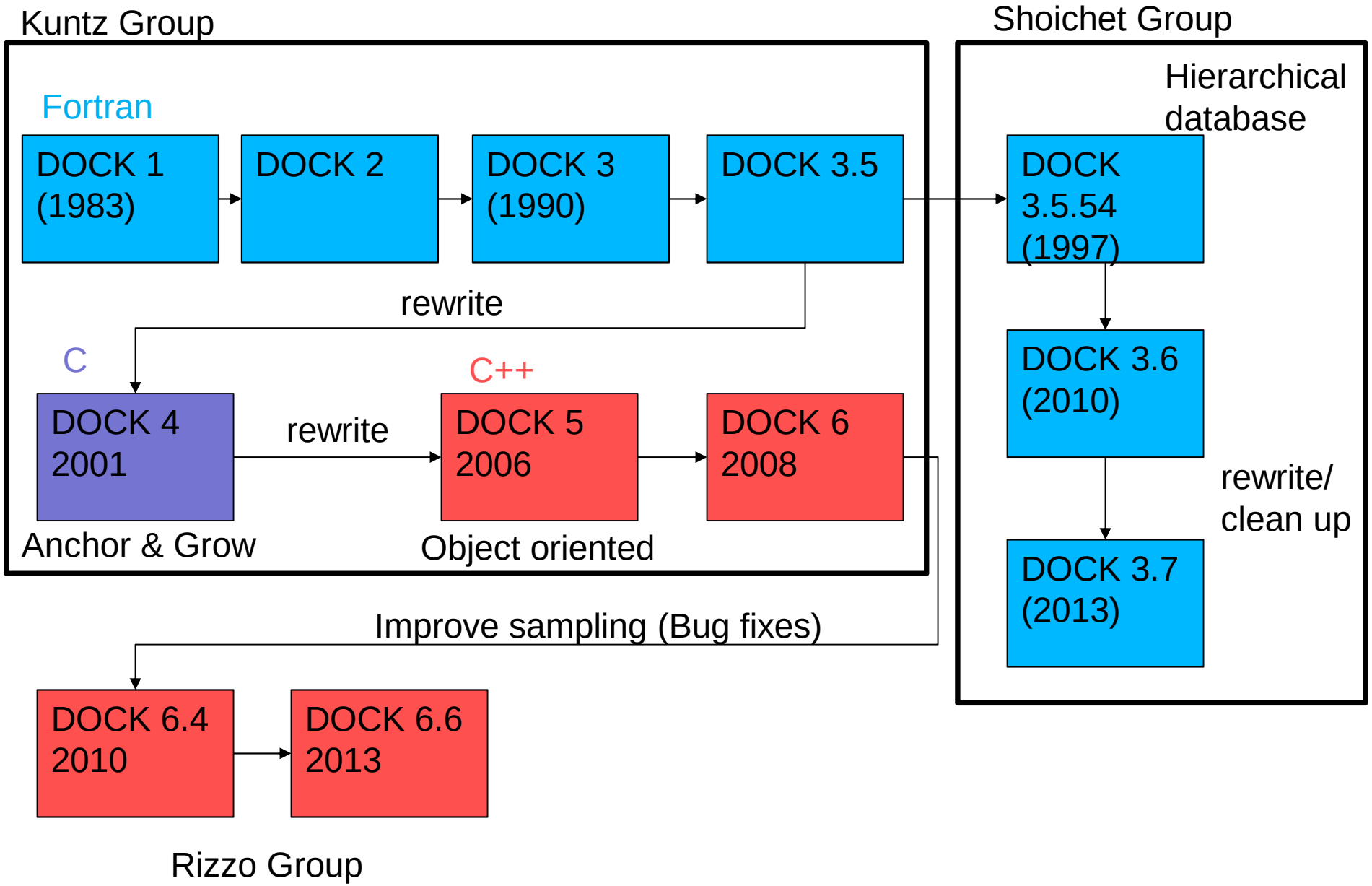
- Tweaking ones docking setups:
  - Make a residue more polar
  - Modify Spheres
  - Rescore the docked poses / molecules

Outline

- DOCK
  - History
  - Differences between 3.6 and 3.7
  - Sampling method
  - Scoring function
  - New Input file
  - New dir structure
  - My work – Vitamin D Receptor; Receptor desolvation
  - Methods of evaluation docking performance

- Tweaking ones docking setups:
  - Make a residue more polar
  - Modify Spheres
  - Rescore the docked poses / molecules

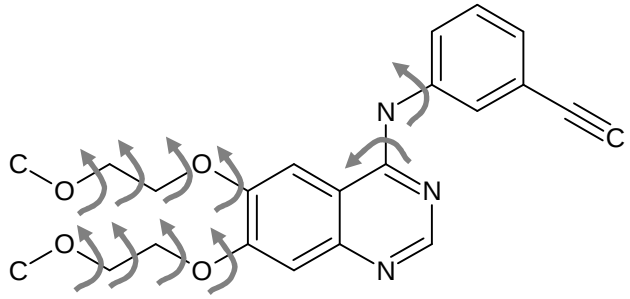
DOCK: A History

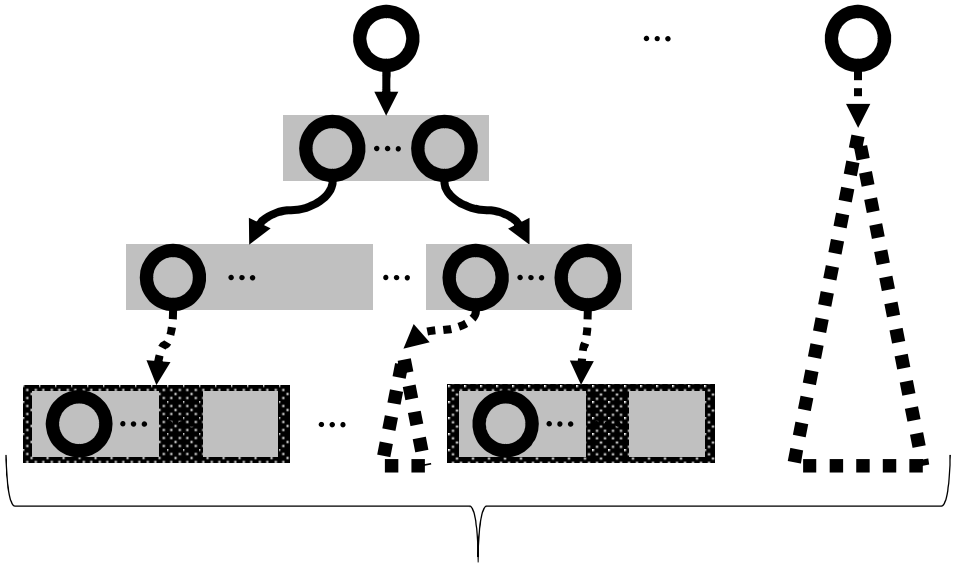


How DOCK 3.6 and 3.7 works

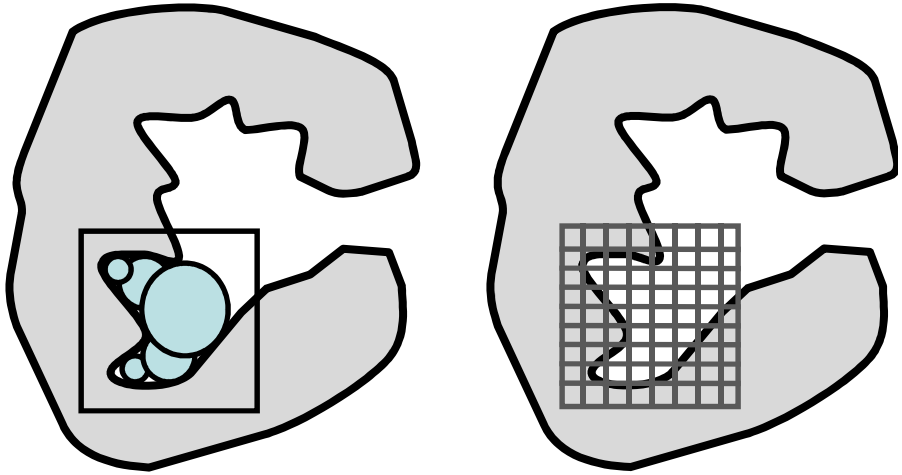
Preparation, Sampling, and Scoring

Ligand Sampling –
database construction





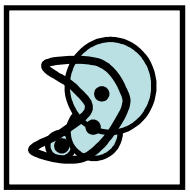
Dockable database file

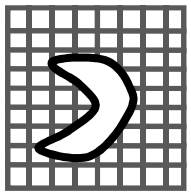


DOCKING

orient

score





Scoring using a grid to speed up
the calculations

DOCK 3.7 is different from, better than? 3.6

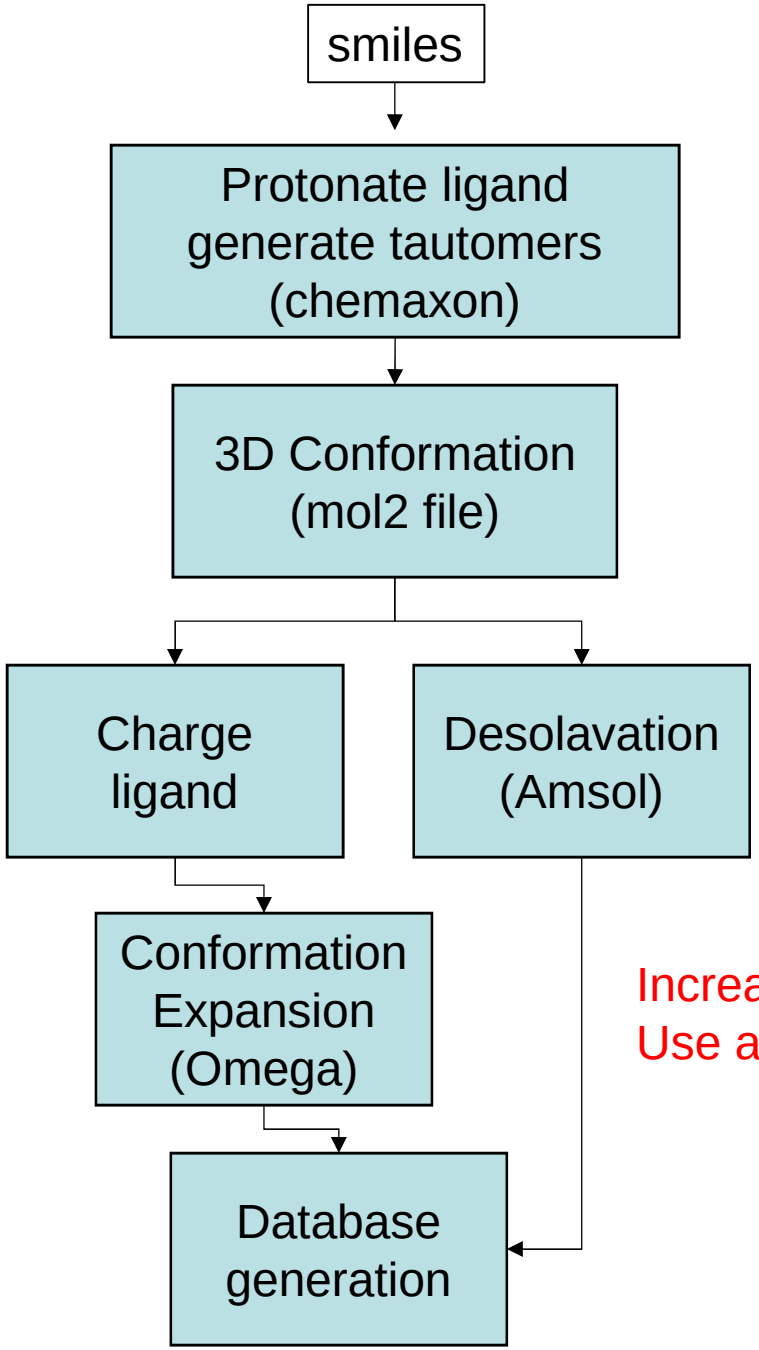
- Ligand Internal Sampling
  - DB format → DB2 format
  - There are far fewer Broken Molecules
    - DOCK 3.6 would combine incompatible branches
- Ligand Orienting
  - Bipartite graph matching – this method differs between version.
  - Minimizer was removed.
- Setup modifications – external program changes
  - DELPHI → QNNIFT
  - Schrodinger → Reduce
  - Epic → Marvin
  - Makefile → blastermaster.py
- Directories restructured
- Now outputs mol2 files

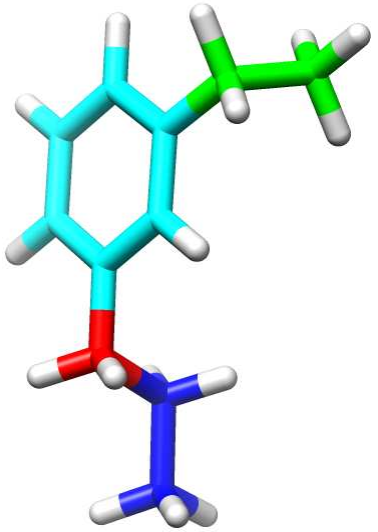
Outline

- DOCK
  - History
  - Differences between 3.6 and 3.7
  - Sampling method
  - Scoring function
  - New Input file
  - New dir structure
  - My work – Vitamin D Receptor; Receptor desolvation
  - Methods of evaluation docking performance

- Tweaking ones docking setups:
  - Make a residue more polar
  - Modify Spheres
  - Rescore the docked poses / molecules

Ligand Preparation

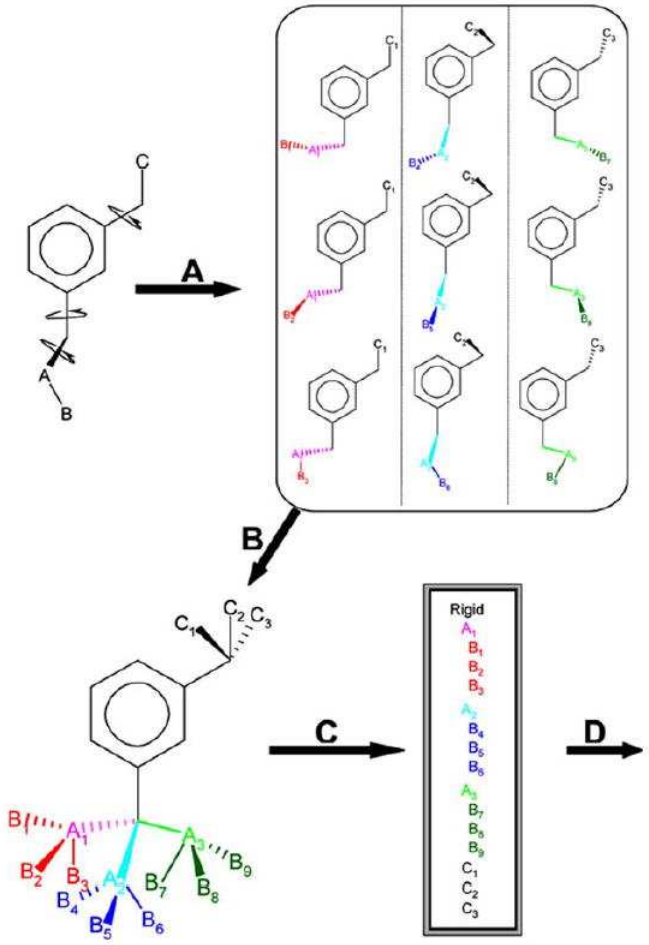




Increase sampling (turn nobles in omega)
Use alternative sampling engines.

