

DOCK: where it is and where its going
Trent Balias, Ph.D.

Postdoctoral Scholar, Shoichet Lab
University of California, San Francisco

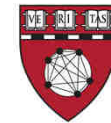
Tuesday, June 20th 12:00pm EDT

Upcoming Webinars:

July 25th: BioXTAS RAW with Jesse Hopkins

October 3rd: SPHIRE with Toshio Moriya

November 4th: EMAN2 with Steven Ludtke



SBGrid
CONSORTIUM

Events are posted on SBGrid Google Calendar (to join visit <http://sbgrid.org/calendar>)


Systems Check

Outline

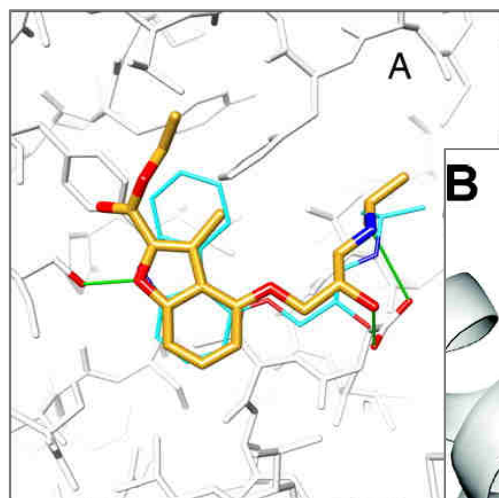
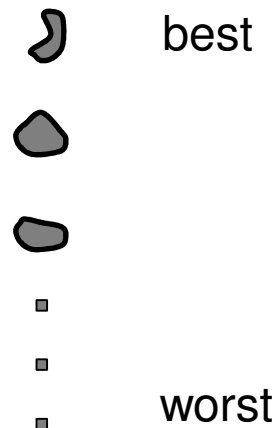
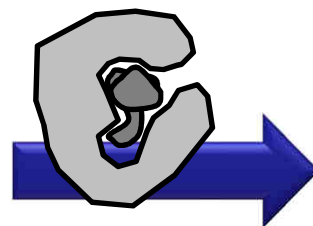
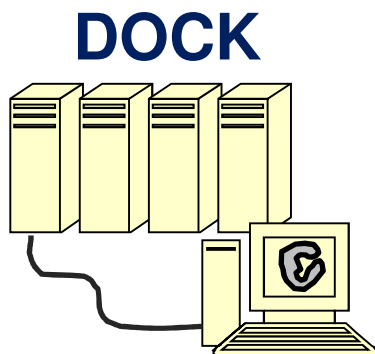
- Introduction to docking and DOCK
- DOCK 3.7
 - Resources
 - How DOCK 3.7 works
 - Tutorial on how to run it
 - Features
- DOCK 6.8
 - Resources
 - Anchor-&-grow algorithm (how DOCK 6.8 works)
 - Many scoring functions
 - Features

Docking is Ligand Discovery Tool

ZINC
~4M molecules



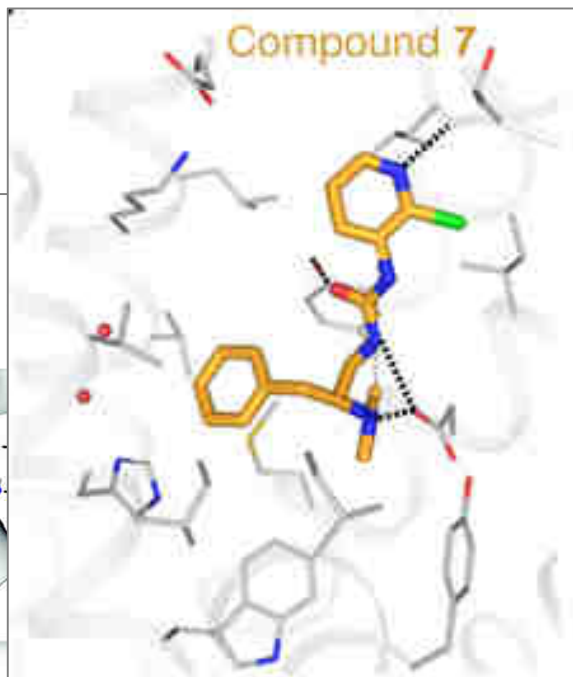
+



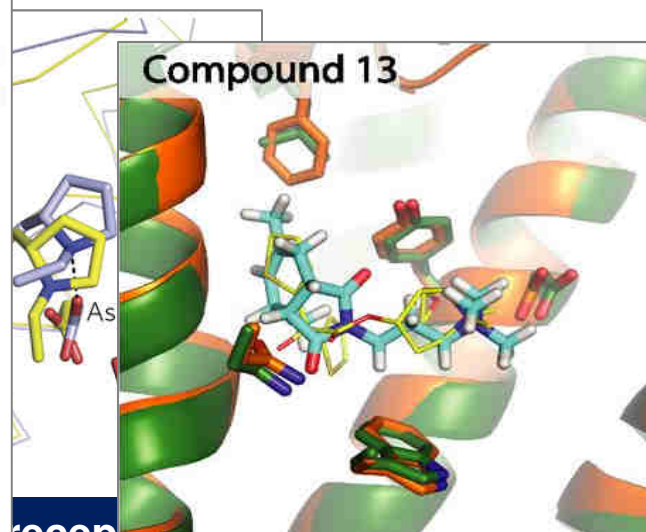
β 2-adrenergic receptor
24% hit rate; 9 nM
Kolb, *PNAS* 2009



A_{2A} adenosine receptor
35% hit rate
Carlsson et al., *J Med Chem* 2005



μ -Opioid receptor
30% hit rate; 2.3 μ M
Manglik et al., *Nature* 2016



M2/M3 muscarinic receptor
58% hit rate; 400 nM
Kruse, *Mol Pharm* 2013

Why use Dock?

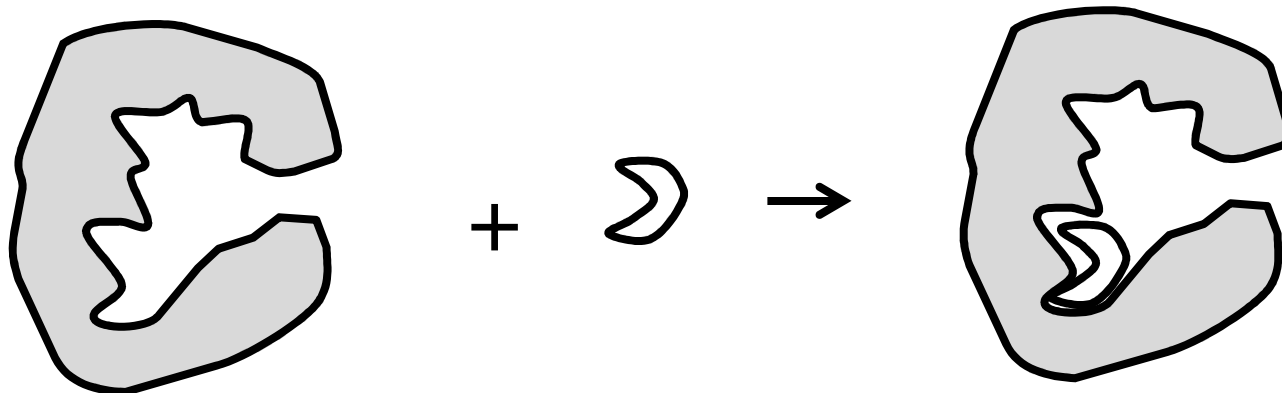
Applications of docking

- Virtual Screening given a protein and database of molecules find those that bind.
- Give a molecule and a protein predict how that molecule binds to the receptor

Docking Tasks

- Sampling
 - Generate all possibilities including the correct geometry
- Scoring
 - Rank order all the possibilities, rank the correct pose first
 - Rank the binders better than decoys
- Balance of speed and accuracy, docking has to be fast.

DOCK 3.7
is among
the fastest
docking
programs

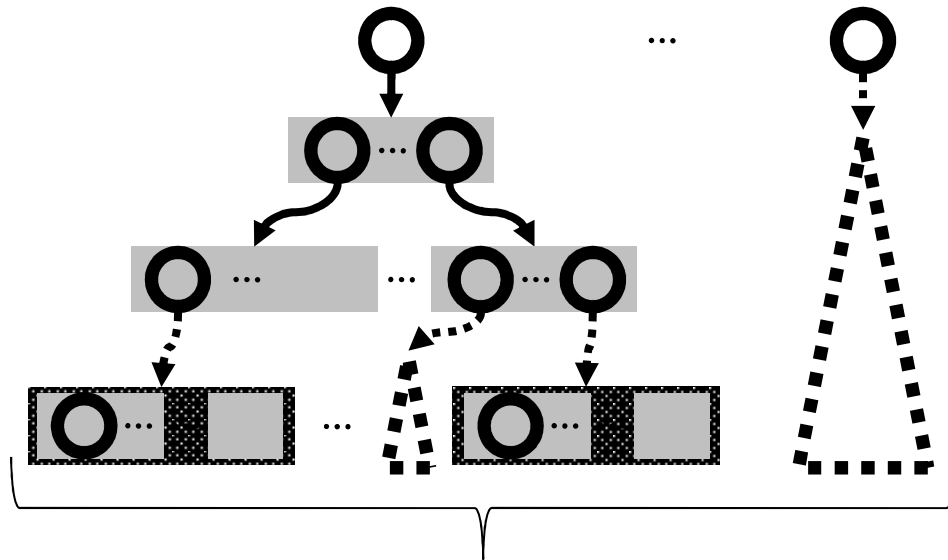
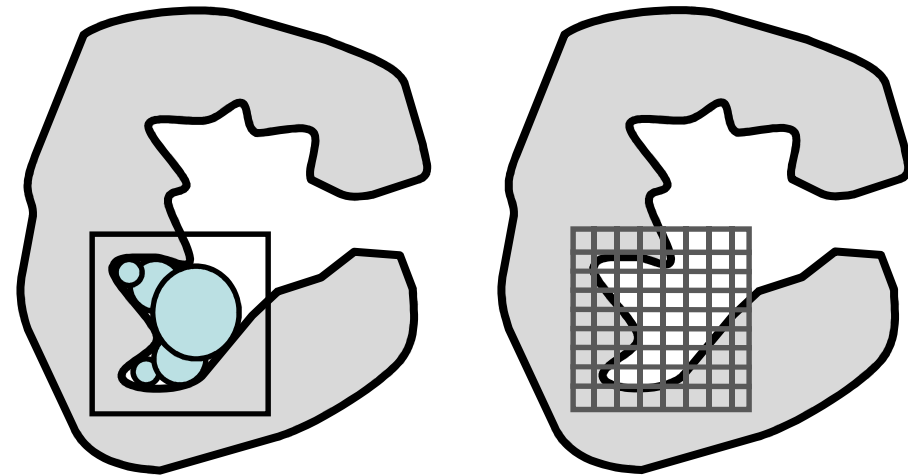
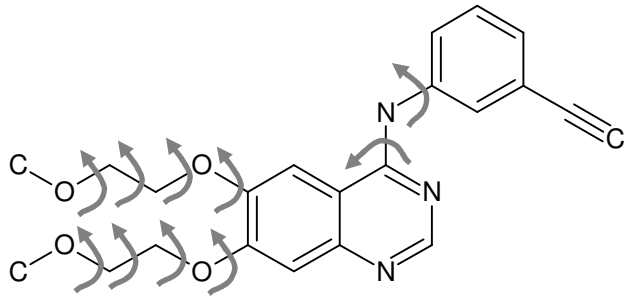


How to use DOCK 3.7

How DOCK 3.7 works

Preparation, Sampling, and Scoring

Ligand Sampling done outside of DOCK 3.7 – database construction



Dockable database file

DOCKING

orient	score
Scoring using a grid to speed up the calculations	

Scoring using a grid to speed up the calculations

DOCK 3.7 Scoring Function

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

- 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

Resources for DOCK 3.7

- BKS lab wiki: http://wiki.bkslab.org/index.php/DOCK_3.7
- DOCK-fans email archive (list to search for solutions to problems or submit questions): <http://mailman.docking.org/pipermail/dock-fans/>
- Database of small molecules:
<http://zinc15.docking.org/>
- DUD-E databases and decoy generator
<http://dude.docking.org/>
<http://dude.docking.org/generate>
<http://autodude.docking.org/>

Tutorial

Scenario 1:

Use docking to predicted how Erlotinib
(an approved drug) binds to the
Epidermal Growth Factor Receptor

Search for Your Molecule in ZINC

The screenshot shows the ZINC15 docking website interface. At the top, the browser address bar displays `zinc15.docking.org/substances/home/`. The navigation menu includes **ZINC**, **Substances**, **Catalogs**, **Tranches**, **Biological**, and **More**. The main heading is **Substances**. Below this, there are tabs for **Help**, **Examples**, **Browse**, **Table**, **Subsets**, and **Shipping List**. A search bar on the right contains the text **Erlotinib** and a **Search** button. The interface is divided into two main sections: **Draw/Search Structure** on the left and **Resolve Substances From Text/File** on the right. The **Draw/Search Structure** section features a text input field for **ZINC ID, SMILES, SMARTS, or InChI** and a chemical drawing toolbar with various icons for drawing and editing. The **Resolve Substances From Text/File** section includes a **Paste SMILES** text area, an **Upload a File** section with a **Browse...** button and the text **No file selected**, and several filter options: **Allow Lookups** (with checkboxes for **ZINC ID**, **Structure**, **Names**, **Suppliers**, and **Analogs Slow!**), **Match Tolerance** (with checkboxes for **Retired IDs**, **Charge**, **Scaffold**, **Full Text**, and **Accept Multiple Results**), and **Where to Search** (with a dropdown menu for **Subset(s) to Check** currently set to **Nothing selected**). At the bottom, there is an **Output Format** dropdown menu set to **Summary Table** and a **Resolve File** button.

Get 2D Files from ZINC

zinc15.docking.org/substances/ZINC00001546066/

ZINC Substances Catalogs Tranches Biological+ More+ About+

/substances/ZINC00001546066

ZINC1546066 (Erlotinib)

In: [bb](#) [fda](#) [for-sale](#) [in-stock](#) [standard](#)

Google [Wikipedia](#) [PubMed](#)

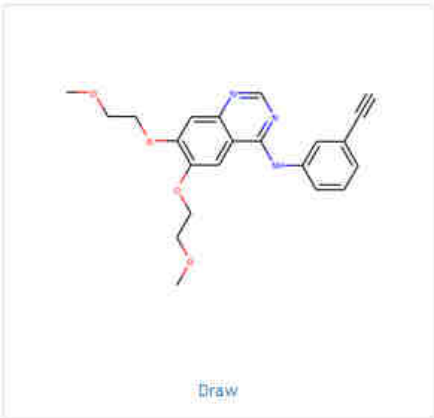
Added	Available	Since	Mwt	logP	Heavy Atoms	Tranche	Download
2004-10-06	In-Stock	2015-08-07	393.443	3.405	29	GGEA	Download

SMILES C#Cc1cccc(Nc2ncnc3cc(OCCOC)c(OCCOC)cc23)c1

InChI InChI=1S/C22H23N3O4/c1-4-16-8-5-7-17(12-16)25-22-18-13-20(28-10-8-26-2)21(29-11-9-27-3)14-19(18)23-15-24-22/h1,5-7,12-15H,8-11H2,2-3H3

InChI Key AAKJLRGGTJKAMG-UHFFFAOYSA-N

Download menu: SMILES, SDF, CSV, XML, JSON



Draw

Available 3D Representations

Find Decoys

pH range	Net charge	H-bond donors	H-bond acceptors	tPSA	Rotatable bonds	Apolar desolvation	Polar desolvation	Download
Reference	0	1	7	74	10	9.61	-15.93	Download

Vendors (46 Total)

80 Items Total

- Aldrich CPR [CDS022564](#)ALDRICH
- KeyOrganica [KS-1202](#)
- Bioactives
- MedChem Express [401-420621](#)

Annotated Catalogs (29 Total)

36 Items Total

- Illuminating the Druggable Genome Screening Library [Prestwick-6-B-11](#), [Selleck-1-D-3](#)
- ML SMR [136849132](#) [164175282](#)

Get Files for Docking from ZINC

zinc15.docking.org/substances/ZINC000001546066/

ZINC Substances Catalogs Tranches Biological+ More+ About+

/substances / ZINC000001546066-

ZINC1546066 (Erlotinib)

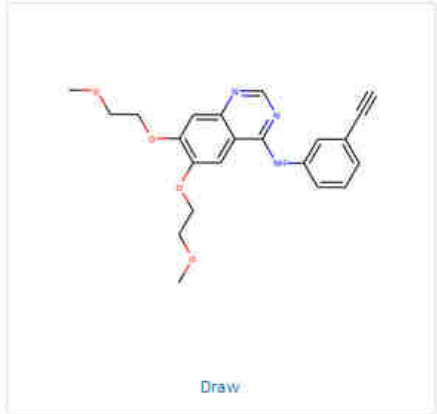
In: [bb](#) [fda](#) [for-sale](#) [in-stock](#) [standard](#)
[Google](#) [Wikipedia](#) [PubMed](#)

Added	Available	Since	Mw	logP	Heavy Atoms	Tranche	Download
2004-10-06	In-Stock	2015-08-07	393.443	3.405	29	GGEA	↓

SMILES: C#Cc1cccc(Nc2ncnc3cc(OCCOC)c(OCCOC)cc23)c1

InChI: InChI=1S/C22H23N3O4/c1-4-16-6-5-7-17(12-16)25-22-18-13-20(28-10-8-26-2)21(29-11-9-27-3)14-19(18)23-15-24-22/h1,5-7,12-15H,8-11H2,2-3H3,(H,23,24,25)

InChI Key: AAKJLRGGTJKAMG-UHFFFAOYSA-N



Draw

Available 3D Representations [Find Decoys](#)

pH range	Net charge	H-bond donors	H-bond acceptors	TPSA	Rotatable bonds	Apolar desolvation	Polar desolvation	Download
Reference	0	1	7	74	10	9.61	-15.93	↓

Vendors (46 Total) [80 Items Total](#)

[Aldrich CPR](#) [CDS022564/ALDRICH](#)
[KeyOrganics](#) [KS-1202](#)
[files.docking.org/protomers/16/89/55/209168955.db2.gz](#)

Annotated Catalogs (29 Total)

[Illuminating the Druggable Genome Screening Library](#) [Prestwick-6-B-11, Selleck-1-D-3](#)
[MILSMR](#) [136949132_164175282](#)

- DB
- DB2
- Mol2
- Solvation

wget http://files.docking.org/protomers/16/89/55/209168955.db2.gz

ls /path/tutorial_for_webinar/dock3.7/209168955.db2.gz > ligands.sdi₁₃

Get Files Needed for Docking

Get the link from the zinc webpage and use wget to download:

```
wget http://files.docking.org/protomers/16/89/55/209168955.db2.gz
```

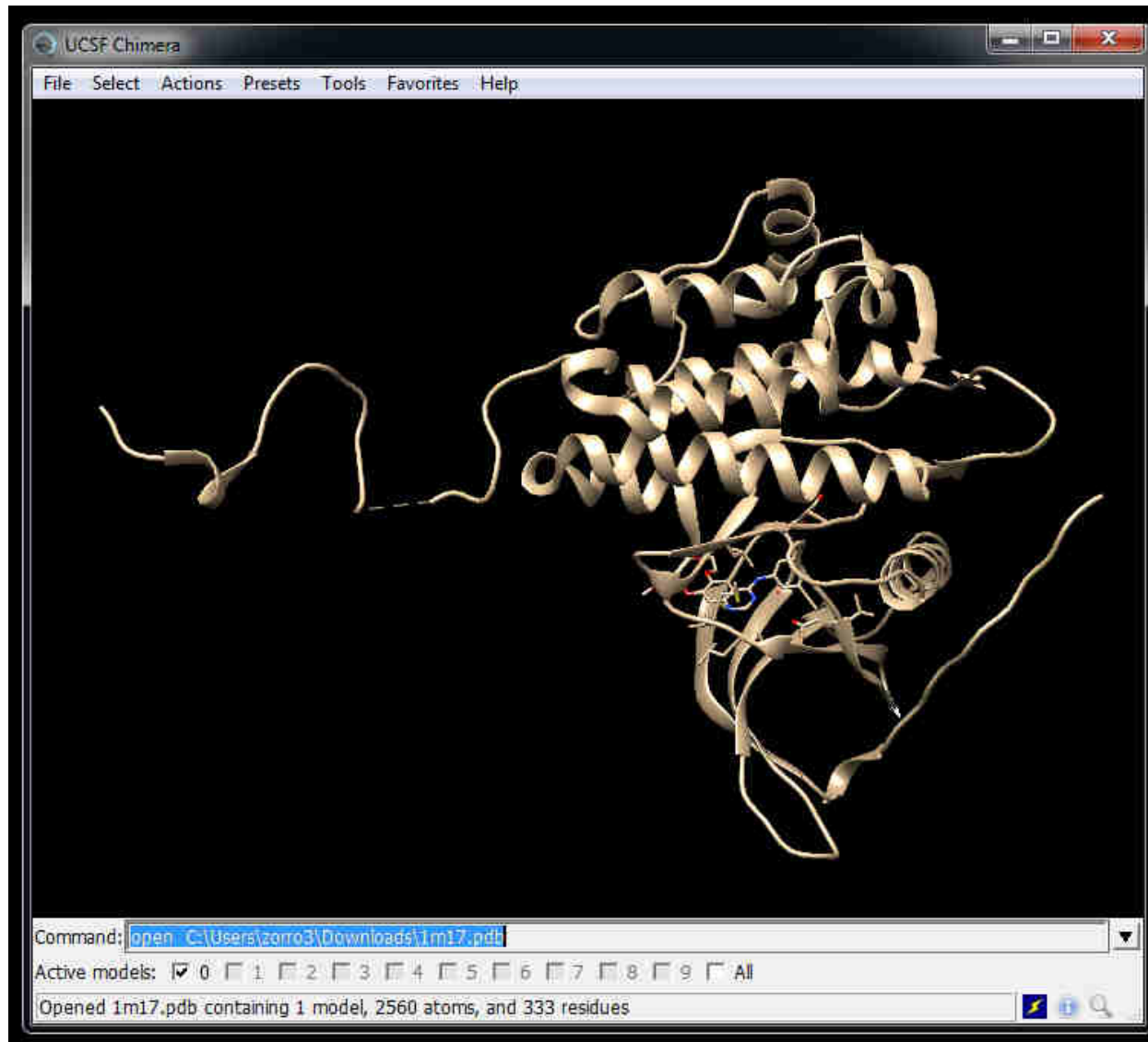
Put the path of the downloaded database into the split database index file (this file usually contains many db2 files):

```
ls /path/tutorial_for_webinar/dock3.7/209168955.db2.gz > ligands.sdi
```

Get the receptor structure from the PDB website

```
wget https://files.rcsb.org/download/1M17.pdb --no-check-certificate
```

Break Xtal into Receptor and Ligand Files



You may use a program like Chimera for this

Receptor file must be called: `rec.pdb`

Ligand file: `xtal-lig.pdb`

What if the crystal does not have a ligand:

Place atoms in the site where you want to dock. One way is to run `sphgen` and selecting spheres near residues in the site convert to `pdb`

Prepare Receptor for Docking

Make the receptor file (remove alternative side chains):

```
grep "^ATOM" 1M17.pdb | grep -v ^.....B > rec.pdb
```

Make ligand file:

```
grep AQ4 1M17.pdb | sed -e 's/HETATM/ATOM /g' > xtal-lig.pdb
```

Run blastermaster: input rec.pdb, xtal-lig.pdb and makes all receptor file need for docking.

```
python $DOCKBASE/proteins/blastermaster/blastermaster.py --addhOptions="-HIS -FLIPs " -v
```

This command may take several minutes to run.

Prepare Receptor for Docking Output

```
-rw-r--r--. 1 tbalius bks 3163 Jun 17 12:27 INDOCK
```

```
dockfiles/:
```

```
total 30388
```

```
-rw-r--r--. 1 tbalius bks 1206051 Jun 17 12:27 ligand.desolv.heavy  
-rw-r--r--. 1 tbalius bks 1206051 Jun 17 12:27 ligand.desolv.hydrogen  
-rw-r--r--. 1 tbalius bks 3376 Jun 17 12:27 matching_spheres.sph  
-rw-r--r--. 1 tbalius bks 908086 Jun 17 12:27 trim.electrostatics.phi  
-rw-r--r--. 1 tbalius bks 3121095 Jun 17 12:27 vdw.bmp  
-rw-r--r--. 1 tbalius bks 1653 Jun 17 12:27 vdw.parms.amb.mindock  
-rw-r--r--. 1 tbalius bks 24660016 Jun 17 12:27 vdw.vdw
```

Modifying INDOCK File

match_goal	5000	How much orienting to do Reduce 1000 if docking takes to long
number_save	1	Number of poses to write out
number_write	1	Consider writing out 100
#	MINIMIZATION	
minimize	no	Minimize the 6 degrees for the poses written out (3 rotation, 3 translation)
sim_itmax	500	
sim_trnstep	0.2	All molecules written out will be minimized
sim_rotstep	5.0	
sim_need_to_restart	1.0	
sim_cnvrge	0.1	
min_cut	1.0e15	
iseed	777	

Run Docking Calculations

Make perpare docking directories.

```
$DOCKBASE/docking/setup/setup_db2_zinc15_file_number.py ./ ligand  
ligand.sdi 500 count
```

Submit jobs to queue (we use SGE queuing system):

```
$DOCKBASE/docking/submit/submit.csh
```

To analyze the results we need to combine the results and then get poses

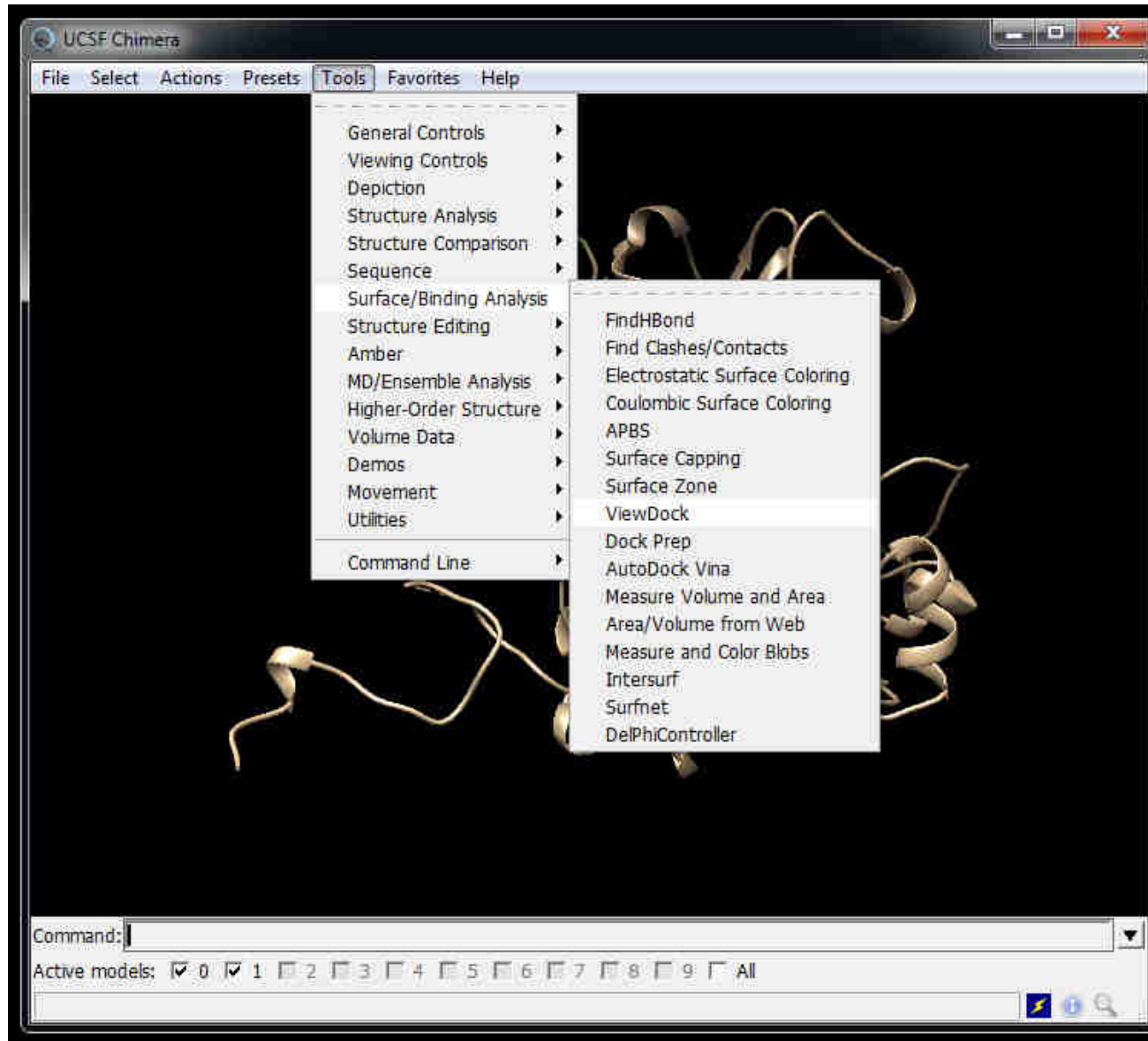
```
$DOCKBASE/analysis/extract_all.py  
$DOCKBASE/analysis/getposes.py
```

It is also possible to run dock locally:

```
$DOCKBASE/docking/DOCK/bin/dock64
```

Output: OUTDOCK and test.mol2.gz

Visualize poses in Chimera with Viewdock



Visualize Poses in Chimera with Viewdock

The image displays the UCSF Chimera software interface with a protein structure and a docked ligand. Two ViewDock windows are open, showing docking results for different poses. The first window shows results for Chimera Model #3.2, and the second window shows results for Chimera Model #2.101.

