

DOCK: where it is and where its going

Trent Balius, 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

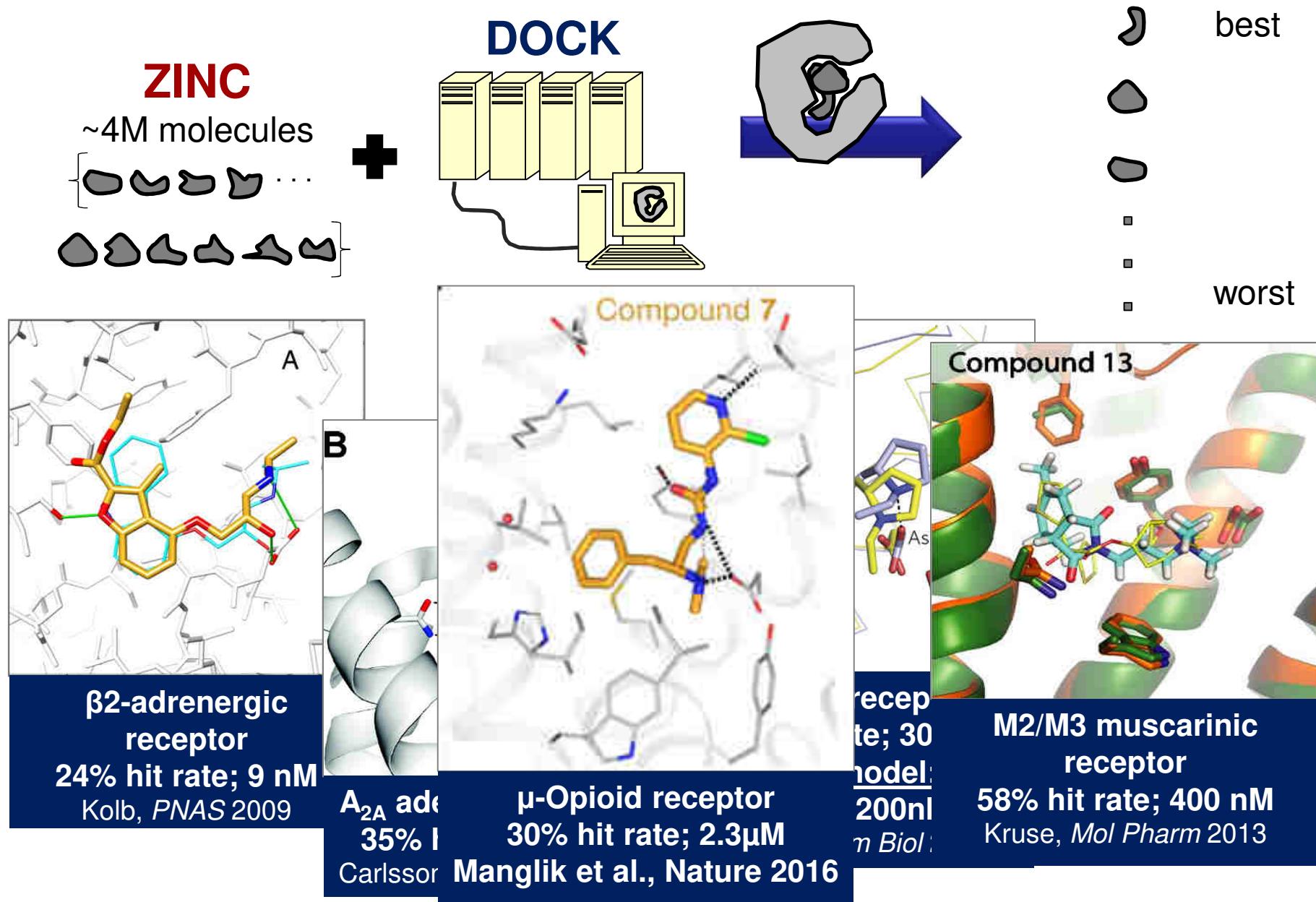
Events are posted on SBGrid Google Calendar (to join visit <http://sbgrid.org/calendar>)

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



Why use Dock?

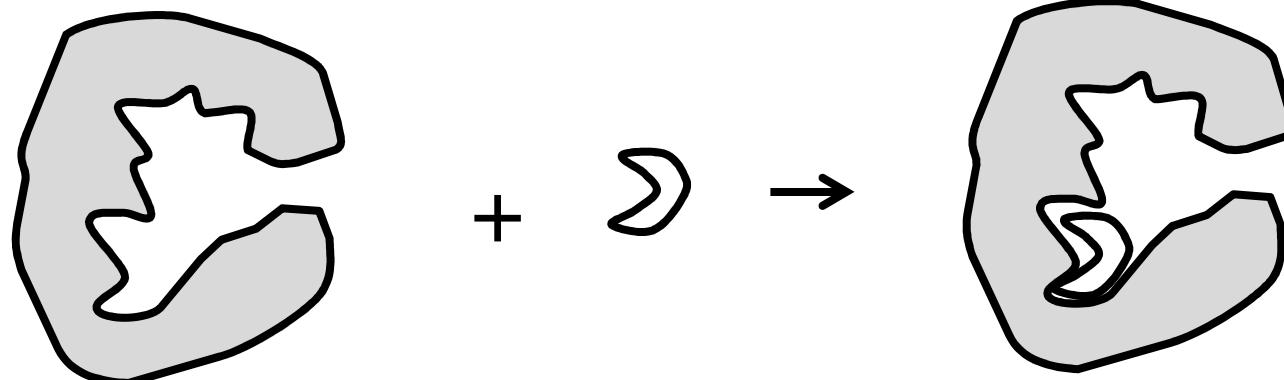
Applications of docking

- 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

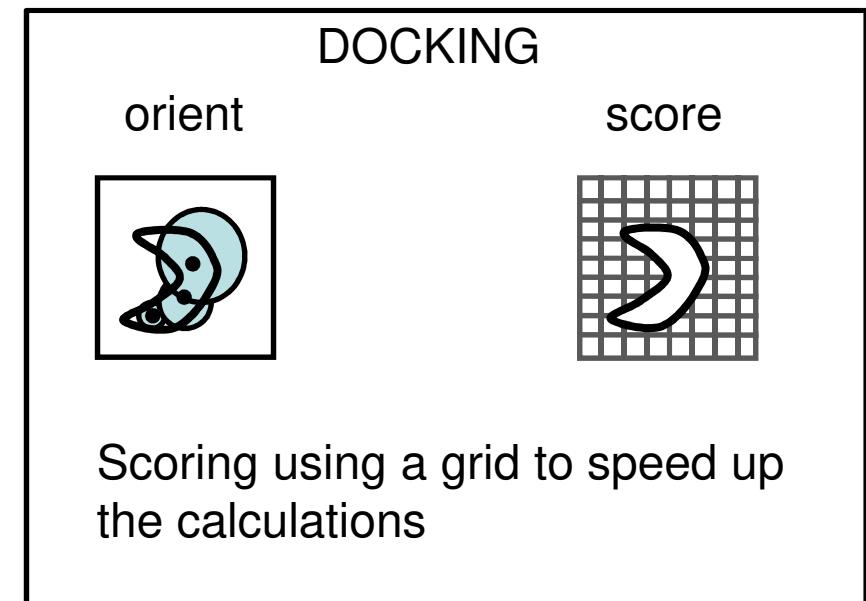
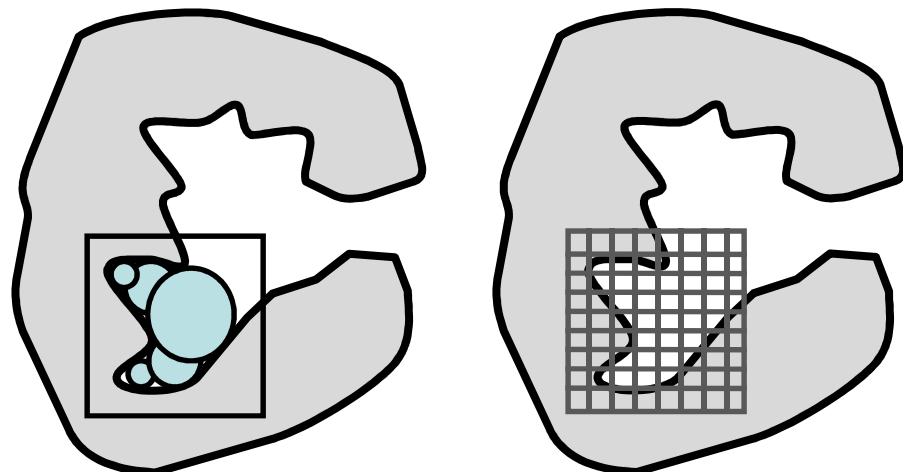
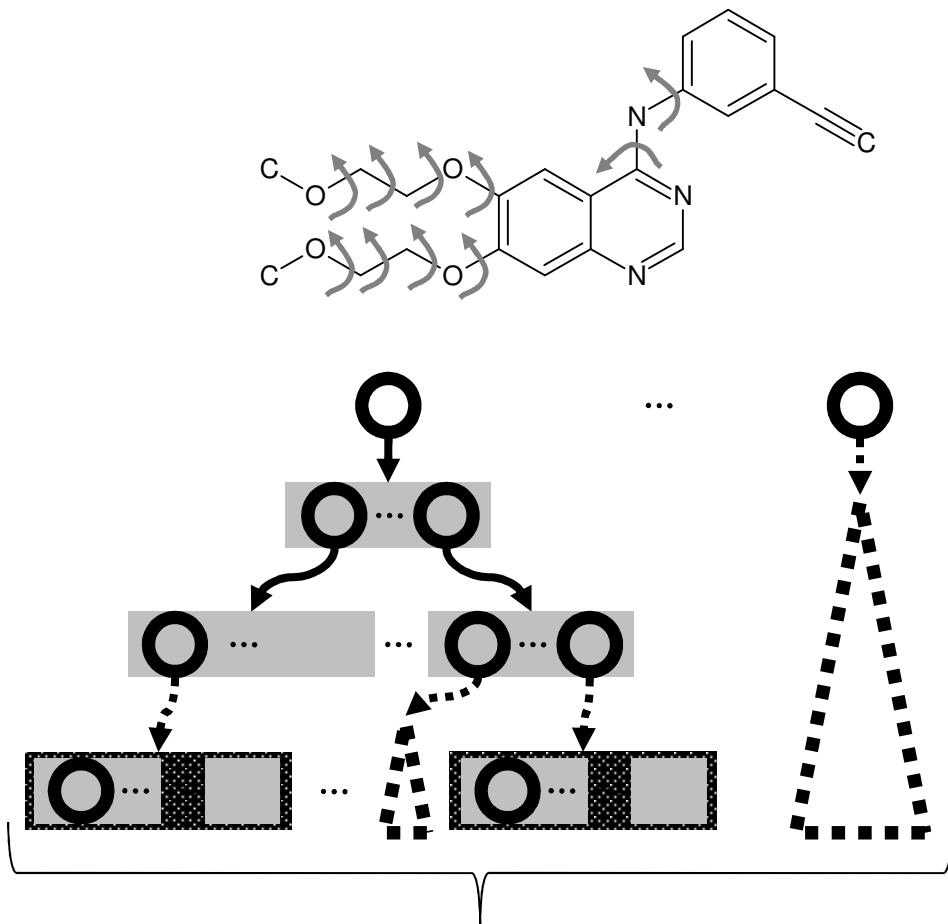