How Sampling Works (Internal)

Hierarchical docking of databases

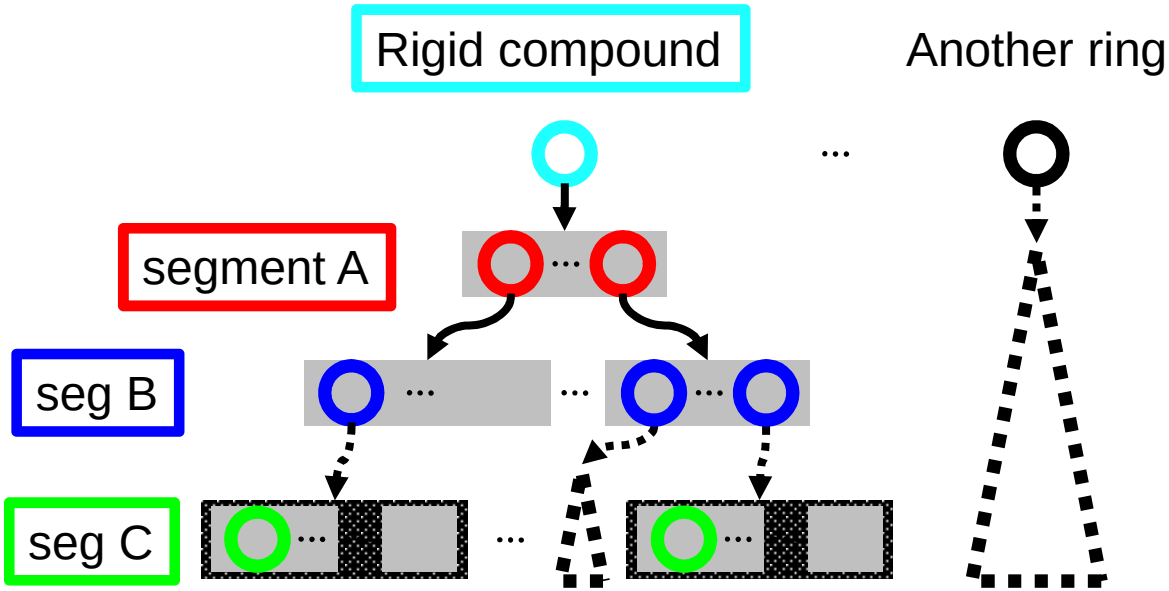


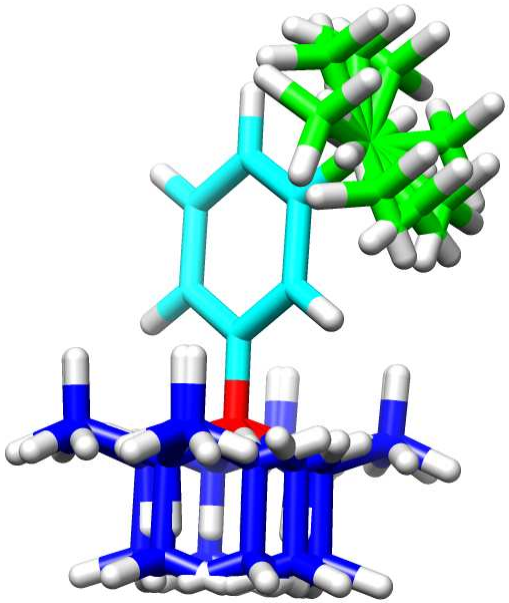
conf #	Atom A	B	C
1	1	1	1
2	1	2	1
3	1	3	1
4	2	4	1
5	2	5	1
6	2	6	1
7	3	7	1
8	3	8	1
9	3	9	1
10	1	1	2
11	1	2	2
12	1	3	2
13	2	4	2
14	2	5	2
15	2	6	2
16	3	7	2
17	3	8	2
18	3	9	2
19	1	1	3
20	1	2	3
21	1	3	3
22	2	4	3
23	2	5	3
24	2	6	3
25	3	7	3
26	3	8	3
27	3	9	3

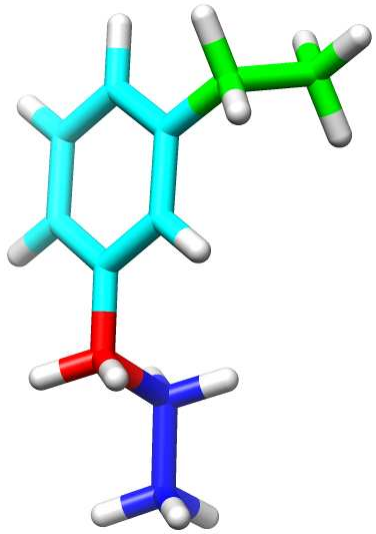
- DOCK 3.5.54 through DOCK 3.6
- Highly compact format
- DOCK 3.7 format has changed.
  - Less compacted
  - Solves some problems with the old format/tools
  - Connectivity is represented

Lorber DM and Shoichet BK., Curr Top Med Chem. 2005;5(8):739-49.

Ligand Sampling – Database Construction





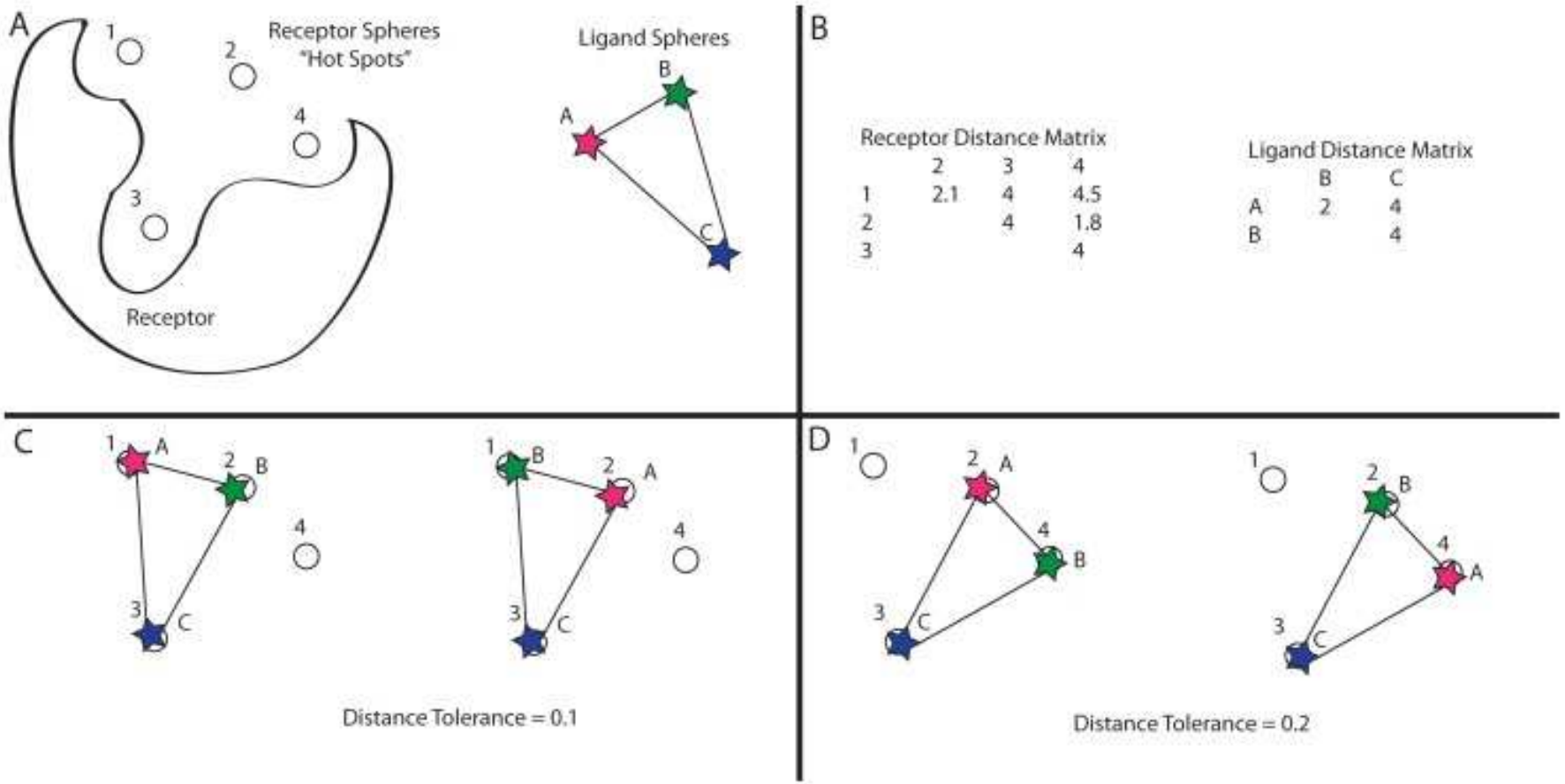


Fully Grown Conformers

Branch
(conformer)

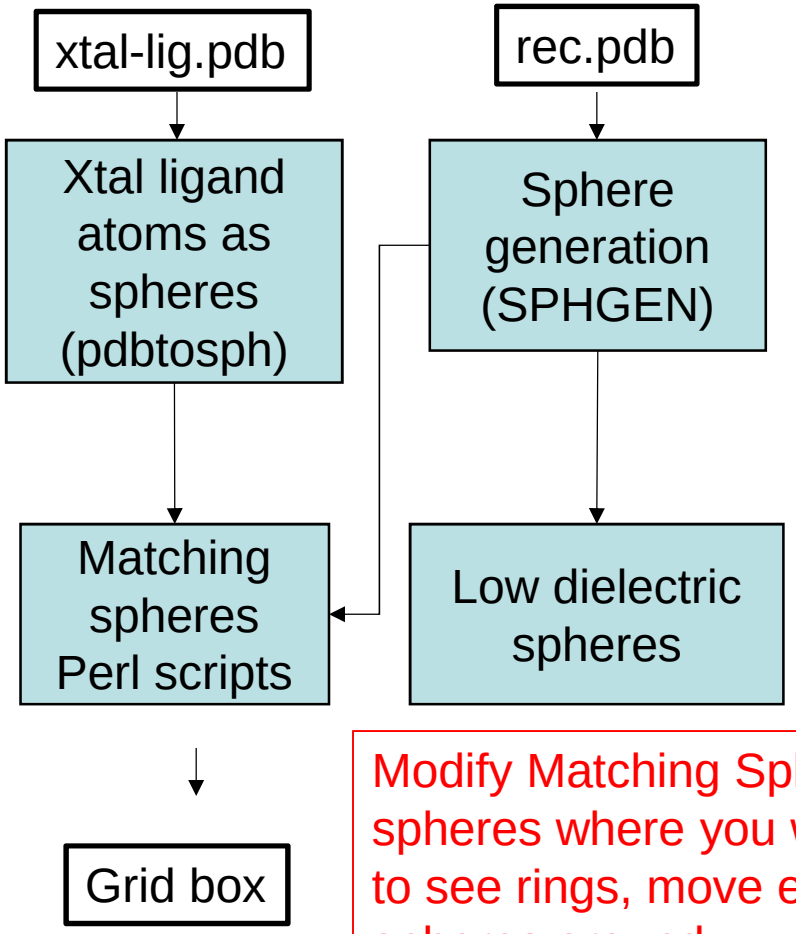
How Sampling Works (Orientational)

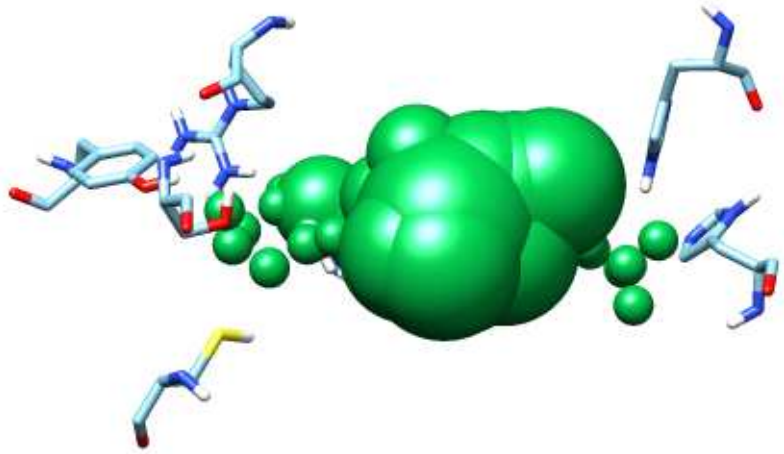
A toy example illustrating the matching sphere orientational matching algorithm



Coleman, RG et al. PLoS One. 2013; 8(10): e75992.

Sphere Generation





Modify Matching Spheres: add spheres where you would like to see rings, move existing spheres around

Add (or remove) Low dielectric Spheres To regions where you would like to have less (more) screening

Outline

- DOCK
  - History
  - Differences between 3.6 and 3.7
  - Sampling method
  - Scoring function
  - New Input file
  - New dir structure
  - My work – Vitamin D Receptor; Receptor desolvation
  - Methods of evaluation docking performance

- Tweaking ones docking setups:
  - Make a residue more polar
  - Modify Spheres
  - Rescore the docked poses / molecules

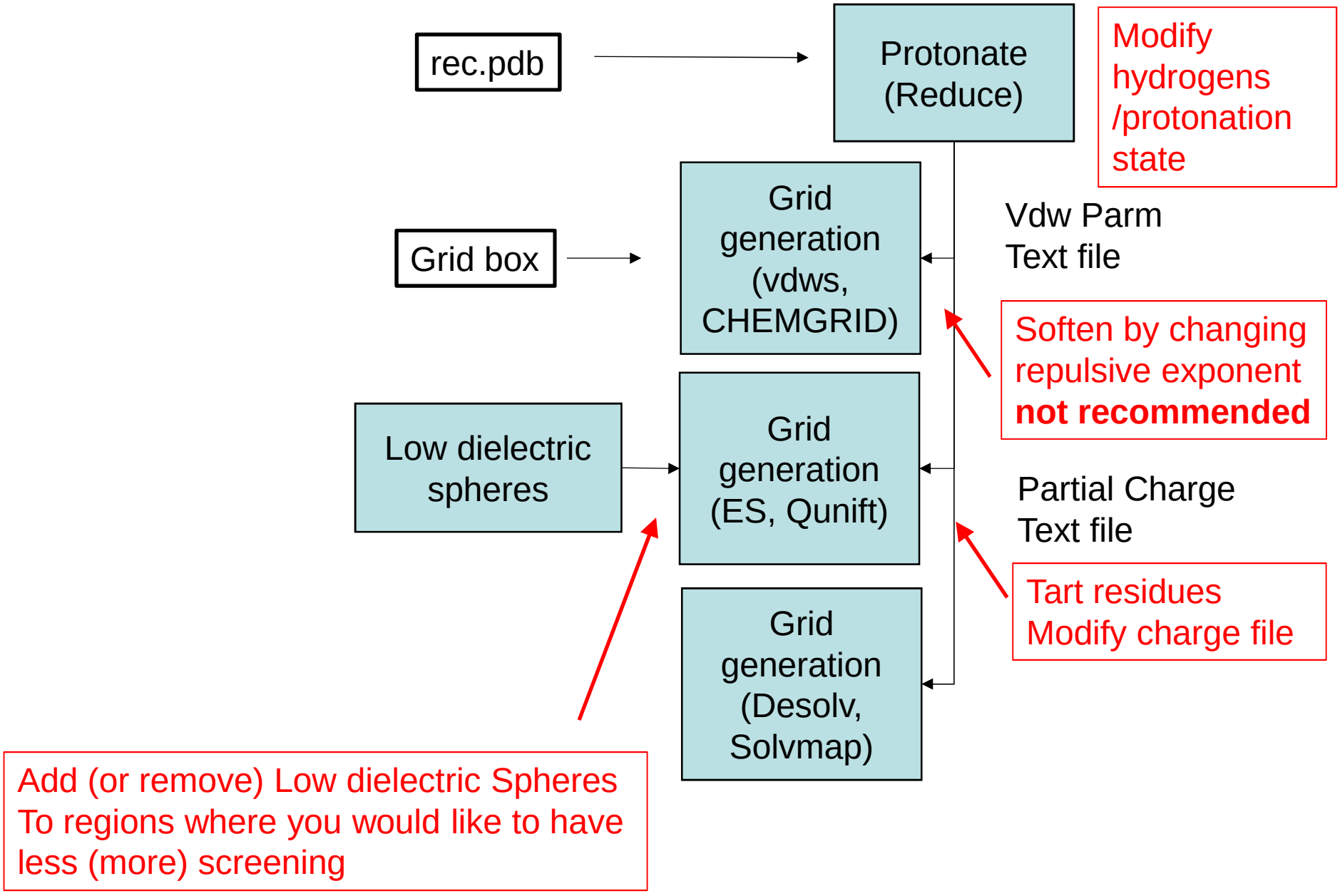
Current Scoring Function (DOCK 3.6 and 3.7)

$$E_{score} = E_{VDW} + E_{ES} + E_{lig,desol}$$

- VDW term is based on the AMBER united-atom force field
- Electrostatics term
  - PB calculation using DELPHI or QNIFFT
  - Binding site has low dielectric by including spheres.
- Ligand Desolvation
  - desolvation grid value times by the polar and nonpolar terms in ligand file
  - General Born approximation
- What's Missing ?