ViewDock - Chimera Model #3.2

S	Number	Name	Total Energy▲	Matchnum	Setnum
V	2	ZINC000001546066	-37.793495	2736	224
V	1	ZINC000001546066	-29.297897	3385	290

ViewDock - Chimera Model #2.101

S	Number	Name	Total Energy▲	Matchnum	Setnum
V	2	ZINC000001546066	-41.285645	2736	260
V	2	ZINC000001546066	-40.456554	1386	224
V	2	ZINC000001546066	-40.06303	2504	494
V	2	ZINC000001546066	-39.777264	2736	494
V	2	ZINC000001546066	-39.10701	966	494
V	2	ZINC000001546066	-38.780209	966	137
V	2	ZINC000001546066	-38.580563	2739	99
V	2	ZINC000001546066	-38.541042	2736	224

Command: |focus #1
Active models: 0 1 2 3 4 5 6 7 8 9 All
Session written

Pose with and without minimization: Energy -37.79 -> -38.54

Visualize Poses in Chimera with Viewdock

The image shows a screenshot of the UCSF Chimera software interface. The main window displays a 3D ribbon representation of a protein structure in tan, with a ligand molecule docked in the binding pocket. The ligand is shown in a stick representation with atoms colored by element: carbon (grey), oxygen (red), nitrogen (blue), and sulfur (yellow). The ViewDock panel on the right shows a table of docking results for a ZINC000001546066 ligand. The table lists 10 poses, with the top pose (Setnum 259) highlighted in blue, indicating it is the best pose. Below the table, the Chimera Model #2.104 details are shown, including the name, protonation state, SMILES string, and long name.

S	Number	Name	Total Energy▲	Matchnum	Setnum
V	2	ZINC000001546066	-37.793495	2736	224
V	1	ZINC000001546066	-29.297897	3385	290

No compound chosen

Change Compound State

Viable Deleted Purged

Hide Quit Help

S	Number	Name	Total Energy▲	Matchnum	Setnum
V	2	ZINC000001546066	-41.315582	2736	259
V	2	ZINC000001546066	-41.285645	2736	260
V	2	ZINC000001546066	-40.456554	1386	224
V	2	ZINC000001546066	-40.06303	2504	494
V	2	ZINC000001546066	-39.777264	2736	494
V	2	ZINC000001546066	-39.10701	966	494
V	2	ZINC000001546066	-38.780209	966	137
V	2	ZINC000001546066	-38.580563	2739	99

Chimera Model #2.104

MinStatus: YES
Name: ZINC000001546066
Protonation: none
SMILES: C#CC1=CC=CC(NC2=C3C=C(OCCOC)C(OCCOC)=CC3=NC=N2)
)=C1
Long Name: NO_LONG_NAME
FlexResCode: 1

Change Compound State

Viable Deleted Purged

Hide Quit Help

Command: focus #1

Active models: 0 1 2 3 4 5 6 7 8 9 All

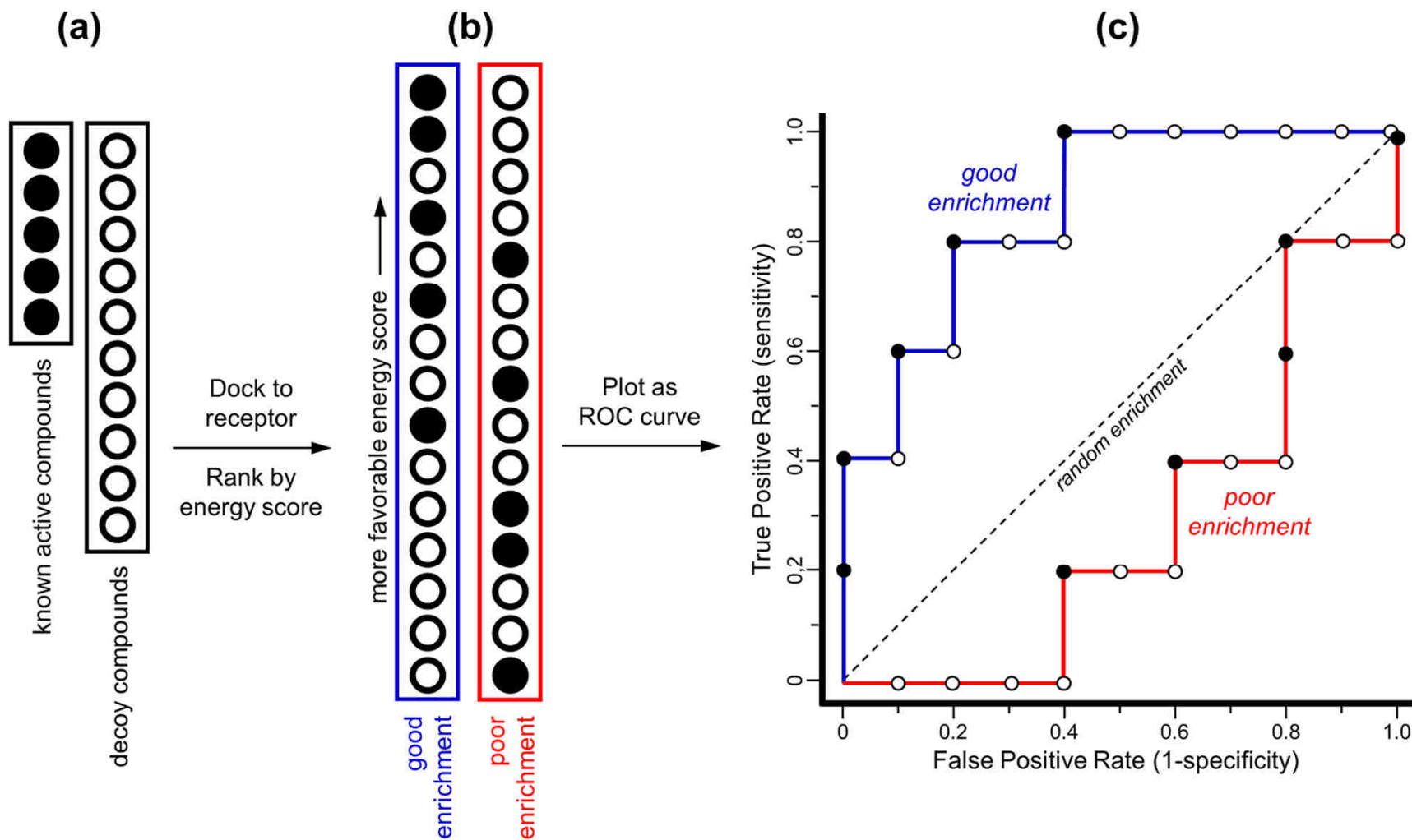
Best pose out of the top 100 after min: Energy -41.31

Tutorial

Scenario 2:

Use docking to test enrichment capabilities of Epidermal Growth Factor Receptor using 12 ligands and DUD-E property matched decoys

Enrichments: ROC Curves



Get Known Ligands for Docking

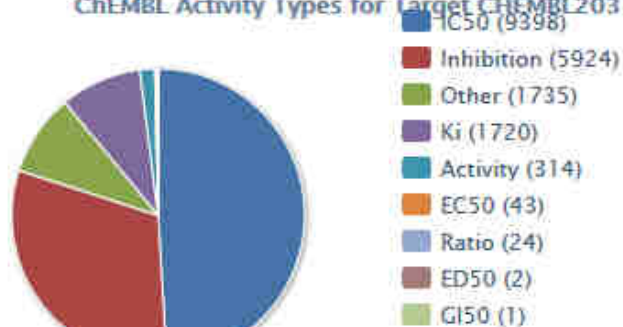
<https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL203>

More...

<https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL203>

Target Associated Bioactivities

ChEMBL Activity Types for Target CHEMBL203



Total: 19166

Target Associated Assays

ChEMBL Assays for Target CHEMBL203



Total: 1550

Target Ligand Efficiencies

Get Known Ligands for Docking

https://www.ebi.ac.uk/chembl/bioactivity/results/1/cmpd_chemblid/asc/tab/display

67%

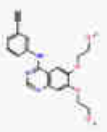
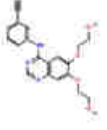
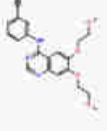

Search ChEMBL

Compounds Targets Assays Documents Cells Tissues Exact Match Active Source Filter

ChEMBL Bioactivity Search Results: 178

10 records per page


Please select...
Please select...
Download All Bioactivity Data (Tab-delimited)
Download All Bioactivity Data (XLS)

Ingredient	Molweight	Canonical SMILES	Standard Type	Relation	Standard Value	Standard Units	pChEMBL Value	Assay Type	Description	Assay Src	Assay Organism	Target Type	Target Name	Target Organism	Reference
 CHEMBL553	393.44	<chem>COCCOc1cc2nnc(Nc3cccc(c3)C#C)c2cc1OCCOC</chem>	Ki	=	3	nM	9.52	E	Inhibition of EGFR L858R mutant (unknown origin)	Scientific Literature	Homo sapiens	PROTEIN FAMILY	Epidermal growth factor receptor	Homo sapiens	Biores. Med. Chem. Lett. (2015) 26:2-534
 CHEMBL553	393.44	<chem>COCCOc1cc2nnc(Nc3cccc(c3)C#C)c2cc1OCCOC</chem>	Ki	=	138	nM	6.87	E	Inhibition of recombinant human EGFR T790M/L858R double mutant by mass spectrometric analysis	Scientific Literature	Homo sapiens	PROTEIN FAMILY	Epidermal growth factor receptor	Homo sapiens	ACS Med. Chem. Lett. (2015) 7:1-100
 CHEMBL553	393.44	<chem>COCCOc1cc2nnc(Nc3cccc(c3)C#C)c2cc1OCCOC</chem>	Ki	=	95	nM	7.02	E	Inhibition of EGFR T790M/del746 (746 to 750 residues) mutant (unknown origin)	Scientific Literature	Homo sapiens	PROTEIN FAMILY	Epidermal growth factor receptor	Homo sapiens	Biores. Med. Chem. Lett. (2015) 26:2-534
 CHEMBL553	393.44	<chem>COCCOc1cc2nnc(Nc3cccc(c3)C#C)c2cc1OCCOC</chem>	Ki	=	95	nM	7.02	E	Inhibition of EGFR del746 to 750 mutant (unknown origin)	Scientific Literature	Homo sapiens	PROTEIN FAMILY	Epidermal growth factor receptor	Homo sapiens	J. Med. Chem. (2015) 58:1000-1010

Generate Decoy Smiles File

<http://dude.docking.org/generate>

UCSF | University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center



[Home](#) | [Targets](#) | [Subsets](#) | [Generate](#) | [Other](#) | [FAQ](#) | [Revisions](#) | [Thanks](#)

Generate DUD•E Decoys

To generate decoys for your active compounds, use our free on-line system below. Here is how:

1. Paste a list of SMILES in the window below or choose a text file to upload. In each case, the format is one SMILES per line, optionally followed by white space and an identifier.
2. Provide your email address to which the results should be sent. We do not retain this information. If you are logged in to docking.org, this will be filled in for you.
3. Anonymous users must complete a CAPTCHA.
4. Click **Generate Decoys**.

Result Notification

Send results to [Login](#) to use your docking.org account

Human Text: How are you today? (anything is fine)

Input SMILES by pasting OR uploading a text file

Upload SMILES No file selected.

As an example, the following lines are accepted formats:

```
catnip[1] (C)C1=CC=C(C=C1)C2=CC=CC=C2C3=CC=CC=C3  
CCCC1=CC=CC=C1C2=CC=CC=C2C3=CC=CC=C3  
CC1=CC=CC=C1C2=CC=CC=C2C3=CC=CC=C3
```

A list of 45 SMILES will take about 10 minutes on average, but can take up to a few hours depending on system load.

Building Databases Locally

If the system is in DUD-E, You may download ready to dock databases here:

<http://autodude.docking.org/>

Here I just used the first 12 ligands from ChEMBL

Generated decoys using DUD-E webserver

Use the link in email:

```
wget http://dude.docking.org/generate/results/4094969748/dude-decoys.tar.gz
tar -xzf dude-decoys.tar.gz
```

```
grep -v ligand dude-decoys/decoys/decoys.P*.picked | awk -F: '{print $2}' | awk '{print
$1 " " $2}' > ! decoys.smi
```

```
${DOCKBASE}/ligand/generate/build_database_ligand.sh -H 7.4
ligands_12_from_chembl.smi
```

```
csh wrapper_queue_build_smiles_ligand_mod_corina.csh decoys.smi
```

http://wiki.docking.org/index.php/Ligand_preparation_-_20170424

Perform Enrichment Calculations

Make a list of all the databases:

```
ls /path/databases/ligands_12_from_chembl/CHEMBL*/*.db2.gz  
/path/databases/decoys/sgejob_*/finished/C*/*.db2.gz > ! ligands_decoys.sdi
```

```
awk '{print $2}' databases/ligands_12_from_chembl.smi >  
databases/ligands_names.txt
```

```
awk '{print $2}' databases/decoys.smi > databases/decoys_names.txt
```

Make directories for docking:

```
$DOCKBASE/docking/setup/setup_db2_zinc15_file_number.py ./ ligands_decoys  
databases/ligands_decoys.sdi 100 count
```

Submit docking jobs:

```
$DOCKBASE/docking/submit/submit.csh
```

Perform Enrichment Calculations

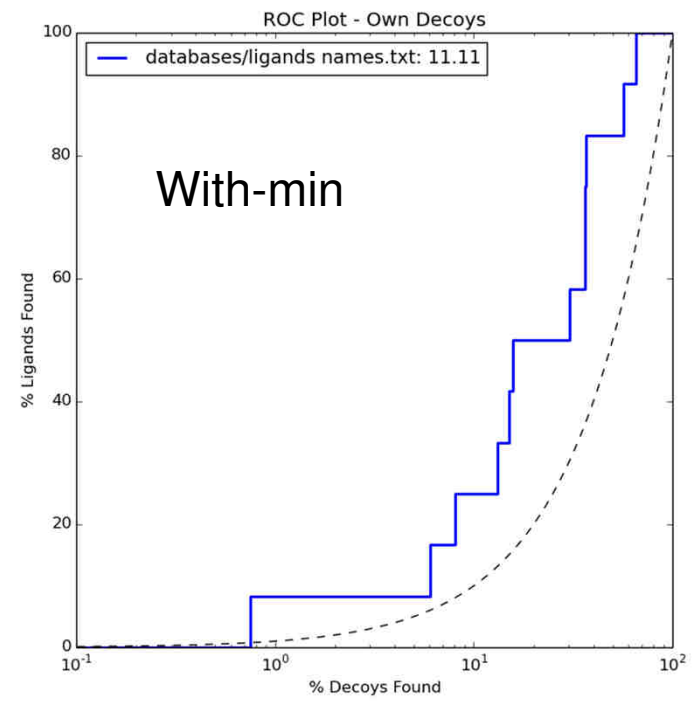
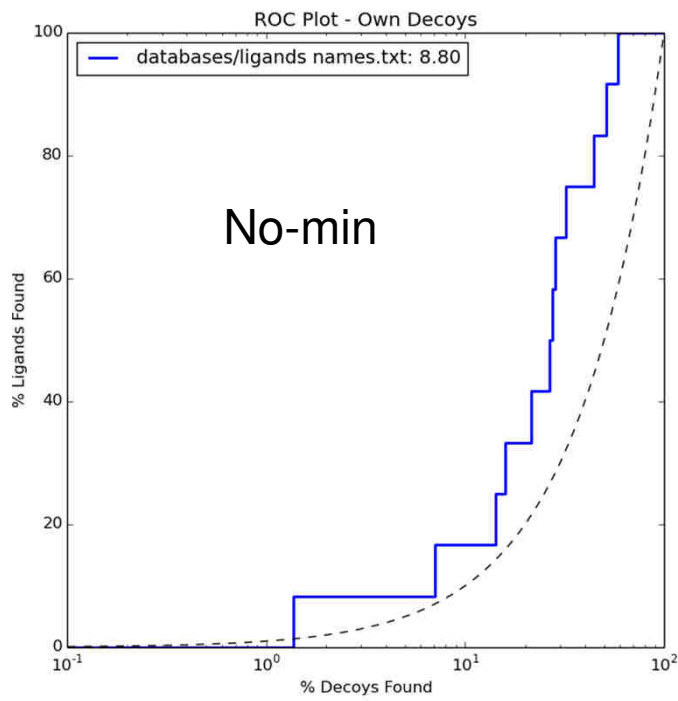
Process results combining results and get the best poses:

```
$DOCKBASE/analysis/extract_all.py  
$DOCKBASE/analysis/getposes.py
```

Calculate enrichments:

```
$DOCKBASE/analysis/enrich.py -i . -l databases/ligands_names.txt -d  
databases/decoys_names.txt  
$DOCKBASE/analysis/plots.py -i . -l databases/ligands_names.txt -d  
databases/decoys_names.txt
```

Perform Enrichment Calculations



Using ZINC for Drug Discovery

Once you are happy with the retrospective sanity checks now we can do screening.

Get Lead-like Molecules to DOCK

zinc15.docking.org/tranches/home/

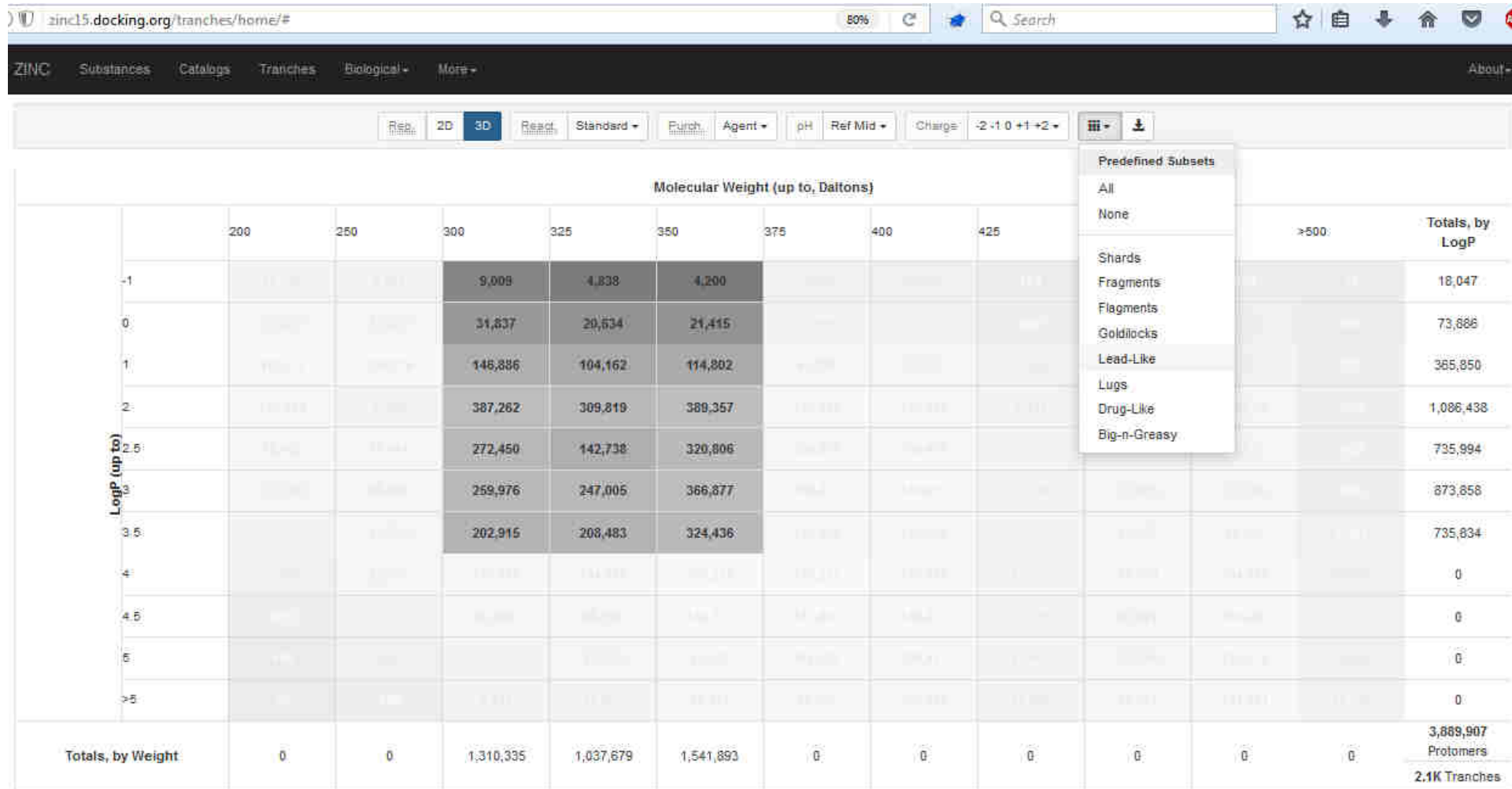
ZINC Substances Catalogs Tranches Biological+ More+ About+

Rep. 2D 3D Reac. Standard+ Purch. Wait OK+ pH N/A+ Charge N/A+ iii+ ↓

Molecular Weight (up to, Daltons)

	200	250	300	325	350	375	400	425	450	500	>500	Totals, by LogP
-1	25,720	50,752	133,898	91,084	99,566	77,162	67,427	29,709	20,474	5,113	6,554	607,259
0	128,914	458,683	1,248,854	838,607	880,249	677,930	582,351	262,117	176,626	23,137	3,432	5,280,900
1	364,947	1,603,156	5,825,851	4,253,289	4,589,207	3,660,337	3,157,643	1,460,639	993,997	124,796	5,710	26,039,572
2	483,333	3,108,015	13,672,003	11,584,885	12,063,732	11,164,701	9,965,279	4,769,007	3,376,617	496,995	13,860	70,698,427
2.5	170,855	1,586,885	8,638,205	8,385,476	10,207,399	9,029,529	8,420,071	4,332,250	3,163,503	526,672	15,087	54,475,932
3	93,943	1,220,939	8,034,742	8,666,780	11,340,608	10,506,090	10,092,542	5,662,871	4,282,350	781,952	25,601	60,708,418
3.5	31,178	745,938	5,943,808	7,372,527	10,435,684	10,476,209	10,451,334	6,541,952	5,087,277	1,039,924	43,736	58,169,567
4	6,586	294,964	3,259,154	4,785,175	7,607,690	8,547,769	9,089,143	6,455,950	5,199,175	1,219,997	65,974	46,531,577
4.5	1,005	43,836	1,096,781	2,182,136	4,072,349	5,374,413	6,238,196	5,173,230	4,373,741	1,214,056	93,577	29,865,321
5	145	4,693	181,753	572,091	1,421,241	2,844,769	4,042,678	3,132,224	2,858,306	1,516,287	111,402	16,685,579
>5	38	1,062	18,400	86,882	305,521	202,823	357,924	1,812,215	1,884,638	692,202	522,371	5,684,076
Totals, by Weight	1,306,667	9,118,923	48,055,249	48,818,932	63,023,246	62,561,722	62,464,587	39,432,164	31,416,704	7,641,130	907,304	374,746,628 Substances 1.9K Tranches

Get Lead-like Molecules to DOCK



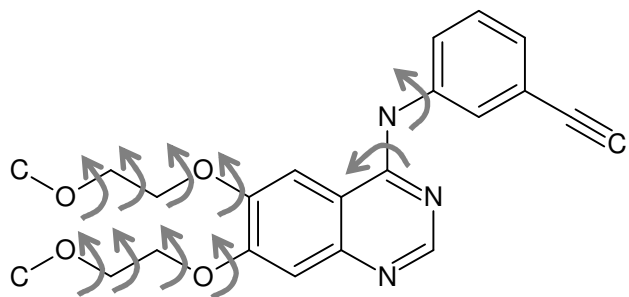
Features in DOCK 3.7

- DOCKovalent – used to screen databases of reactive molecules
- Receptor Desolvation – incorporated
- Minimization
- DOCK 3.7 is among the fastest docking software
- Purchasable chemical space continues to grow (Dock 100 million molecules)

Information About DOCK 6.8

Resources for DOCK 6.8

- Manual: http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm
- Tutorials: http://dock.compbio.ucsf.edu/DOCK_6/tutorials/index.htm
- Rizzo lab wiki:
http://ringo.ams.sunysb.edu/index.php/DOCK_Tutorials
- DOCK-fans email archive (list to search for solutions to problems or submit questions): <http://mailman.docking.org/pipermail/dock-fans/>
- Email dock-fans when you need advise or assistance:
dock-fans@docking.org