How to use DOCK 3.7

How DOCK 3.7 works

Preparation, Sampling, and Scoring

Ligand Sampling done outside of
DOCK 3.7 – database construction



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 QNFFT
 - 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 predict how Erlotinib
(an approved drug) binds to the
Epidermal Growth Factor Receptor

Search for Your Molecule in ZINC

zinc15.docking.org/substances/horne/ 80%

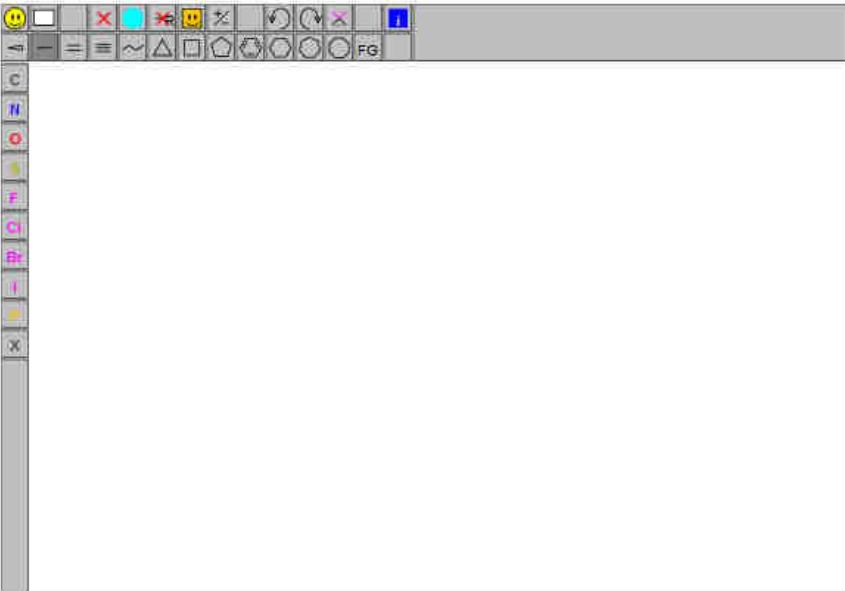
ZINC Substances Catalogs Tranches Biological More About

Substances

Help Examples Browse Table Subsets Shopping List Eritorinib Search

Draw/Search Structure

ZINC ID, SMILES, SMARTS, or InChI



Resolve Substances From Text/File

Paste SMILES

SMILES

Upload a File

Browse... No file selected

Allow Lookups

ZINC ID Structure Names Suppliers Analogs Slow!

Match Tolerance

Retired IDs Charge Scaffold Full Text Accept Multiple Results

Where to Search

Subset(s) to Check

Nothing selected

Results

Output Format Summary Table Resolve File

Get 2D Files from ZINC

Screenshot of the ZINC15 docking.org substance page for ZINC000001546066 (Erlotinib).

Header: zinc15.docking.org/substances/ZINC000001546066/

Navigation: ZINC, Substances, Catalogs, Tranches, Biological, More, About

Breadcrumbs: / substances / ZINC000001546066

Substance Name: 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	GSEA	Download

SMILES: C#Cc1ccccc(Nc2ncnc3cc(OCCOC)c(OCCOC)cc23)c1

InChI: InChI=1S/C22H23N3O/c1-4-16-5-7-17(12-18)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: AAKJLROGTJKAMG-UHFFFAOYSA-N

Chemical Structure:

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

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
KeyOrganics	KS-1202
Bioactives	
MedChem Exchange	HOV-470093

Annotated Catalogs (29 Total): 36 Items Total

Illuminating the Druggable Genome Screening Library	ML-SMR	Prestwick-6-B-11, Selleck-1-D-3
		126949132_164175282

Get Files for Docking from ZINC

Screenshot of the ZINC1546066 (Erlotinib) page on zinc15.docking.org.

The page shows the following details:

- ZINC1546066 (Erlotinib)**
- In:** bb fda for-sale in-stock standard
- Google Wikipedia PubMed**
- Added:** 2004-10-06 | **Available:** In-Stock | **Since:** 2015-08-07 | **Mwt:** 393.443 | **logP:** 3.405 | **Heavy Atoms:** 29 | **Tranche:** GGEA | **Download:** [Link](#)
- SMILES:** C#Cc1ccccc(Nc2ncnc3cc(OCCOC)c(OCCOC)cc23)c1
- InChI:** InChI=1S/C22H23N3O4/c1-4-16-5-7-17(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:** AAKJLRRGGTJKAMG-UHFFFAOYSA-N
- Chemical Structure:** A 2D chemical structure of Erlotinib is displayed.
- Available 3D Representations:** A table showing various properties:

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	Link
- Vendors (46 Total):** Aldrich CPR, CDS022564(ALDRICH), KeyOrganics, KS-1202
- Annotated Catalogs (29 Total):** Illuminating the Druggable Genome Screening Library, MLSMR, Prestwick-6-B-11, Selleck-1-D-3
- Links:** files.docking.org/protomers/16/89/55/209168955.db2.gz, [wget http://files.docking.org/protomers/16/89/55/209168955.db2.gz](http://files.docking.org/protomers/16/89/55/209168955.db2.gz)