Meng, et al J. Comput. Chem. 1992, 13, 505– 524
Mysinger and Shoichet J Chem Inf Model. 2010, 50(9):1561-73

Receptor preparation



Blastermaster.py vs Makefile

- Both are to prepare a pdb / protein ligand pair for docking
- All work done in working directory
- All docking files are copied to dockingfiles
- Blastermaster_mod.py allows one to tweak receptor file for protonation state, tarting, etc.:

    --addNOhydrogensflag

DOCK Input File

```
DOCK 3.7 parameter
#######################################################
# NOTE: split_database_index is reserved to specify a list of files
ligand_atom_file               split_database_index
#######################################################
#                               OUTPUT
output_file_prefix             test.
#######################################################
#                               MATCHING
match_method                   2
distance_tolerance             0.05
match_goal                     5000
distance_step                  0.05
distance_maximum               0.5
timeout                        10.0
nodes_maximum                  4
nodes_minimum                  4
bump_maximum                   50.0
bump_rigid                     50.0
#######################################################
#                               COLORING
chemical_matching              no
case_sensitive                 no
#######################################################
#                               SEARCH MODE
atom_minimum                   4
atom_maximum                   100
number_save                    1
molecules_maximum              100000
check_clashes                  yes
do_premax                      no
do_clusters                    no
```

Increase for more sampling

Max is 4. To go higher you have to change the code.

Allowing clashes can help sampling

DOCK Input File (continued)

```
#####################################################
#                               SCORING
ligand_desolvation              volume
vdw_maximum                     1.0e10
electrostatic_scale             1.0
vdw_scale                       1.0
internal_scale                  0.0
#####################################################
# INPUT FILES / THINGS THAT CHANGE
receptor_sphere_file            ../dockfiles/matching_spheres.sph
vdw_parameter_file              ../dockfiles/vdw.parms.amb.mindock
delphi_nsize                    81
flexible_receptor               no
total_receptors                 1
############## grids/data for one receptor
rec_number                      1
rec_group                       1
rec_group_option                1
solvmap_file                    ../dockfiles/ligand.desolv.heavy
hydrogen_solvmap_file           ../dockfiles/ligand.desolv.hydrogen
delphi_file                     ../dockfiles/trim.electrostatics.phi
chemgrid_file                   ../dockfiles/vdw.vdw
bumpmap_file                    ../dockfiles/vdw.bmp
############## end of INDOCK
```

Delphi_nsize is electrostatic grid size

Code Reorganization

- We have moved from SVN to Github repository.
- Directory structure is very different.
- Far easier to install, test and use.

Things are structured differently in Github

```
ls -l ~/zzz.github/DOCK/
total 32
-rw-r--r--.  1 tbalius bks 2737 Aug 28 14:09 README.md
drwxr-xr-x.  2 tbalius bks 4096 Aug 27 11:22 analysis
drwxr-xr-x.  2 tbalius bks 4096 Aug 28 14:50 common
drwxr-xr-x.  7 tbalius bks 4096 Sep 11 16:53 docking
drwxr-xr-x.  3 tbalius bks 4096 Aug 27 11:22 install
drwxr-xr-x.  9 tbalius bks 4096 Aug 28 14:09 ligand
drwxr-xr-x. 19 tbalius bks 4096 Aug 27 11:22 proteins
drwxr-xr-x.  3 tbalius bks 4096 Aug 27 11:23 test
```

Work done by Teague and Ryan.

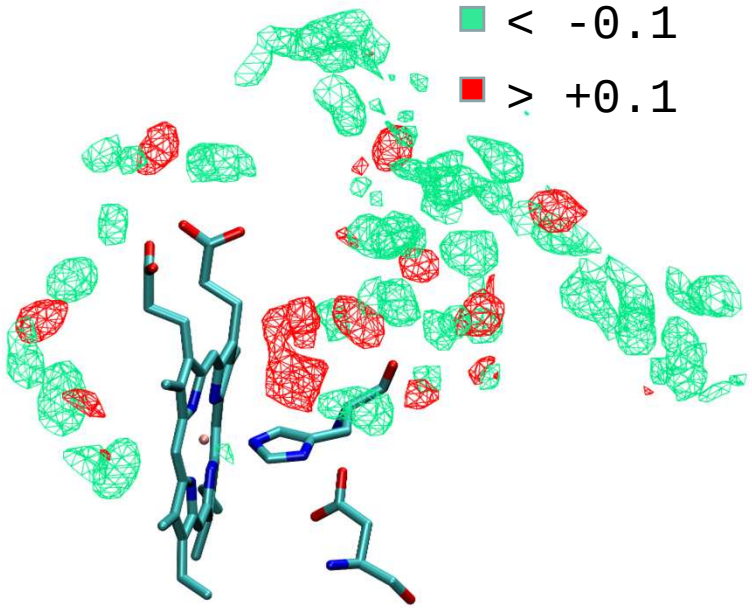
Outline

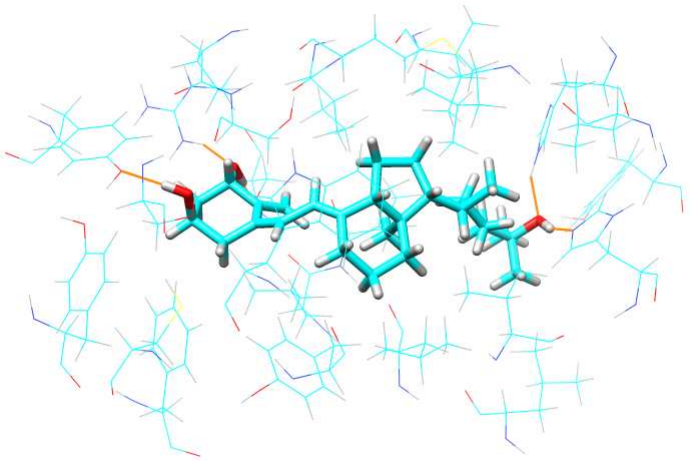
- DOCK
  - History
  - Differences between 3.6 and 3.7
  - Sampling method
  - Scoring function
  - New Input file
  - New dir structure
  - My work – Vitamin D Receptor; Receptor desolvation
  - Methods of evaluation docking performance

- Tweaking ones docking setups:
  - Make a residue more polar
  - Modify Spheres
  - Rescore the docked poses / molecules

My Projects as Illustrations

- Water Project – Incorporating GIST into the DOCK scoring functions
- Vitamin D Receptor – Finding new ligands

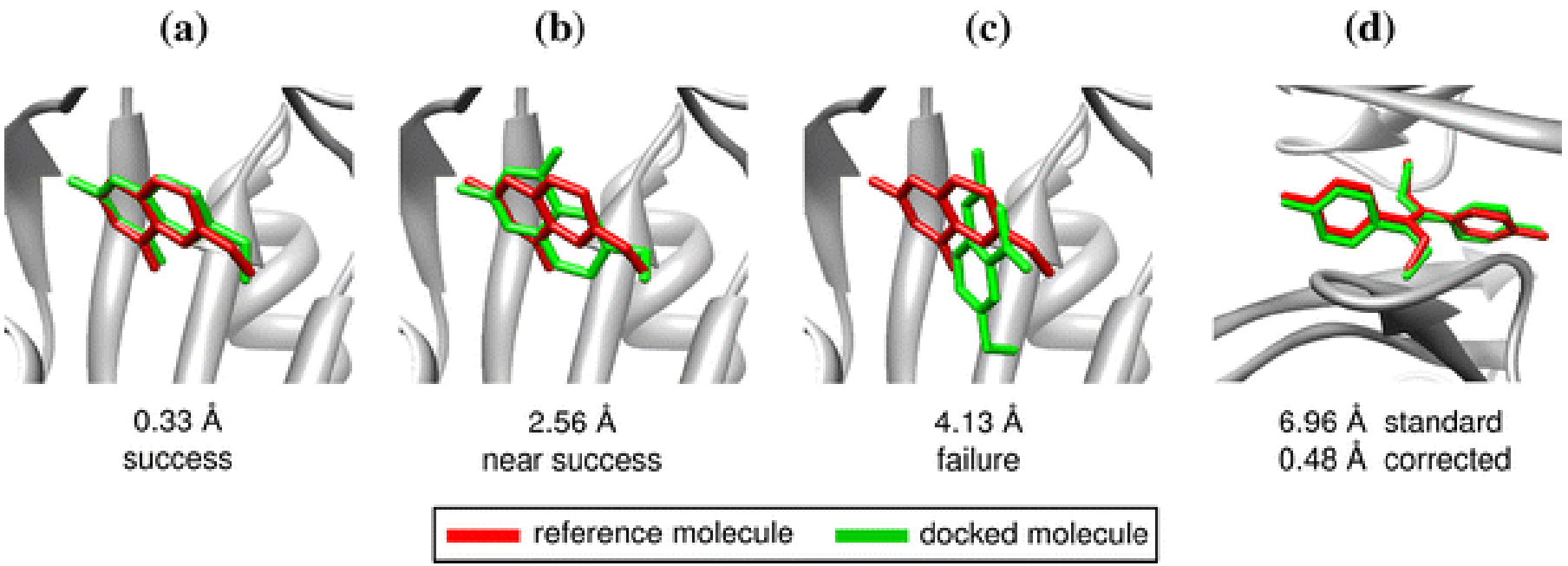




Evaluation Methods

- Pose Reproduction (cognate docking, cross docking)
- Enrichment calculations
- Look at fragment screens and see if things seem interesting
  - Look for promising / expect interaction
  - If not, back to the drawing board
  - This is subjective
- Prospective testing of predictions

Pose Reproduction: RMSD Calculations



We can correct for molecular symmetry using the Hungarian algorithm

It is important that we are obtaining the correct binding mode.

We can calculate RMSDs for DOCK 3.7 runs use DOCK 6.6

Use RMSD Calc. to Compare Scoring

lig pdb	rec conf	standard	gist_Esw_plus_Eww_ref2
4JM6	CcP_A	2.1668	1.5725
4JMA	CcP-A	1.9497	1.9497
4JMS	CcP-A	1.5665	1.5665
4JMT	CcP-A	0.7843	3.2321
4JMV	CcP-A	0.3990	0.3990
4JMW	CcP-A	1.4758	0.7236
4JMZ	CcP-A	1.3876	1.3876
4JN0	CcP-A	1.0397	1.0397
4JPT	CcP-A	2.6364	0.7402
4JQK	CcP-A	1.4696	1.4696
4JQM	CcP-A	4.0098	0.3250
4JQN	CcP-A	3.0806	3.0867