Sampling using Anchor-&-Grow

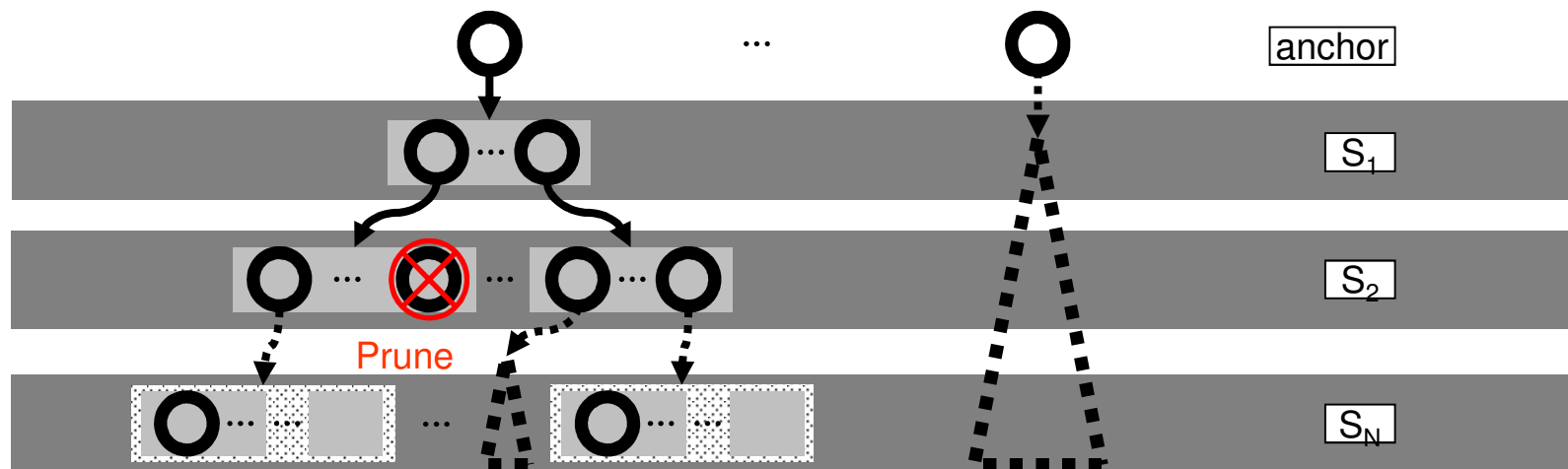
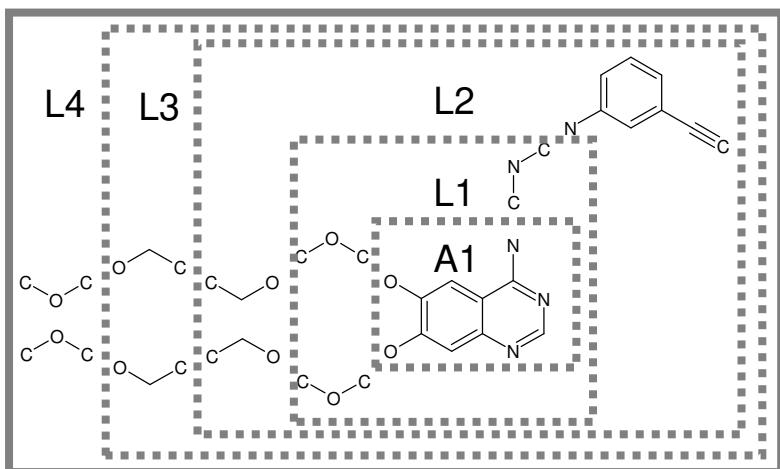
Start from anchor orient

Arrange segments by layer about anchor center

Grow each segment one at a time

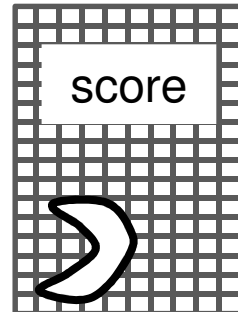
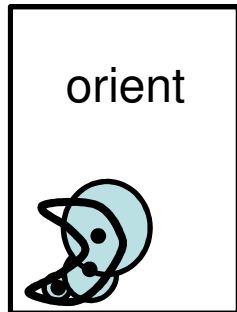
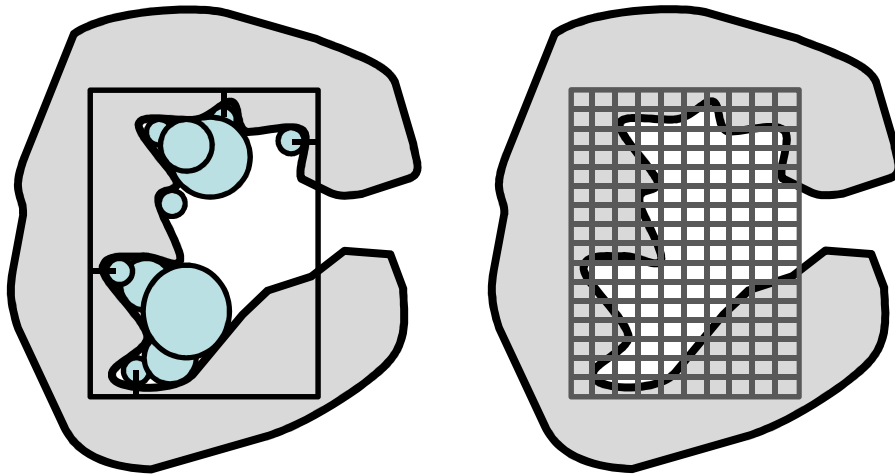
Grow segments in Layer 1

Grow segments in Layer 2 and so on

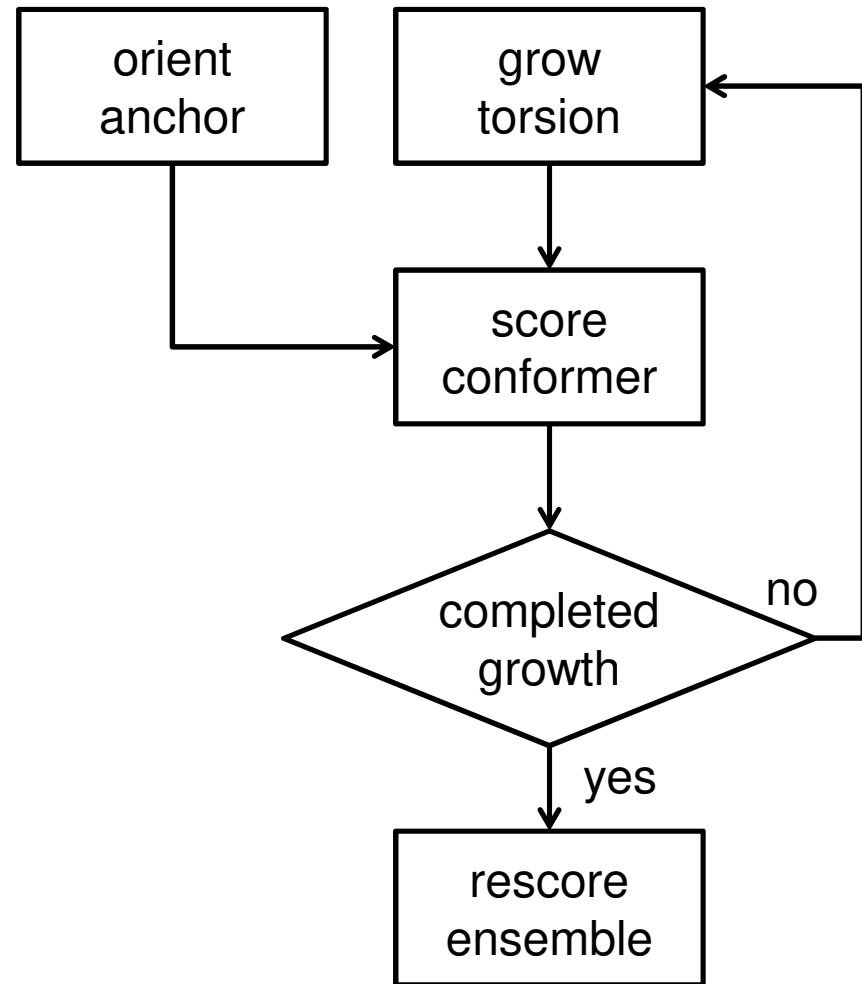


Sampling and Scoring

DOCK preparation



Scoring using a grid to speed up the calculations



DOCK 6 is a parallel program using MPI

DOCK6 Has Many Scoring Functions

- Bump Filter
- Contact Score
- **Grid-Based Score**
$$E_{Grid} = E_{VDW} + E_{ES}$$
- DOCK 3.5 Score
- Continuous Score
- Zou GB/SA Score
- Hawkins GB/SA Score
- PB/SA Score
- AMBER Score
- Footprint Score
- MultiGrid FPS Score
- Pharmacophore Matching Similarity Score
- SASA Score
- Descriptor Score
 - Tanimoto Score
 - Hungarian Matching Similarity Score
 - Volume Overlap Score

Features in DOCK 6.8

- Array of Scoring Function
 - New is Descriptor Score, Pharmacophore Score
- *De novo* design (will be released in future versions)
- Modular
- MPI parallel implementation

DOCK: where it is and where its going
Trent Balias, Ph.D.

Postdoctoral Scholar, Shoichet Lab
University of California, San Francisco

Tuesday, June 20th 12:00pm EDT

Upcoming Webinars:

July 25th: BioXTAS RAW with Jesse Hopkins

October 3rd: SPHIRE with Toshio Moriya

November 4th: EMAN2 with Steven Ludtke



SBGrid
CONSORTIUM

Events are posted on SBGrid Google Calendar (to join visit <http://sbgrid.org/calendar>)

Extra Slides

Rescoring: Footprint Similarity

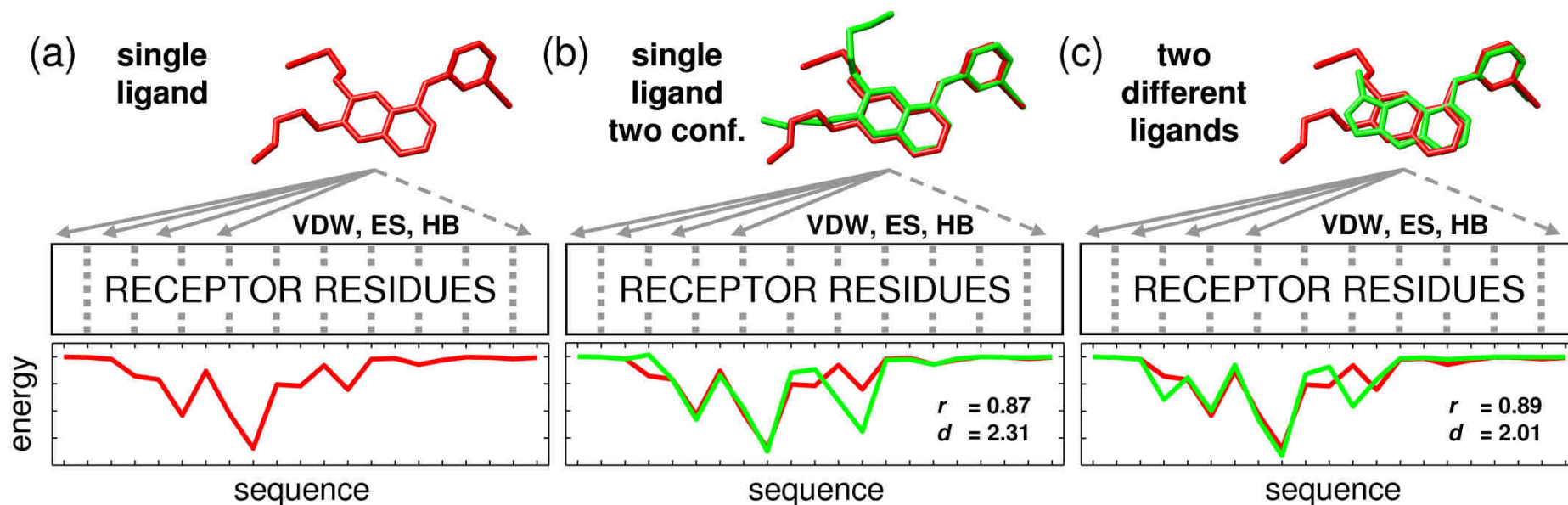


Table I. Examples of possible reference types to derive molecular footprints.

Reference Types	Description
Known inhibitor	FDA-approved drug or experimental inhibitor validated to bind
Natural substrate	Native peptide or cofactor
Transition state	Predicted transition state geometry for a chemical reaction
Modified structure	Key functionality/substructure (side-chain mediating protein-protein interactions)
Text file footprint	Modified entries to increase/decrease importance of select residues (resistance mutations)
Ensemble-weighted	Averaged footprints derived from MD/MC simulations

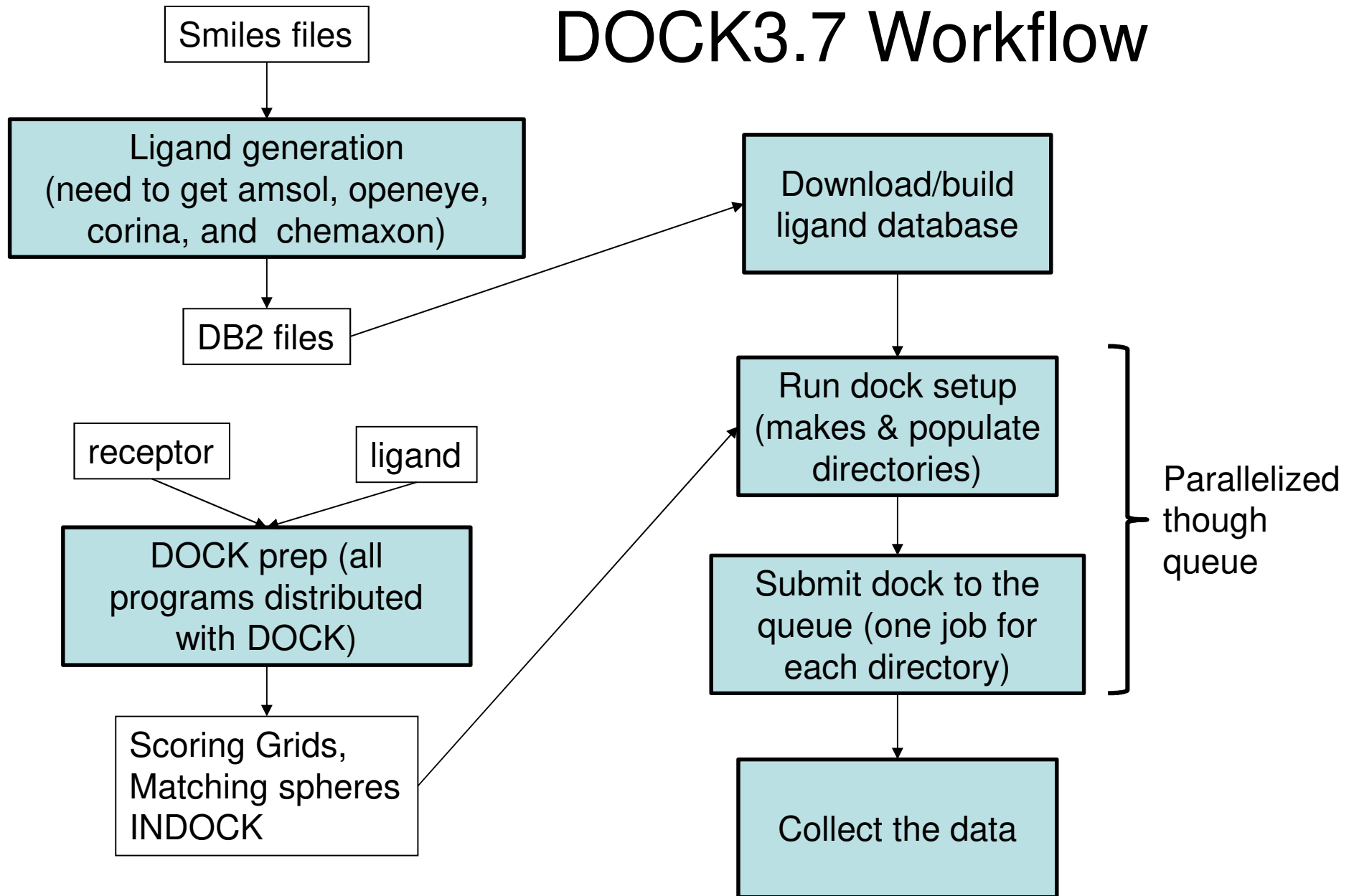
DOCK: Where it is and where its going

SBGrid Lunch Break

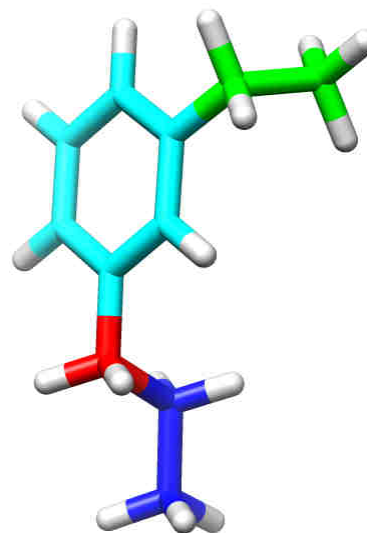
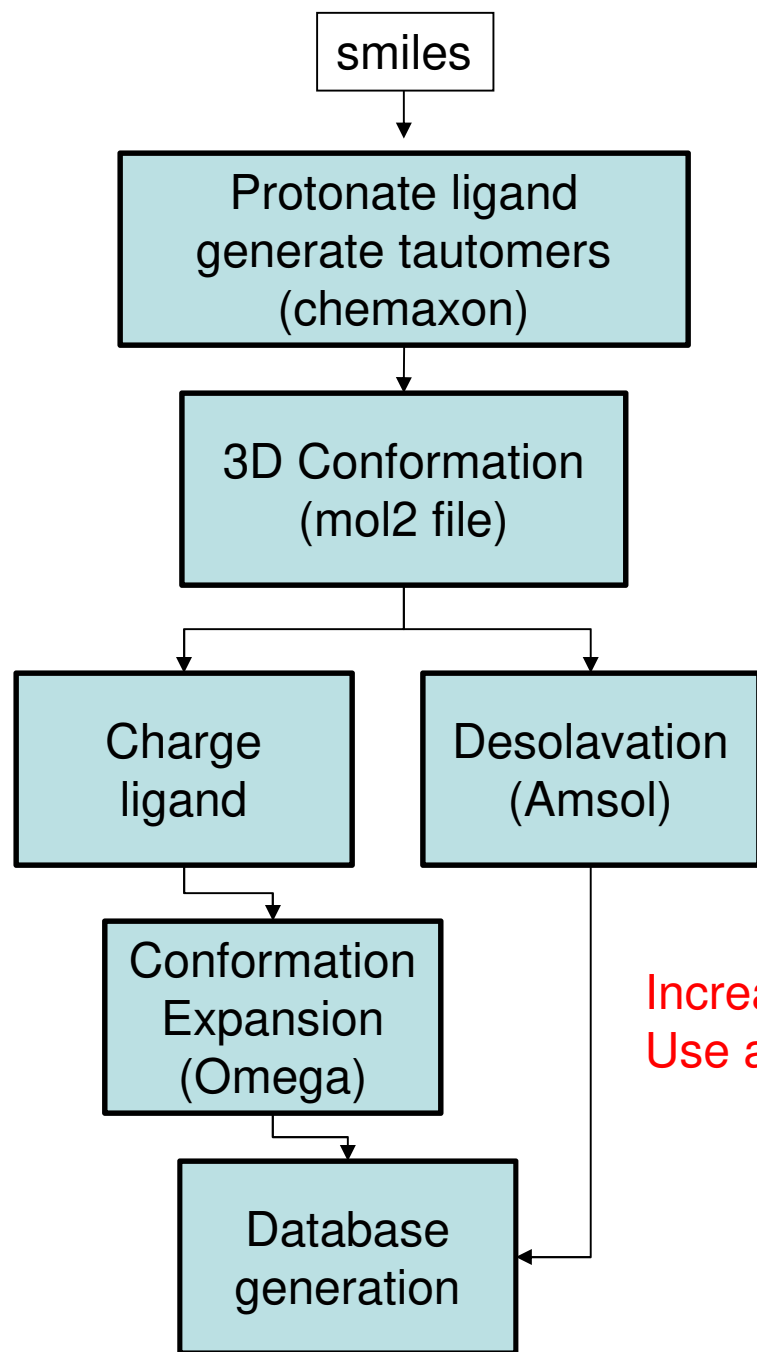
Trent Balius

June 20, 2017

DOCK3.7 Workflow



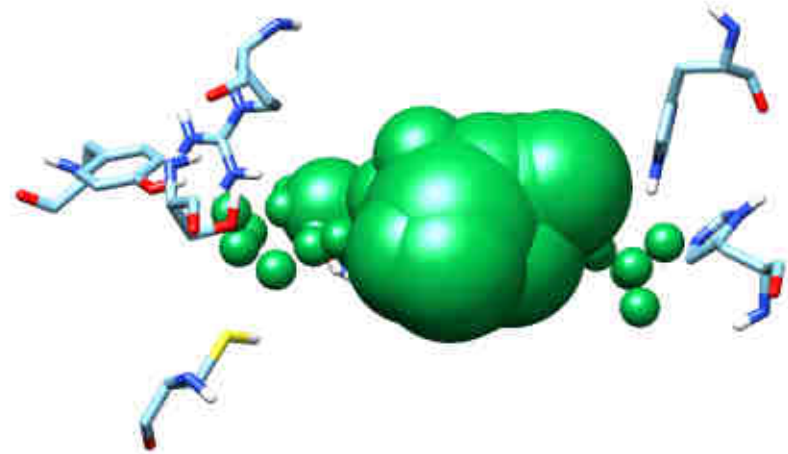
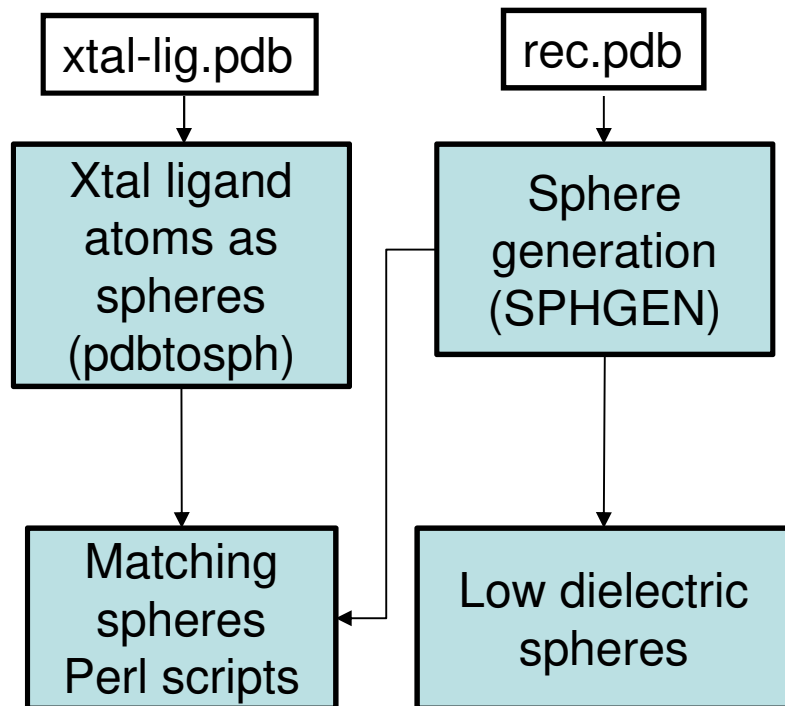
Ligand Preparation



DOCKable Ligand databases are available through **ZINC15**

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

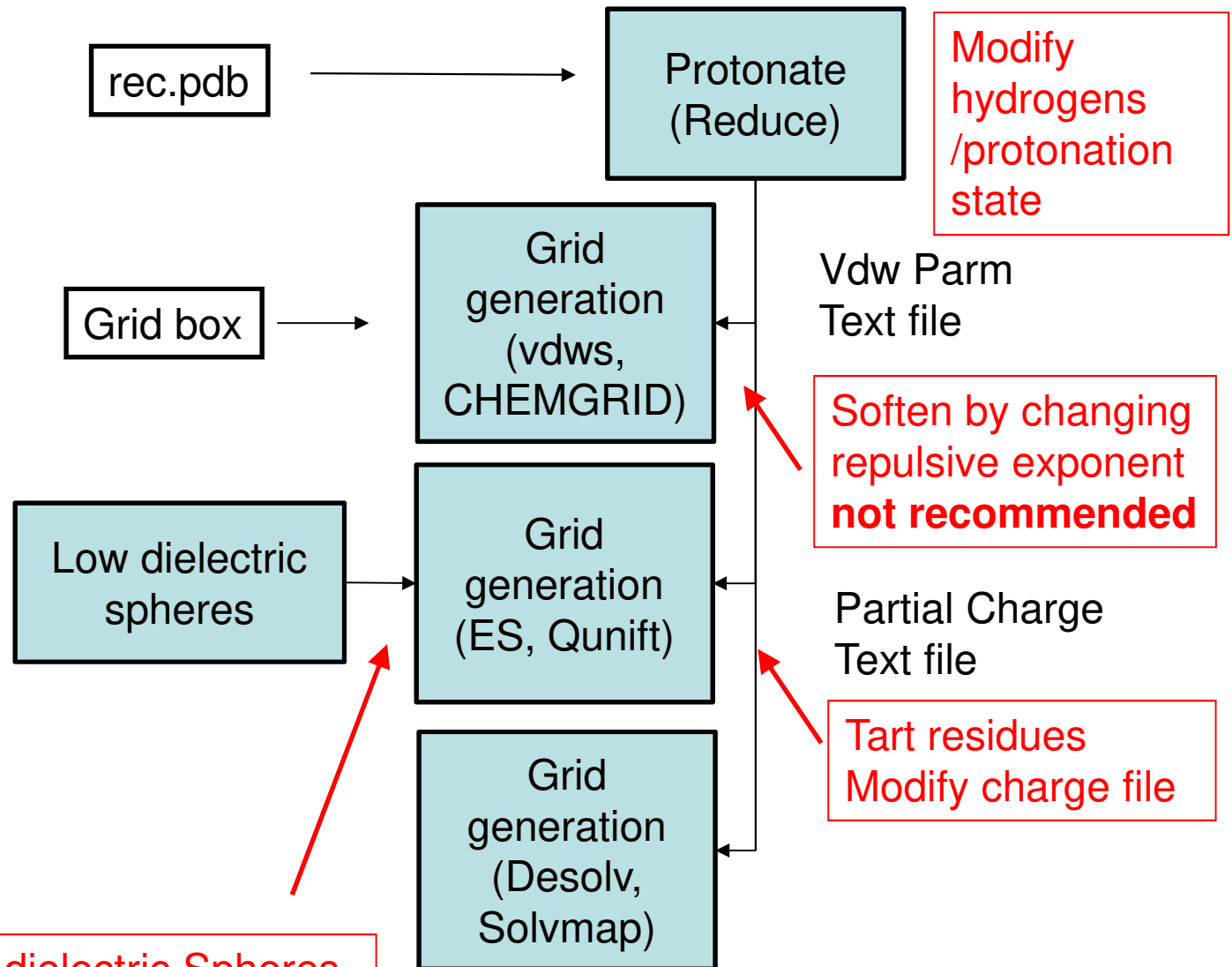
Sphere Generation: Receptor Prep.