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

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

Get Files Needed for Docking

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

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

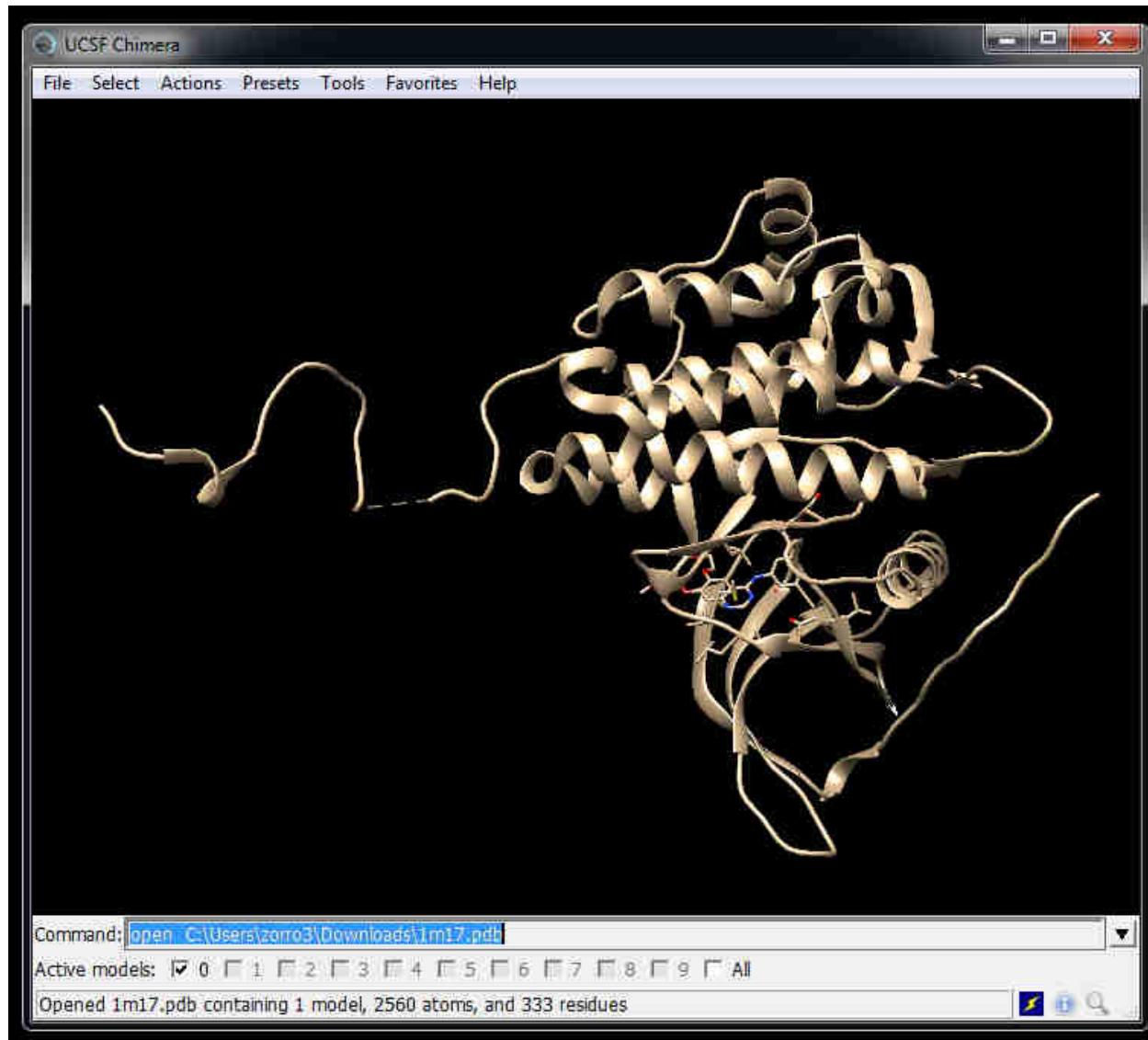
Put the path of the downloaded database into the split database index file (this file usually contain many db2 file):

```
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 were 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	
sim_rotstep	5.0	All molecules written out will be minimized
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

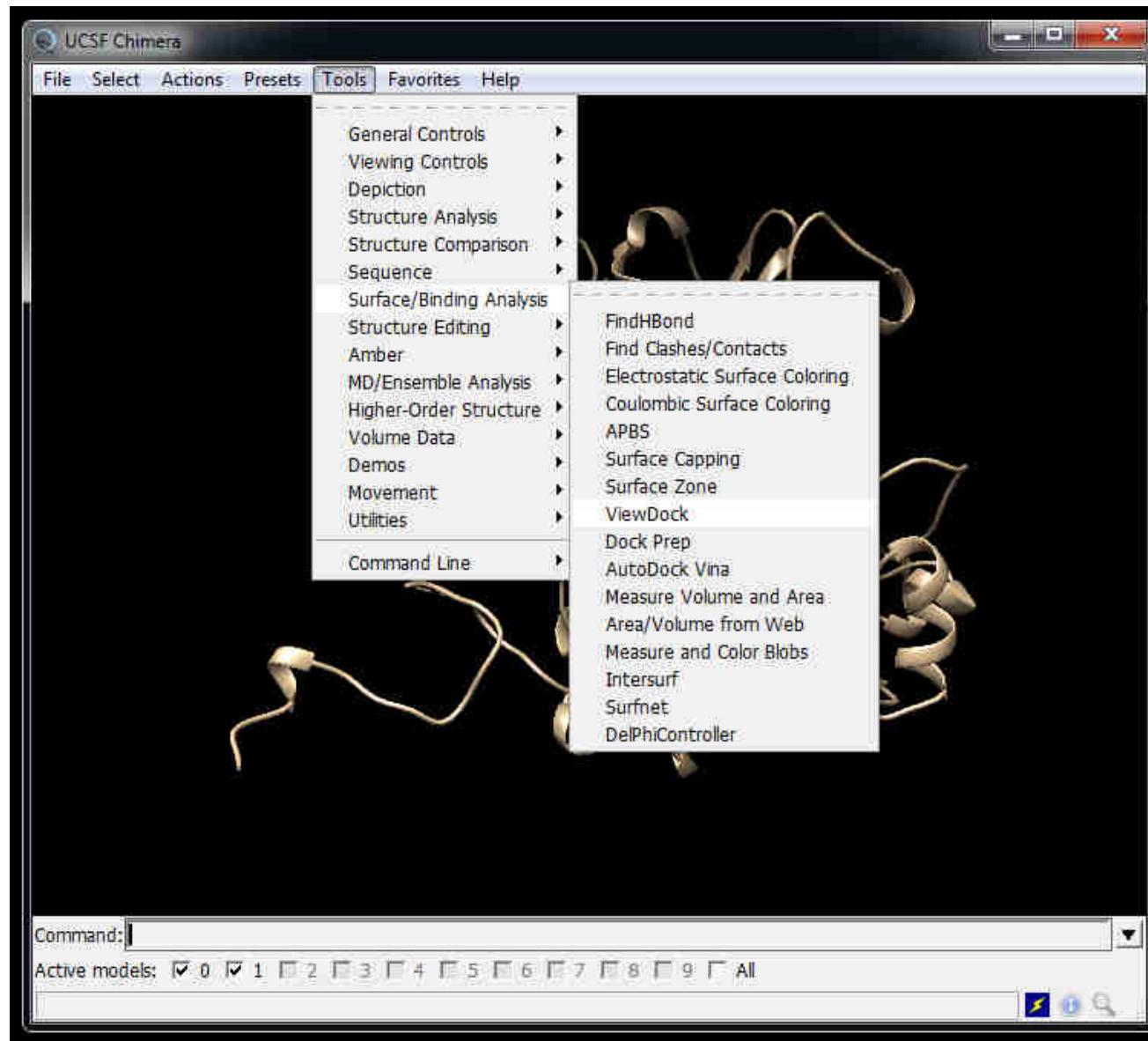
```
$DOCKBASE/analysis/extract_all.py  
$DOCKBASE/analysis/getposes.py
```

It is also possible to run dock locally:

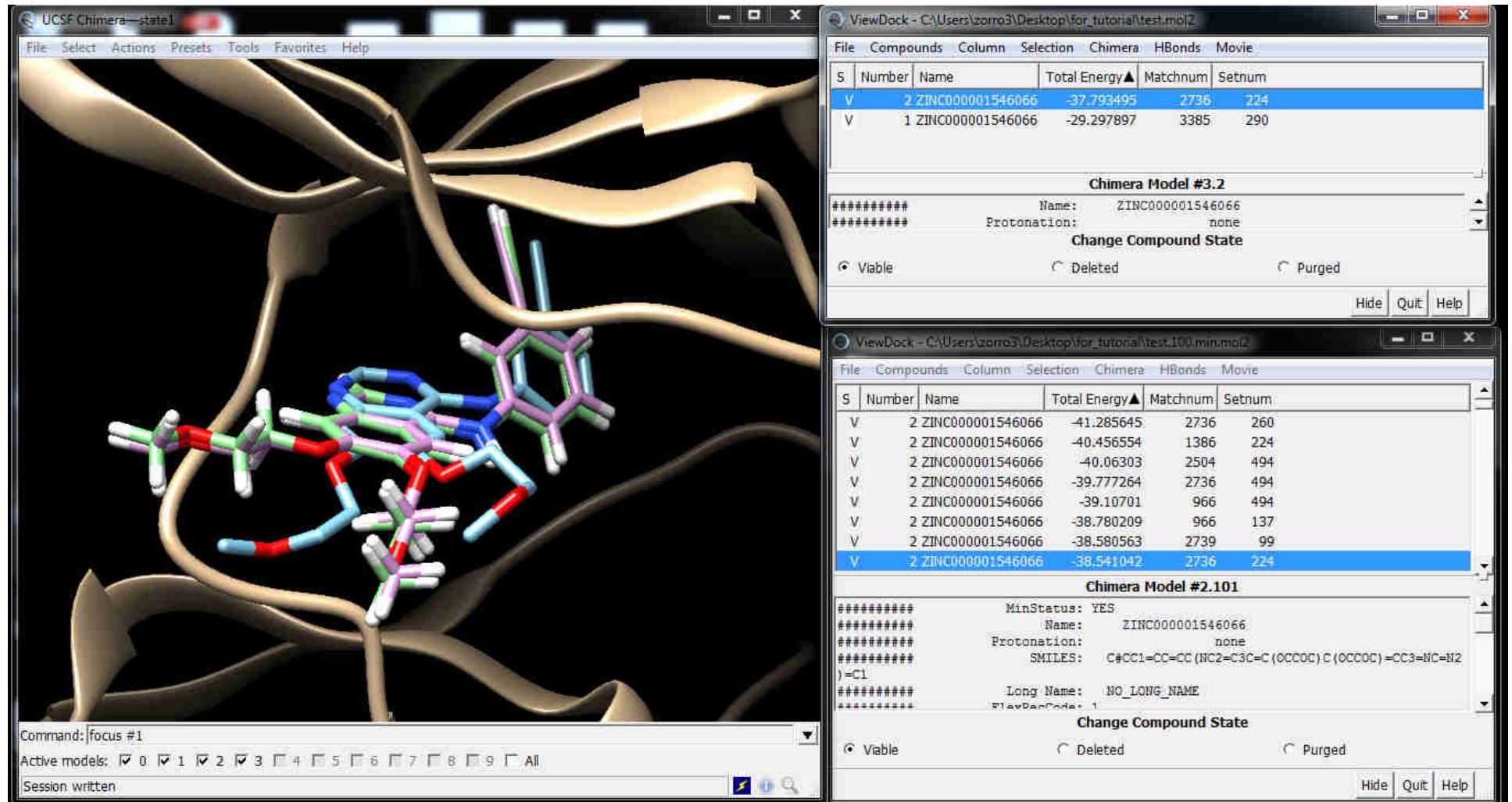
```
$DOCKBASE/docking/DOCK/bin/dock64
```

Output: OUTDOCK and test.mol2.gz

Visualize poses in Chimera with Viewdock

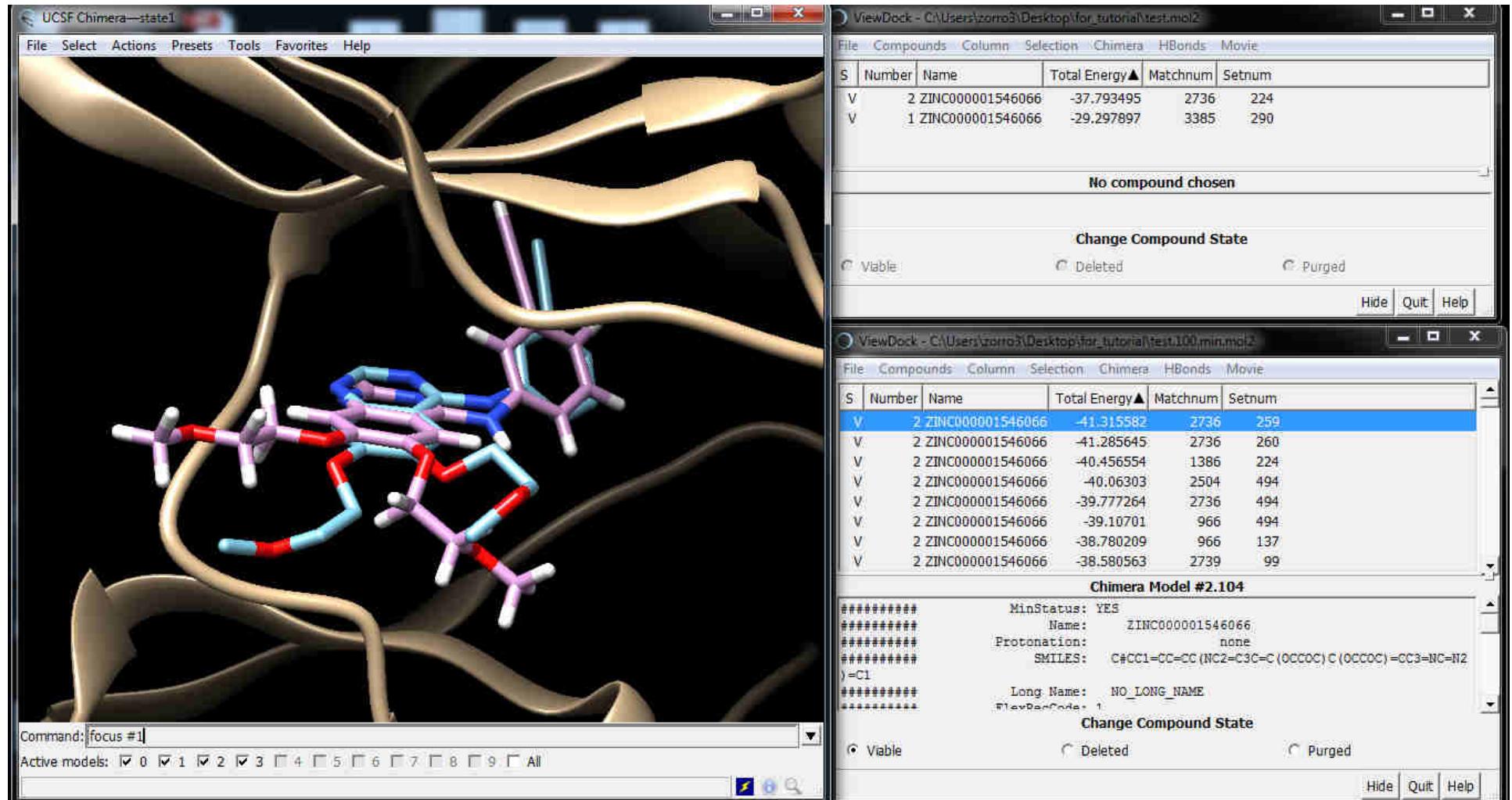


Visualize Poses in Chimera with Viewdock



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

Visualize Poses in Chimera with Viewdock

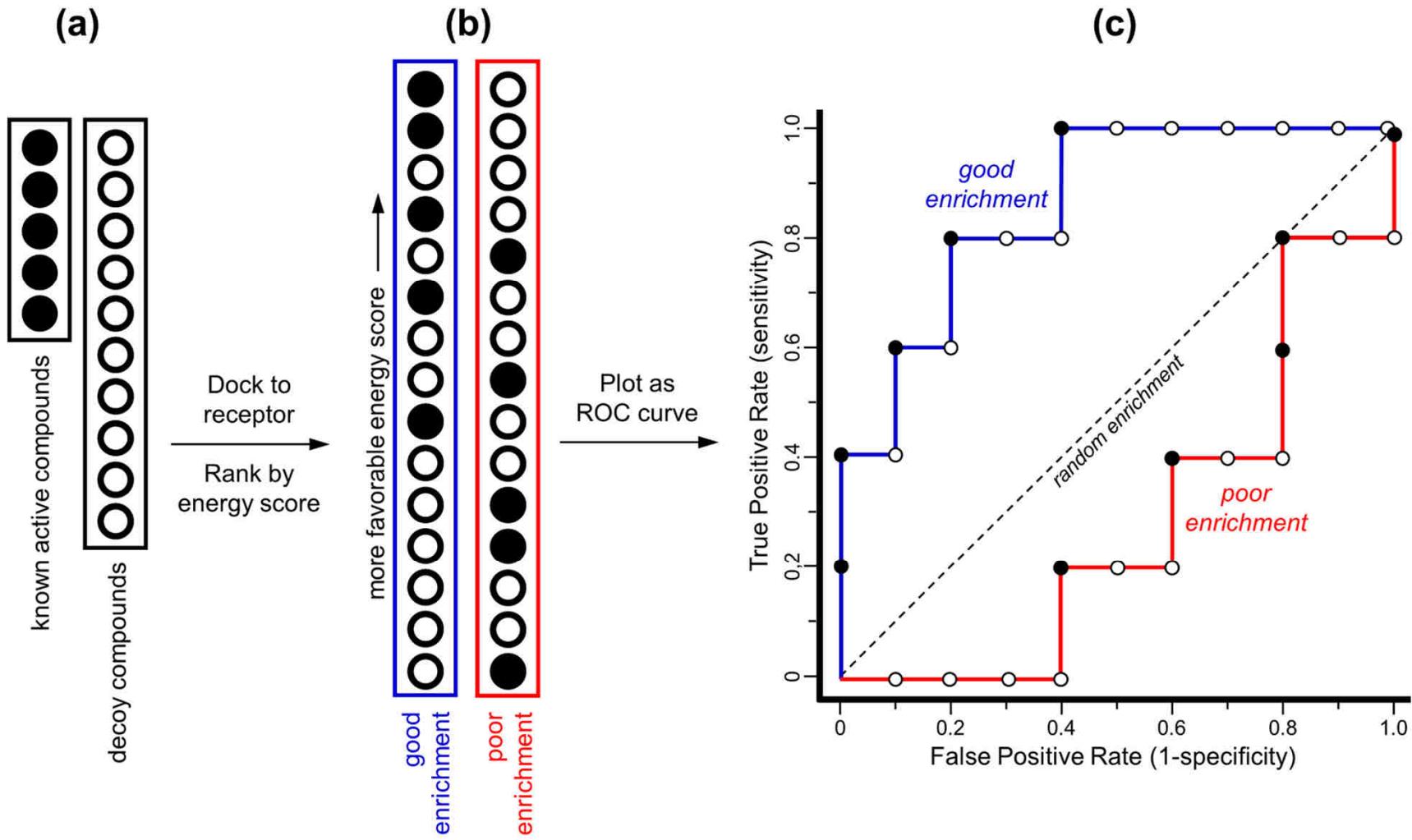


Best pose out of the top 100 after min: Energy -41.31

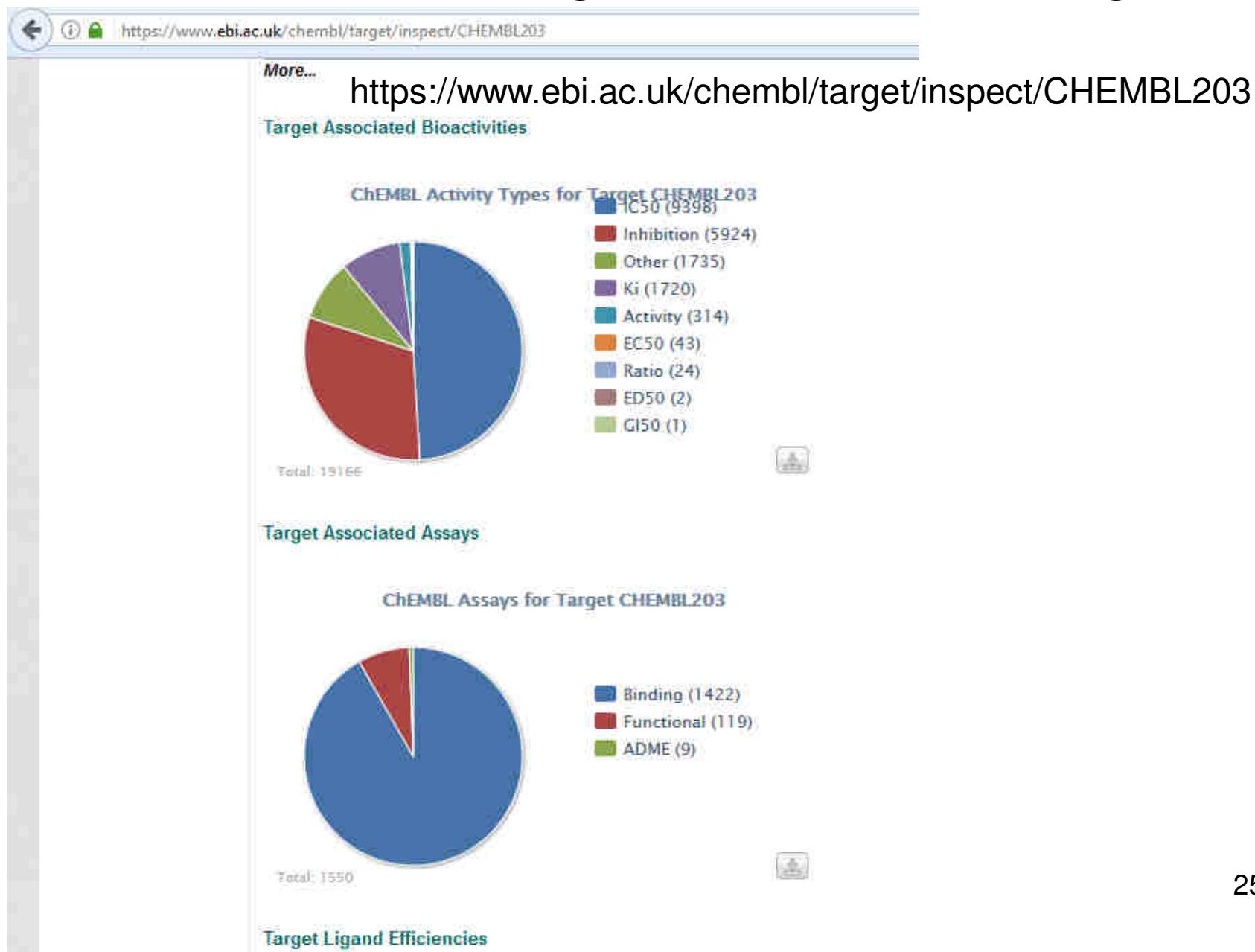
Tutorial

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



Get Known Ligands for Docking