Four cases, RMSDs get better

One case, RMSD gets worse

Seven cases, RMSDs stay the same

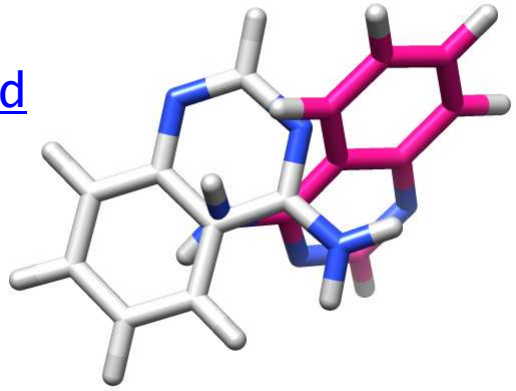
Use RMSD Calc. to Compare Scoring

CcP conformation A

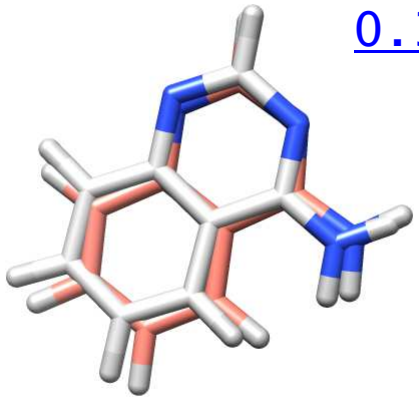
Standard

With GIST

4JQM ligand

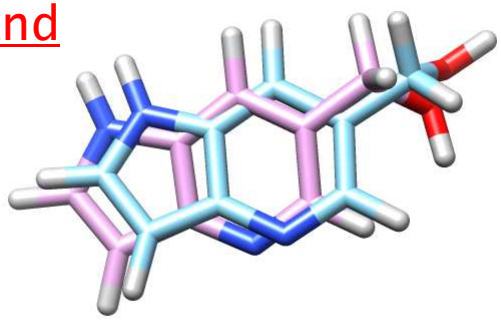


4.0098

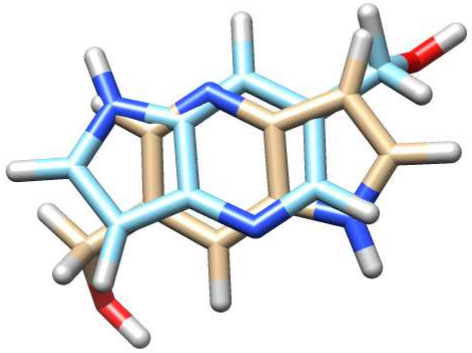


0.3250

4JMT ligand

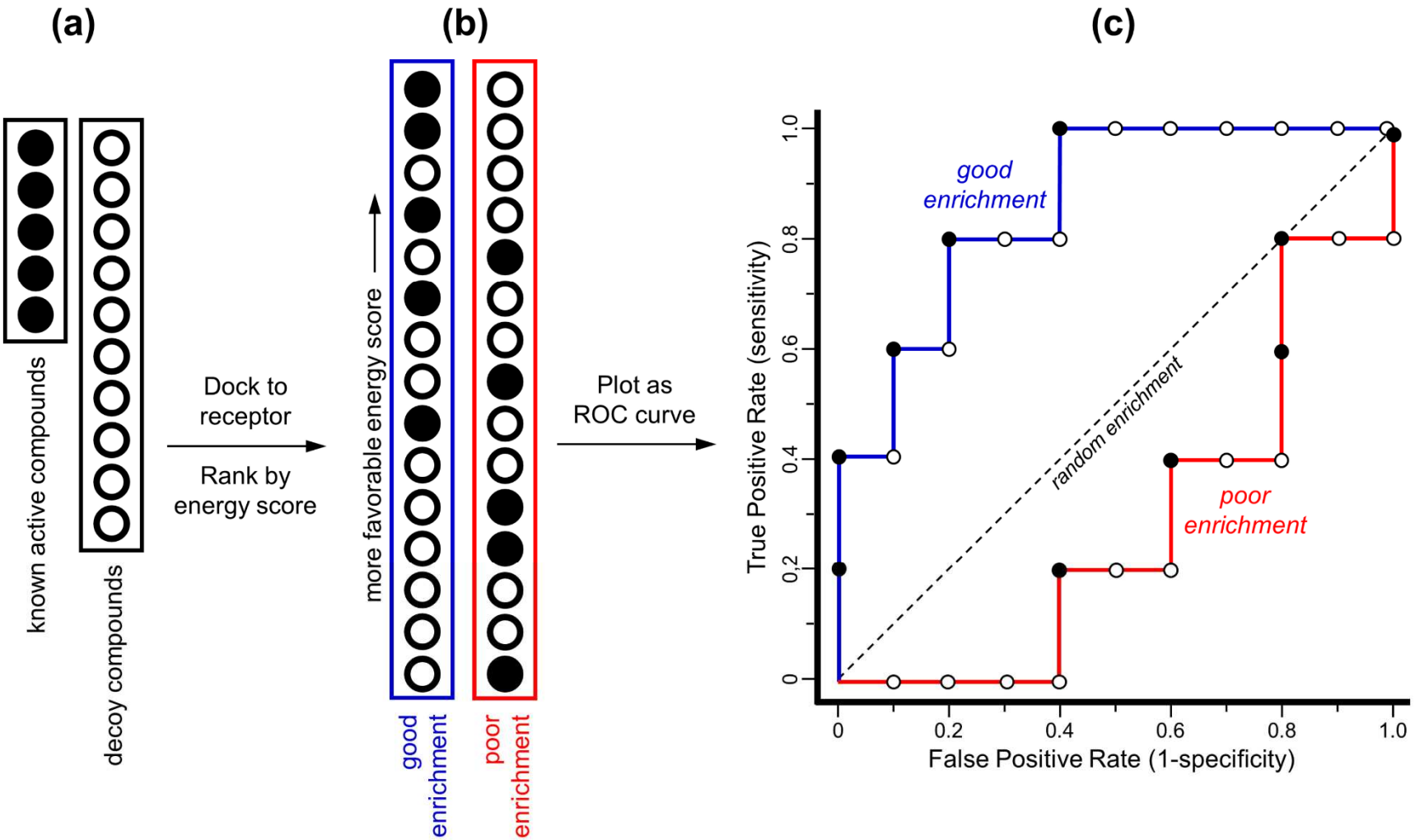


0.7843



3.2321

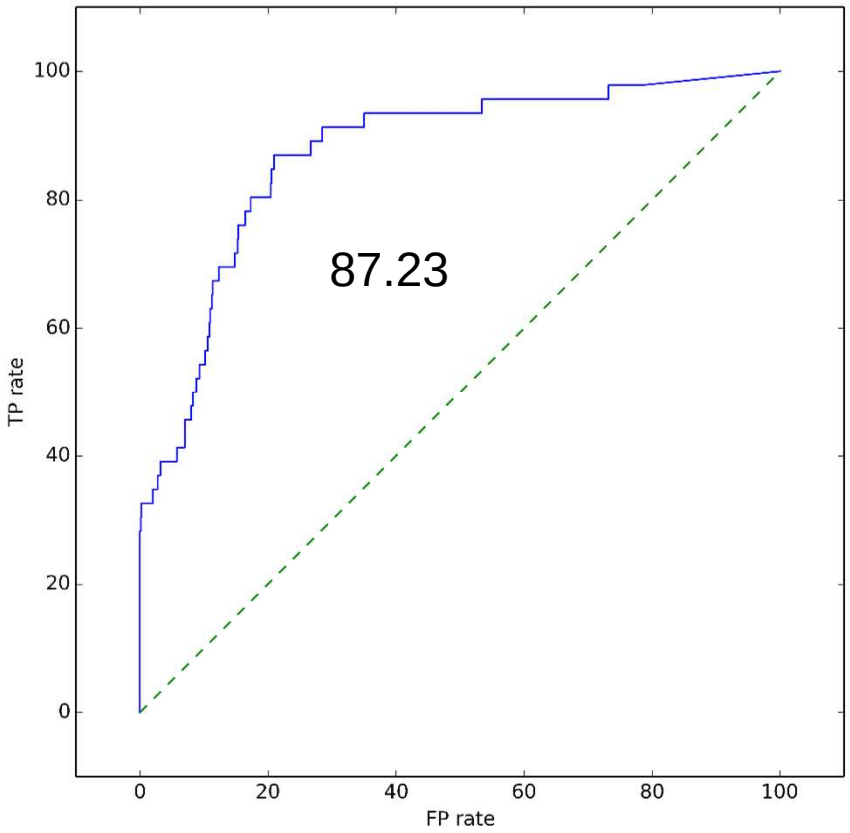
Enrichments: ROC Cuves



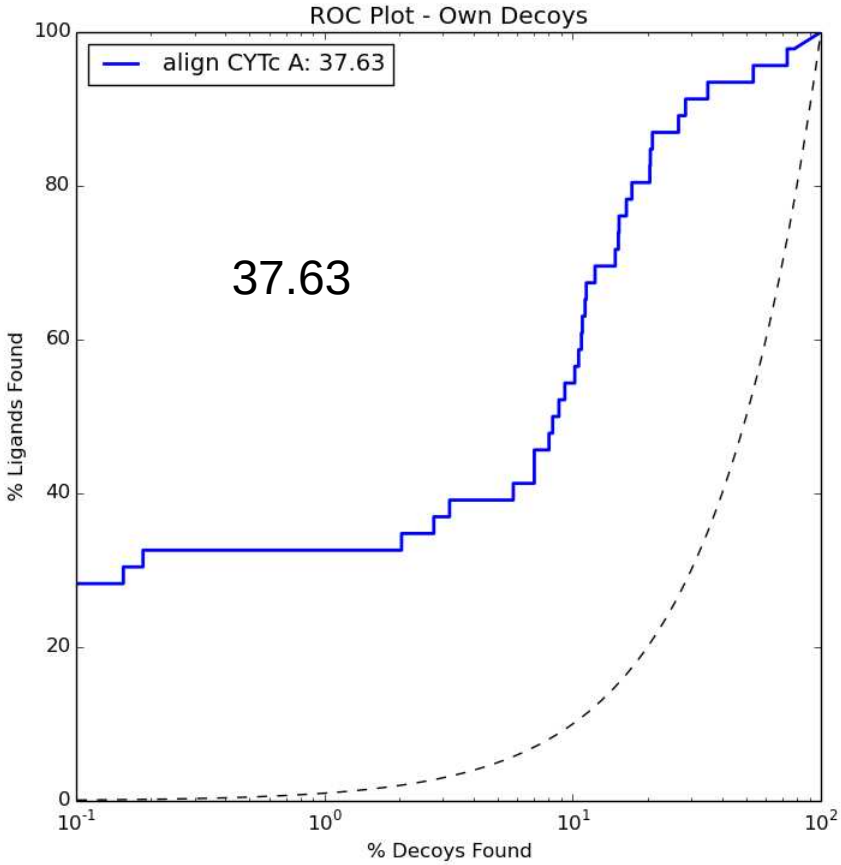
Log-adjusted AUC: Early Enrichment Weighted More

CcP conformation A

ROC curve

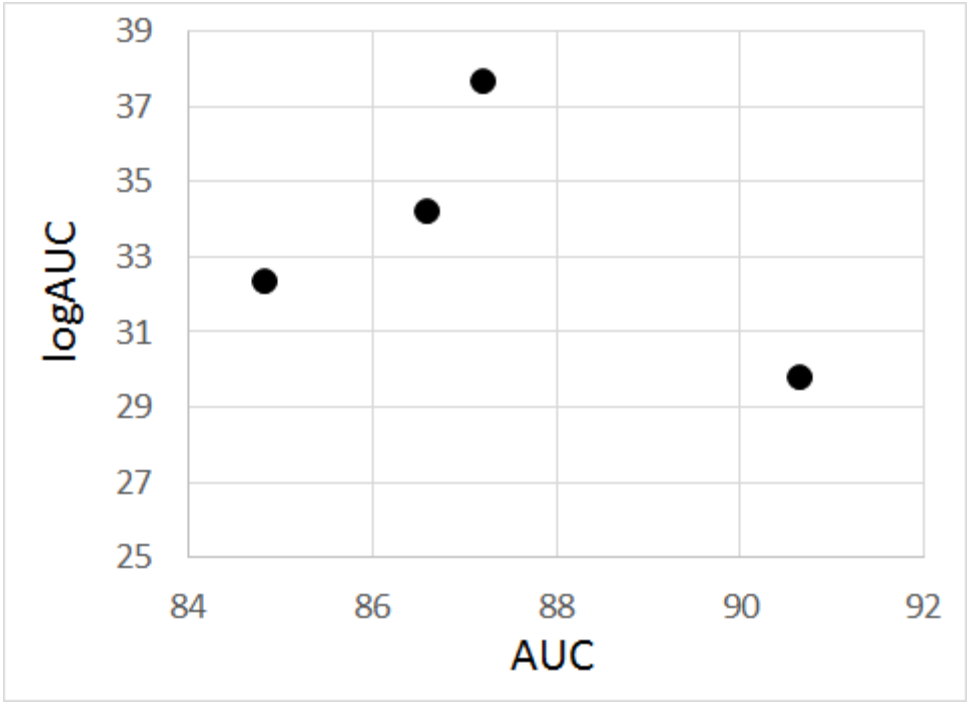


Log adjusted ROC curve



Log-adjusted AUC:
Early Enrichment Weighted More

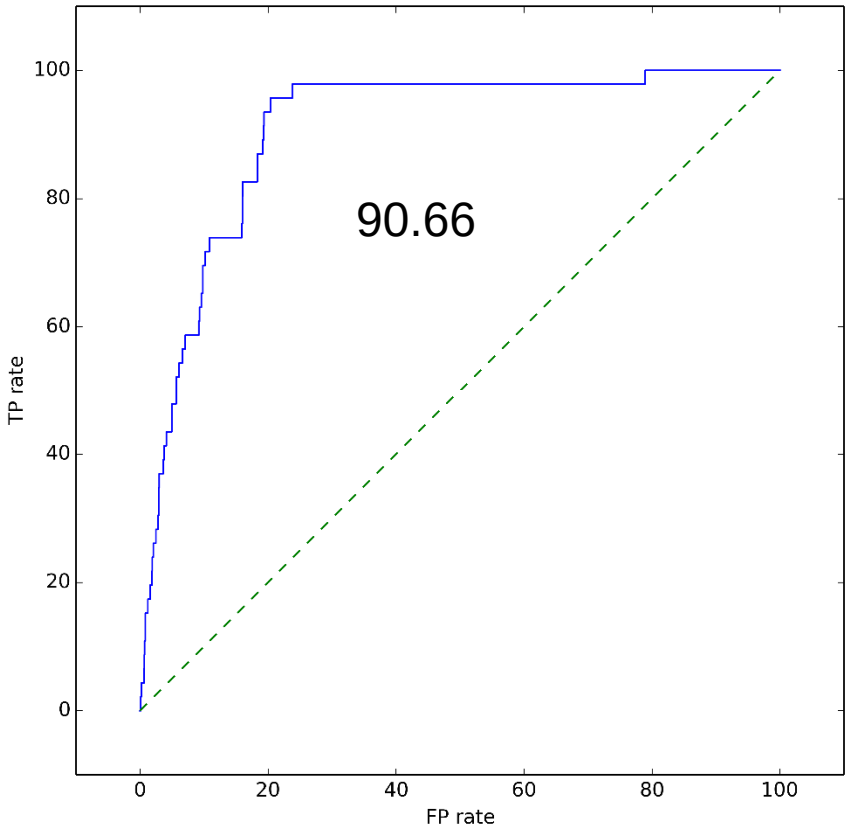
	AUC	logAUC
CcP_A:	87.23	37.63
CcP_B:	84.85	32.30
CcP_C:	86.61	34.14
CcP_D:	90.66	29.73



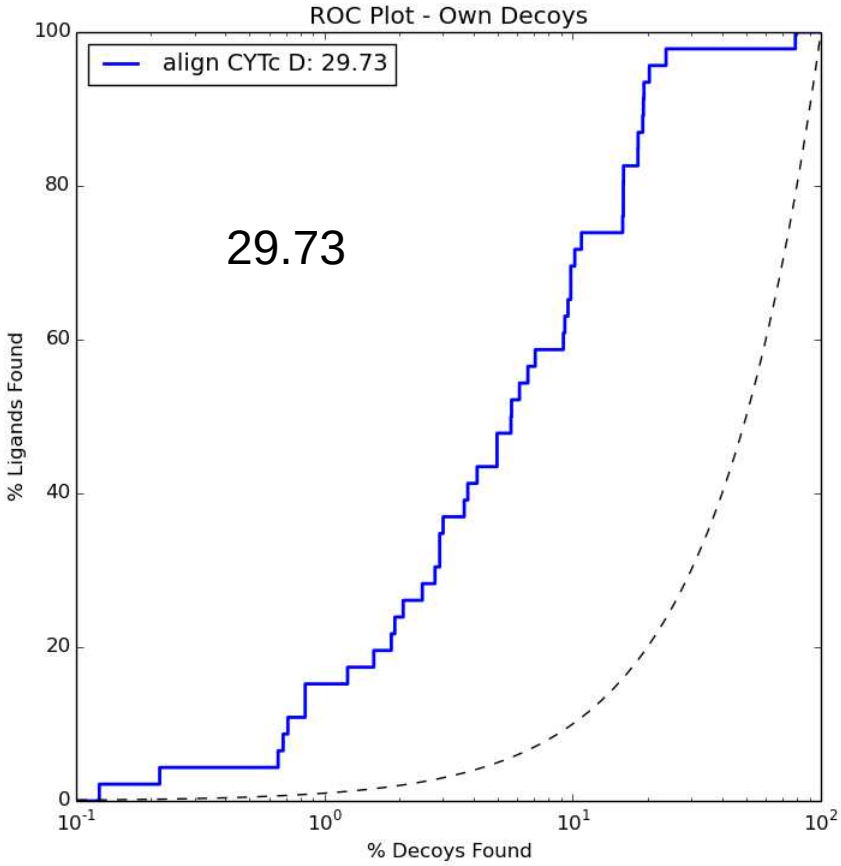
Log-adjusted AUC:
Early Enrichment Weighted More