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

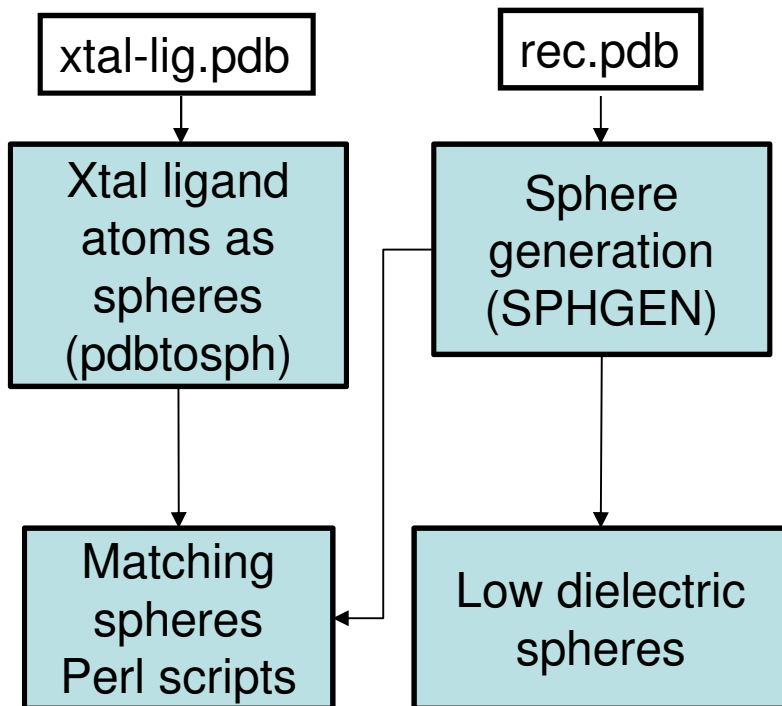
Add (or remove) Low dielectric Spheres
To regions where you would like to have less (more) screening

Grid Generation: Receptor preparation



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

Sphere Generation

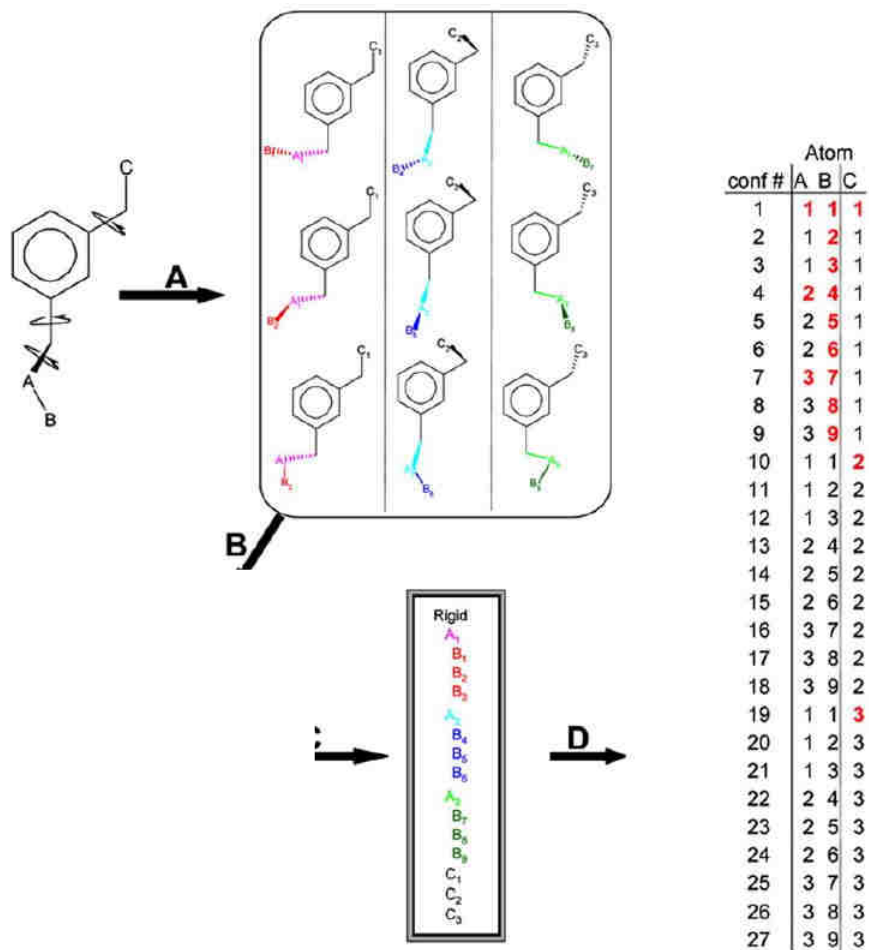


Modify Matching Spheres: add spheres where you would like to see rings, move existing spheres around

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

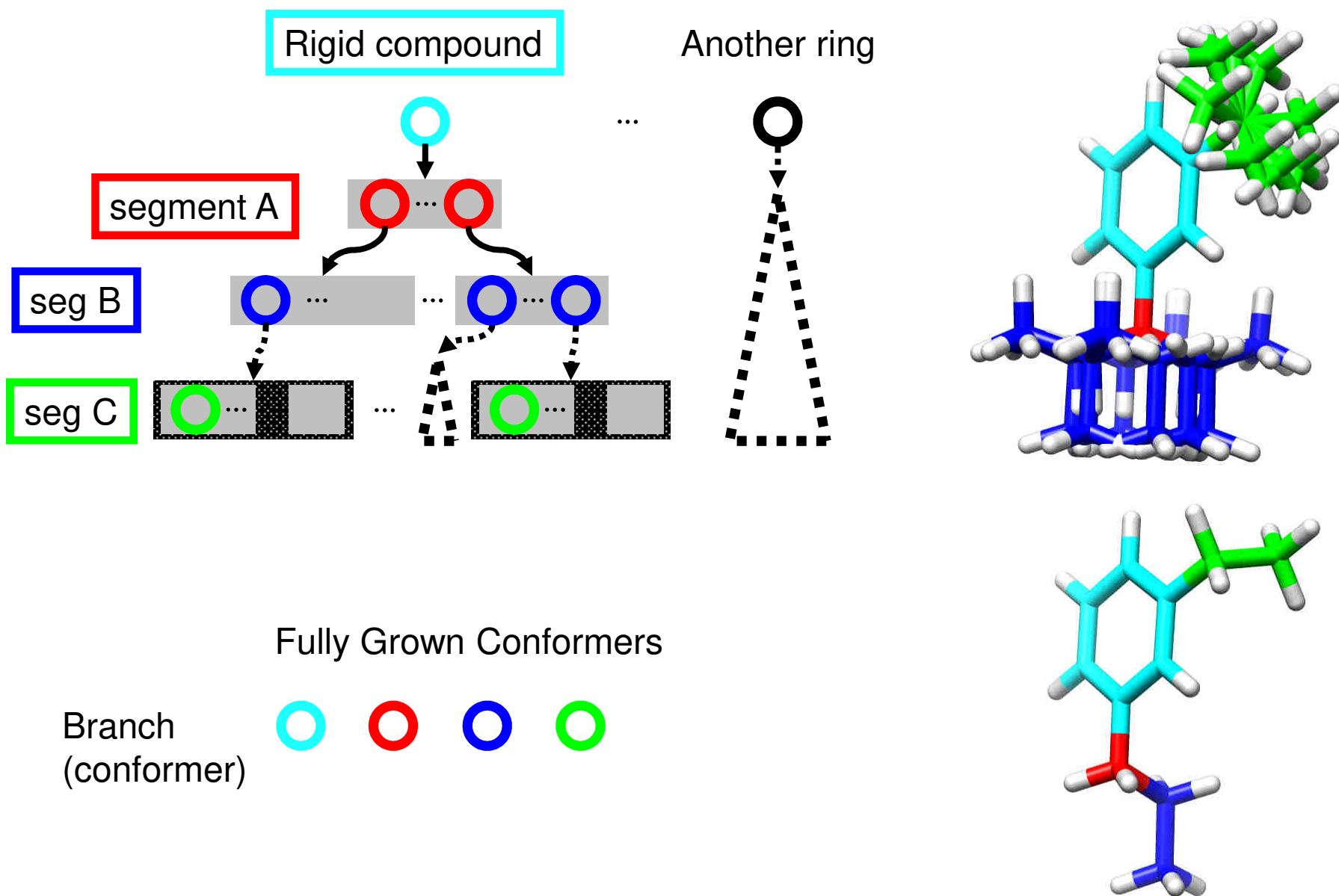
How Sampling Works (Internal)

Hierarchical docking of databases



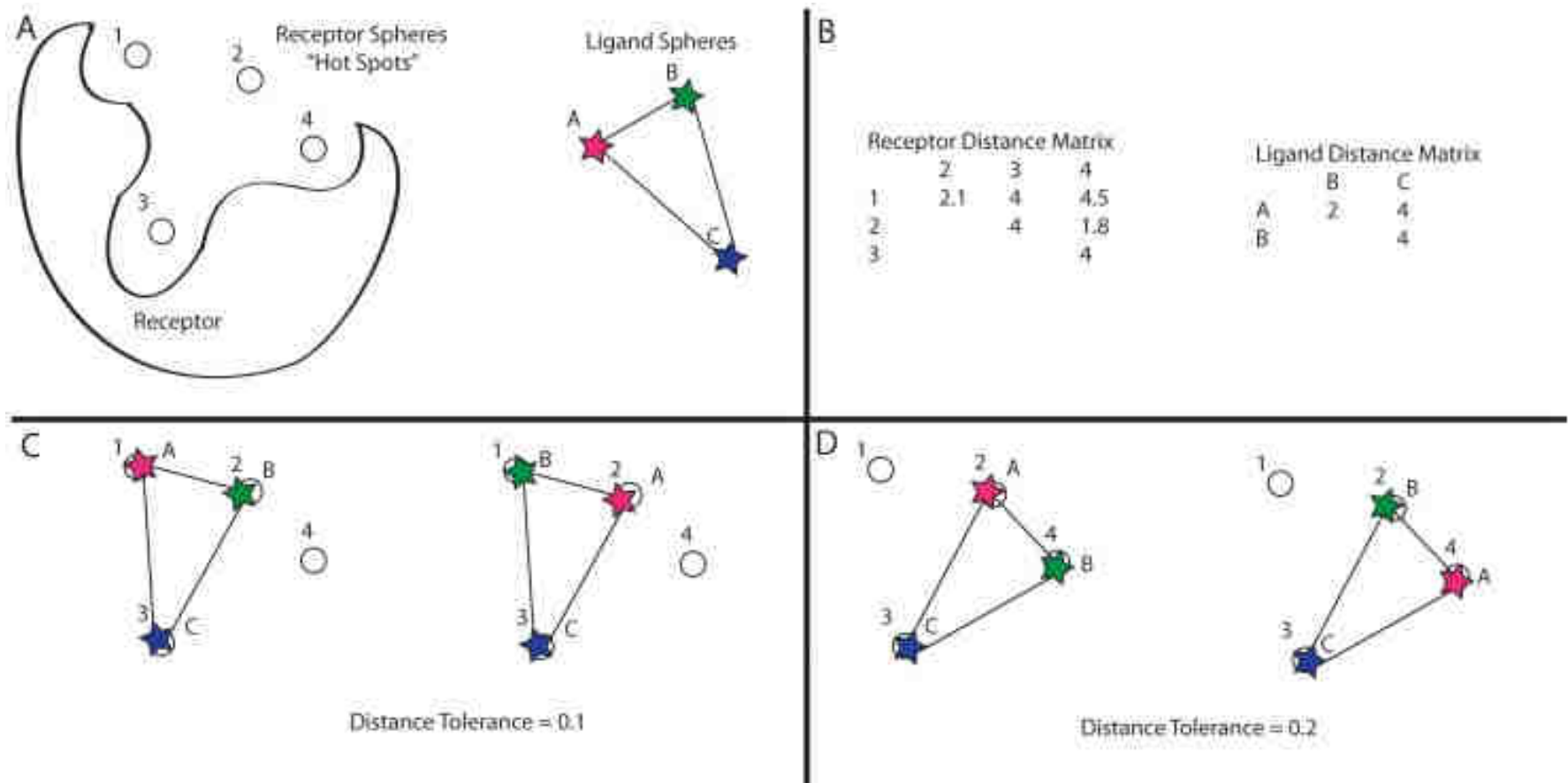
Lorber DM and Shoichet BK., Curr Top Med Chem. 2005;5(8):739-49.

Ligand Sampling – Database Construction

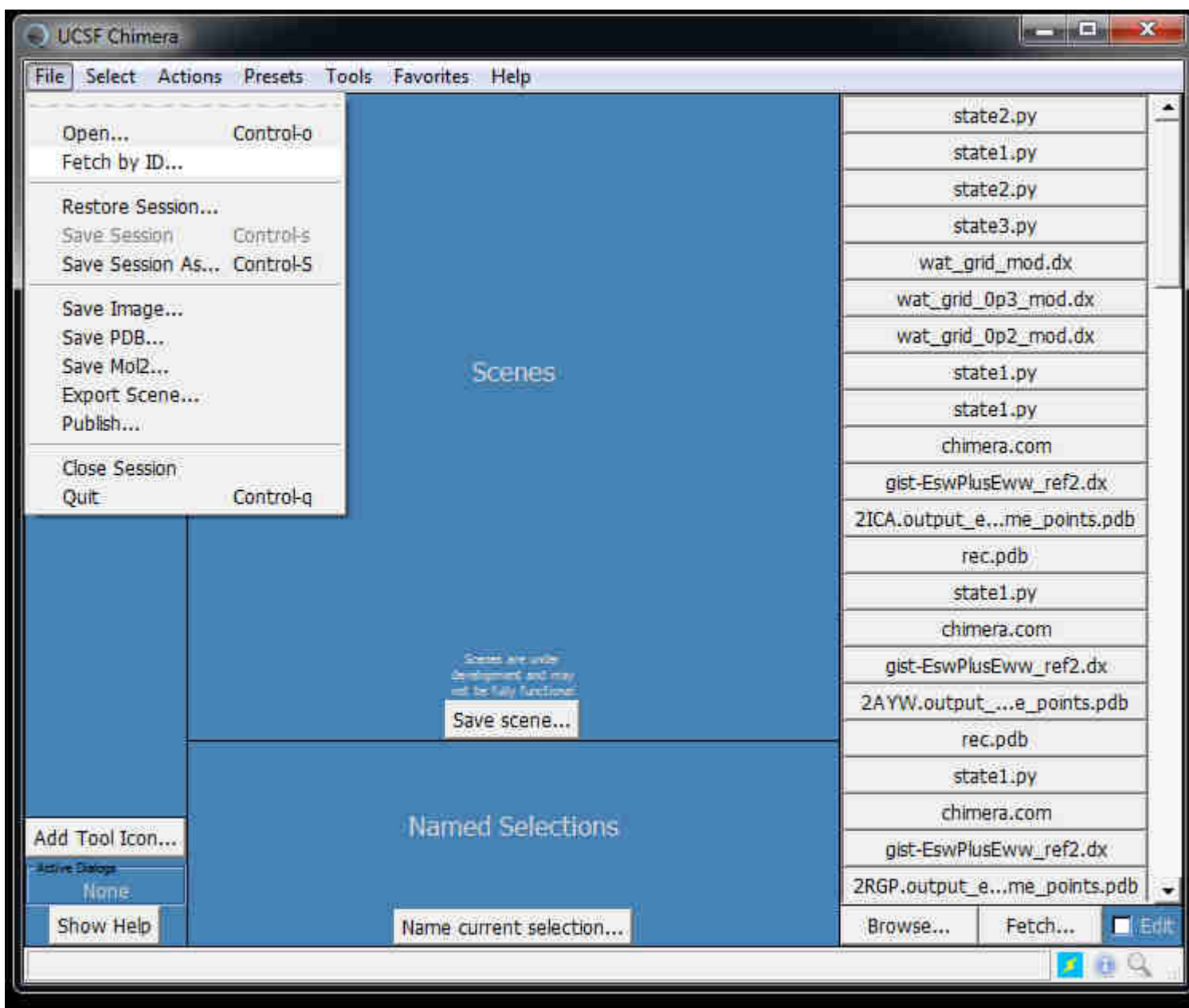


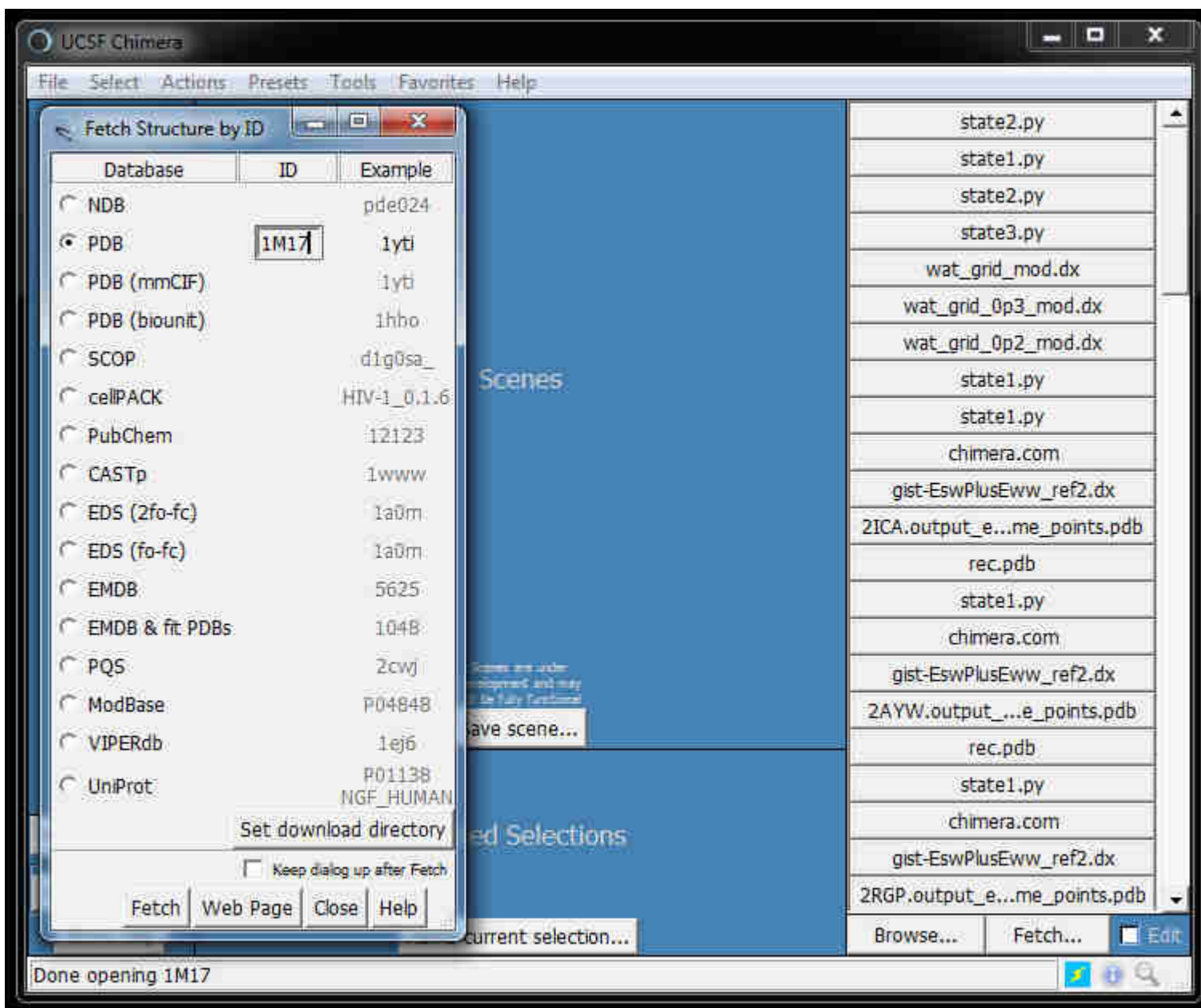
How Sampling Works (Orientational)

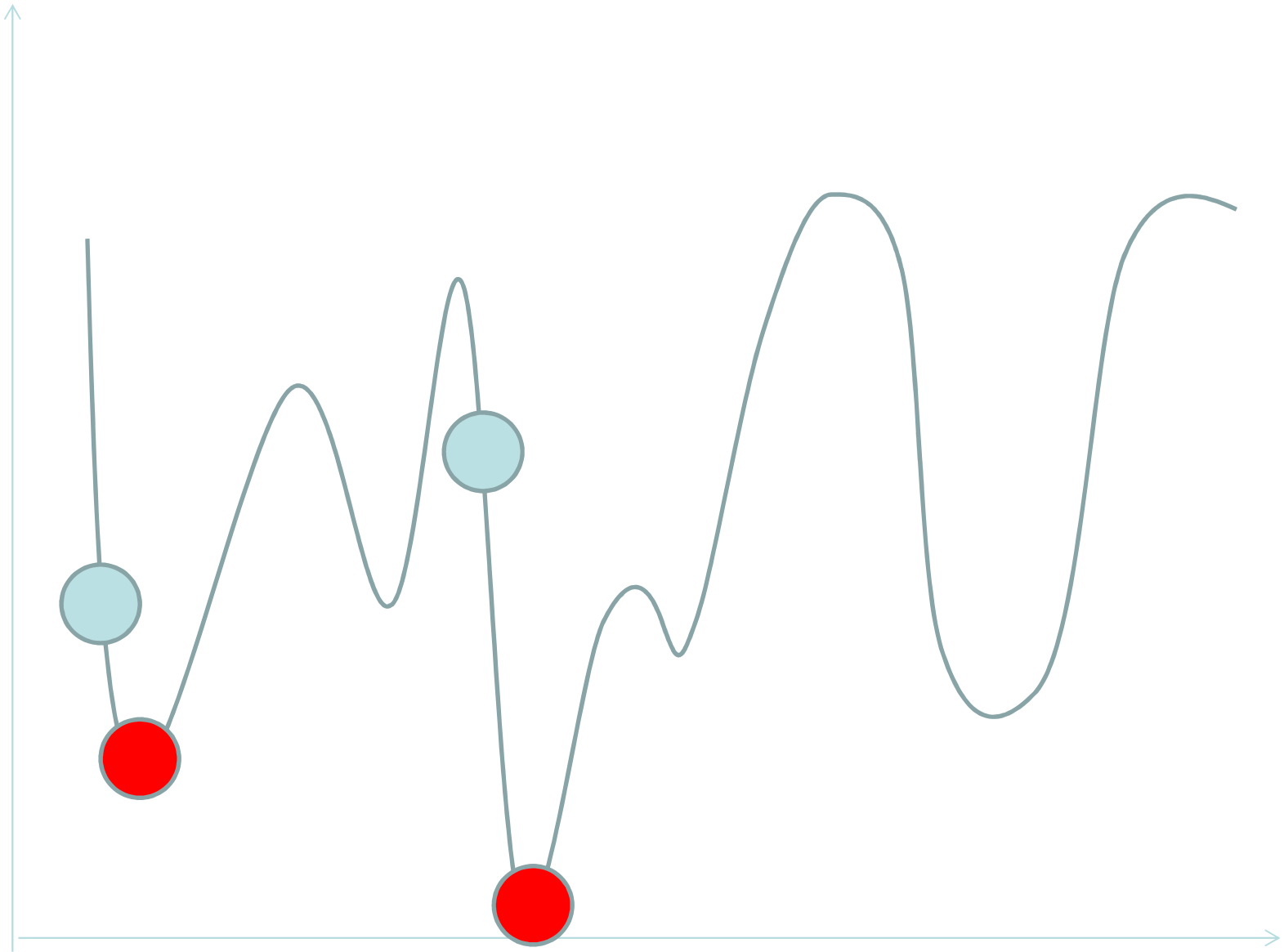
A toy example illustrating the matching sphere orientational matching algorithm



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

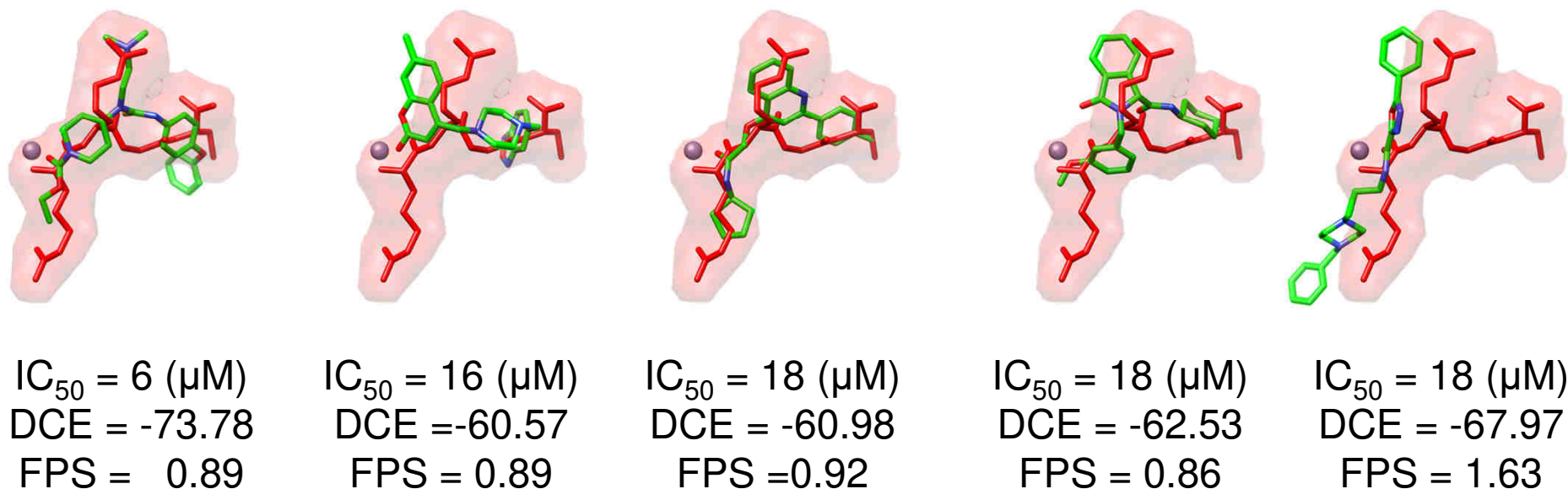






Lab Projects Employing FPS: Bot NT/A

~900,000 screened → 99 purchased → 5 active (picked using FPS)

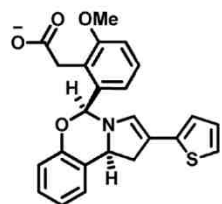
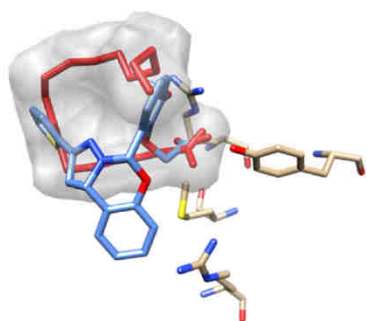


Preliminary results, work in progress: Bill Berger, Trent Balias, Eduard Melief, Peter Tonge, Subramanyam Swaminathan, Iwao Ojima, and Robert Rizzo

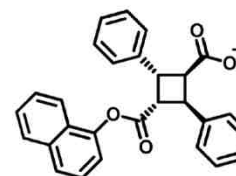
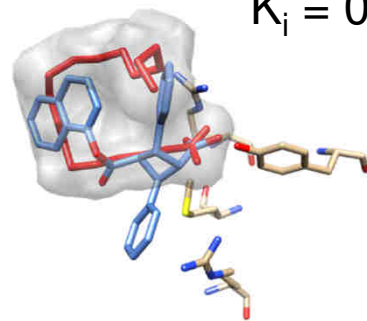
Y.-H. G Teng et al. *Bioorg. Med. Chem.* 23 (2015) 5489–5495