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

this website uses cookies by continuing to browse this site, you are agreeing to the use of our cookies. to find out more, see our [Terms of Use](#).

EMBL-EBI Services Research Training About us

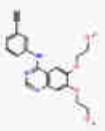
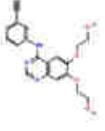
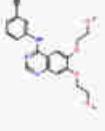
ChEMBL wellcome trust

Search ChEMBL... Compounds Targets Assays Documents Cells Tissues Exact Match Activity Source Filter

ChEMBL Bioactivity Search Results: 178

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

10 records per page

Ingredient	Molweight	Canonical SMILES	Standard Type	Relation	Standard Value	Standard Units	pChEMBL Value	Assay Type	Description	Assay Src Description	Assay Organism	Target Type	Target Name	Target Organism	Reference
	393.44	COCCOc1cc2ncnc(Nc3cccc(c3)C#Cc2cc1OCCOC)	Ki	=	3	nM	9.52	E	Inhibition of EGFR L850R mutant (unknown origin)	Scientific Literature	Homo sapiens	PROTEIN FAMILY	Epidermal growth factor receptor	Homo sapiens	J. Bios. Med. Chem. Lett. (2018) 26:2:514
	393.44	COCCOc1cc2ncnc(Nc3cccc(c3)C#Cc2cc1OCCOC)	Ki	=	135	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. (2018) 2:1:100
	393.44	COCCOc1cc2ncnc(Nc3cccc(c3)C#Cc2cc1OCCOC)	Ki	=	95	nM	7.02	E	Inhibition of EGFR T790M deletion (746 to 750 residues) mutant (unknown origin)	Scientific Literature	Homo sapiens	PROTEIN FAMILY	Epidermal growth factor receptor	Homo sapiens	J. Bios. Med. Chem. Lett. (2018) 26:2:514
	393.44	COCCOc1cc2ncnc(Nc3cccc(c3)C#Cc2cc1OCCOC)	Ki	=	95	nM	7.02	E	Inhibition of EGFR del746 to	Scientific Literature	Homo sapiens	PROTEIN FAMILY	Epidermal growth	Homo sapiens	J. Med. Chem. (2018)

Generate Decoy Smiles File

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

The screenshot shows the DUD•E website interface. At the top, there's a navigation bar with links for Home, Targets, Subsets, Generate, Other, FAQ, Revisions, and Thanks. Below the navigation is a large logo featuring the letters D, U, D, •, and E in a stylized font, with the text "A Database of Useful Decoys: Enhanced" underneath. The main section is titled "Generate DUD•E Decoys". It contains instructions for generating decoys, a result notification section with fields for email and CAPTCHA, and a large input area for pasting SMILES strings or uploading files. Below this input area, there's a note about accepted formats and an example of valid SMILES code. A note at the bottom states that generating 45 SMILES will take about 10 minutes on average.

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 Test: How are you today? (anything is fine)

Input SMILES by pasting OR uploading a text file

Uploaded SMILES [Browse](#) No file selected.

As an example, the following lines are accepted formats:

```
CC(C)(C)c1ccccc1c2ccccc2c3ccccc3c4ccccc4c5ccccc5c6ccccc6c7ccccc7c8ccccc8c9ccccc9c10ccccc10c11ccccc11c12ccccc12c13ccccc13c14ccccc14c15ccccc15c16ccccc16c17ccccc17c18ccccc18c19ccccc19c20ccccc20c21ccccc21c22ccccc22c23ccccc23c24ccccc24c25ccccc25c26ccccc26c27ccccc27c28ccccc28c29ccccc29c30ccccc30c31ccccc31c32ccccc32c33ccccc33c34ccccc34c35ccccc35c36ccccc36c37ccccc37c38ccccc38c39ccccc39c40ccccc40c41ccccc41c42ccccc42c43ccccc43c44ccccc44c45ccccc45c46ccccc46c47ccccc47c48ccccc48c49ccccc49c50ccccc50c51ccccc51c52ccccc52c53ccccc53c54ccccc54c55ccccc55c56ccccc56c57ccccc57c58ccccc58c59ccccc59c60ccccc60c61ccccc61c62ccccc62c63ccccc63c64ccccc64c65ccccc65c66ccccc66c67ccccc67c68ccccc68c69ccccc69c70ccccc70c71ccccc71c72ccccc72c73ccccc73c74ccccc74c75ccccc75c76ccccc76c77ccccc77c78ccccc78c79ccccc79c80ccccc80c81ccccc81c82ccccc82c83ccccc83c84ccccc84c85ccccc85c86ccccc86c87ccccc87c88ccccc88c89ccccc89c90ccccc90c91ccccc91c92ccccc92c93ccccc93c94ccccc94c95ccccc95c96ccccc96c97ccccc97c98ccccc98c99ccccc99c100ccccc100c101ccccc101c102ccccc102c103ccccc103c104ccccc104c105ccccc105c106ccccc106c107ccccc107c108ccccc108c109ccccc109c110ccccc110c111ccccc111c112ccccc112c113ccccc113c114ccccc114c115ccccc115c116ccccc116c117ccccc117c118ccccc118c119ccccc119c120ccccc120c121ccccc121c122ccccc122c123ccccc123c124ccccc124c125ccccc125c126ccccc126c127ccccc127c128ccccc128c129ccccc129c130ccccc130c131ccccc131c132ccccc132c133ccccc133c134ccccc134c135ccccc135c136ccccc136c137ccccc137c138ccccc138c139ccccc139c140ccccc140c141ccccc141c142ccccc142c143ccccc143c144ccccc144c145ccccc145c146ccccc146c147ccccc147c148ccccc148c149ccccc149c150ccccc150c151ccccc151c152ccccc152c153ccccc153c154ccccc154c155ccccc155c156ccccc156c157ccccc157c158ccccc158c159ccccc159c160ccccc160c161ccccc161c162ccccc162c163ccccc163c164ccccc164c165ccccc165c166ccccc166c167ccccc167c168ccccc168c169ccccc169c170ccccc170c171ccccc171c172ccccc172c173ccccc173c174ccccc174c175ccccc175c176ccccc176c177ccccc177c178ccccc178c179ccccc179c180ccccc180c181ccccc181c182ccccc182c183ccccc183c184ccccc184c185ccccc185c186ccccc186c187ccccc187c188ccccc188c189ccccc189c190ccccc190c191ccccc191c192ccccc192c193ccccc193c194ccccc194c195ccccc195c196ccccc196c197ccccc197c198ccccc198c199ccccc199c200ccccc200c201ccccc201c202ccccc202c203ccccc203c204ccccc204c205ccccc205c206ccccc206c207ccccc207c208ccccc208c209ccccc209c210ccccc210c211ccccc211c212ccccc212c213ccccc213c214ccccc214c215ccccc215c216ccccc216c217ccccc217c218ccccc218c219ccccc219c220ccccc220c221ccccc221c222ccccc222c223ccccc223c224ccccc224c225ccccc225c226ccccc226c227ccccc227c228ccccc228c229ccccc229c230ccccc230c231ccccc231c232ccccc232c233ccccc233c234ccccc234c235ccccc235c236ccccc236c237ccccc237c238ccccc238c239ccccc239c240ccccc240c241ccccc241c242ccccc242c243ccccc243c244ccccc244c245ccccc245c246ccccc246c247ccccc247c248ccccc248c249ccccc249c250ccccc250c251ccccc251c252ccccc252c253ccccc253c254ccccc254c255ccccc255c256ccccc256c257ccccc257c258ccccc258c259ccccc259c260ccccc260c261ccccc261c262ccccc262c263ccccc263c264ccccc264c265ccccc265c266ccccc266c267ccccc267c268ccccc268c269ccccc269c270ccccc270c271ccccc271c272ccccc272c273ccccc273c274ccccc274c275ccccc275c276ccccc276c277ccccc277c278ccccc278c279ccccc279c280ccccc280c281ccccc281c282ccccc282c283ccccc283c284ccccc284c285ccccc285c286ccccc286c287ccccc287c288ccccc288c289ccccc289c290ccccc290c291ccccc291c292ccccc292c293ccccc293c294ccccc294c295ccccc295c296ccccc296c297ccccc297c298ccccc298c299ccccc299c299ccccc299c300ccccc300c301ccccc301c302ccccc302c303ccccc303c304ccccc304c305ccccc305c306ccccc306c307ccccc307c308ccccc308c309ccccc309c310ccccc310c311ccccc311c312ccccc312c313ccccc313c314ccccc314c315ccccc315c316ccccc316c317ccccc317c318ccccc318c319ccccc319c320ccccc320c321ccccc321c322ccccc322c323ccccc323c324ccccc324c325ccccc325c326ccccc326c327ccccc327c328ccccc328c329ccccc329c330ccccc330c331ccccc331c332ccccc332c333ccccc333c334ccccc334c335ccccc335c336ccccc336c337ccccc337c338ccccc338c339ccccc339c340ccccc340c341ccccc341c342ccccc342c343ccccc343c344ccccc344c345ccccc345c346ccccc346c347ccccc347c348ccccc348c349ccccc349c350ccccc350c351ccccc351c352ccccc352c353ccccc353c354ccccc354c355ccccc355c356ccccc356c357ccccc357c358ccccc358c359ccccc359c360ccccc360c361ccccc361c362ccccc362c363ccccc363c364ccccc364c365ccccc365c366ccccc366c367ccccc367c368ccccc368c369ccccc369c370ccccc370c371ccccc371c372ccccc372c373ccccc373c374ccccc374c375ccccc375c376ccccc376c377ccccc377c378ccccc378c379ccccc379c380ccccc380c381ccccc381c382ccccc382c383ccccc383c384ccccc384c385ccccc385c386ccccc386c387ccccc387c388ccccc388c389ccccc389c390ccccc390c391ccccc391c392ccccc392c393ccccc393c394ccccc394c395ccccc395c396ccccc396c397ccccc397c398ccccc398c399ccccc399c3100ccccc3100c3101ccccc3101c3102ccccc3102c3103ccccc3103c3104ccccc3104c3105ccccc3105c3106ccccc3106c3107ccccc3107c3108ccccc3108c3109ccccc3109c3110ccccc3110c3111ccccc3111c3112ccccc3112c3113ccccc3113c3114ccccc3114c3115ccccc3115c3116ccccc3116c3117ccccc3117c3118ccccc3118c3119ccccc3119c3120ccccc3120c3121ccccc3121c3122ccccc3122c3123ccccc3123c3124ccccc3124c3125ccccc3125c3126ccccc3126c3127ccccc3127c3128ccccc3128c3129ccccc3129c3130ccccc3130c3131ccccc3131c3132ccccc3132c3133ccccc3133c3134ccccc3134c3135ccccc3135c3136ccccc3136c3137ccccc3137c3138ccccc3138c3139ccccc3139c3140ccccc3140c3141ccccc3141c3142ccccc3142c3143ccccc3143c3144ccccc3144c3145ccccc3145c3146ccccc3146c3147ccccc3147c3148ccccc3148c3149ccccc3149c3150ccccc3150c3151ccccc3151c3152ccccc3152c3153ccccc3153c3154ccccc3154c3155ccccc3155c3156ccccc3156c3157ccccc3157c3158ccccc3158c3159ccccc3159c3160ccccc3160c3161ccccc3161c3162ccccc3162c3163ccccc3163c3164ccccc3164c3165ccccc3165c3166ccccc3166c3167ccccc3167c3168ccccc3168c3169ccccc3169c3170ccccc3170c3171ccccc3171c3172ccccc3172c3173ccccc3173c3174ccccc3174c3175ccccc3175c3176ccccc3176c3177ccccc3177c3178ccccc3178c3179ccccc3179c3180ccccc3180c3181ccccc3181c3182ccccc3182c3183ccccc3183c3184ccccc3184c3185ccccc3185c3186ccccc3186c3187ccccc3187c3188ccccc3188c3189ccccc3189c3190ccccc3190c3191ccccc3191c3192ccccc3192c3193ccccc3193c3194ccccc3194c3195ccccc3195c3196ccccc3196c3197ccccc3197c3198ccccc3198c3199ccccc3199c3199ccccc3199c3200ccccc3200c3201ccccc3201c3202ccccc3202c3203ccccc3203c3204ccccc3204c3205ccccc3205c3206ccccc3206c3207ccccc3207c3208ccccc3208c3209ccccc3209c3210ccccc3210c3211ccccc3211c3212ccccc3212c3213ccccc3213c3214ccccc3214c3215ccccc3215c3216ccccc3216c3217ccccc3217c3218ccccc3218c3219ccccc3219c3220ccccc3220c3221ccccc3221c3222ccccc3222c3223ccccc3223c3224ccccc3224c3225ccccc3225c3226ccccc3226c3227ccccc3227c3228ccccc3228c3229ccccc3229c3230ccccc3230c3231ccccc3231c3232ccccc3232c3233ccccc3233c3234ccccc3234c3235ccccc3235c3236ccccc3236c3237ccccc3237c3238ccccc3238c3239ccccc3239c3240ccccc3240c3241ccccc3241c3242ccccc3242c3243ccccc3243c3244ccccc3244c3245ccccc3245c3246ccccc3246c3247ccccc3247c3248ccccc3248c3249ccccc3249c3250ccccc3250c3251ccccc3251c3252ccccc3252c3253ccccc3253c3254ccccc3254c3255ccccc3255c3256ccccc3256c3257ccccc3257c3258ccccc3258c3259ccccc3259c3260ccccc3260c3261ccccc3261c3262ccccc3262c3263ccccc3263c3264ccccc3264c3265ccccc3265c3266ccccc3266c3267ccccc3267c3268ccccc3268c3269ccccc3269c3270ccccc3270c3271ccccc3271c3272ccccc3272c3273ccccc3273c3274ccccc3274c3275ccccc3275c3276ccccc3276c3277ccccc3277c3278ccccc3278c3279ccccc3279c3280ccccc3280c3281ccccc3281c3282ccccc3282c3283ccccc3283c