CcP conformation D

ROC curve

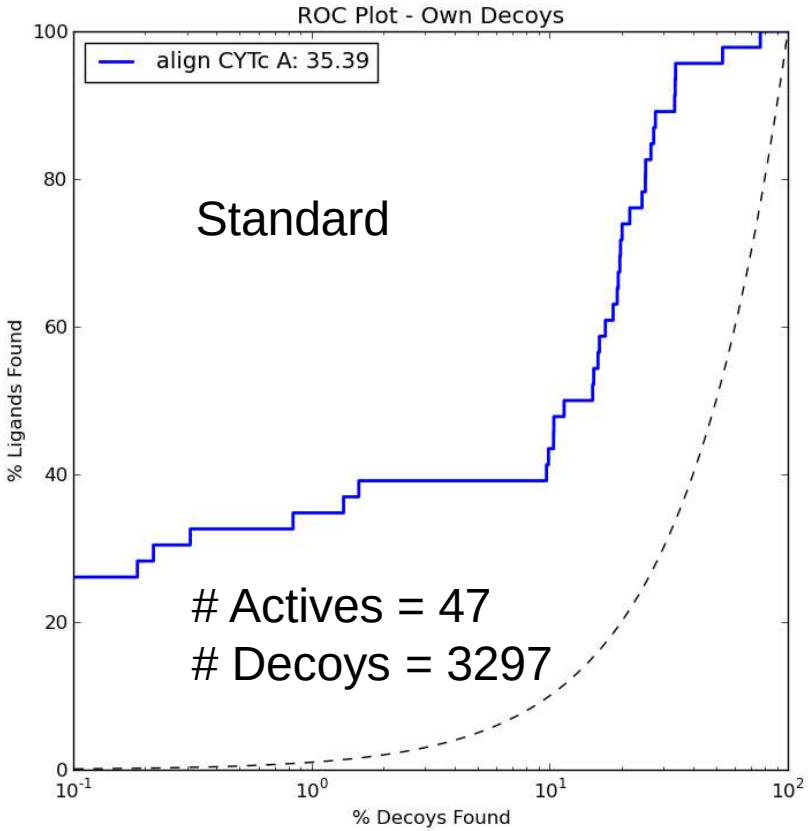


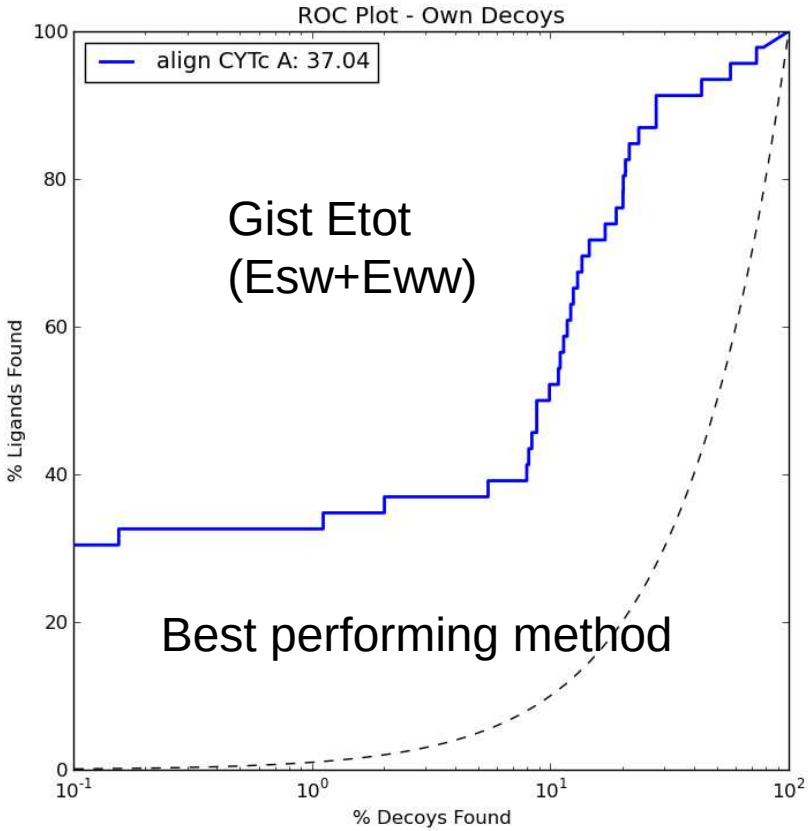
Log adjusted AUC curve



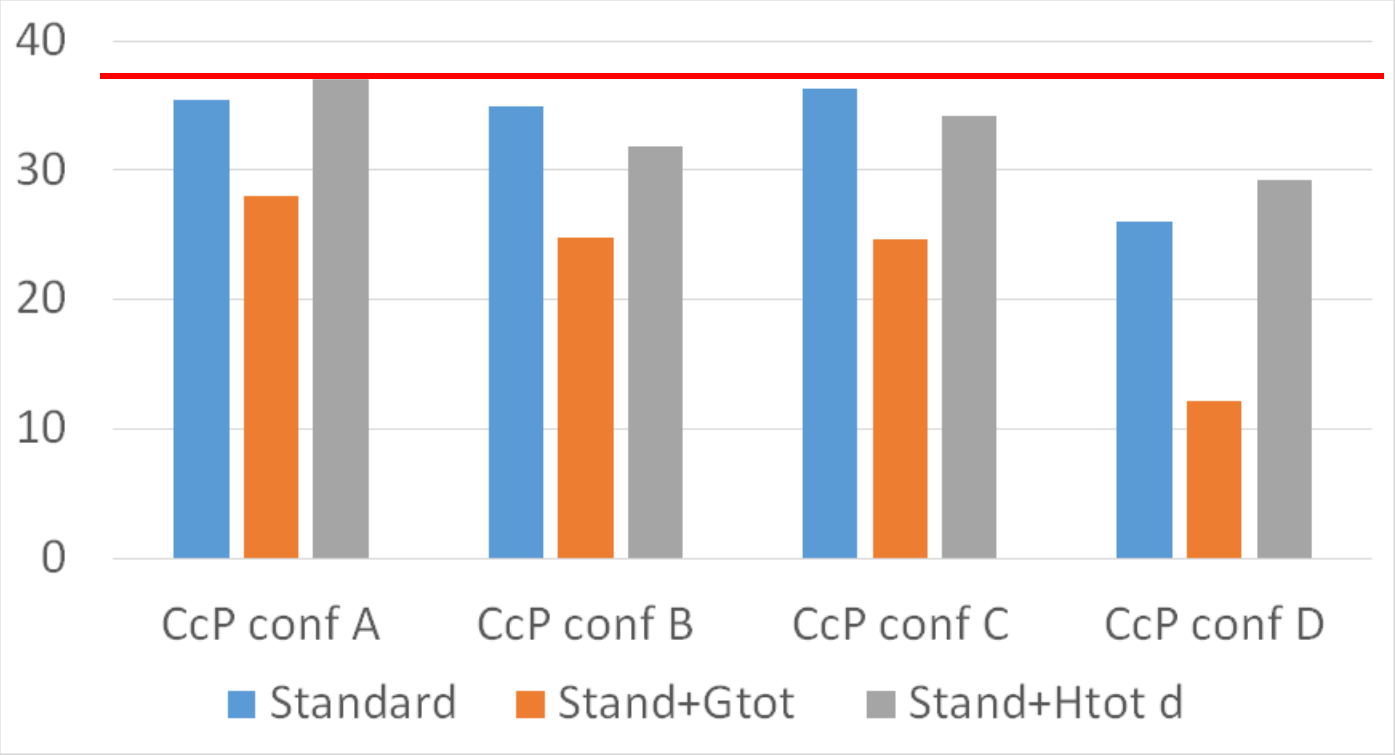
Uses Enrichments to Compare Scoring

CcP Conformation A





Uses Enrichments to Compare Scoring



CcP gateless Enrichment results[a] for 4 different loop conformations.				
	CcP conf A	CcP conf B	CcP conf C	CcP conf D
Standard[b]	35.39	34.89	36.35	26.01
Stand+Gtot[c]	28.04	24.75	24.7	12.19
Stand+Htot[d]	37.04	31.81	34.19	29.28

[a] Reporting the log adjusted AUC

[b] standard is: $E_{score} = E_{vdw} + E_{Es} + E_{lig\text{-}des}$

[c] standard score plus the enthalpy and entropy (G_{tot}).

[d] standard score plus only the Enthalpy ($H = E_{sw}+E_{ww}$).

Outline

- DOCK
  - History
  - Differences between 3.6 and 3.7
  - Sampling method
  - Scoring function
  - New Input file
  - New dir structure
  - My work – Vitamin D Receptor; Receptor desolvation
  - Methods of evaluation docking performance

- Tweaking ones docking setups:
  - Make a residue more polar
  - Modify Spheres
  - Rescore the docked poses / molecules

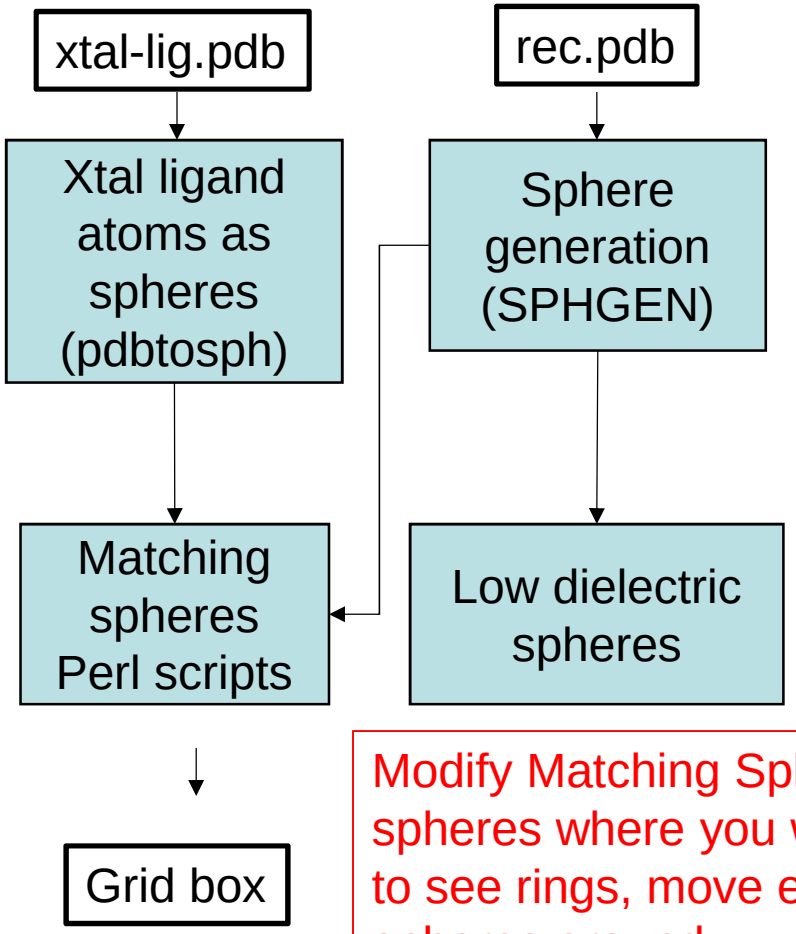
Useful Scripts to Run DOCK 3.7

```
ls -l ~fischer/work/VDR/masterscriptsTEB/
total 100
-rw-r--r--. 1 fischer bks 2158 Apr 14  2014 0001.be_balsti_py.csh
-rw-r--r--. 1 fischer bks 1609 Apr 14  2014 0002.blastermaster.csh
-rw-r--r--. 1 fischer bks 2664 Apr 15  2014 0002a.blastermaster_manualProt.csh
-rw-r--r--. 1 fischer bks 1642 Oct  3 13:30 0002b.alignwithchimera.movespheres.csh
-rw-r--r--. 1 fischer bks 3399 Apr 14  2014 0003.lig-decoy_enrichment.csh
-rw-r--r--. 1 fischer bks 3430 Oct  6 16:08 0003a.lig-decoy_enrich_NotOnQueue.csh
-rw-r--r--. 1 fischer bks 1039 Oct  7 10:43 0003b.addFORTRANSTOP.csh
-rw-r--r--. 1 fischer bks 2255 Apr 14  2014 0004.combineScoresAndPoses.csh
-rw-r--r--. 1 fischer bks 3586 Apr 16  2014 0005.AUCplot_of-lig-decoys.csh
-rw-r--r--. 1 fischer bks 1593 Oct  3 14:07 0006.VS_db2.csh
-rw-r--r--. 1 fischer bks  466 Apr 11  2014 0006a.VS_db2_restart.csh
-rw-r--r--. 1 fischer bks  924 Oct  7 11:25 0007.VS_combineScoresAndPoses.csh
-rw-r--r--. 1 fischer bks 1037 Jun 13 12:45 0007.VS_combineScoresAndPoses.qsub.csh
-rw-r--r--. 1 fischer bks  920 Oct 15 16:28 0021.rescore_DOCK6_preproc_addHcrg.csh
-rw-r--r--. 1 fischer bks 1106 Oct 15 15:49 0021_chimera_dock6prep.py
-rw-r--r--. 1 fischer bks 4058 Oct 15 16:13 0022.rescore_DOCK6_make_grid_sph.csh
-rw-r--r--. 1 fischer bks 2346 Oct 15 16:14 0023.rescore_DOCK6_hopt_ligmin_onGrid.csh
-rw-r--r--. 1 fischer bks 4417 Oct 15 17:45 0024.rescore_DOCK6_FP_for3.7.csh
-rw-r--r--. 1 fischer bks 1467 Oct 15 17:26 README_covalent_temp
-rw-r--r--. 1 fischer bks  574 Oct  3 13:46 README_mod_sph
-rw-r--r--. 1 fischer bks  328 Oct  3 14:13 README_overview
-rw-r--r--. 1 fischer bks 1551 Oct  3 13:22 README_tart
-rw-r--r--. 1 fischer bks  308 Oct 14 11:17 chimera.startup.com
drwxr-xr-x. 2 fischer bks   28 Oct  3 14:04 for_old_cluster
drwxr-xr-x. 2 fischer bks 4096 Sep  4 20:08 zzz.inputs
```

Tweaking your set up – Spheres

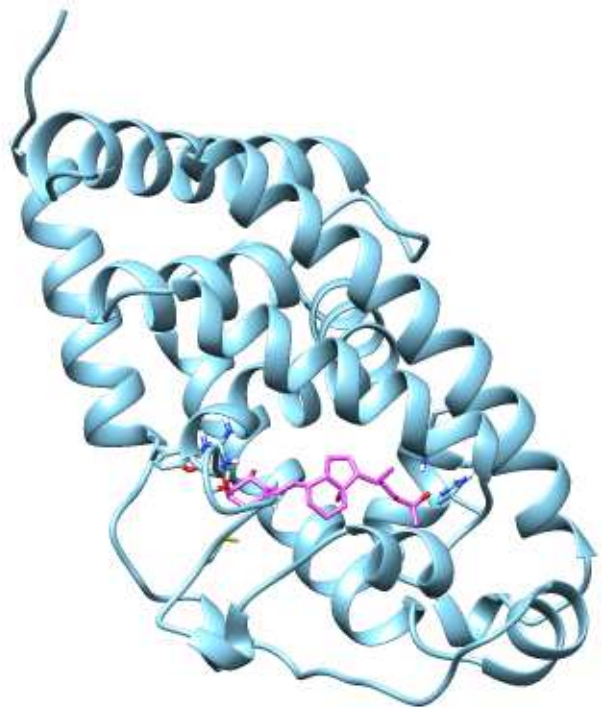
- Place spheres were you want your rings to go
- Use previous screens for ideas

Sphere Generation

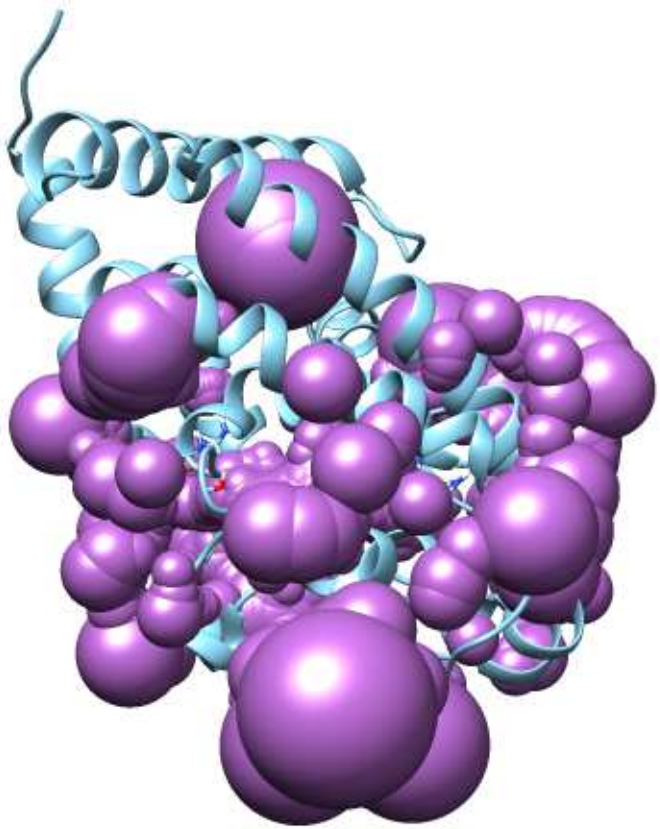


Modify Matching Spheres: add spheres where you would like to see rings, move existing spheres around

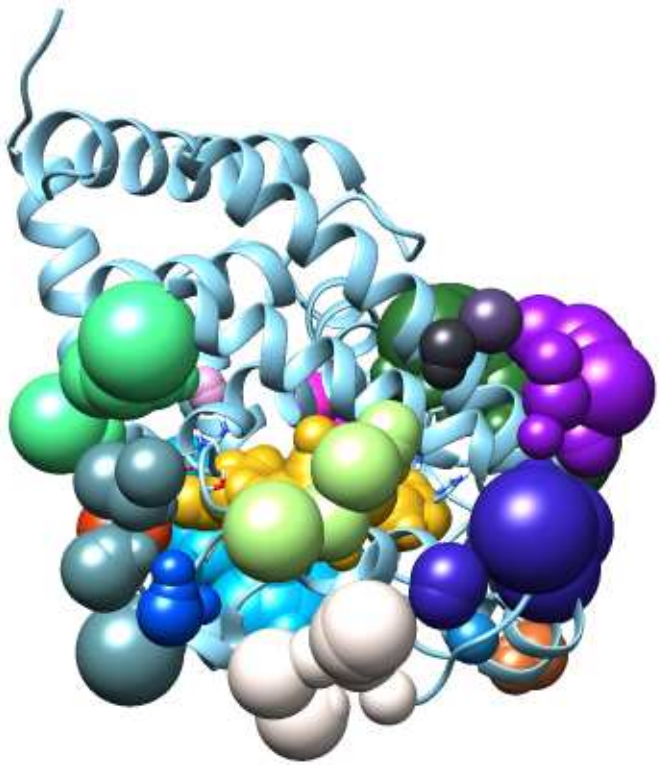
Add (or remove) Low dielectric Spheres To regions where you would like to have less (more) screening