Lab Projects Employing FPS: FABP

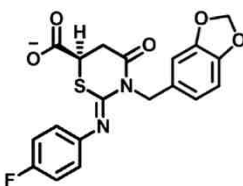
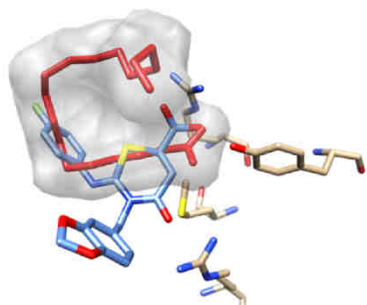
$K_i = 0.93 \pm 0.08 \mu\text{M}$



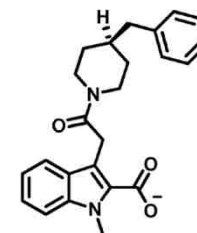
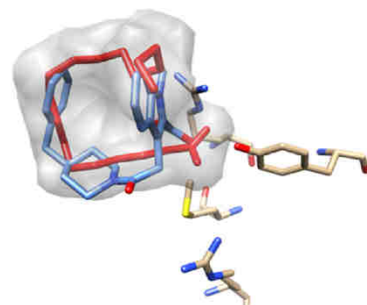
SB-FI-19
(ChemDiv 5511-0235)



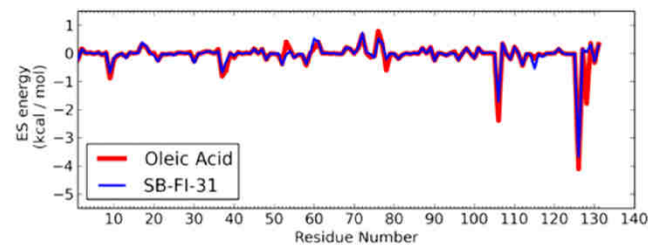
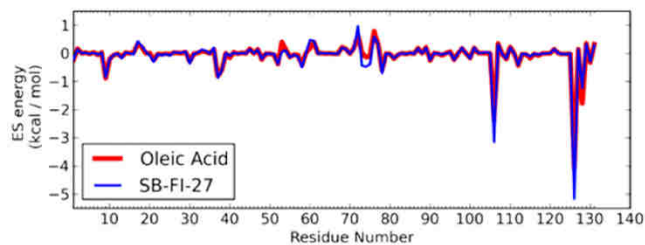
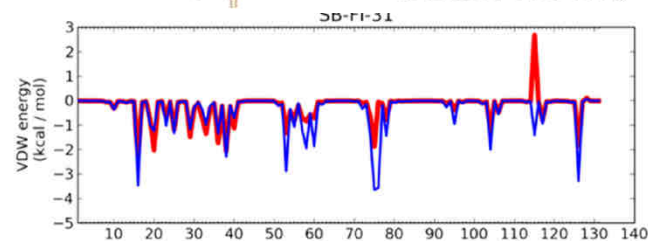
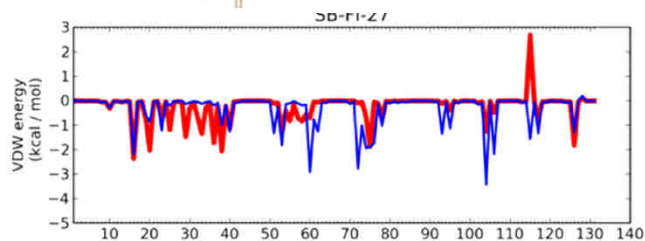
SB-FI-26
(ChemDiv 8009-2334)



SB-FI-27
(ChemDiv 8009-7646)

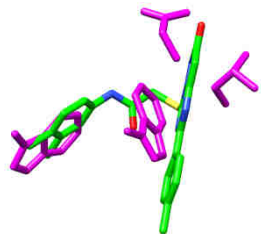


SB-FI-31
(ChemDiv C075-0064)



Lab Projects Employing FPS: HIVgp41

~500,000 screened → 115 purchased → 7 active (3 / 7 picked using FPS)



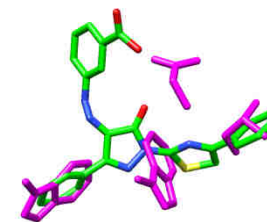
sum_fp
Ki = 0.46 (μM)



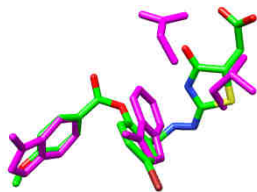
es_fp
Ki = 4.45 (μM)



tots
Ki = 7.1 (μM)



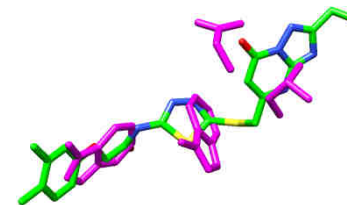
tots
Ki = 0.20 (μM)



tots
Ki = 1.6 (μM)



es_fp
Ki = 13 (μM)

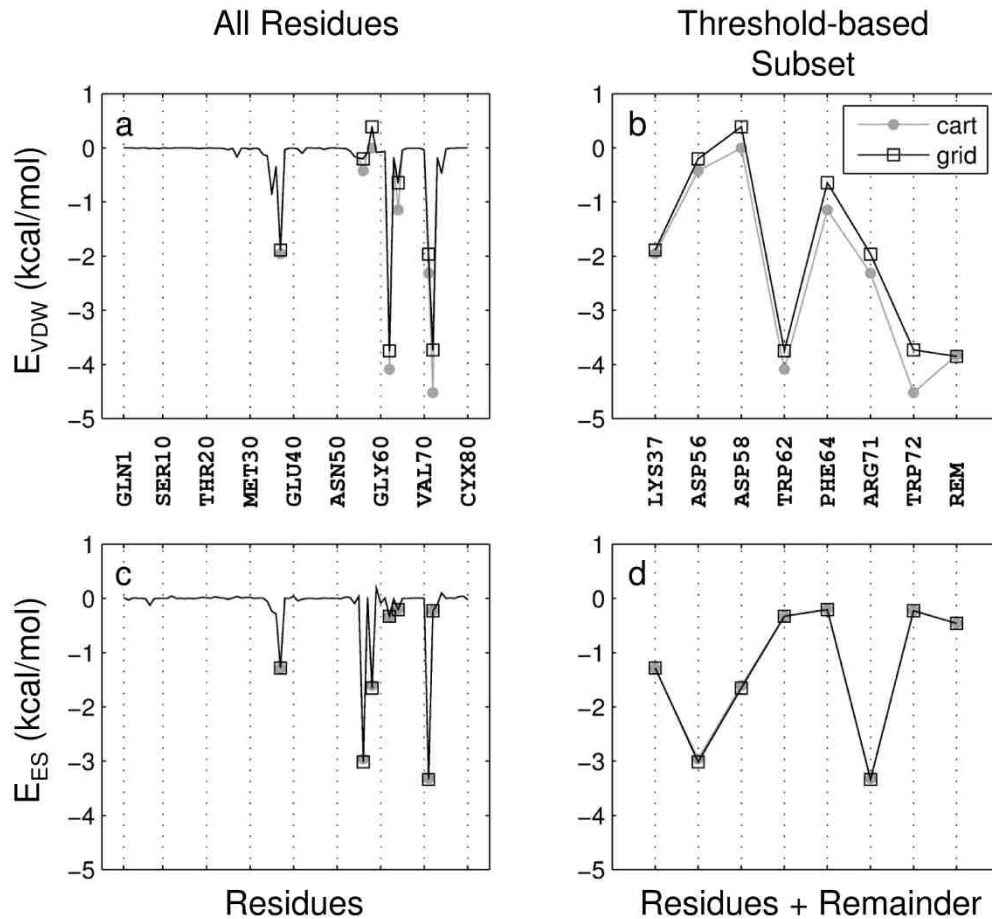


tots
Ki = 2.55 (μM)

Patrick M. Holden, Harmeet Kaur, Miriam Gochin, and Robert C. Rizzo

Holden et al., . Bioorg. Med. Chem. (2012) doi: 10.1016/j.bmcl.2012.02.017

Grid-Based Footprints



Per-residues may be abstracted to a grid

Grid-based and Cartesian footprints show good agreement

Plasminogen kringle-4 protein which is an important protein in blood clotting (pdb code 2PK4)

2PK4 consists of 80 residues

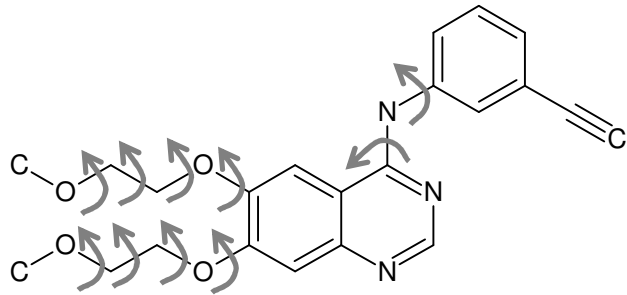
All 80 shown on the left, Subset shown on right

VDW on top, ES on bottom

A remainder grid is calculated to insure that the FP sum = total

Grid are faster than Cart

a



Sampling using Anchor-&-Grow

Start from anchor orient

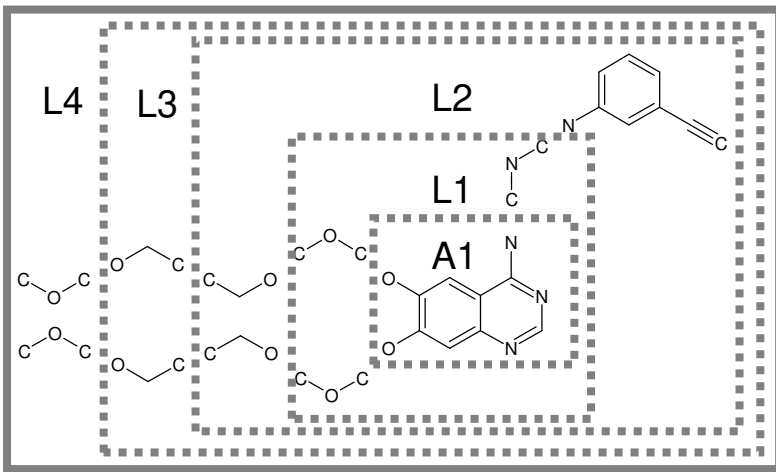
Arrange segments by layer about anchor center

Grow each segment one at a time

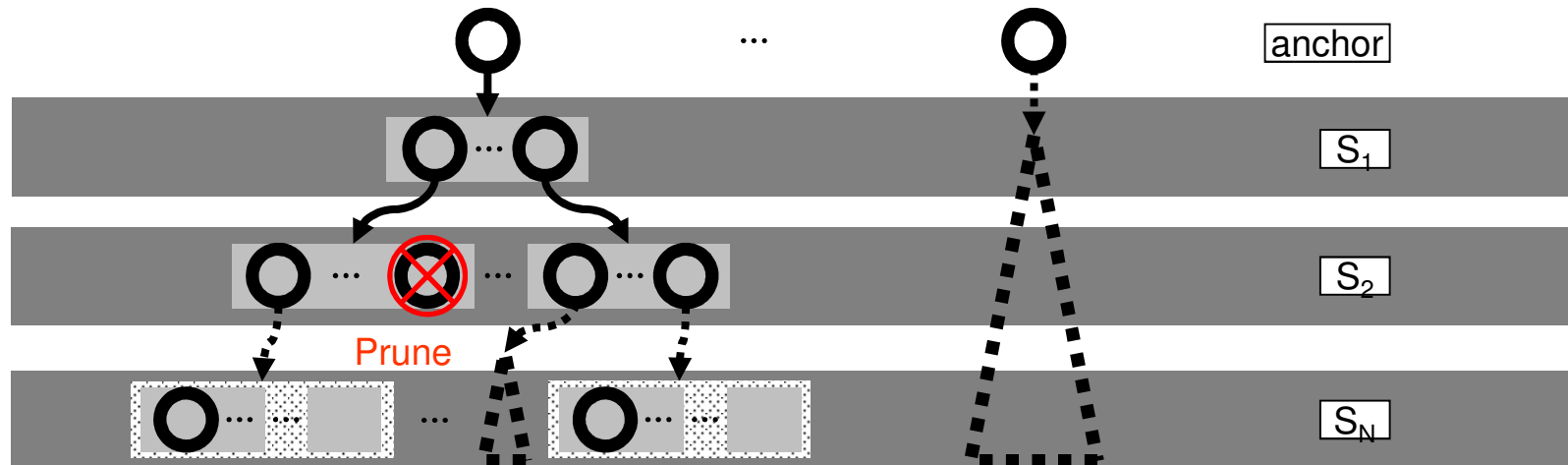
Grow segments in Layer 1

Grow segments in Layer 2 and so on

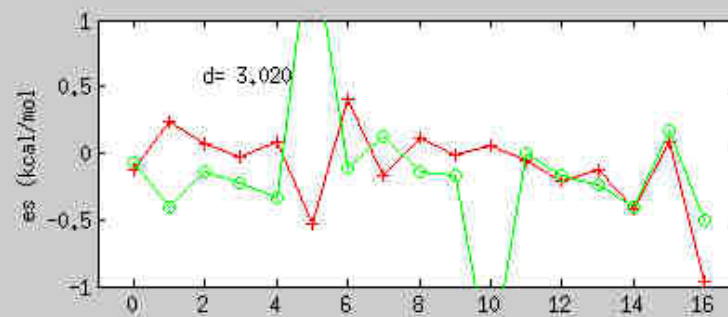
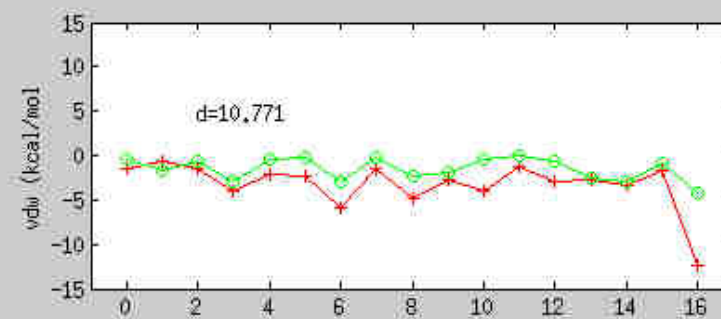
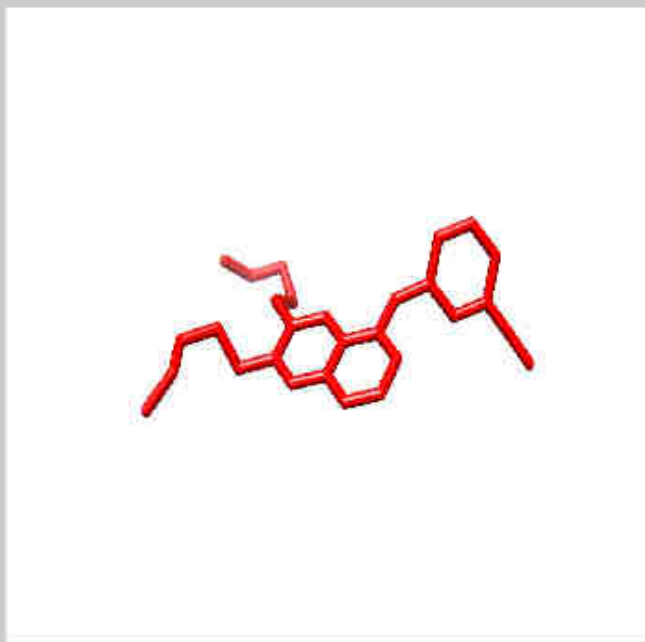
b



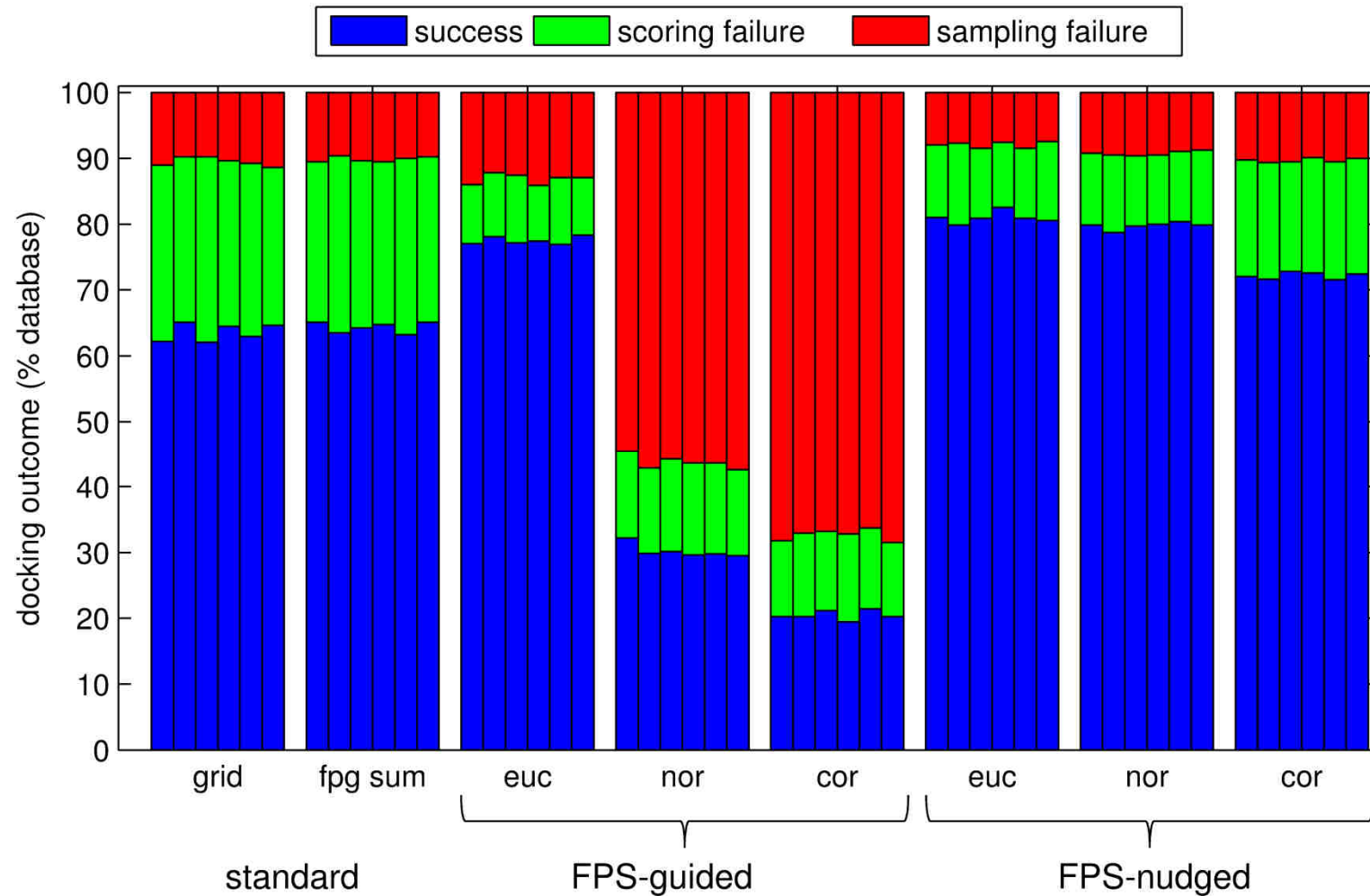
c



Footprint-Guided Growth



Grid-Based Footprints

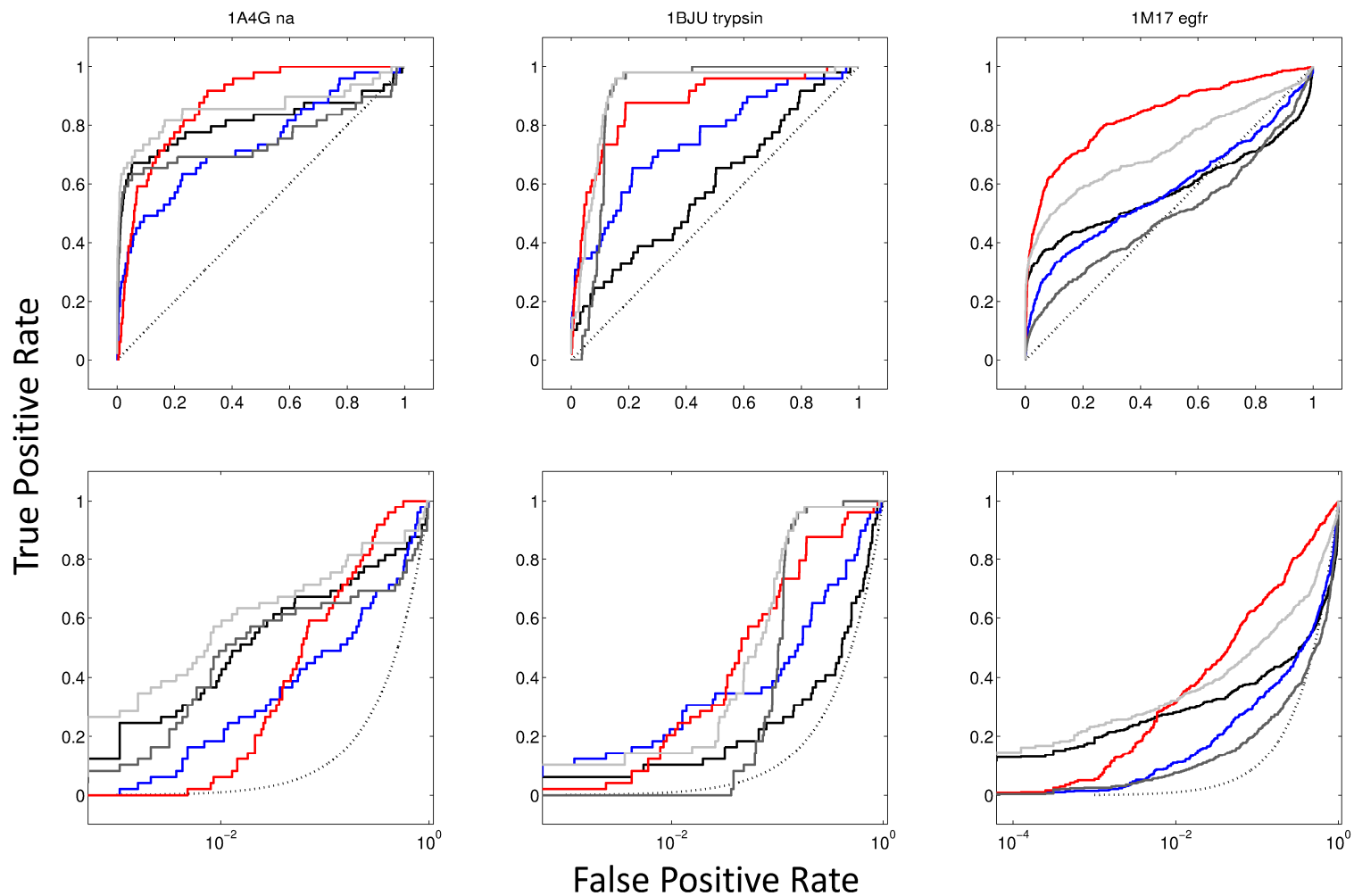


Grid is $O(N)$, Multi-grid is $O(N*M)$, Cartesian $O(N*K)$ $M \ll K$
 Multi-grid is about 5 times slower

Effect of scoring function on sampling

d = driven; s = stated; r = rescored

..... Rand — DGE — FPS d euc — FPS s norm — FPS r euc — FPS r norm



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

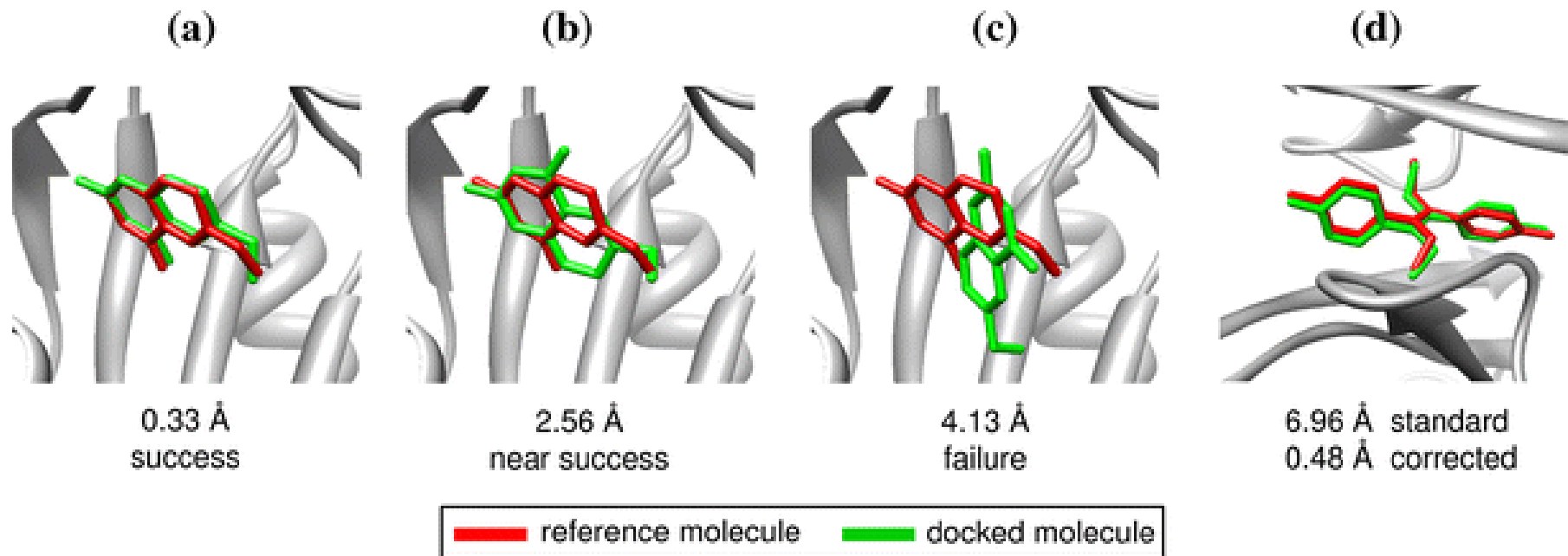
Things are structured differently in Github

```
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
```

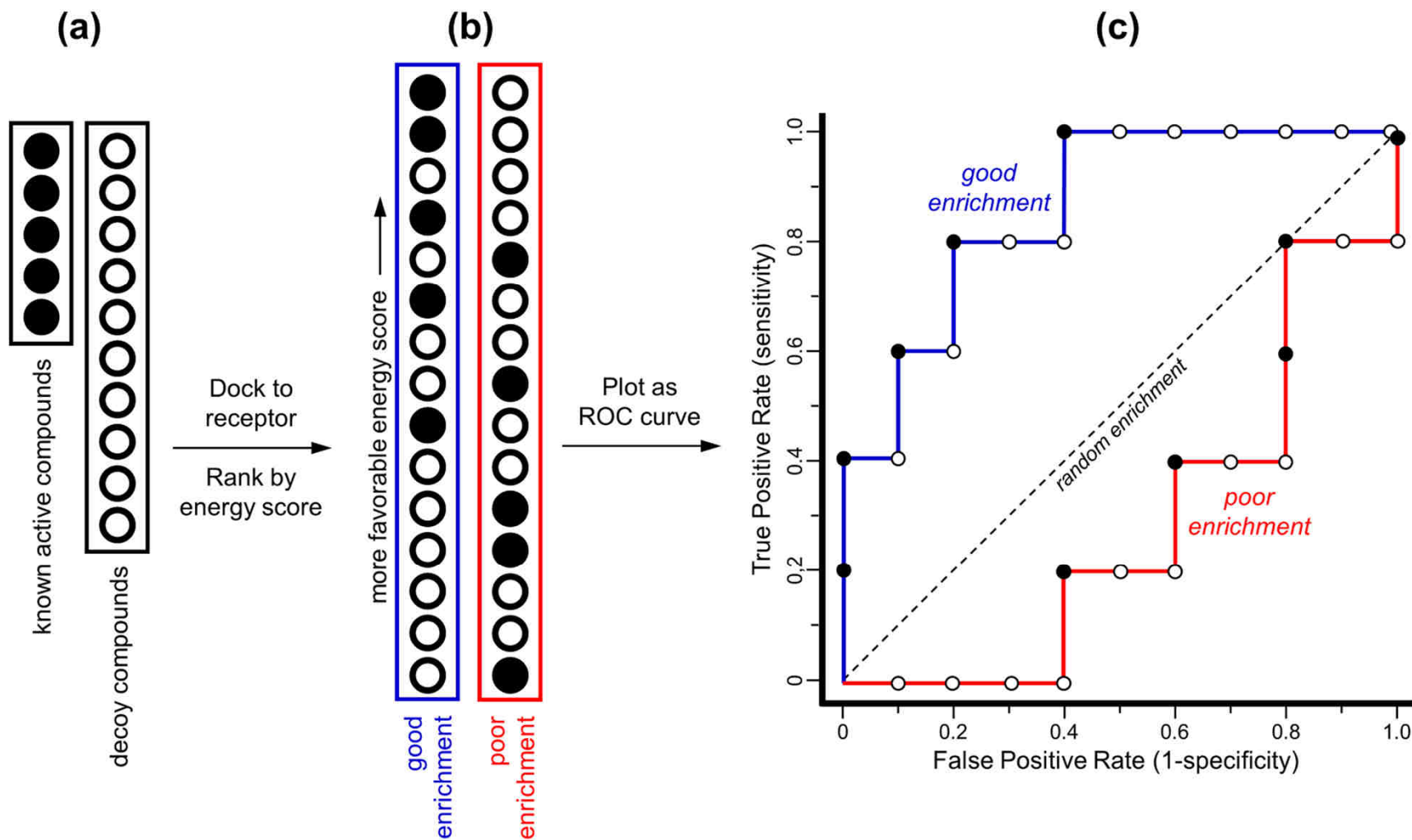
Work done by Teague and Ryan.