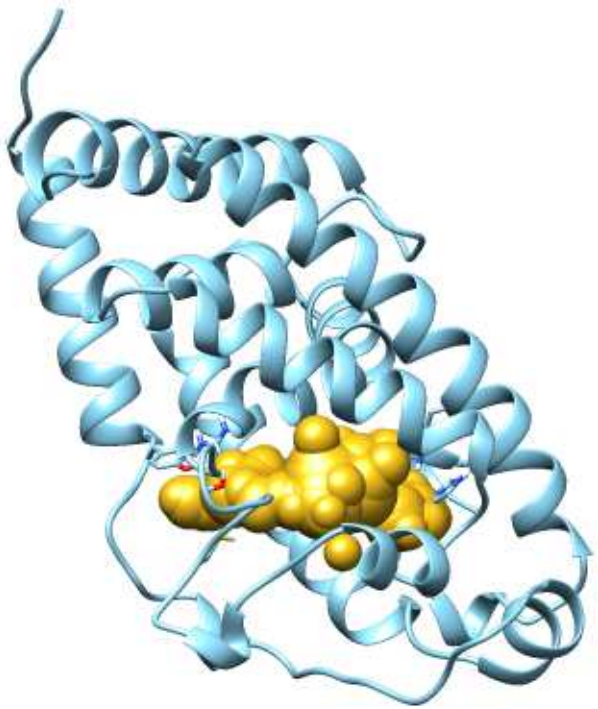
SPHGEN Spheres



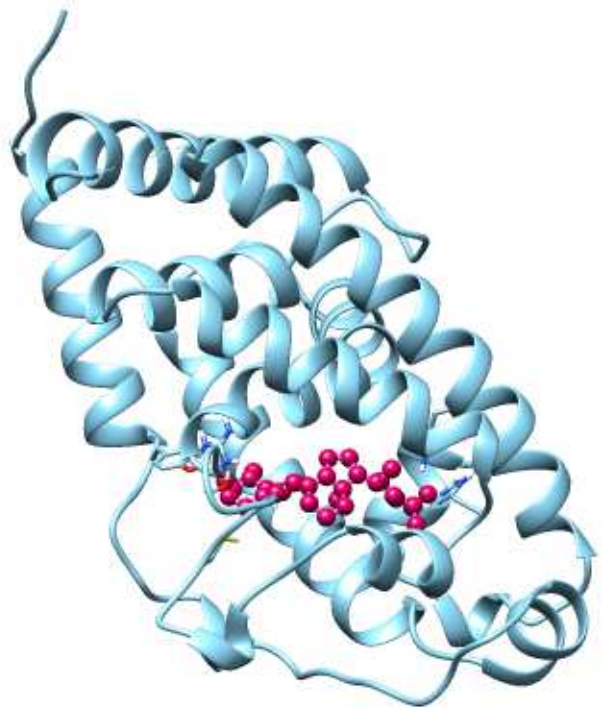
SPHGEN Sphere Clusters



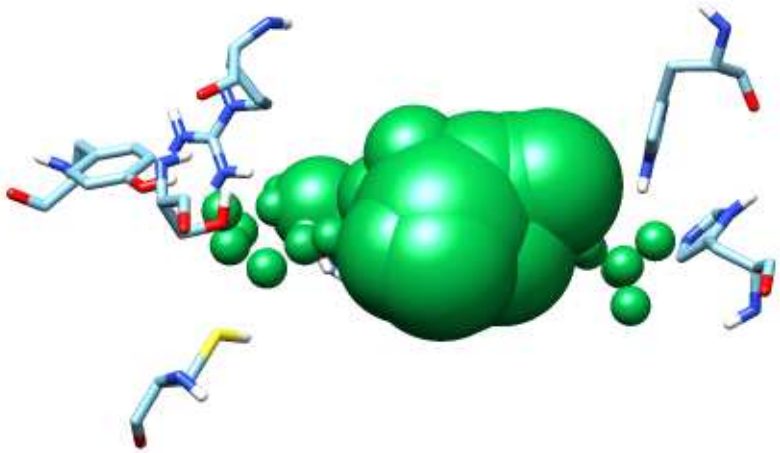
SPHGEN Cluster1 Spheres



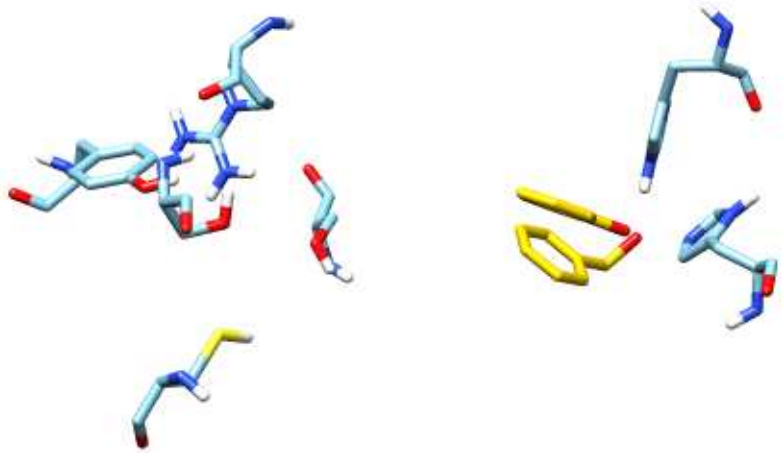
Ligand Atoms Spheres



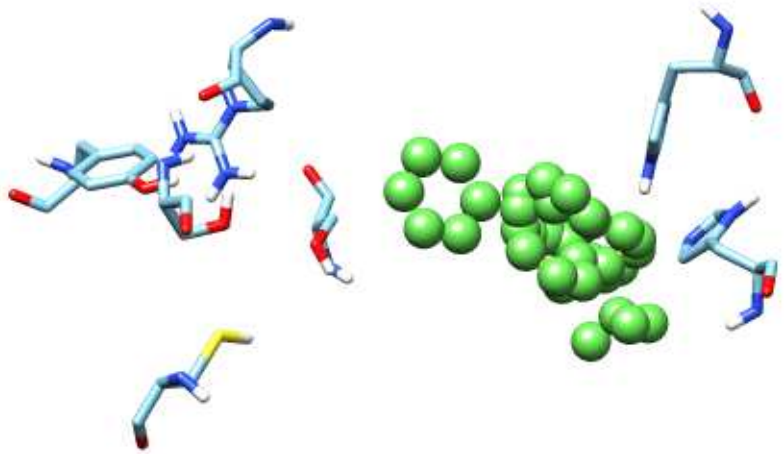
Matching Spheres



Idea for Where to Put Spheres



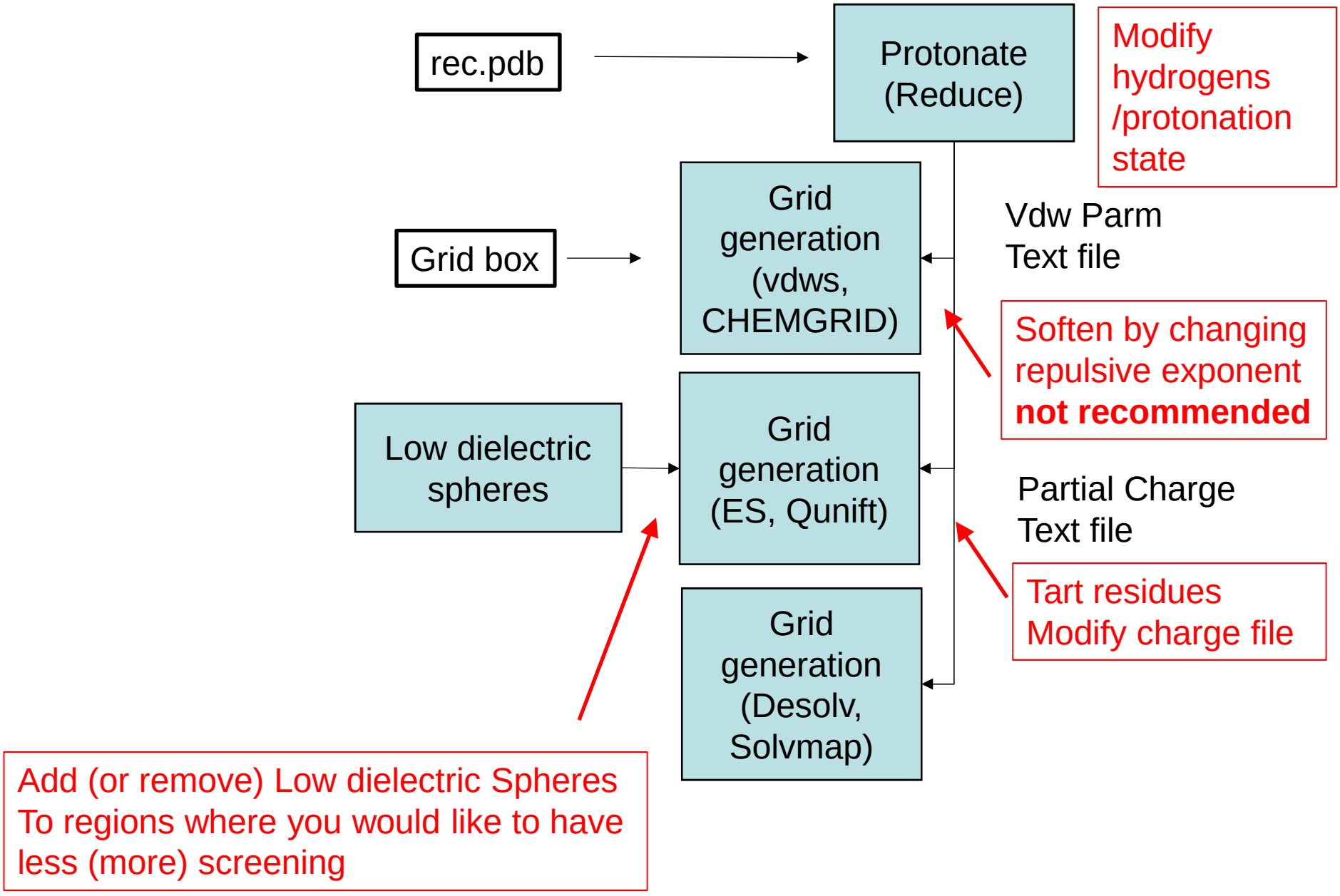
Right Side Spheres



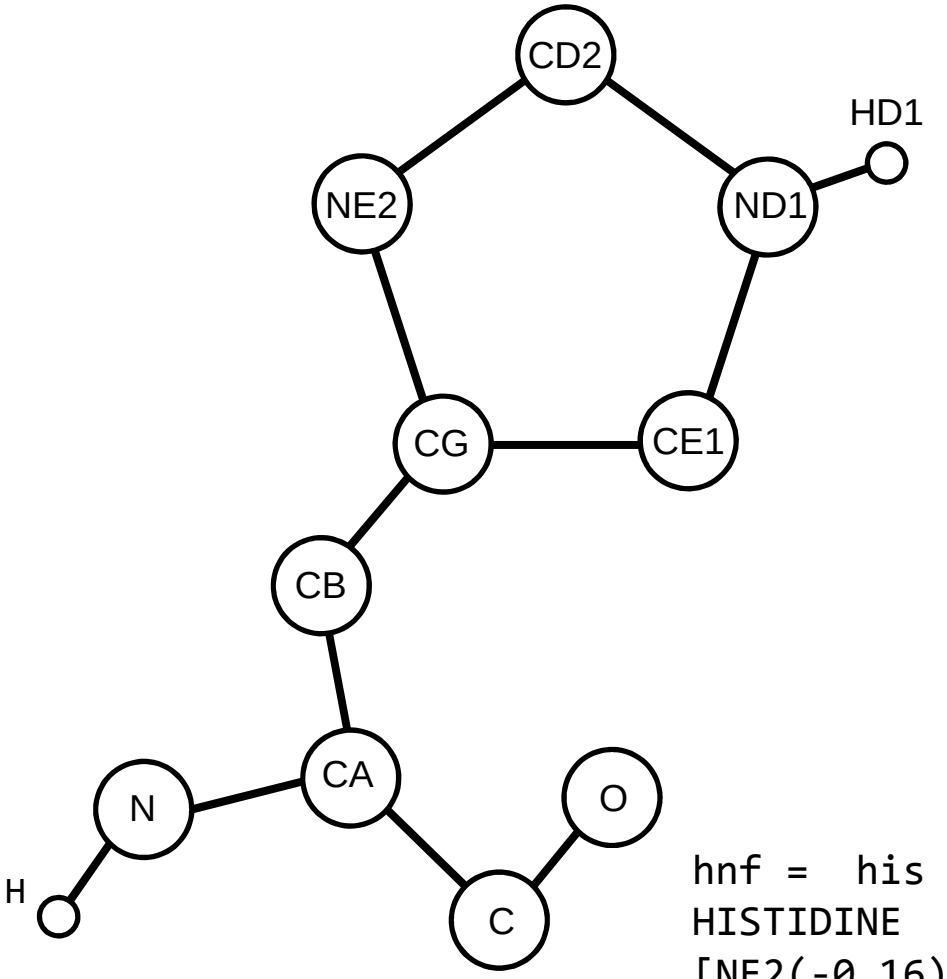
Tweaking your set up – Tarting

- Making residues more polar

Receptor preparation



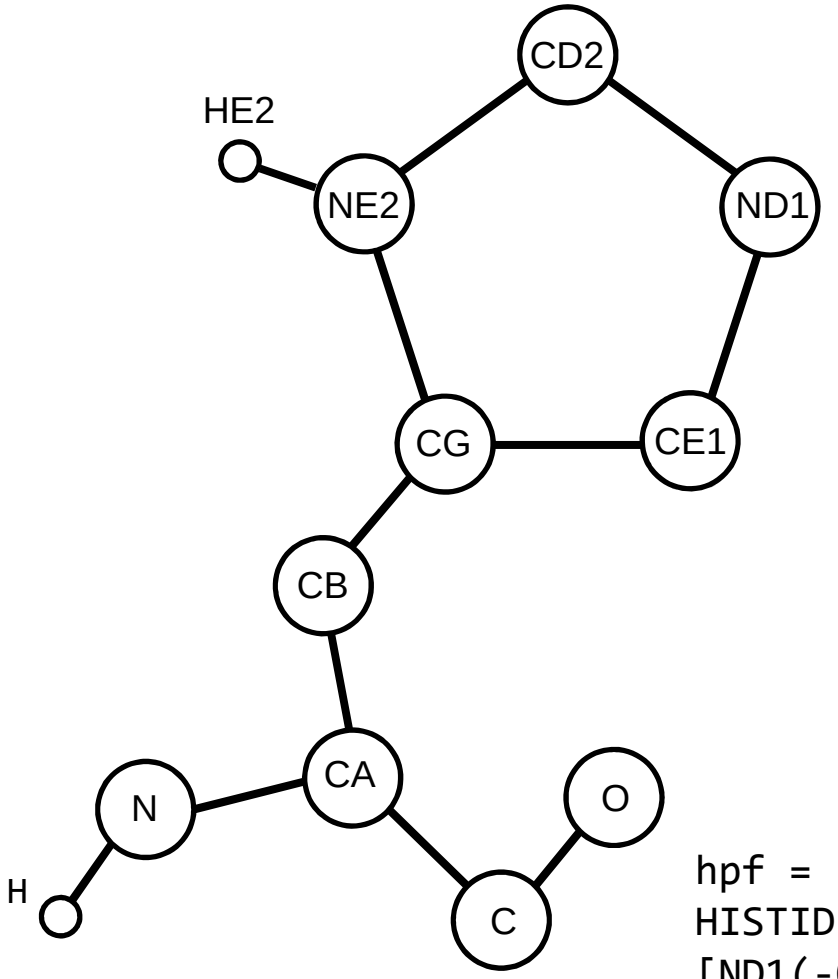
HISTIDINE neutral delta proton [tart2]



N	-0.520	N	-0.520	
C	0.526	C	0.526	
O	-0.500	O	-0.500	
CA	0.219	CA	0.219	
CB	0.060	CB	0.060	
CG	0.089	CG	0.089	
CD2	0.145	CD2	0.145	
CE1	0.384	CE1	0.384	
ND1	-0.444	ND1	-0.364	+0.08
NE2	-0.527	NE2	-0.687	-0.16
H	0.248	H	0.248	
HD1	0.320	HD1	0.400	+0.08

hnf = his negative epsilon nitrogen
HISTIDINE neutral delta proton, but more polar
[NE2(-0.16) --> ND1 (+0.08), HD1(+0.08)]

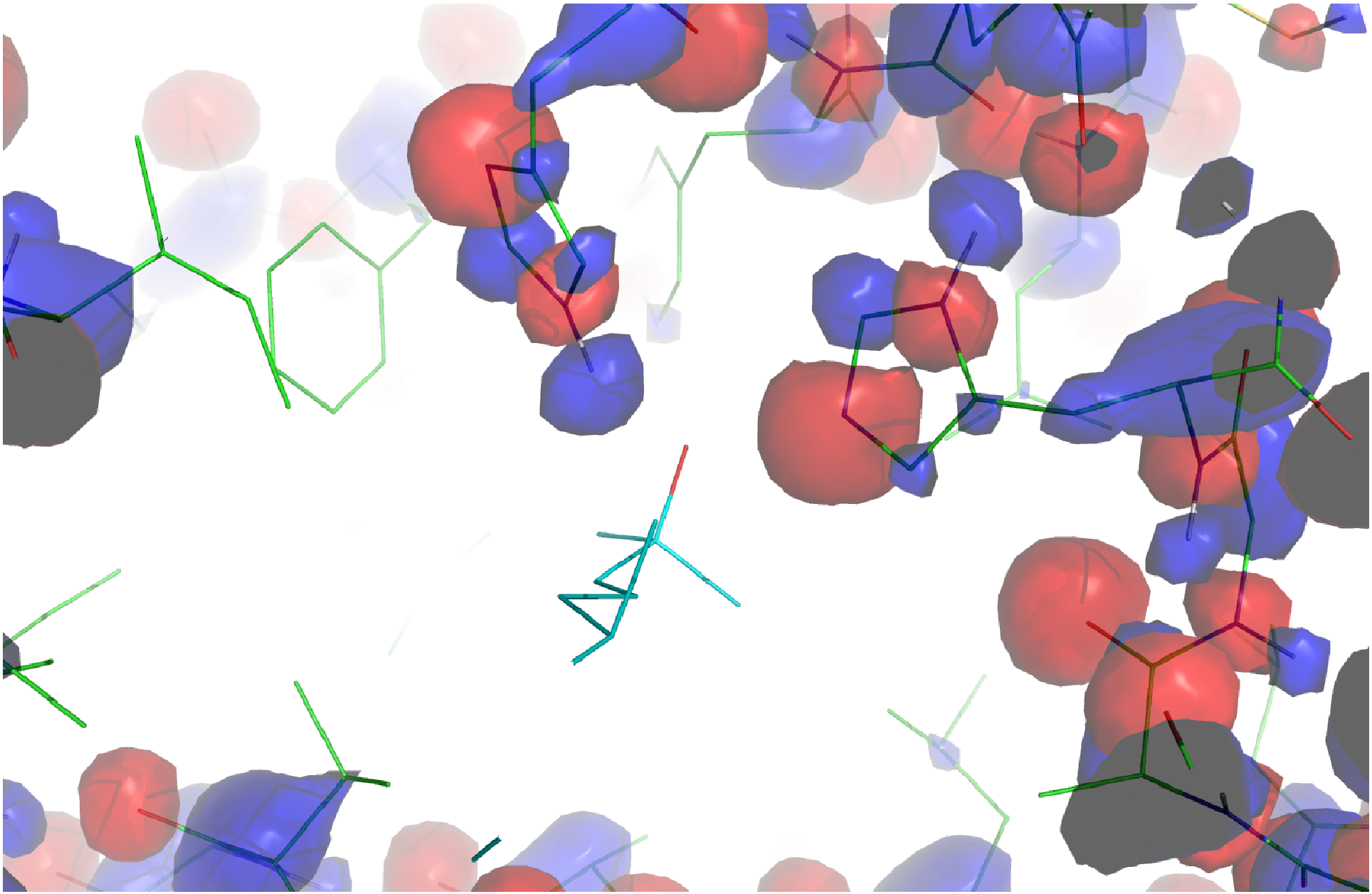
HISTIDINE neutral epsilon proton [tart2]



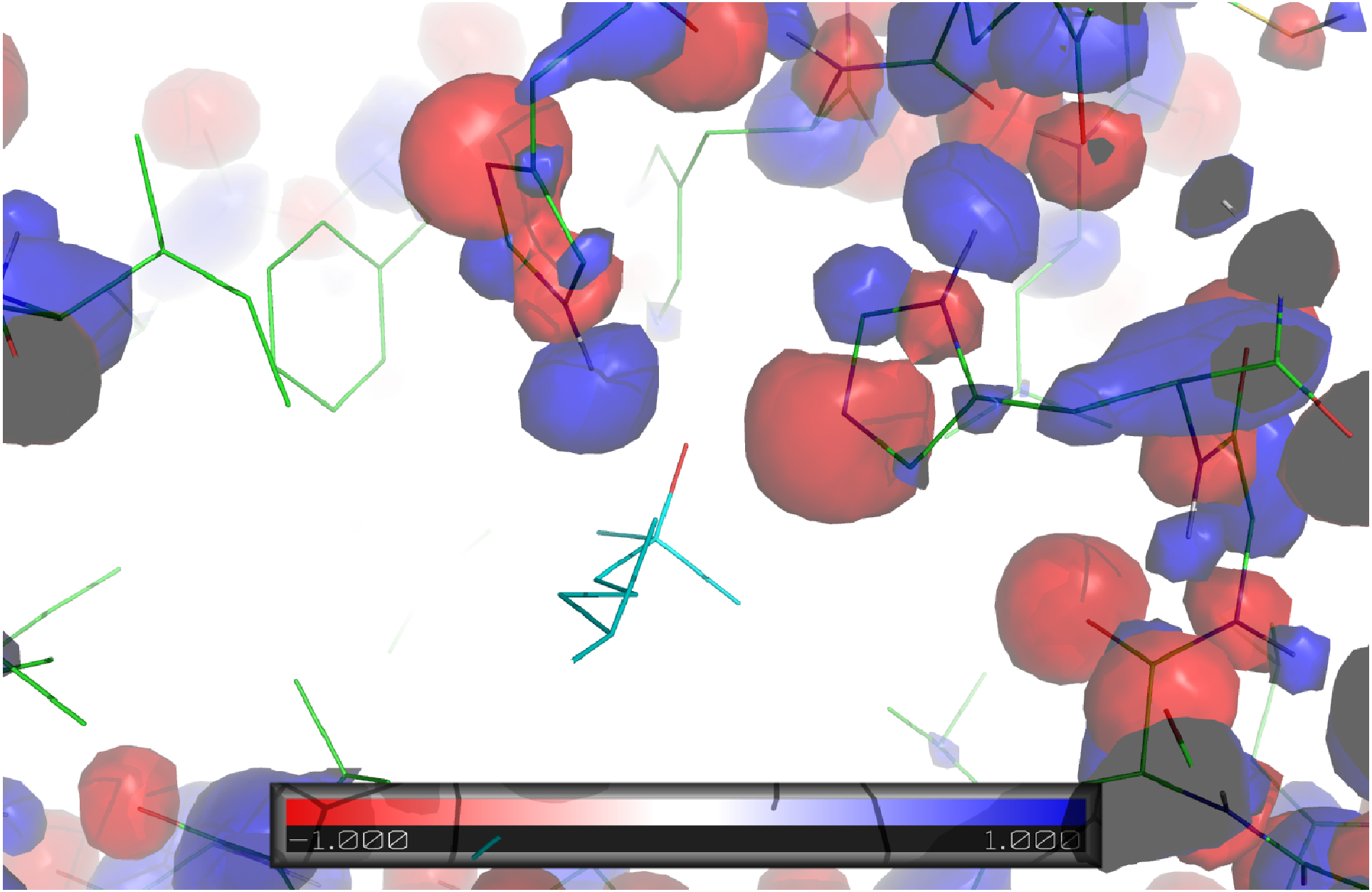
N	-0.520	N	-0.520	
C	0.526	C	0.526	
O	-0.500	O	-0.500	
CA	0.219	CA	0.219	
CB	0.060	CB	0.060	
CG	0.112	CG	0.112	
CD2	0.122	CD2	0.122	
CE1	0.384	CE1	0.384	
ND1	-0.527	ND1	-0.607	-0.08
NE2	-0.444	NE2	-0.524	-0.08
H	0.248	H	0.248	
HE2	0.320	HE2	0.480	+0.16

hpf = his positive epsilon hydrogen
HISTIDINE neutral epsilon proton, but more polar
[ND1(-0.08), NE2 (-0.08) --> HE2(+0.16)]

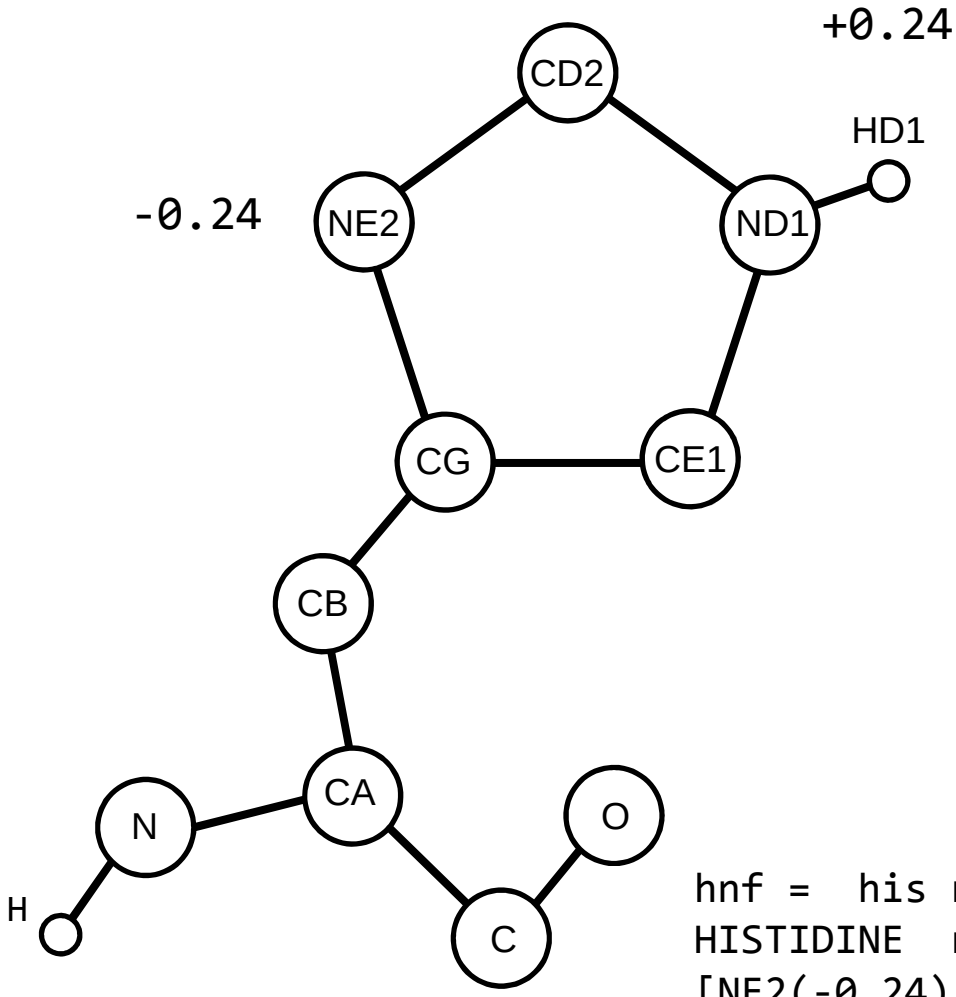
Original partial charges



More polar (tart2) partial charges



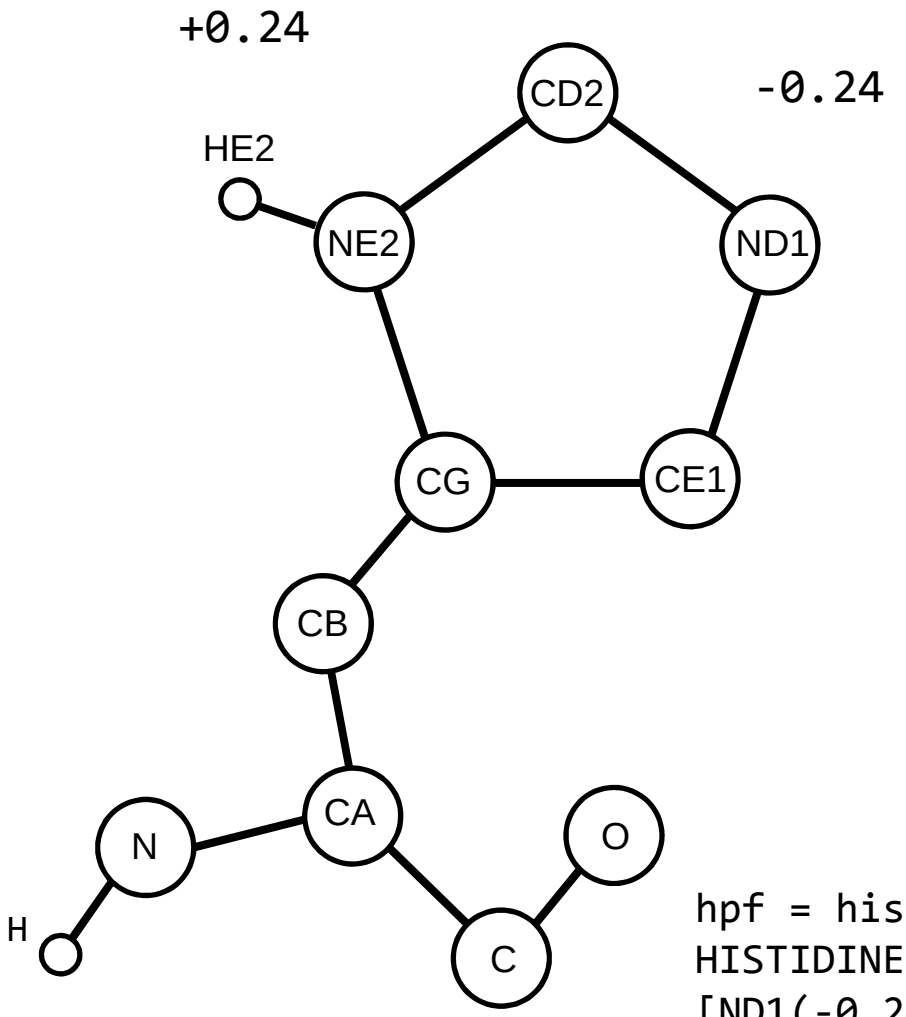
HISTIDINE neutral delta proton [tart3]



N	-0.520	N	-0.520	
C	0.526	C	0.526	
O	-0.500	O	-0.500	
CA	0.219	CA	0.219	
CB	0.060	CB	0.060	
CG	0.089	CG	0.089	
CD2	0.145	CD2	0.145	
CE1	0.384	CE1	0.384	
ND1	-0.444	ND1	-0.444	+0.0
NE2	-0.527	NE2	-0.767	-0.24
H	0.248	H	0.248	
HD1	0.320	HD1	0.560	+0.24

hnf = his negative epsilon nitrogen
HISTIDINE neutral delta proton, but more polar
[NE2(-0.24) --> ND1 (+0.0), HD1(+0.24)]

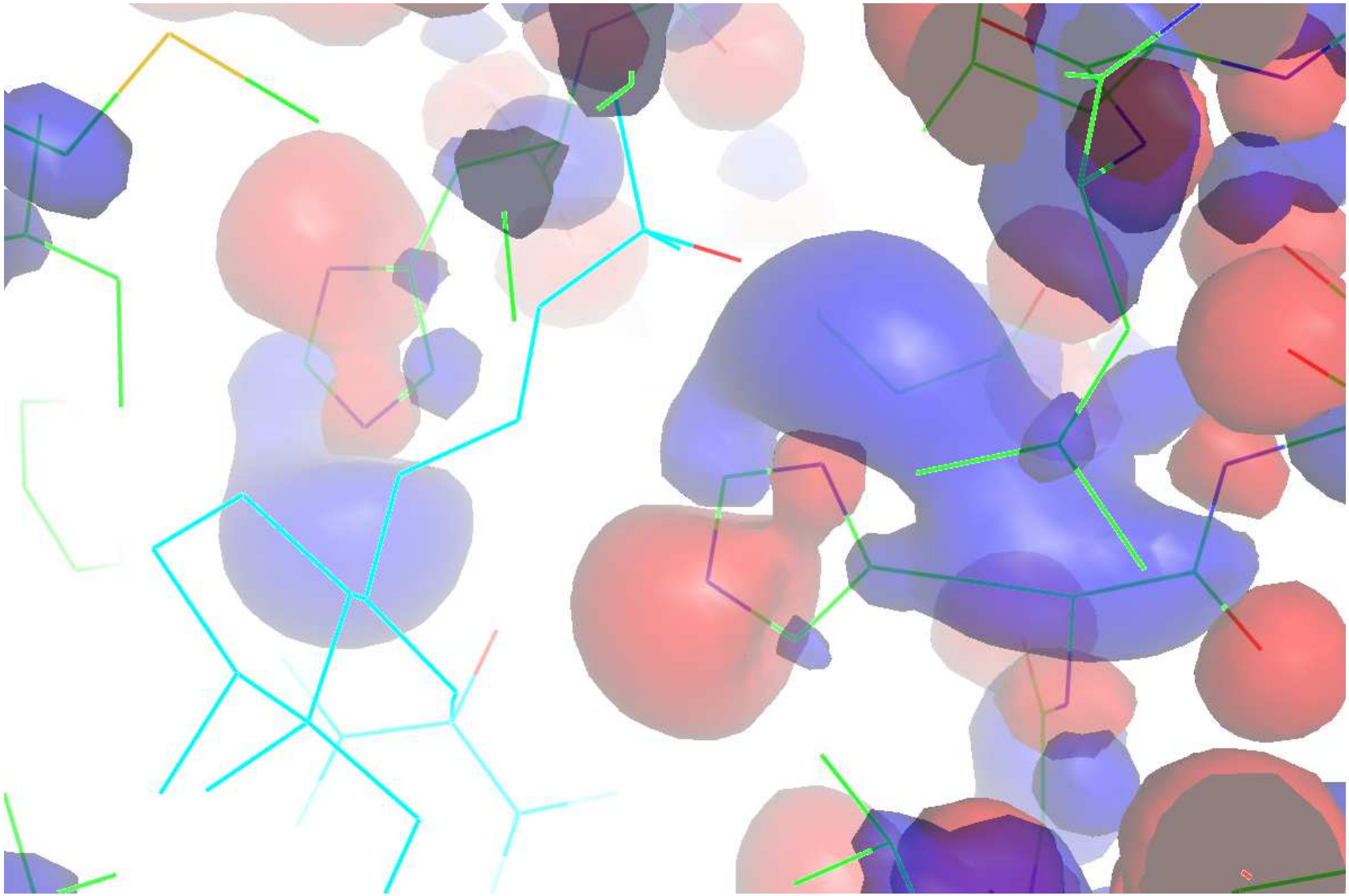
HISTIDINE neutral epsilon proton [tart3]