Evaluation Methods

- Pose Reproduction (cognate docking, cross docking)
- Enrichment calculations
- Prospective testing of predictions



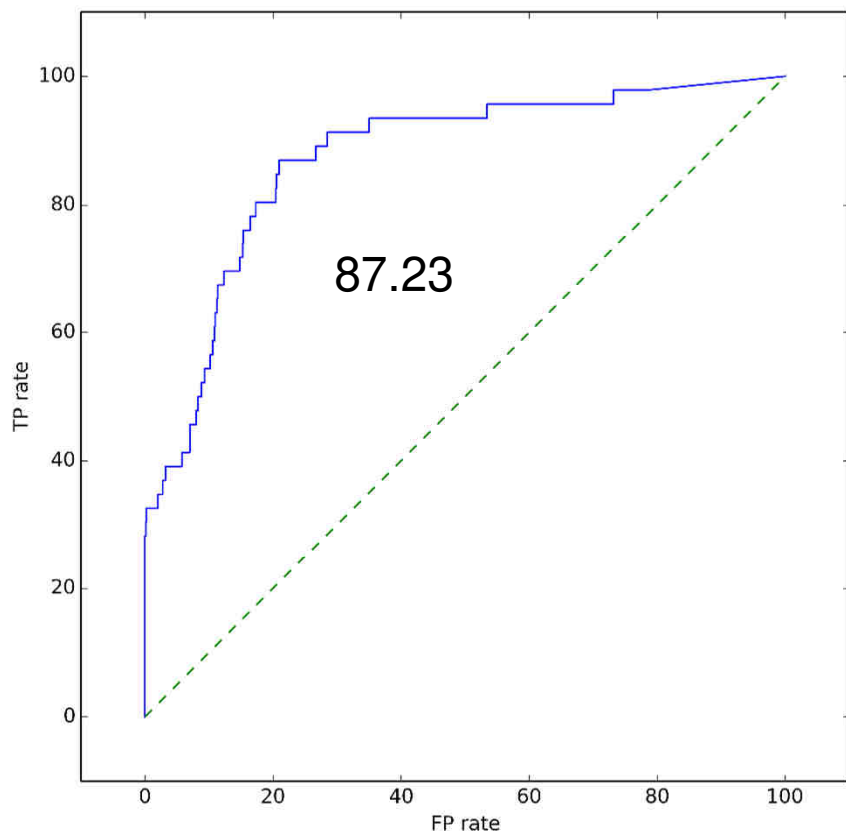
Enrichments: ROC Curves



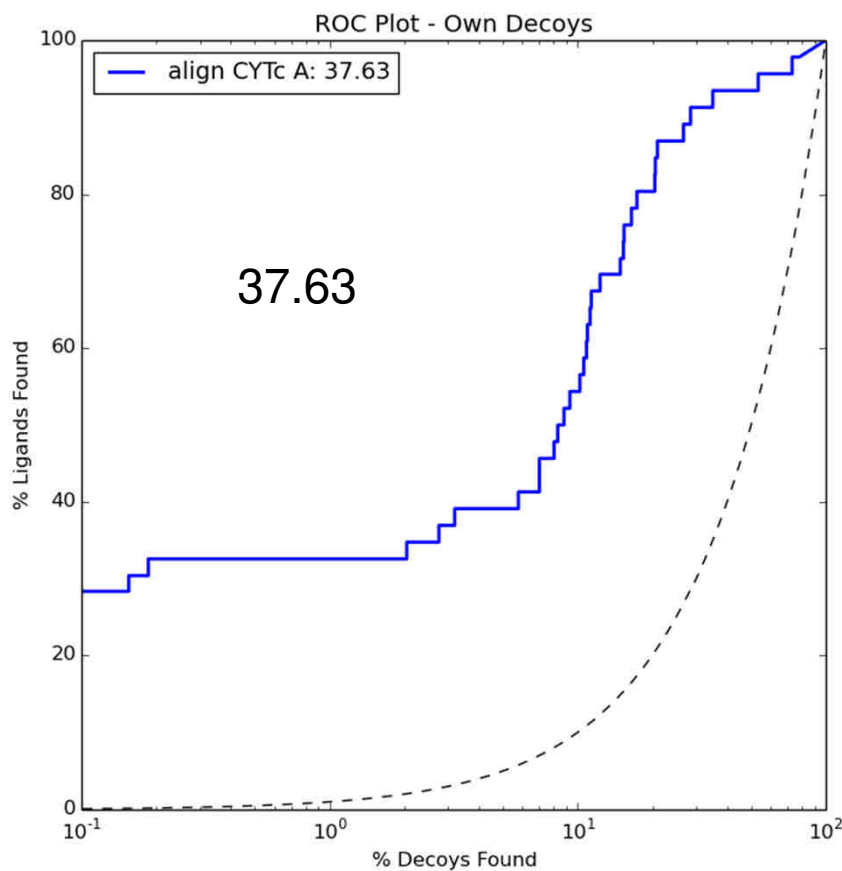
Log-adjusted AUC: Early Enrichment Weighted More

CcP conformation A

ROC curve

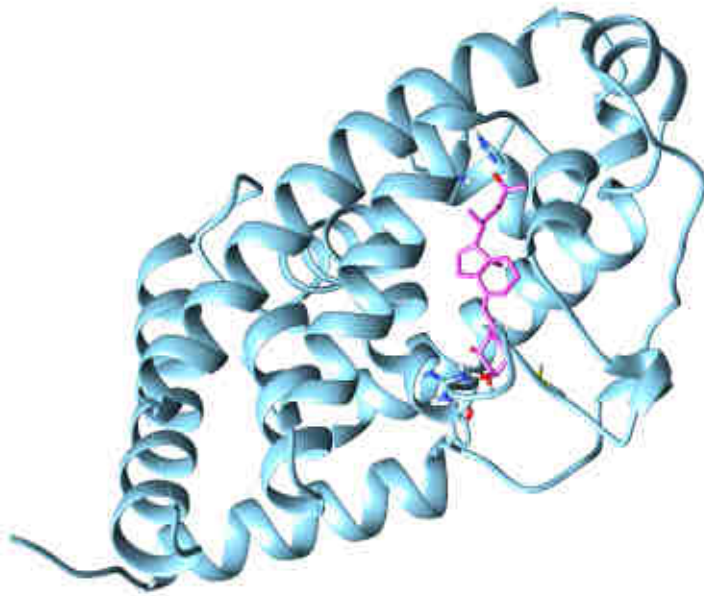


Log adjusted ROC curve

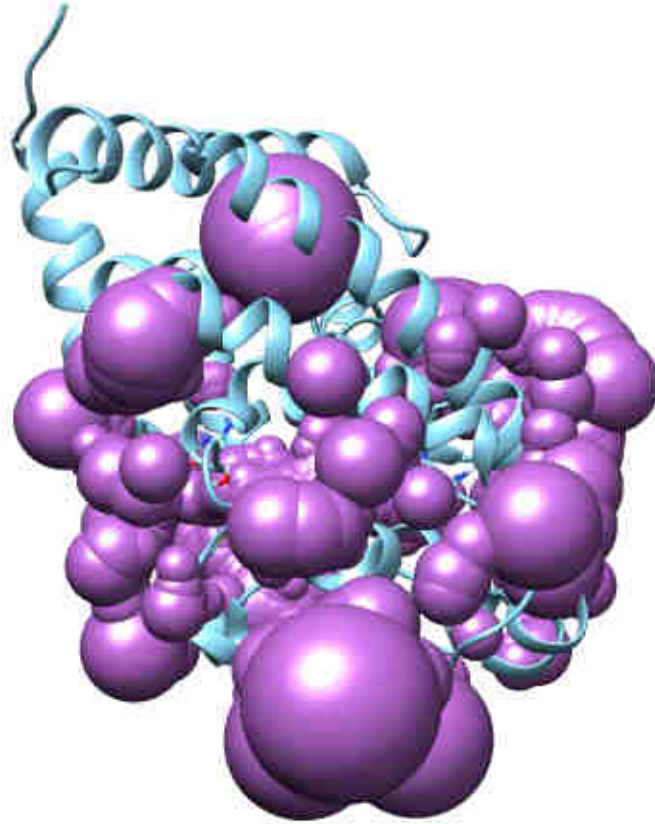


Tweaking your set up – Spheres

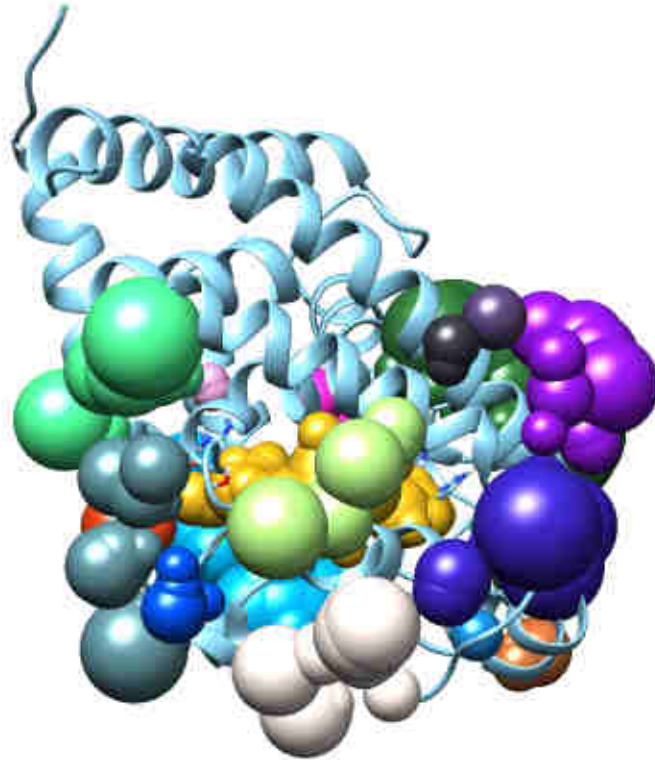
- Place spheres where you want your rings to go
- Use previous screens for ideas



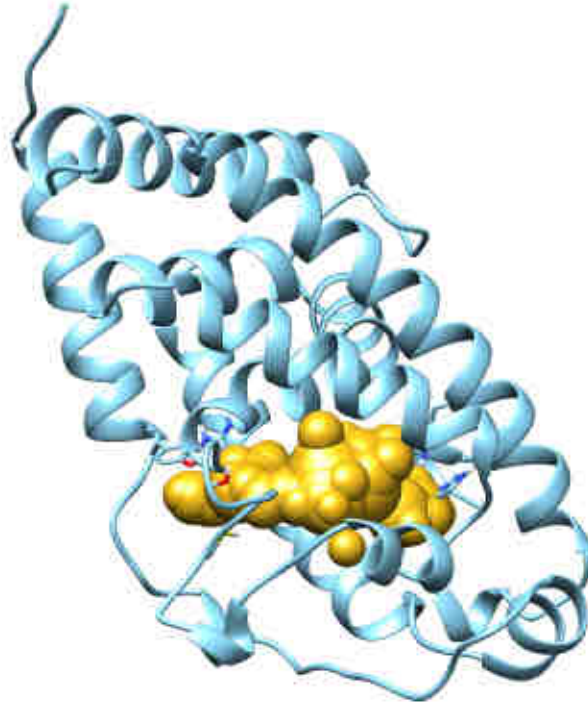
SPHGEN Spheres



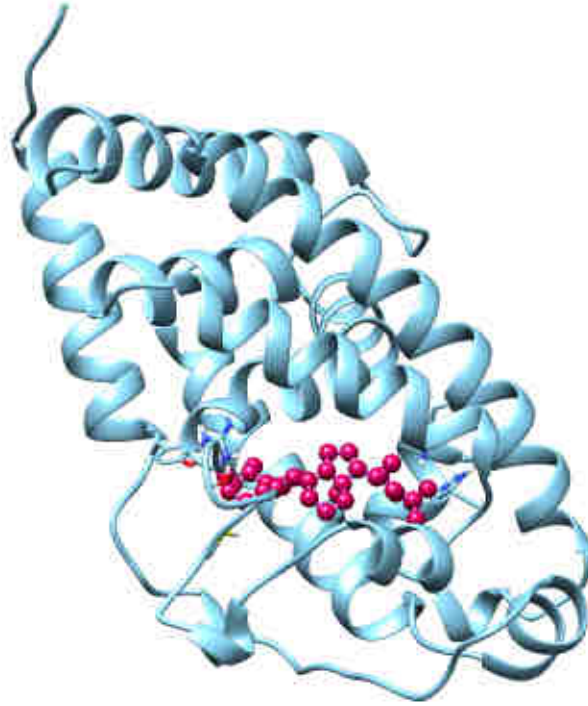
SPHGEN Sphere Clusters



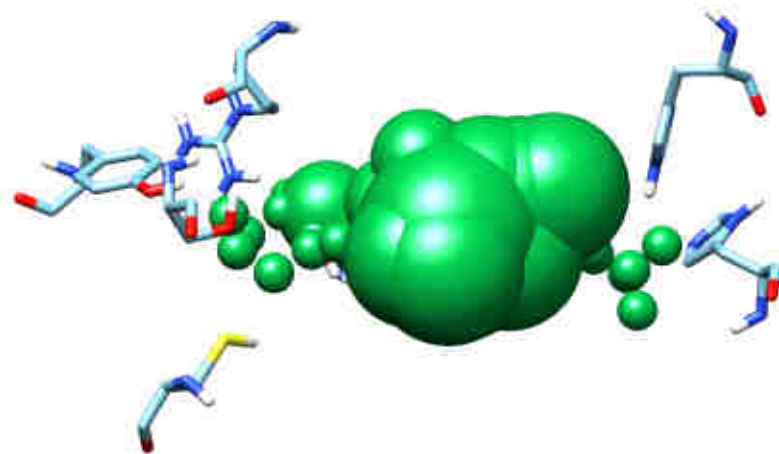
SPHGEN Cluster1 Spheres



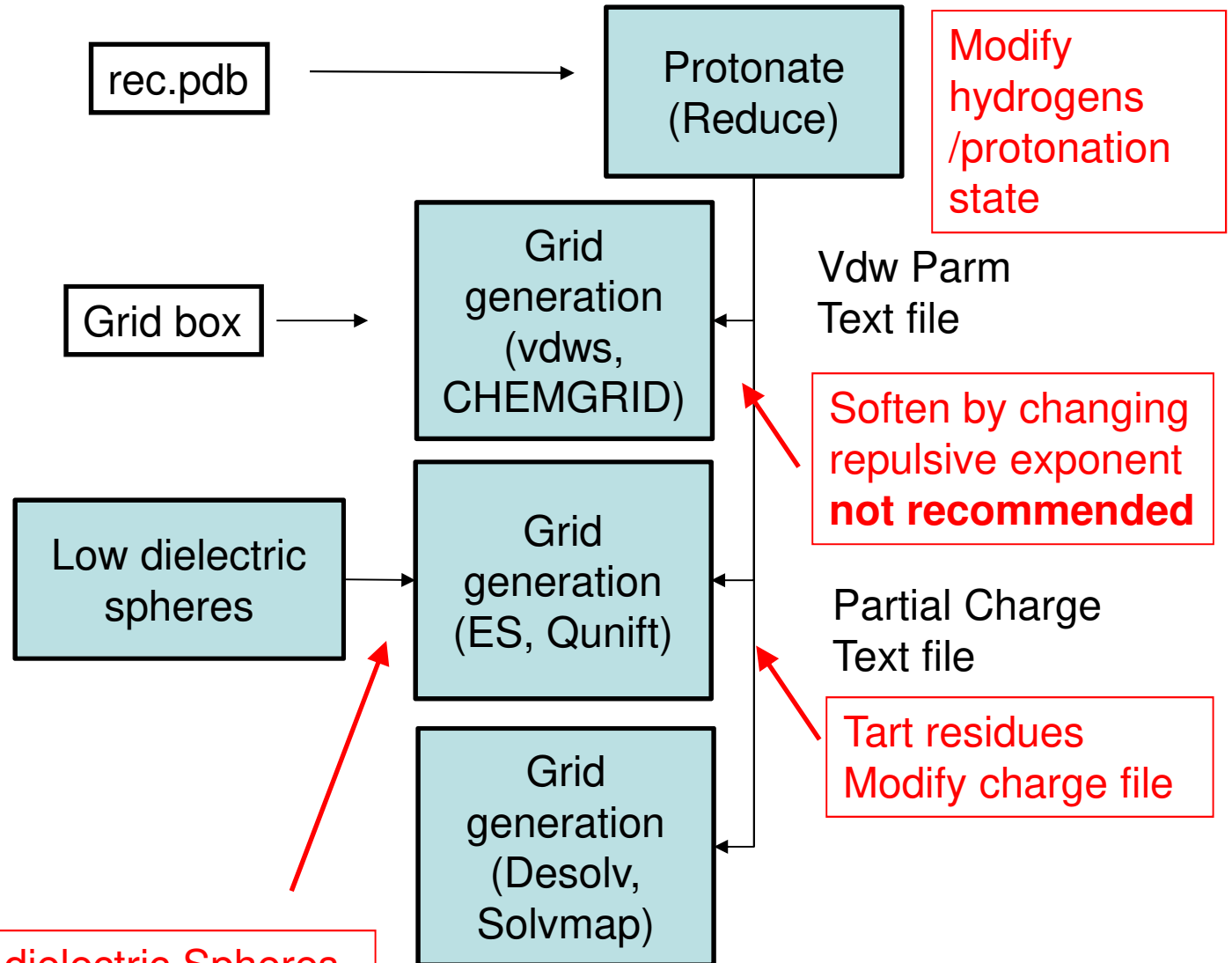
Ligand Atoms Spheres



Matching Spheres

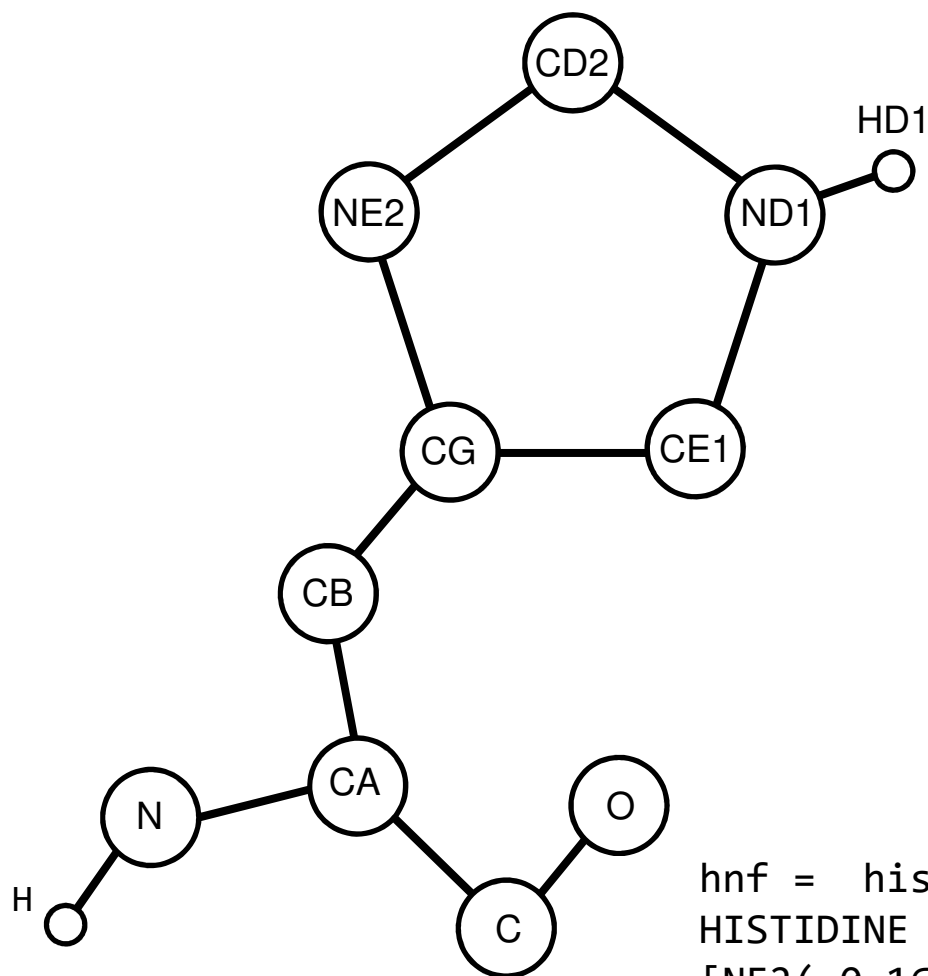


Receptor preparation



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

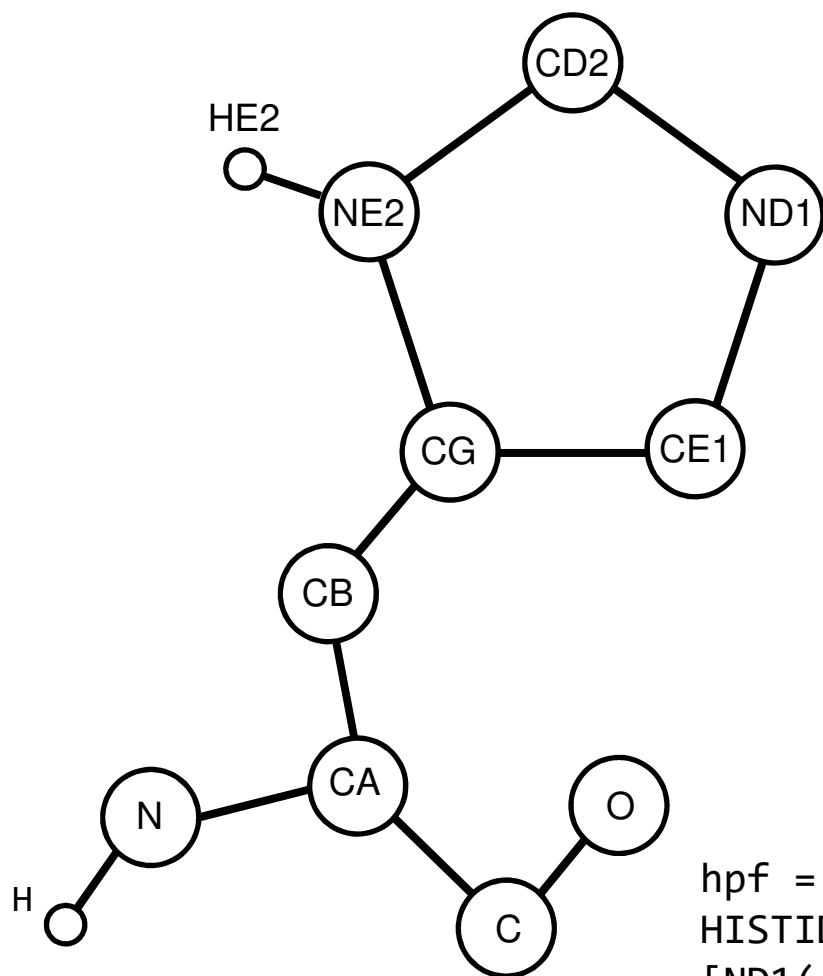
HISTIDINE neutral delta proton [tart2]



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)]

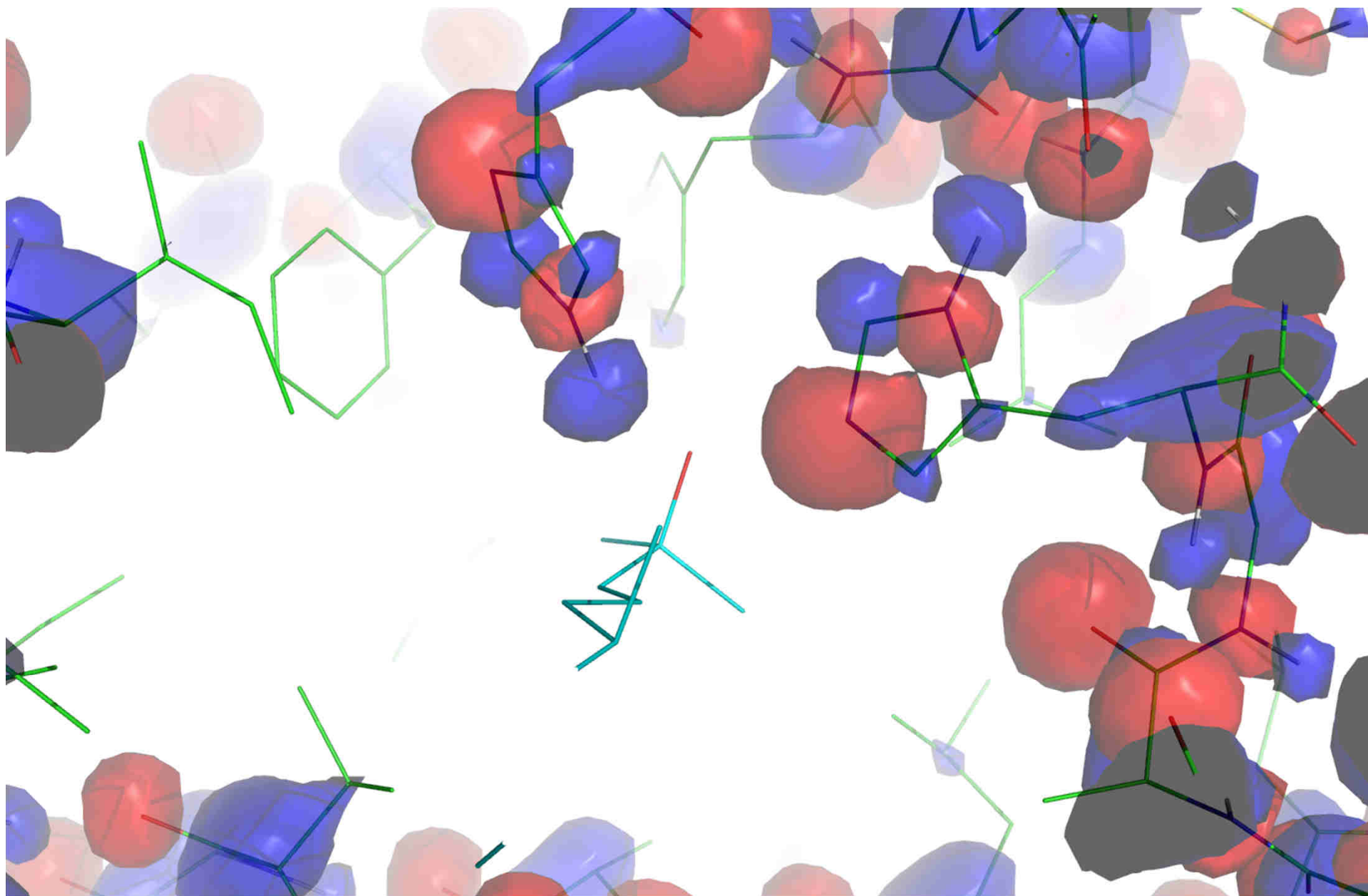
HISTIDINE neutral epsilon proton [tart2]