N	-0.520	N	-0.520	
C	0.526	C	0.526	
O	-0.500	O	-0.500	
CA	0.219	CA	0.219	
CB	0.060	CB	0.060	
CG	0.112	CG	0.112	
CD2	0.122	CD2	0.122	
CE1	0.384	CE1	0.384	
ND1	-0.527	ND1	-0.767	-0.24
NE2	-0.444	NE2	-0.444	-0.0
H	0.248	H	0.248	
HE2	0.320	HE2	0.560	+0.24

hpf = his positive epsilon hydrogen
HISTIDINE neutral epsilon proton, but more polar
[ND1(-0.24), NE2 (-0.0) --> HE2(+0.24)]

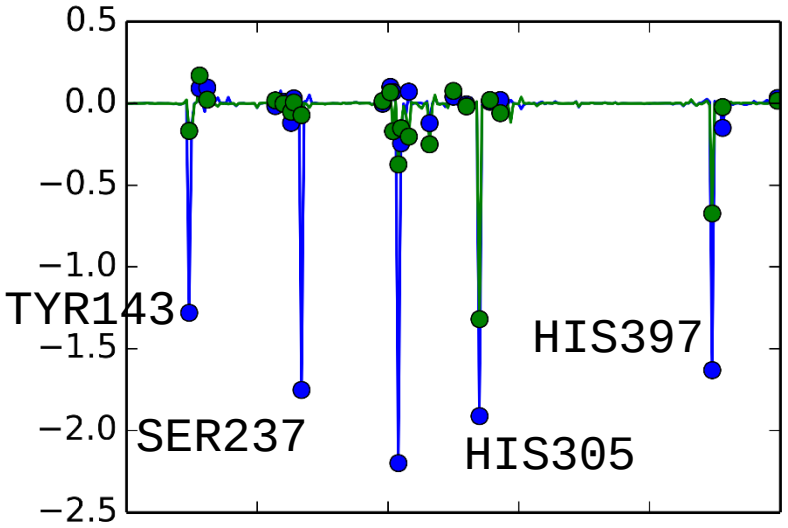
More polar (tart3) partial charges (symmetric)

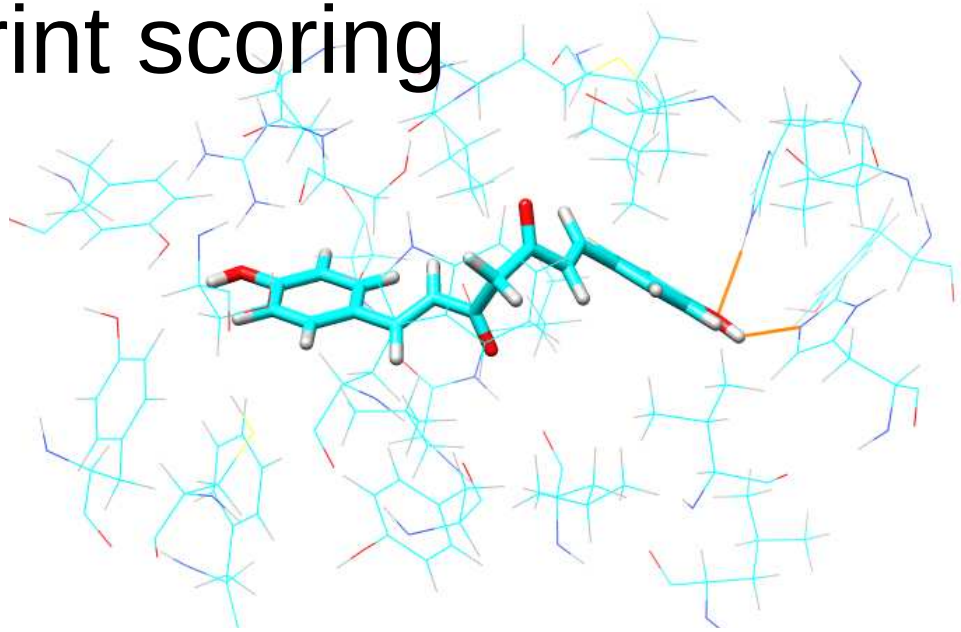


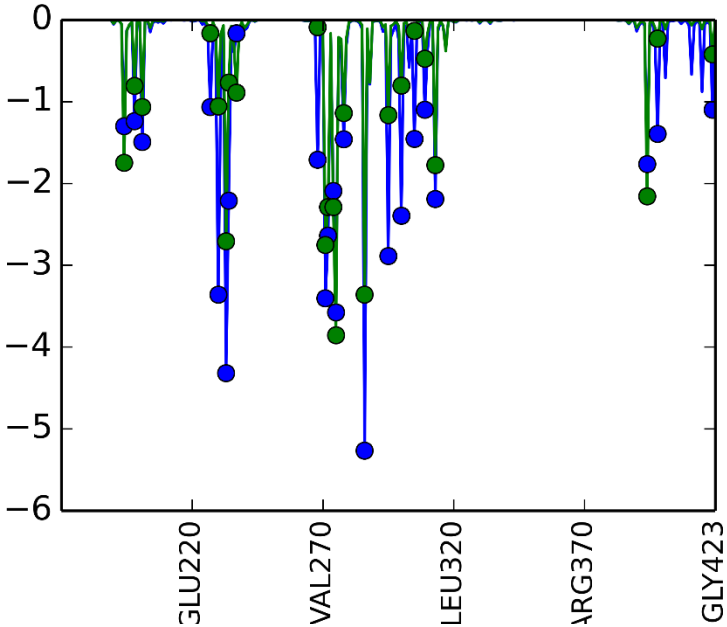
Rescoring Is Usefule

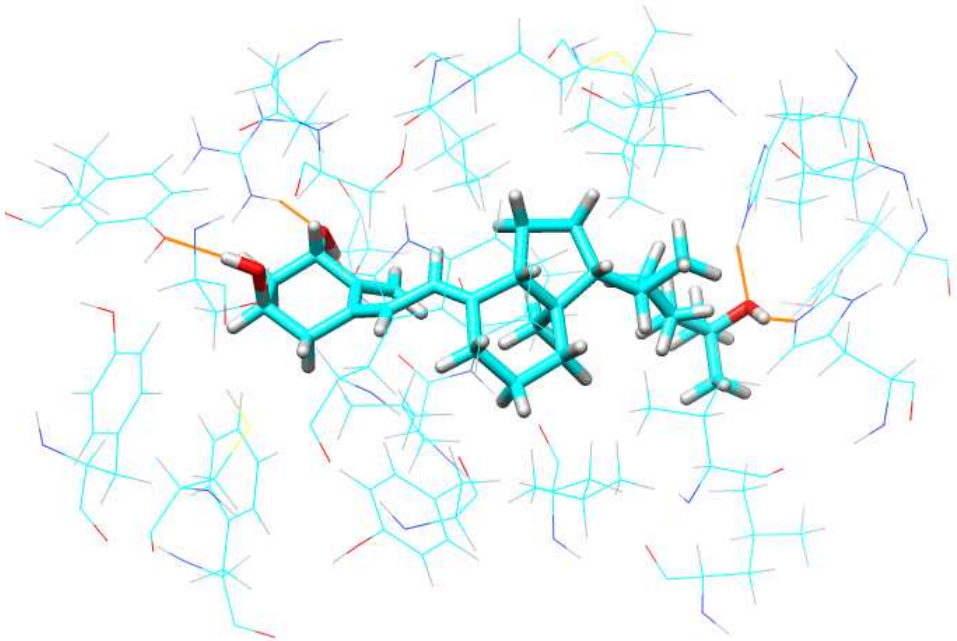
- Footprint scoring (rescore with DOCK 6)
- Can Rescore in my DOCK version (With GIST)

Footprint scoring









Conclusions

- DOCK 3.7 is new and improved!
  - Code is better behaved – fewer broken molecules
  - Switched to free software (more usable for people outside of the group)
  - Restructuring of the directories facilitates installation and usability

- There are several ways to tweak docking set up:
  - Modify Spheres
  - Make residues more polar (e.g. HIS)
  - Rescore results

Questions?

Happy Halloween!

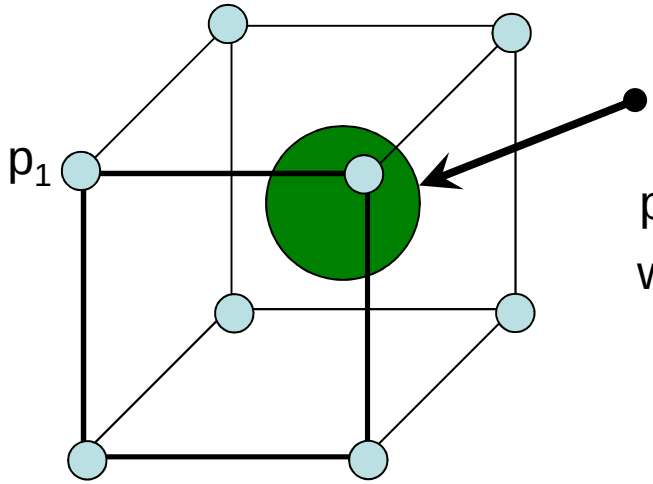
Extra slides

$$E = \sum_{i \in L} \left(\sqrt{A_{i,i}} \sum_{j \in R} \frac{\sqrt{A_{j,j}}}{r_{i,j}^{\,a}} - \sqrt{B_{i,i}} \sum_{j \in R} \frac{\sqrt{B_{j,j}}}{r_{i,j}^{\,b}} + 332 q_i \sum_{j \in R} \frac{q_j}{D r_{i,j}} \right)$$

$$G_{av}(p) = \sum_{l \in R} \frac{\sqrt{A_{l,l}}}{r_{p,l}^{\,a}}$$

$$G_{rv}(p) = \sum_{l \in R} \frac{\sqrt{B_{l,l}}}{r_{p,l}^{\,b}}$$

$$G_{es}(p) = 332 \sum_{l \in R} \frac{q_l}{D r_{p,l}}$$



p = (x,y,z)
where e = f(p)

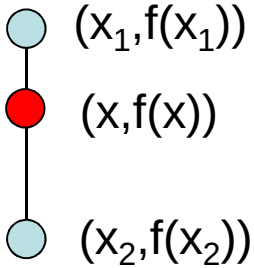
http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm#Grid

$$E \approx \sum_{i \in L} \begin{pmatrix} \sqrt{A_{i,i}}\,\mathrm{interp}\left[G_{av}(p_1), \cdots, G_{av}(p_8)\right] \\ -\sqrt{B_{i,i}}\,\mathrm{interp}\left[G_{rv}(p_1), \cdots, G_{rv}(p_8)\right] \\ +332 q_i\,\mathrm{interp}\left[G_{es}(p_1), \cdots, G_{es}(p_8)\right] \end{pmatrix}$$

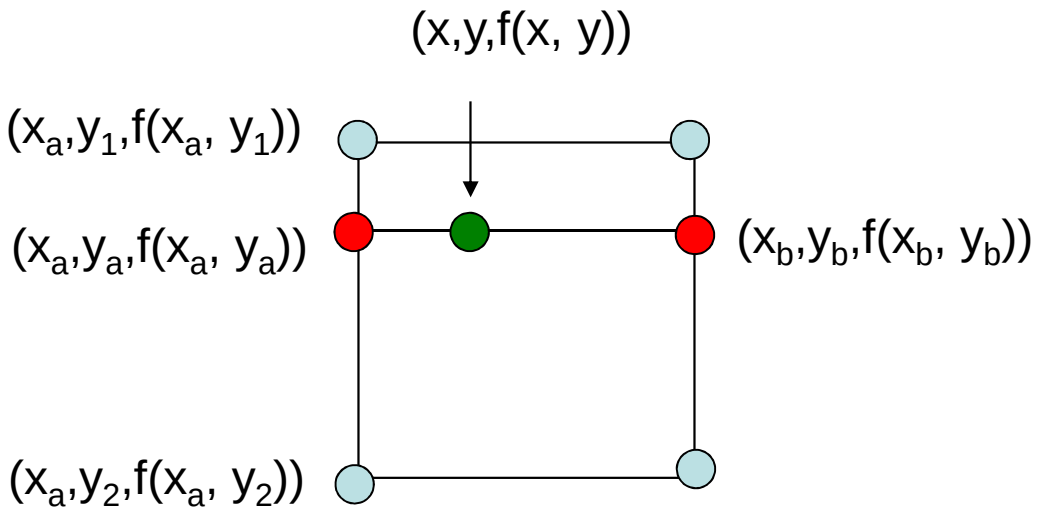
Interpolation

linear

$$f(x) \approx \frac{(x - x_1) f(x_2) + (x_2 - x) f(x_1)}{(x_2 - x_1)}$$

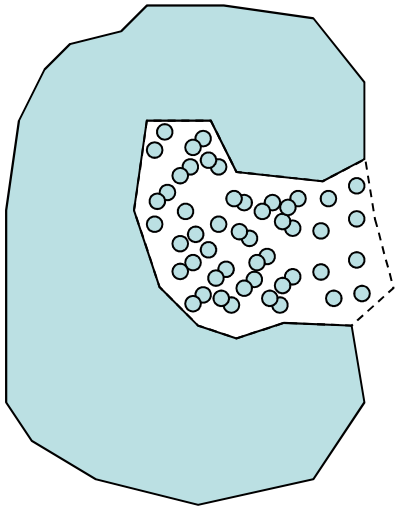


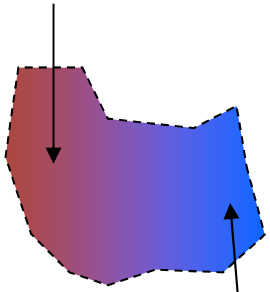
bilinear: Perform 3 linear Interpolations: 2 to calculate red (from cyan); and 1 to calculate green (from red)



Trilinear: for a cube, perform 7 linear interpolations: 4 to calculate red (from the cyan); 2 to calculate green (from red); and 1 to calculate the atomic approximation (from green)

http://en.wikipedia.org/wiki/Trilinear_interpolation



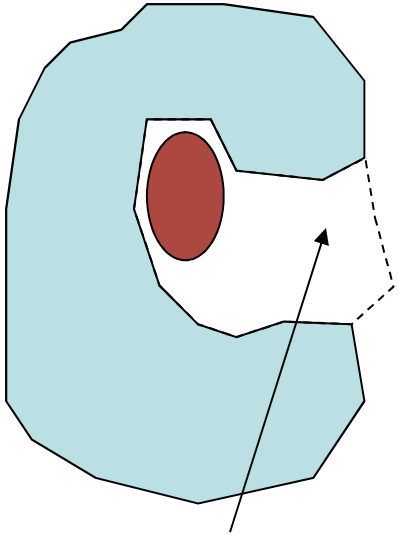


hydrophopic

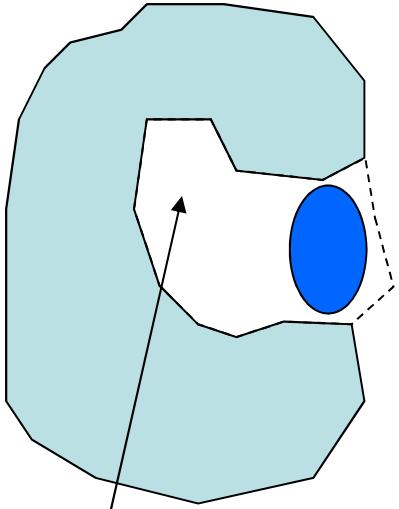
hydrophilic

Thought experiment:

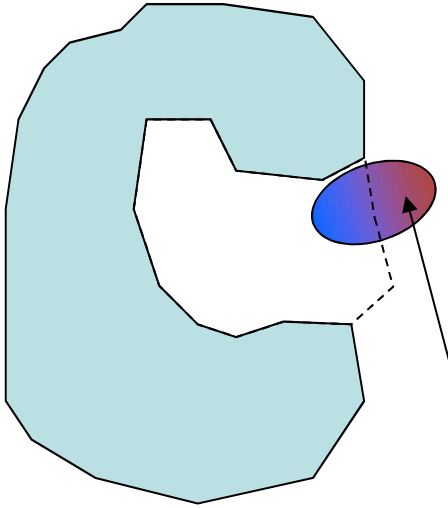
Simple examples of cases
Where the approximation to
Hydration might fail
Using dock 3.5 score
This is considering the desolvation
of the receptor only



Desolvated
unnecessarily
-- Penalty



Desolvated
unnecessarily
-- Benefited



Desolvation
necessarily
-- Benefited