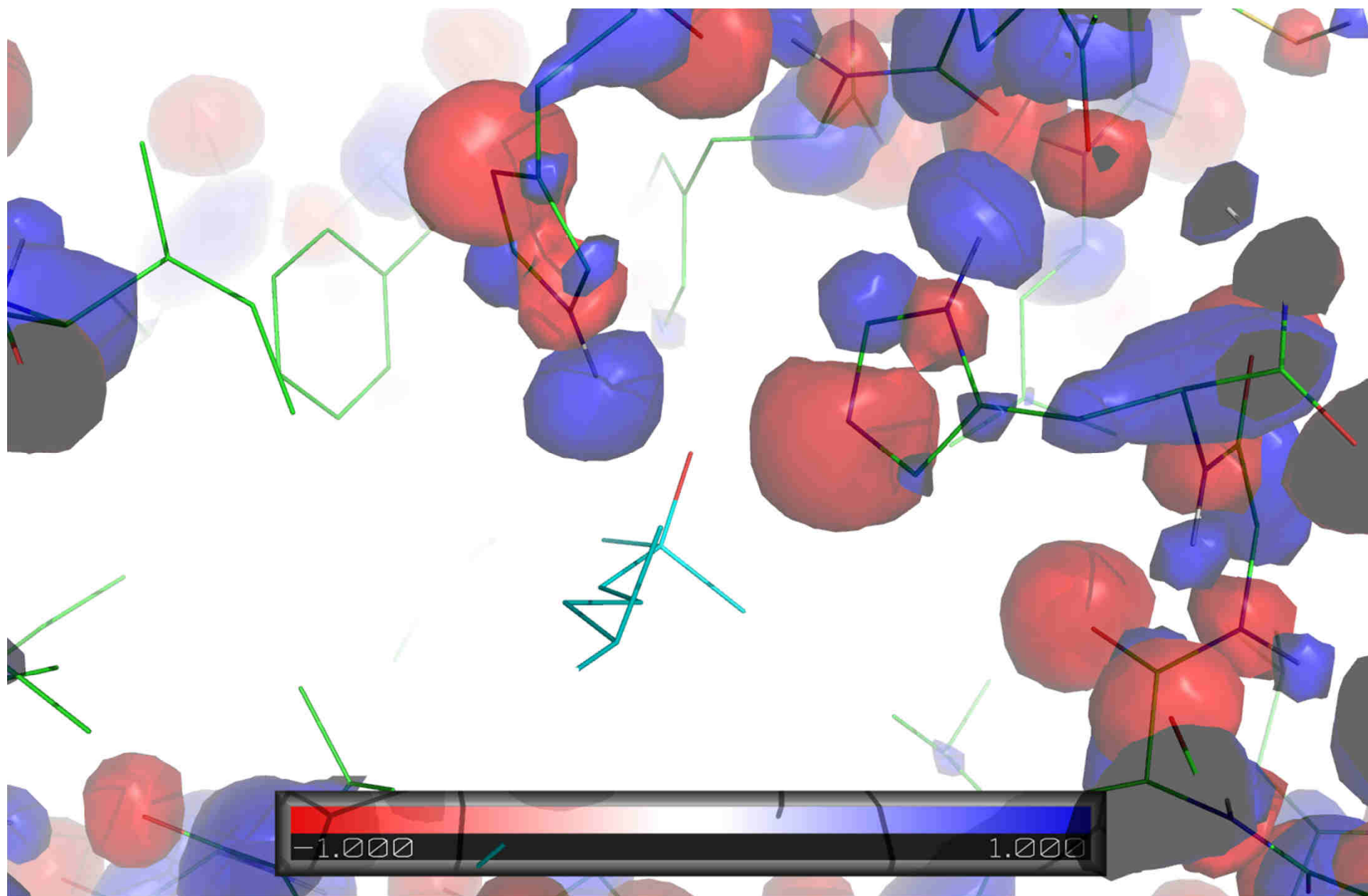
N	-0.520	N	-0.520	
C	0.526	C	0.526	
O	-0.500	O	-0.500	
CA	0.219	CA	0.219	
CB	0.060	CB	0.060	
CG	0.112	CG	0.112	
CD2	0.122	CD2	0.122	
CE1	0.384	CE1	0.384	
ND1	-0.527	ND1	-0.607	-0.08
NE2	-0.444	NE2	-0.524	-0.08
H	0.248	H	0.248	
HE2	0.320	HE2	0.480	+0.16

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

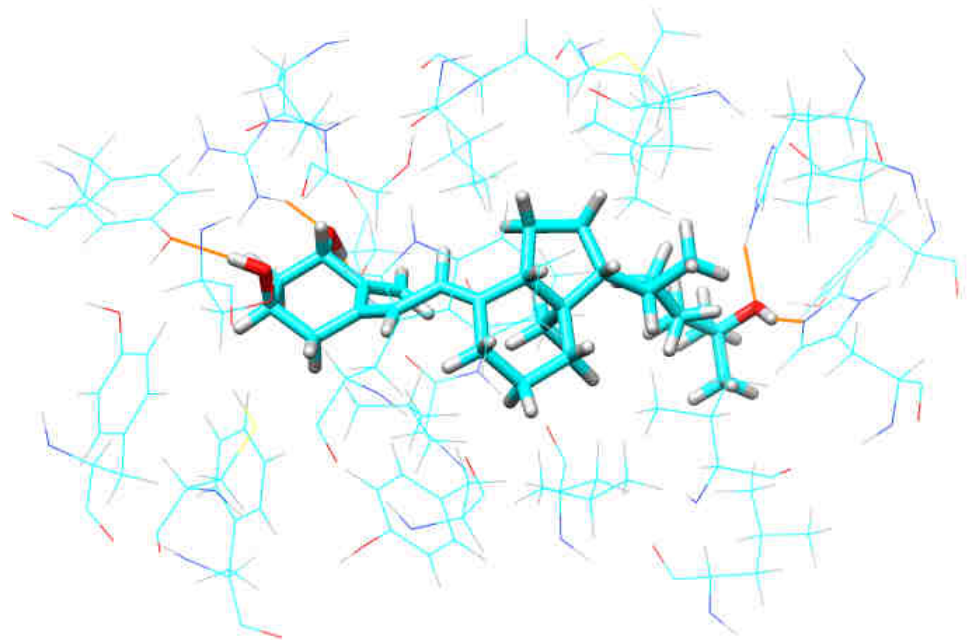
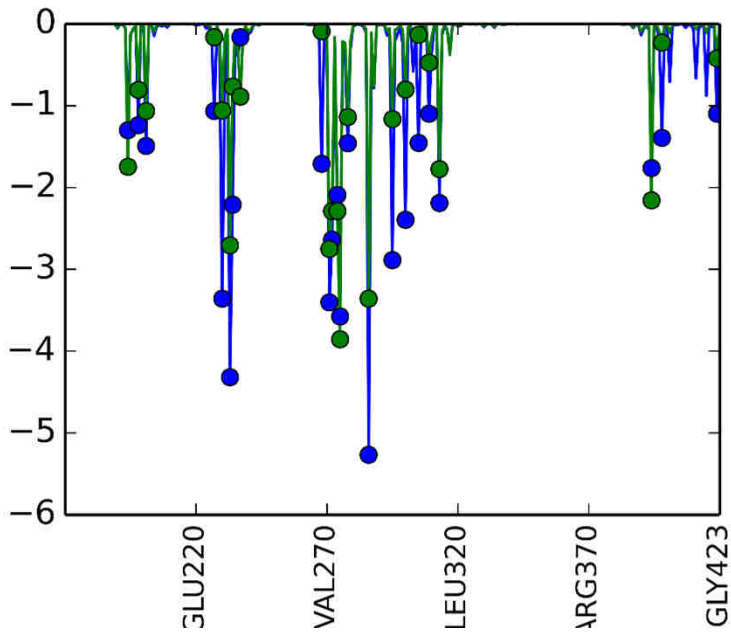
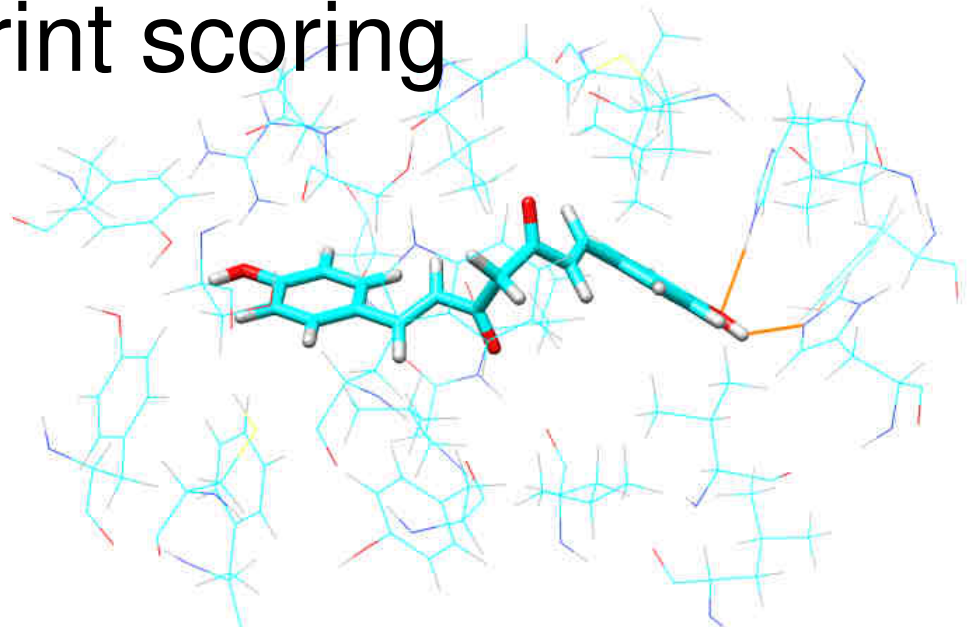
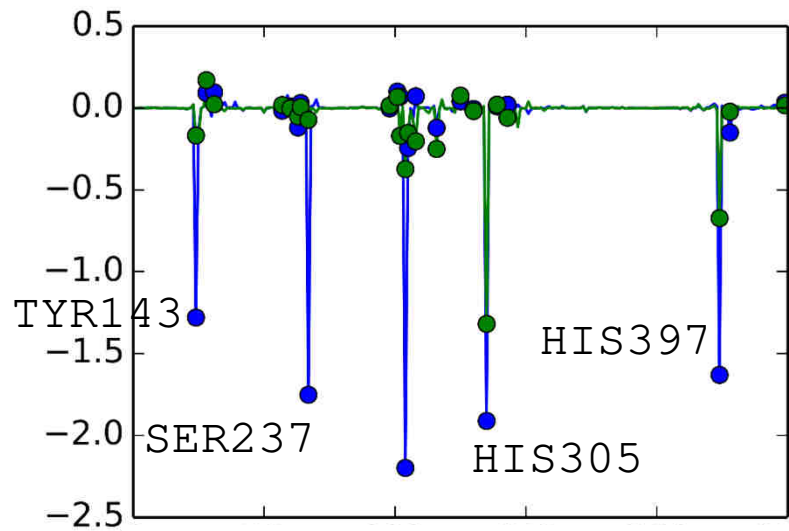
Original partial charges



More polar (tart2) partial charges



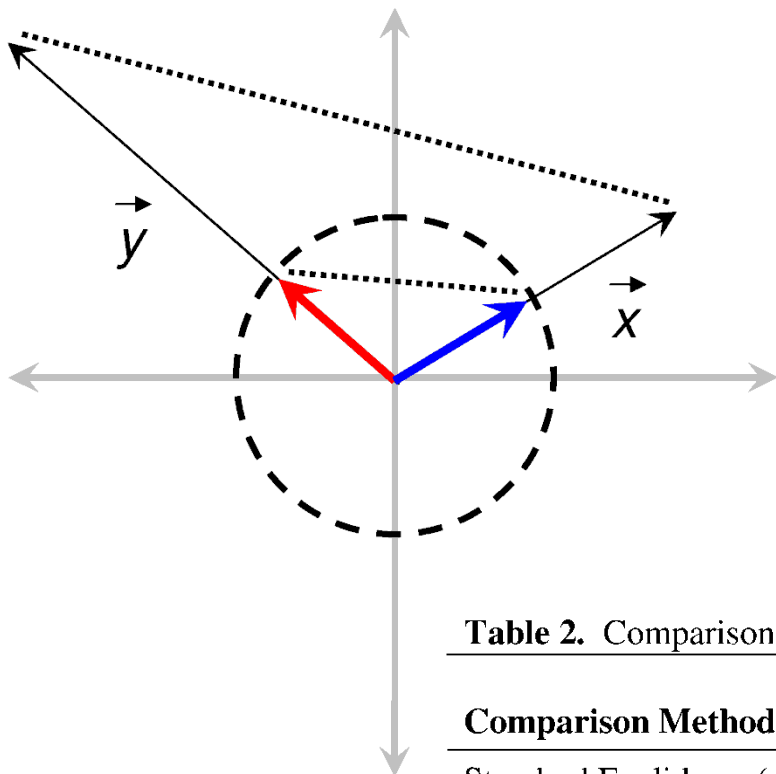
Footprint scoring



Footprint Similarity Nomenclature

- DCE -- Dock Cartesian Energy (Standard Energy function)
- Cartesian energy decompositions of VDW, ES and H-bonds
- FPS_{VDW} -- quantifies the similarity between 2 VDW footprints
- FPS_{ES} -- quantifies the similarity between 2 ES footprints
- FPS_{HB} -- quantifies the similarity between 2 H-bond footprints
- $FPS_{VDW+ES} = FPS_{VDW} + FPS_{ES}$ (consensus score)

Methods for Quantifying Similarity



$$d = \|\vec{x} - \vec{y}\| = \sqrt{\sum (x_i - y_i)^2}$$

$$d_{norm} = \|\vec{X} - \vec{Y}\| \quad \vec{X} = \vec{x}/\|\vec{x}\| \quad \vec{Y} = \vec{y}/\|\vec{y}\|$$

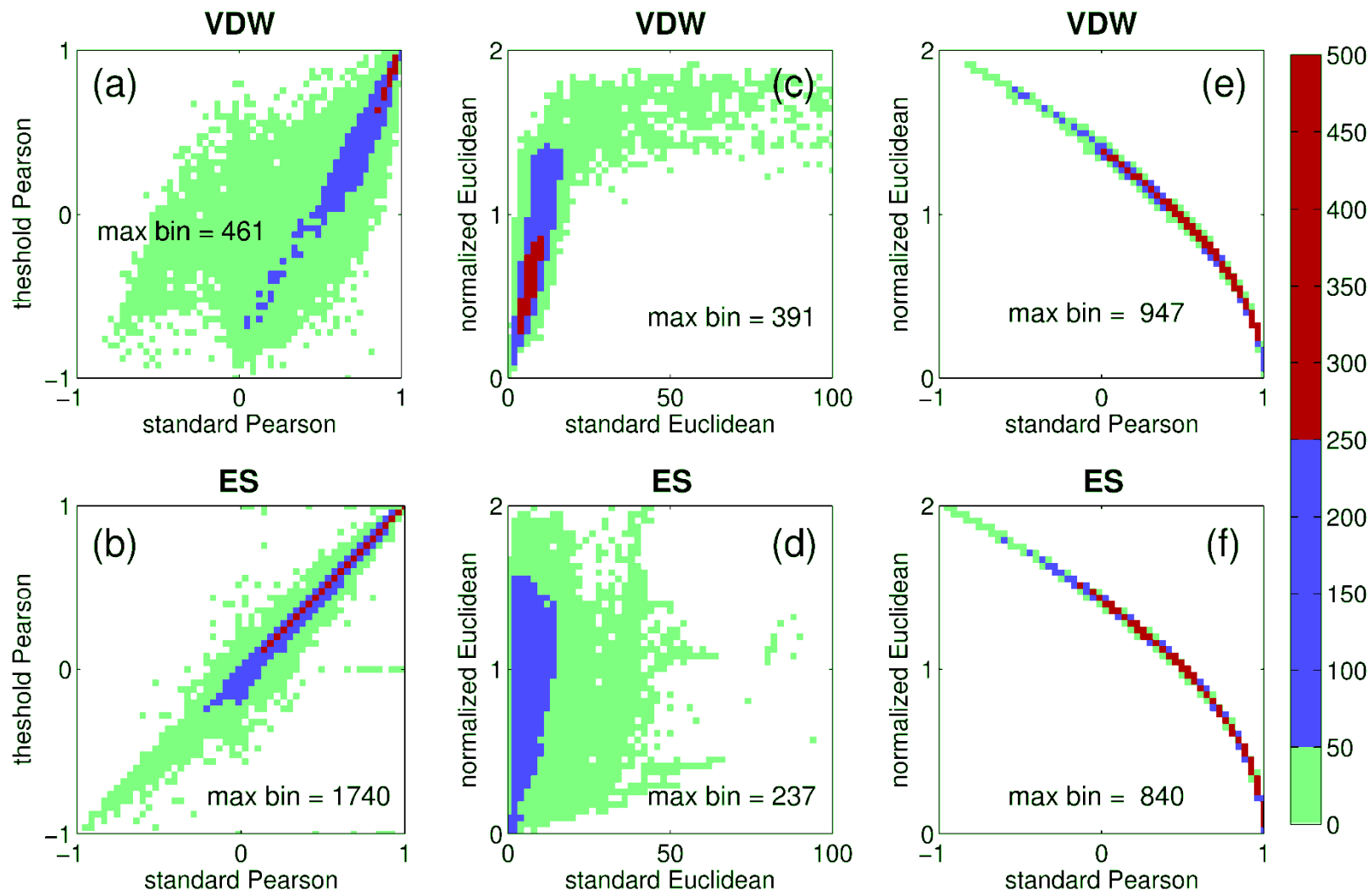
$$r = \frac{\text{cov}(\vec{x}, \vec{y})}{\sqrt{\text{var}(\vec{x})}\sqrt{\text{var}(\vec{y})}}$$

Table 2. Comparison methods and corresponding ranges for footprint similarity (FPS) scores.

Comparison Method	Ranges ^a		
	FPS _{VDW} , FPS _{ES} , FPS _{HB}	FPS _{VDW+ES}	FPS _{VDW+ES+HB}
Standard Euclidean (d)	[<u>0</u> , ∞)	[<u>0</u> , ∞)	[<u>0</u> , ∞)
Normalized Euclidean (d_{norm})	[<u>0</u> , 2]	[<u>0</u> , 4]	[<u>0</u> , 6]
Standard Pearson (r)	[-1, <u>1</u>]	[-2, <u>2</u>]	[-3, <u>3</u>]
Threshold Pearson (r_{thresh})	[-1, <u>1</u>]	[-2, <u>2</u>]	[-3, <u>3</u>]

^aThe most favorable score possible for each method is underlined.

Functional Relationships



$$r = \cos(\theta) = \frac{\vec{x}^\mu \cdot \vec{y}^\mu}{\|\vec{x}^\mu\| \|\vec{y}^\mu\|} \approx \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|} = \cos(\theta^*) \quad d_{norm} = \sqrt{2(1 - \cos(\theta^*))} \approx \sqrt{2(1 - r)} \quad 87$$

Functional Relationships

$$r = \cos(\theta) = \frac{\vec{x}^\mu \cdot \vec{y}^\mu}{\|\vec{x}^\mu\| \|\vec{y}^\mu\|} \approx \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|} = \cos(\theta^*) \quad d_{norm} = \sqrt{2(1 - \cos(\theta^*))} \approx \sqrt{2(1 - r)}$$

Corollary 1: if \vec{u} and \vec{v} are unit vectors, then $\|\vec{u} - \vec{v}\| = \sqrt{2(1 - \cos(\theta))}$, where θ is the angle between \vec{u} and \vec{v} .

Proof:

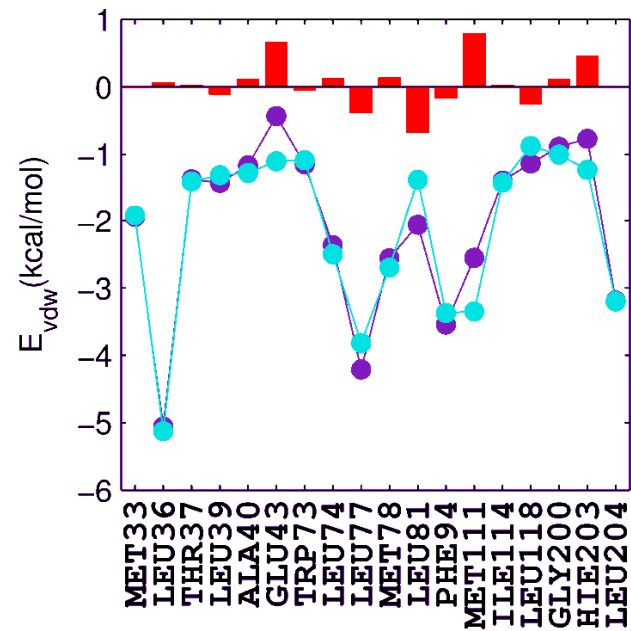
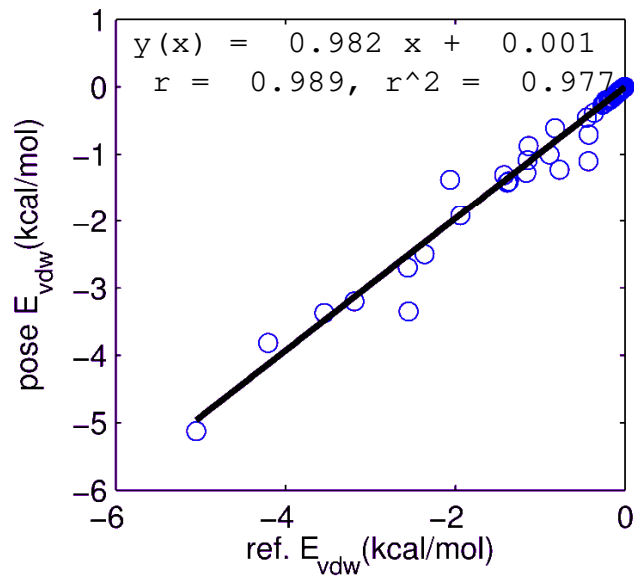
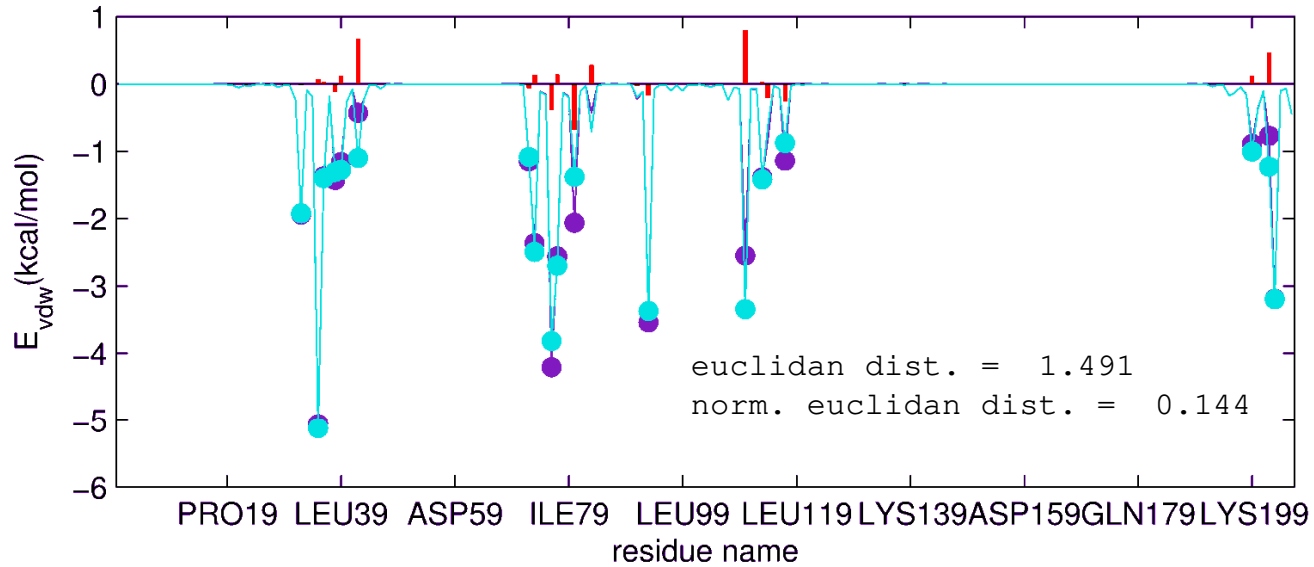
Let \vec{u} and \vec{v} be unit vectors. Then,

$$\cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|} = \vec{u} \cdot \vec{v} \quad (\because \|\vec{u}\| = 1)$$

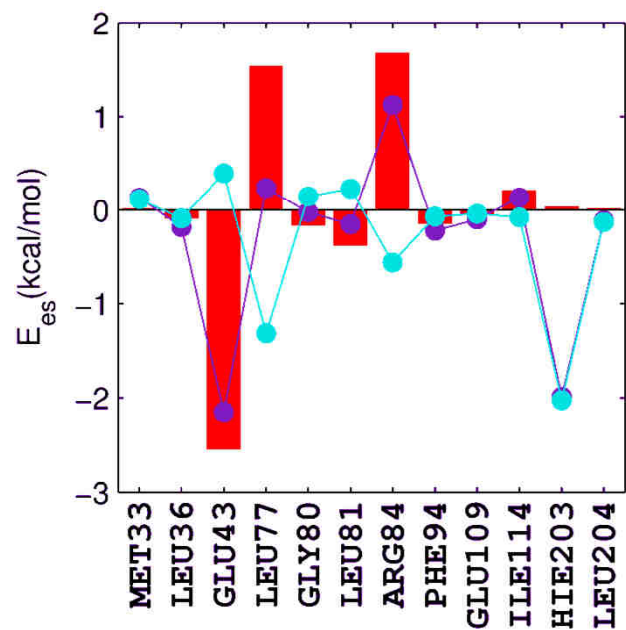
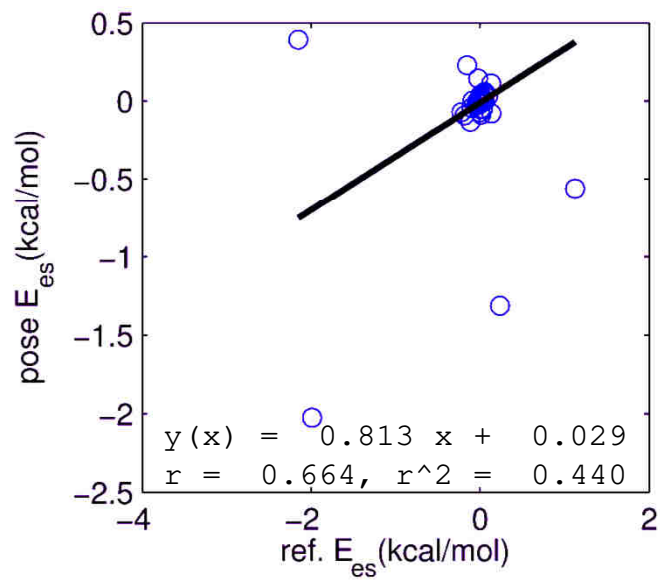
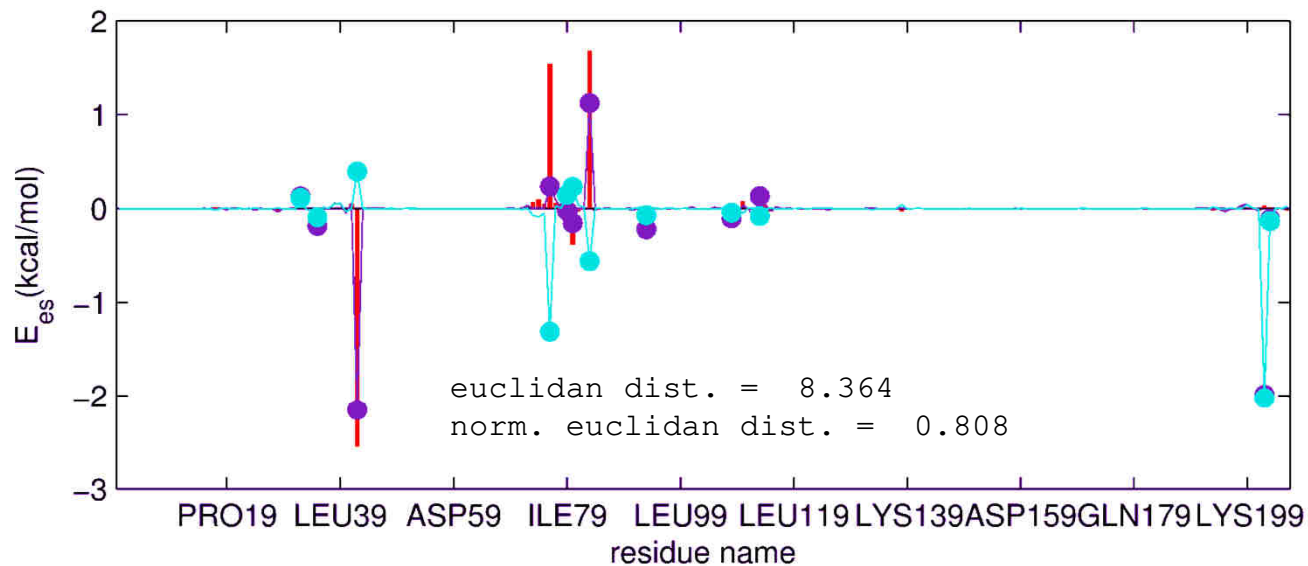
$$\begin{aligned} \|\vec{u} - \vec{v}\| &= \sqrt{\sum (u_i - v_i)^2} = \sqrt{\sum u_i^2 + \sum v_i^2 - 2\sum u_i v_i} \\ &= \sqrt{1+1-2\sum u_i v_i} \quad (\because \|\vec{u}\|^2 = \sum u_i^2 = 1) \\ &= \sqrt{2(1 - \vec{u} \cdot \vec{v})} \end{aligned}$$

Therefore, $\|\vec{u} - \vec{v}\| = \sqrt{2(1 - \cos(\theta))}$

2QE4 van der Waals Footprint (Good Comparison)



2QE4 Electrostatics Footprint (Bad Comparison)

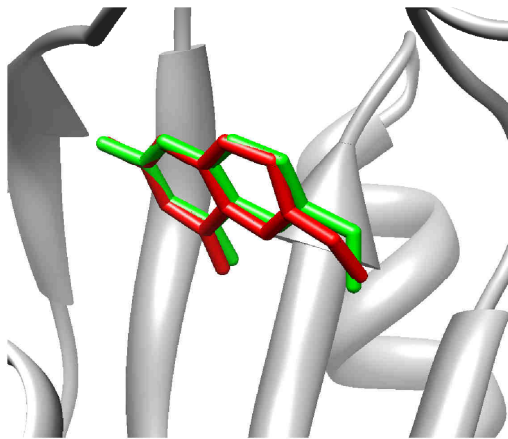


Structural Comparisons using rmsd

— reference molecule

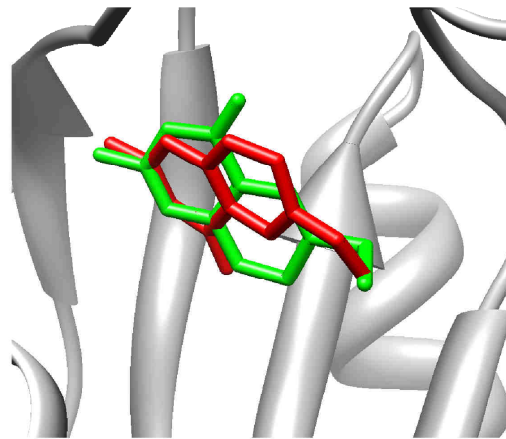
— docked molecule

(a)



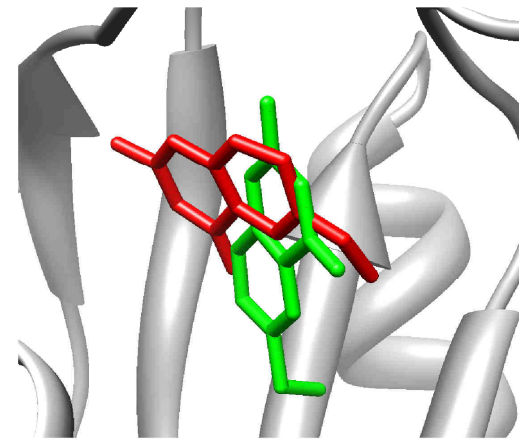
0.33 Å
success

(b)



2.56 Å
near success

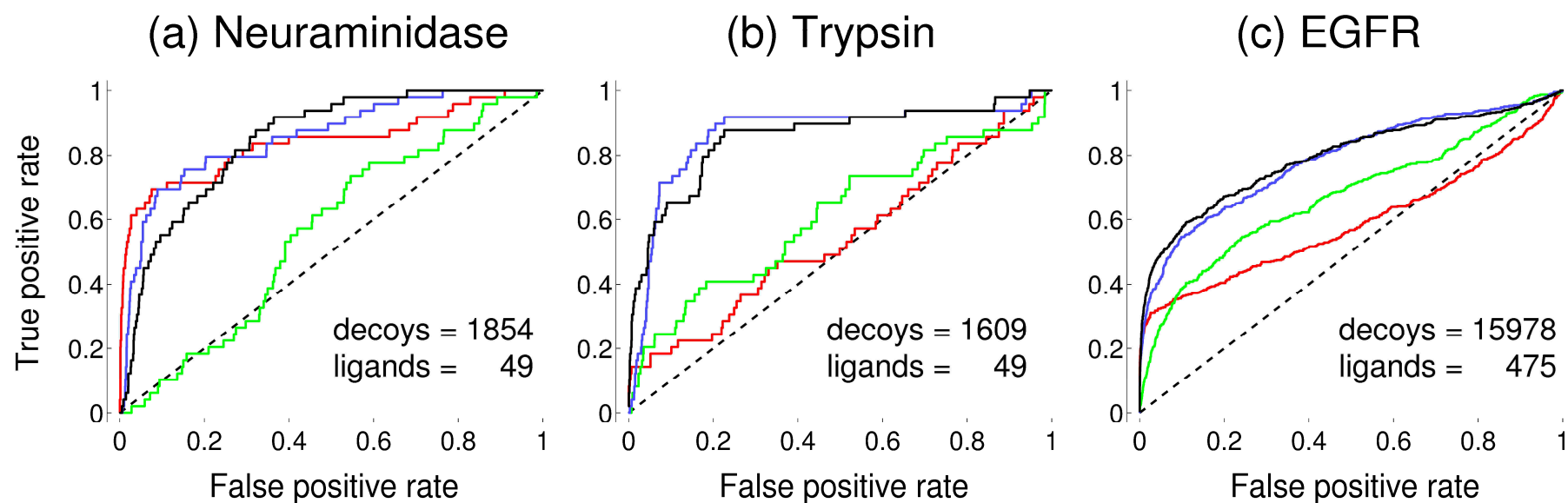
(c)



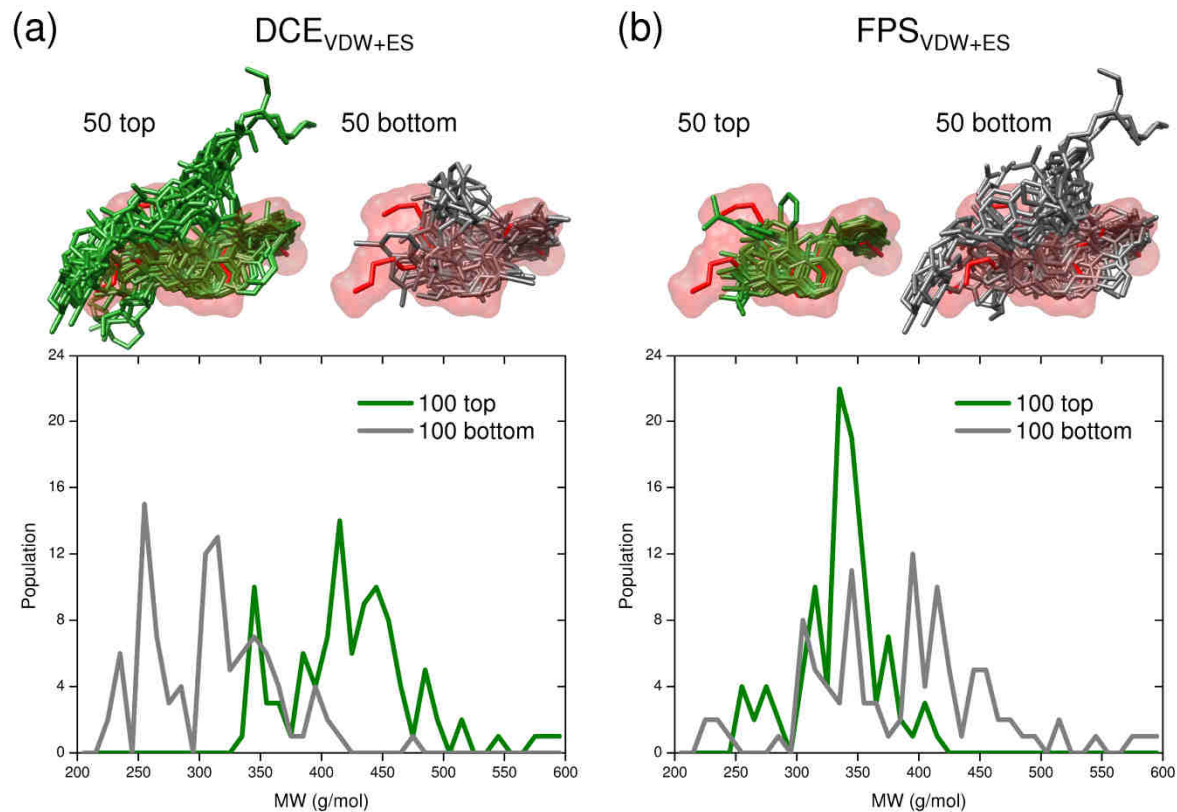
4.13 Å
failure

Enrichment (ROC Curves)

---- random **—** DCE_{VDW+ES} **—** FPS_{VDW+ES} **—** FPS_{VDW} **—** FPS_{ES}

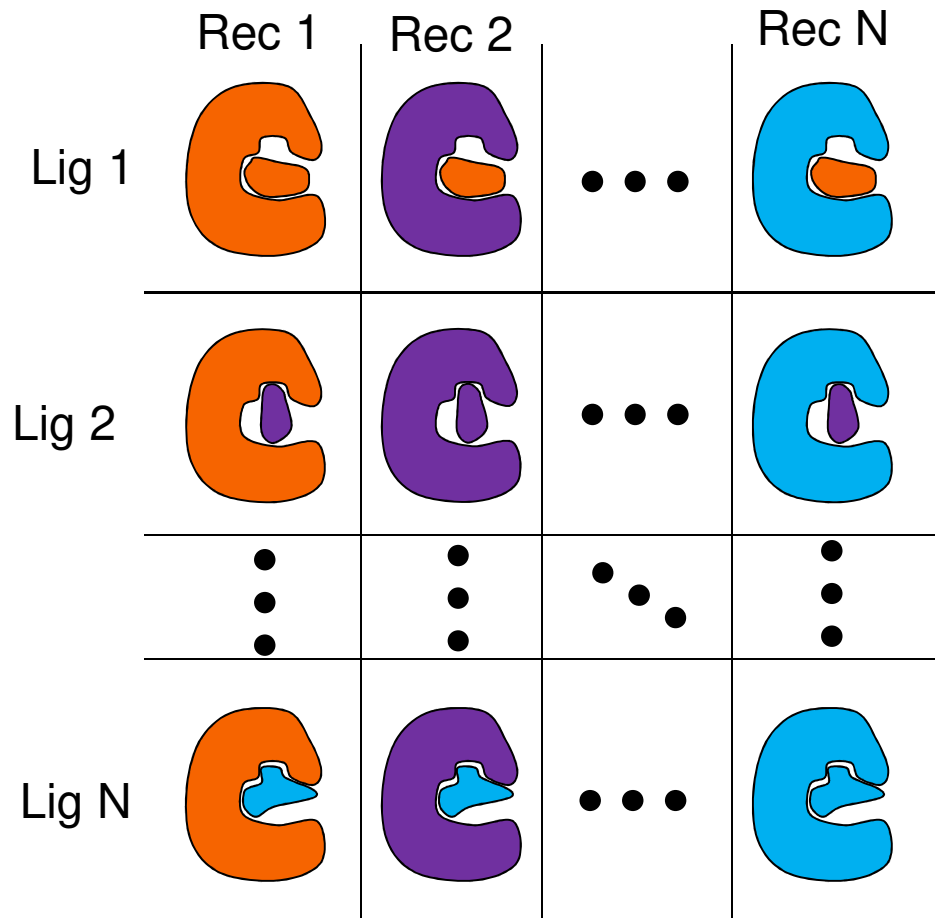


Enrichment (Molecular Weight Bias)



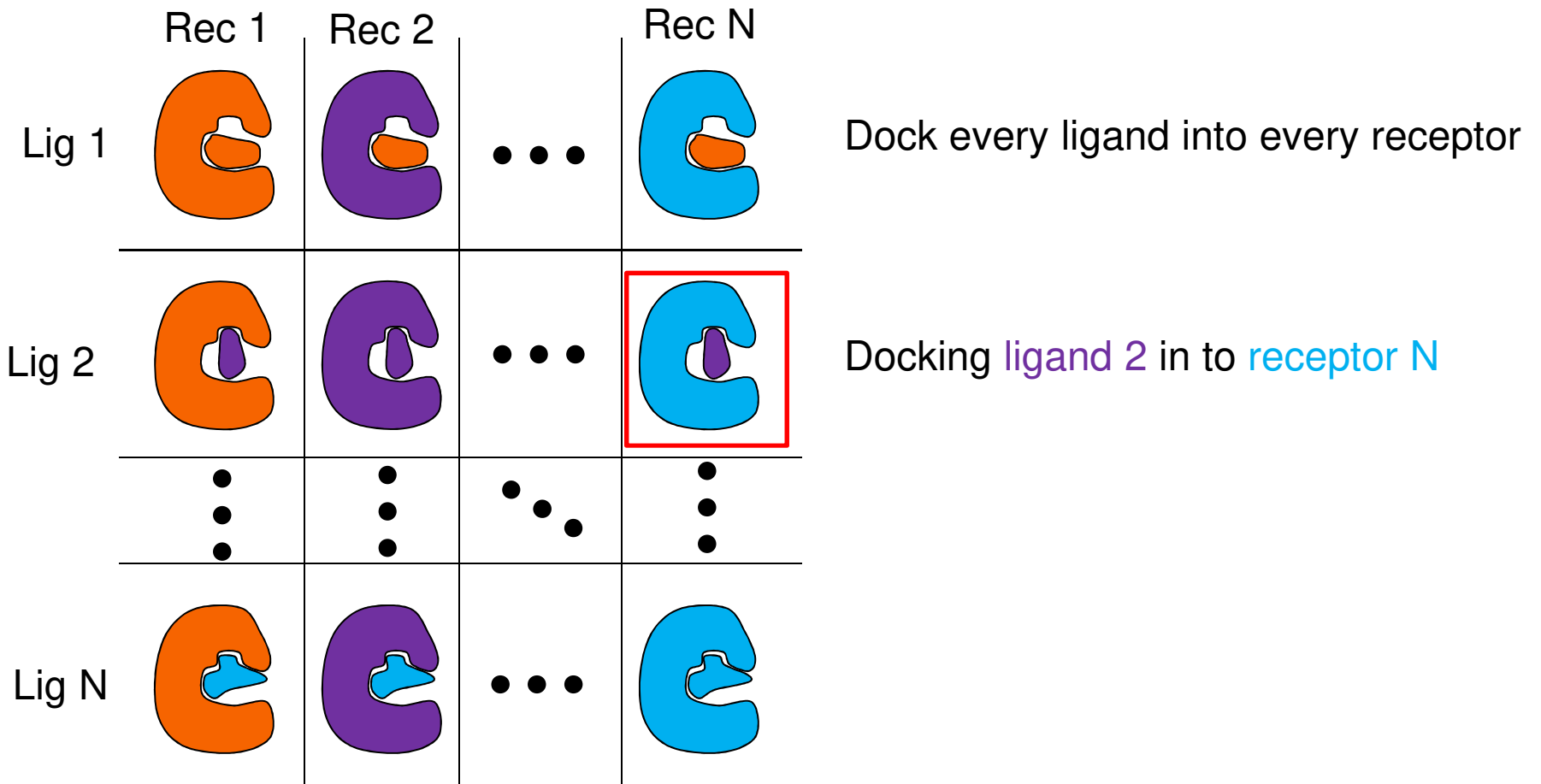
- DCE prefers larger ligands
- FPS prefers molecular weight close to that of erlotinib (393.44 g/mol)

Cross-Docking



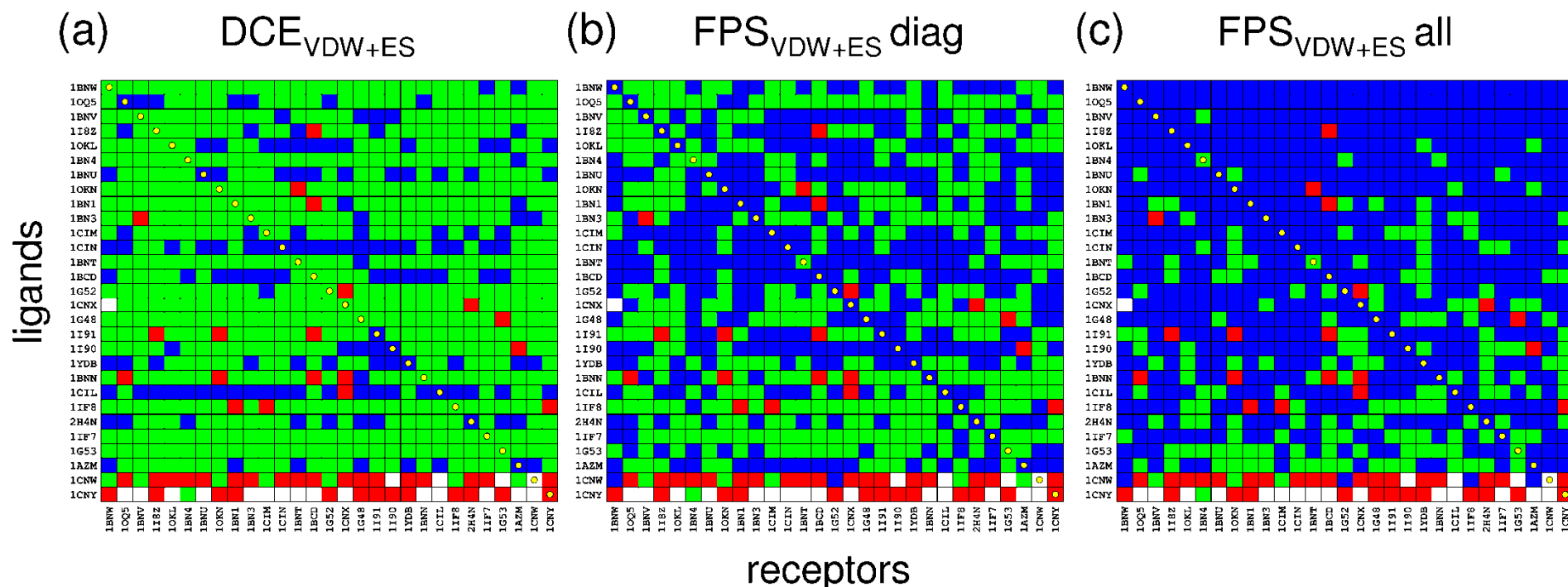
Dock every ligand into every receptor

Cross-Docking



[dock lig i in rec j]

Carbonic Anhydrase Crossdocking



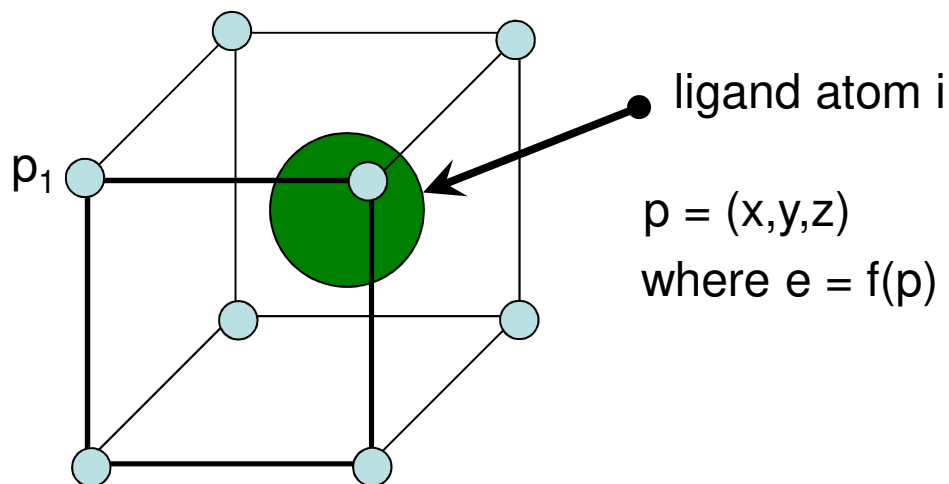
- success, blue; scoring failure, green; sampling failure, red; incomplete growth, white
- (a) standard DCE_{VDW+ES} ,
- (b) FPS_{VDW+ES} in which cognate ligands (diagonals) were used as the footprint-reference corresponding to each receptor
- (c) FPS_{VDW+ES} in which footprint-references were derived by minimizing each ligand in each receptor and every matrix element used a unique reference

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

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

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

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



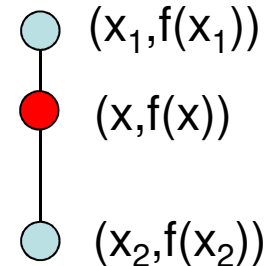
http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm#Grid

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

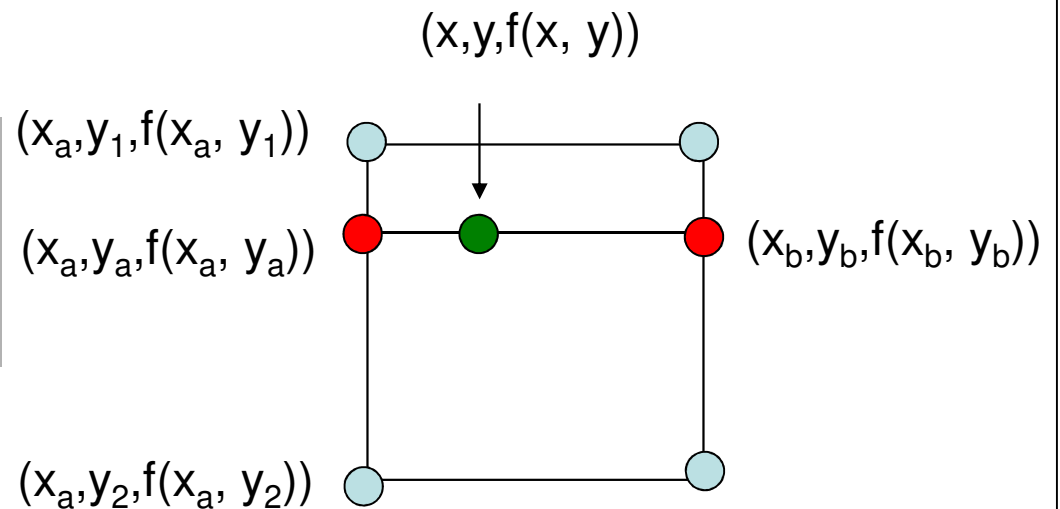
Interpolation

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**)

Decomposition for the grid

grid point

$$\begin{aligned} G_X(p) &= \sum_{l \in R} E_X(p, l) \\ &= \sum_{l \in SC_1} E_X(p, l) + \sum_{l \in SC_2} E_X(p, l) + \cdots + \sum_{l \in SC_M} E_X(p, l) \\ &= \sum_{k \in [1, M]} \sum_{l \in SC_k} E_X(p, l) \\ &= \sum_{k \in [1, M]} G_{X, SC_k}(p) \end{aligned}$$

$$R = [1, N]$$

set of receptor atoms

receptor has N atoms

receptor has M residues

DOCK program

- DOCK develop in 1980's in group of Irwin "Tack" D. Kuntz
Kuntz et al., *Journal of molecular biology* **161** (2): 269–88 (1982).
- Anchor-&-grow for ligand sampling (DOCK 4, 5, and 6)
Ewing et al., *Journal of computer-aided molecular design* **15** (5): 411–28 (2001) .
Moustakas et al., *Journal of computer-aided molecular design* **20** (10–11): 601–19 (2006) .

DOCK: A History