3284ccccc3284c3285ccccc3285c3286ccccc3286c3287ccccc3287c3288ccccc3288c3289ccccc3289c3290ccccc3290c3291ccccc3291c3292ccccc3292c3293ccccc3293c3294ccccc3294c3295ccccc3295c3296ccccc3296c3297ccccc3297c3298ccccc3298c3299ccccc3299c3299ccccc3299c3300ccccc3300c3301ccccc3301c3302ccccc3302c3303ccccc3303c3304ccccc3304c3305ccccc3305c3306ccccc3306c3307ccccc3307c3308ccccc3308c3309ccccc3309c3310ccccc3310c3311ccccc3311c3312ccccc3312c3313ccccc3313c3314ccccc3314c3315ccccc3315c3316ccccc3316c3317ccccc3317c3318ccccc3318c3319ccccc3319c3320ccccc3320c3321ccccc3321c3322ccccc3322c3323ccccc3323c3324ccccc3324c3325ccccc3325c3326ccccc3326c3327ccccc3327c3328ccccc3328c3329ccccc3329c3330ccccc3330c3331ccccc3331c3332ccccc3332c3333ccccc3333c3334ccccc3334c3335ccccc3335c3336ccccc3336c3337ccccc3337c3338ccccc3338c3339ccccc3339c3340ccccc3340c3341ccccc3341c3342ccccc3342c3343ccccc3343c3344ccccc3344c3345ccccc3345c3346ccccc3346c3347ccccc3347c3348ccccc3348c3349ccccc3349c3350ccccc3350c3351ccccc3351c3352ccccc3352c3353ccccc3353c3354ccccc3354c3355ccccc3355c3356ccccc3356c3357ccccc3357c3358ccccc3358c3359ccccc3359c3360ccccc3360c3361ccccc3361c3362ccccc3362c3363ccccc3363c3364ccccc3364c3365ccccc3365c3366ccccc3366c3367ccccc3367c3368ccccc3368c3369ccccc3369c3370ccccc3370c3371ccccc3371c3372ccccc3372c3373ccccc3373c3374ccccc3374c3375ccccc3375c3376ccccc3376c3377ccccc3377c3378ccccc3378c3379ccccc3379c3380ccccc3380c3381ccccc3381c3382ccccc3382c3383ccccc3383c3384ccccc3384c3385ccccc3385c3386ccccc3386c3387ccccc3387c3388ccccc3388c3389ccccc3389c3390ccccc3390c3391ccccc3391c3392ccccc3392c3393ccccc3393c3394ccccc3394c3395ccccc3395c3396ccccc3396c3397ccccc3397c3398ccccc3398c3399ccccc3399c3399ccccc3399c3400ccccc3400c3401ccccc3401c3402ccccc3402c3403ccccc3403c3404ccccc3404c3405ccccc3405c3406ccccc3406c3407ccccc3407c3408ccccc3408c3409ccccc3409c3410ccccc3410c3411ccccc3411c3412ccccc3412c3413ccccc3413c3414ccccc3414c3415ccccc3415c3416ccccc3416c3417ccccc3417c3418ccccc3418c3419ccccc3419c3420ccccc3420c3421ccccc3421c3422ccccc3422c3423ccccc3423c3424ccccc3424c3425ccccc3425c3426ccccc3426c3427ccccc3427c3428ccccc3428c3429ccccc3429c3430ccccc3430c3431ccccc3431c3432ccccc3432c3433ccccc3433c3434ccccc3434c3435ccccc3435c3436ccccc3436c3437ccccc3437c3438ccccc3438c3439ccccc3439c3440ccccc3440c3441ccccc3441c3442ccccc3442c3443ccccc3443c3444ccccc3444c3445ccccc3445c3446ccccc3446c3447ccccc3447c3448ccccc3448c3449ccccc3449c3450ccccc3450c3451ccccc3451c3452ccccc3452c3453ccccc3453c3454ccccc3454c3455ccccc3455c3456ccccc3456c3457ccccc3457c3458ccccc3458c3459ccccc3459c3460ccccc3460c3461ccccc3461c3462ccccc3462c3463ccccc3463c3464ccccc3464c3465ccccc3465c3466ccccc3466c3467ccccc3467c3468ccccc3468c3469ccccc3469c3470ccccc3470c3471ccccc3471c3472ccccc3472c3473ccccc3473c3474ccccc3474c3475ccccc3475c3476ccccc3476c3477ccccc3477c3478ccccc3478c3479ccccc3479c3480ccccc3480c3481ccccc3481c3482ccccc3482c3483ccccc3483c3484ccccc3484c3485ccccc3485c3486ccccc3486c3487ccccc3487c3488ccccc3488c3489ccccc3489c3490ccccc3490c3491ccccc3491c3492ccccc3492c3493ccccc3493c3494ccccc3494c3495ccccc3495c3496ccccc3496c3497ccccc3497c3498ccccc3498c3499ccccc3499c3499ccccc3499c3500ccccc3500c3501ccccc3501c3502ccccc3502c3503ccccc3503c3504ccccc3504c3505ccccc3505c3506ccccc3506c3507ccccc3507c3508ccccc3508c3509ccccc3509c3510ccccc3510c3511ccccc3511c3512ccccc3512c3513ccccc3513c3514ccccc3514c3515ccccc3515c3516ccccc3516c3517ccccc3517c3518ccccc3518c3519ccccc3519c3520ccccc3520c3521ccccc3521c3522ccccc3522c3523ccccc3523c3524ccccc3524c3525ccccc3525c3526ccccc3526c3527ccccc3527c3528ccccc3528c3529ccccc3529c3530ccccc3530c3531ccccc3531c3532ccccc3532c3533ccccc3533c3534ccccc3534c3535ccccc3535c3536ccccc3536c3537ccccc3537c3538ccccc3538c3539ccccc3539c3540ccccc3540c3541ccccc3541c3542ccccc3542c3543ccccc3543c3544ccccc3544c3545ccccc3545c3546ccccc3546c3547ccccc3547c3548ccccc3548c3549ccccc3549c3550ccccc3550c3551ccccc3551c3552ccccc3552c3553ccccc3553c3554ccccc3554c3555ccccc3555c3556ccccc3556c3557ccccc3557c3558ccccc3558c3559ccccc3559c3560ccccc3560c3561ccccc3561c3562ccccc3562c3563ccccc3563c3564ccccc3564c3565ccccc3565c3566ccccc3566c3567ccccc3567c3568ccccc3568c3569ccccc3569c3570ccccc3570c3571ccccc3571c3572ccccc3572c3573ccccc3573c3574ccccc3574c3575ccccc3575c3576ccccc3576c3577ccccc3577c3578ccccc3578c3579ccccc3579c3580ccccc3580c3581ccccc3581c3582ccccc3582c3583ccccc3583c3584ccccc3584c3585ccccc3585c3586ccccc3586c3587ccccc3587c3588ccccc3588c3589ccccc3589c3590ccccc3590c3591ccccc3591c3592ccccc3592c3593ccccc3593c3594ccccc3594c3595ccccc3595c3596ccccc3596c3597ccccc3597c3598ccccc3598c3599ccccc3599c3599ccccc3599c3600ccccc3600c3601ccccc3601c3602ccccc3602c3603ccccc3603c3604ccccc3604c3605ccccc3605c3606ccccc3606c3607ccccc3607c3608ccccc3608c3609ccccc3609c3610ccccc3610c3611ccccc3611c3612ccccc3612c3613ccccc3613c3614ccccc3614c3615ccccc3615c3616ccccc3616c3617ccccc3617c3618ccccc3618c3619ccccc3619c3620ccccc3620c3621ccccc3621c3622ccccc3622c3623ccccc3623c3624ccccc3624c3625ccccc3625c3626ccccc3626c3627ccccc3627c3628ccccc3628c3629ccccc3629c3630ccccc3630c3631ccccc3631c3632ccccc3632c3633ccccc3633c3634ccccc3634c3635ccccc3635c3636ccccc3636c3637ccccc3637c3638ccccc3638c3639ccccc3639c3640ccccc3640c3641ccccc3641c3642ccccc3642c3643ccccc3643c3644ccccc3644c3645ccccc3645c3646ccccc3646c3647ccccc3647c3648ccccc3648c3649ccccc3649c3650ccccc3650c3651ccccc3651c3652ccccc3652c3653ccccc3653c3654ccccc3654c3655ccccc3655c3656ccccc3656c3657ccccc3657c3658ccccc3658c3659ccccc3659c3660ccccc3660c3661ccccc3661c3662ccccc3662c3663ccccc3663c3664ccccc3664c3665ccccc3665c3666ccccc3666c3667ccccc3667c3668ccccc3668c3669ccccc3669c3670ccccc3670c3671ccccc3671c3672ccccc3672c3673ccccc3673c3674ccccc3674c3675ccccc3675c3676ccccc3676c3677ccccc3677c3678ccccc3678c3679ccccc3679c3680ccccc3680c3681ccccc3681c3682ccccc3682c3683ccccc3683c3684ccccc3684c3685ccccc3685c3686ccccc3686c3687ccccc3687c3688ccccc3688c3689ccccc3689c3690ccccc3690c3691ccccc3691c3692ccccc3692c3693ccccc3693c3694ccccc3694c3695ccccc3695c3696ccccc3696c3697ccccc3697c3698ccccc3698c3699ccccc3699c3699ccccc3699c3700ccccc3700c3701ccccc3701c3702ccccc3702c3703ccccc3703c3704ccccc3704c3705ccccc3705c3706ccccc3706c3707ccccc3707c3708ccccc3708c3709ccccc3709c3710ccccc3710c3711ccccc3711c3712ccccc3712c3713ccccc3713c3714ccccc3714c3715ccccc3715c3716ccccc3716c3717ccccc3717c3718ccccc3718c3719ccccc3719c3720ccccc3720c3721ccccc3721c3722ccccc3722c3723ccccc3723c3724ccccc3724c3725ccccc3725c3726ccccc3726c3727ccccc3727c3728ccccc3728c3729ccccc3729c3730ccccc3730c3731ccccc3731c3732ccccc3732c3733ccccc3733c3734ccccc3734c3735ccccc3735c3736ccccc3736c3737ccccc3737c3738ccccc3738c3739ccccc3739c3740ccccc3740c3741ccccc3741c3742ccccc3742c3743ccccc3743c3744ccccc3744c3745ccccc3745c3746ccccc3746c3747ccccc3747c3748ccccc3748c3749ccccc3749c3750ccccc3750c3751ccccc3751c3752ccccc3752c3753ccccc3753c3754ccccc3754c3755ccccc3755c3756ccccc3756c3757ccccc3757c3758ccccc3758c3759ccccc3759c3760ccccc3760c3761ccccc3761c3762ccccc3762c3763ccccc3763c3764ccccc3764c3765ccccc3765c3766ccccc3766c3767ccccc3767c3768ccccc3768c3769
```

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 -xzvf 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 wraper_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

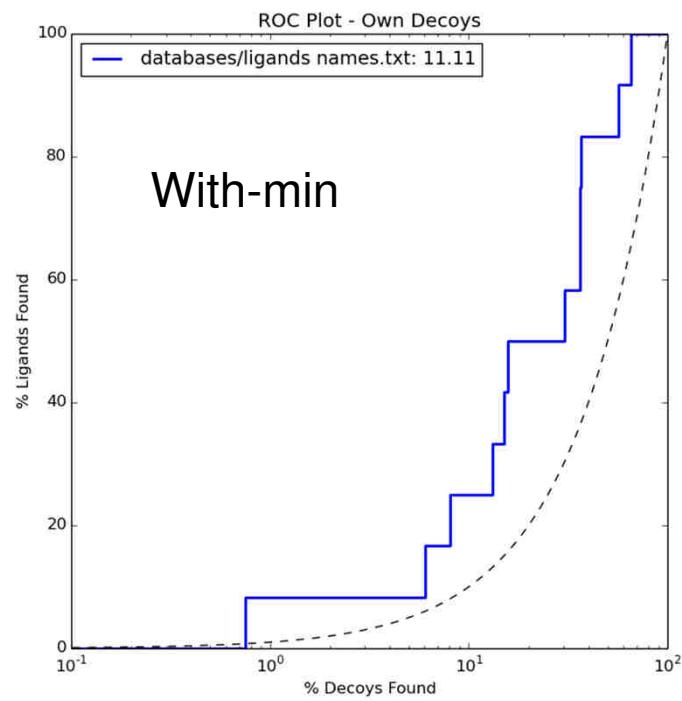
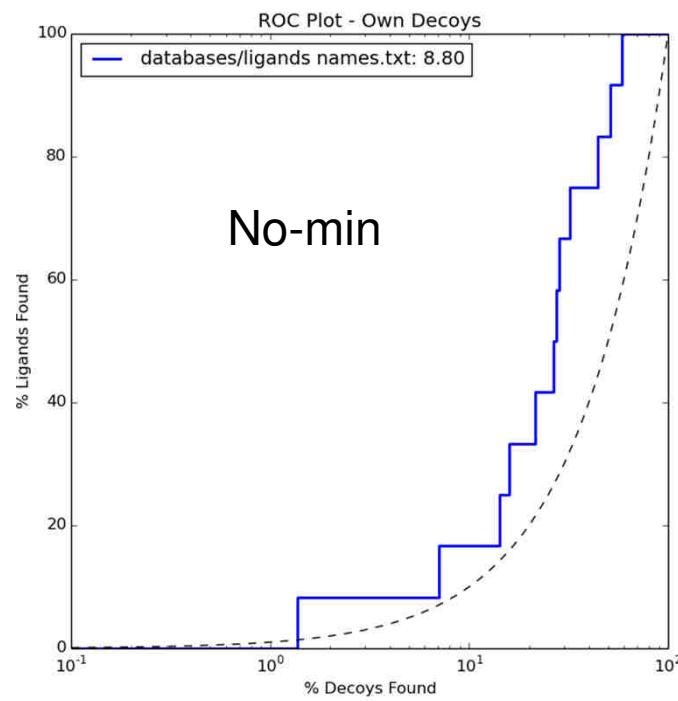
Process results combining results and get the best poses:

```
$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

Molecular Weight (up to, Daltons)												Totals, by LogP	
	200	250	300	325	350	375	400	425	450	500	>500		
-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	458,663	1,248,854	838,607	880,249	677,936	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	124,795	5,719	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	8,838,205	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	5,007,277	1,039,924	43,736	58,169,567	
4	8,586	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	4,002	181,753	572,091	1,421,241	2,844,759	4,042,678	3,132,224	2,858,306	1,516,287	111,402	18,685,579	
>5	38	1,062	18,400	88,882	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

Screenshot of the ZINC15 Tranches interface showing a grid of molecular weights and LogP values.

The grid displays data for various molecular weights (200, 250, 300, 325, 350, 375, 400, 425) across different LogP ranges (from -1 to >5). The counts for each entry are shown in the grid cells.

A dropdown menu titled "Predefined Subsets" is open, listing categories: All, None, Shards, Fragments, Fragments, Goldilocks, Lead-Like, Lugs, Drug-Like, and Big-n-Greasy. The "Lead-Like" subset is currently selected.

Below the grid, "Totals, by Weight" and "2.1K Tranches" are displayed.

Molecular Weight (up to, Daltons)										
	200	250	300	325	350	375	400	425		
-1	1,100	1,100	9,009	4,838	4,200	1,100	1,100	1,100		
0	1,100	1,100	31,837	20,634	21,415	1,100	1,100	1,100		
1	1,100	1,100	146,886	104,162	114,802	1,100	1,100	1,100		
2	1,100	1,100	387,262	309,819	389,357	1,100	1,100	1,100		
2.5	1,100	1,100	272,450	142,738	320,806	1,100	1,100	1,100		
3	1,100	1,100	259,976	247,005	366,877	1,100	1,100	1,100		
3.5	1,100	1,100	202,915	203,483	324,436	1,100	1,100	1,100		
4	1,100	1,100	1,100	1,100	1,100	1,100	1,100	1,100		0
4.5	1,100	1,100	1,100	1,100	1,100	1,100	1,100	1,100		0
5	1,100	1,100	1,100	1,100	1,100	1,100	1,100	1,100		0
>5	1,100	1,100	1,100	1,100	1,100	1,100	1,100	1,100		0
Totals, by Weight	0	0	1,310,335	1,037,679	1,541,893	0	0	0	0	3,889,907
										Protomers
										2.1K Tranches

Get Lead-like Molecules to DOCK

The screenshot shows a web browser window for the ZINC15 Docking website (zinc15.docking.org/tranches/home/#). The main interface displays a histogram of molecules by LogP (up to 6) and a table of totals by LogP. A modal dialog box titled "Download Tranches" is overlaid on the page. The dialog box contains the following information:

- 2,110 (Non-Empty) Tranches Selected (3,889,907 Protomers)**
- A scrollable list of protomer SMILES strings, starting with:

```
CAAAML CAAAMM CAAAMN CAAAMO CAAAMP CAAARL CAAARM CAAARN  
CAAARO CAAARP CAABML CAABMM CAABMN CAABMO CAABMP CAABRL  
CAABRM CAABRN CAABRO CAABRP CAACML CAACMM CAACMN CAACMO  
CAACMP CAACRL CAACRM CAACRN CAACRO CAACRP CABARM CABARN  
CABARO CABBRM CABBRN CABBRP CABCRM CABCRN CACAML CACAMM  
CACAMN CACAMO CACARL CACARM CACARN CACARO CACBMM CACBMN  
CACBRL CACBRM CACBRN CACGML CACGMM CACCMN CACCMO CACCRL
```
- Only If Modified Since**: An empty input field.
- Download Options**: Buttons for "DOCK37 (*.db2.gz)" and "Raw URLs".
- Download Button**: A blue button with a download icon labeled "Download".

The background of the main interface shows a histogram with x-axis ticks at 200, 250, 300, 450, 500, and >500, and a y-axis labeled "LogP (up to 6)". Below the histogram is a table with columns for "LogP" ranges and "Totals, by Weight". The table includes the following data:

LogP Range	Totals, by Weight
<1	18,047
1	73,886
2	365,850
2.5	1,086,438
3	735,994
3.5	873,858
4	735,834
4.5	0
5	0
>5	0
Totals, by Weight	3,889,907 Protomers
	2.1K Tranches

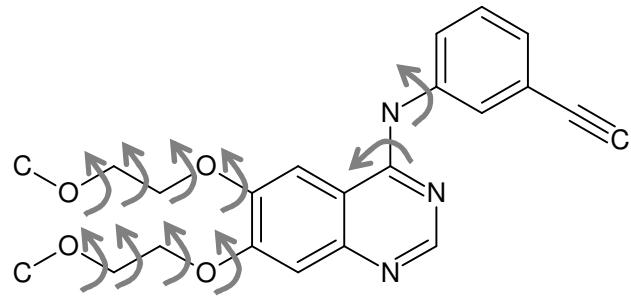
Features in DOCK 3.7

- 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

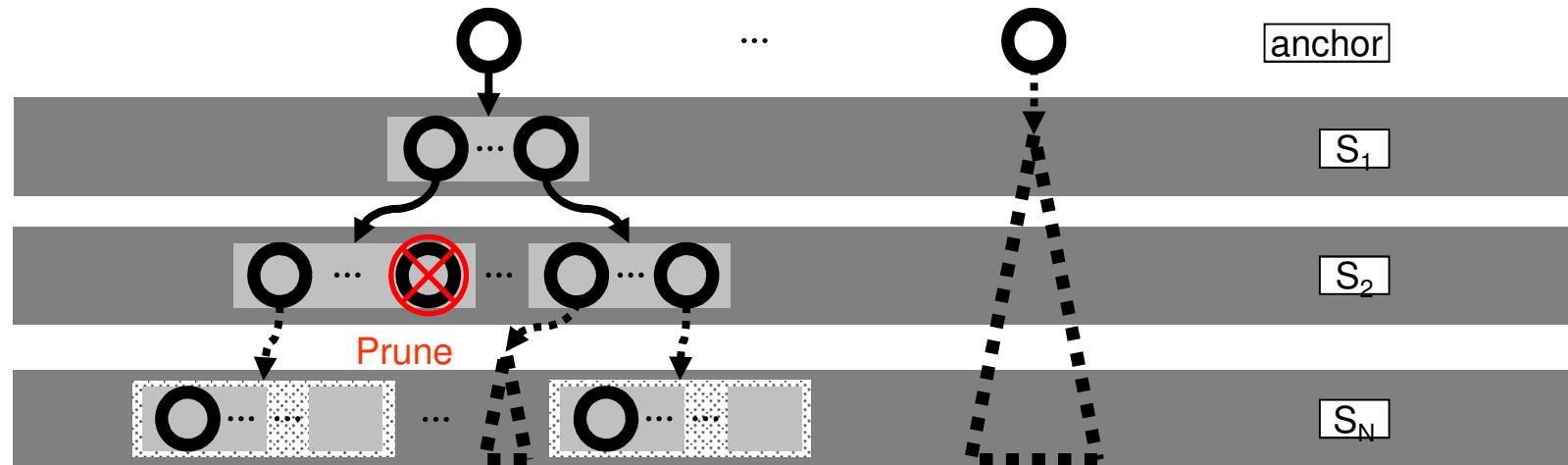
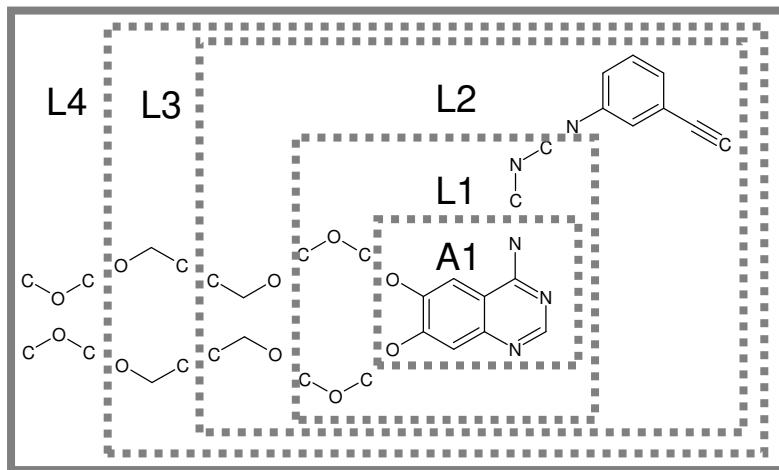
Start from anchor orient

Arrange segments by layer about anchor center

Grow each segment one at a time

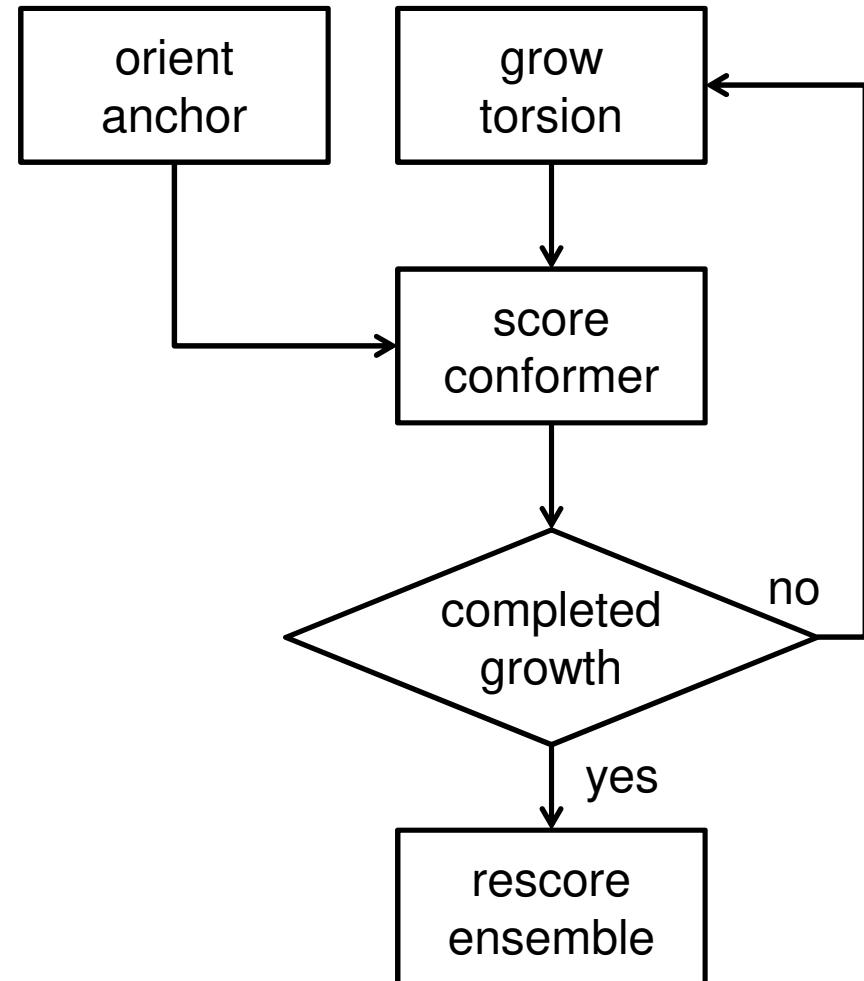
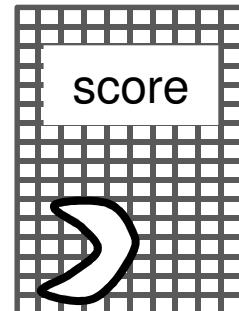
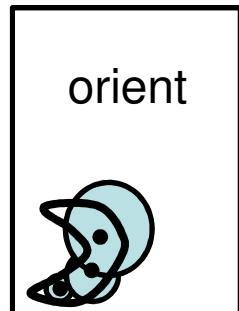
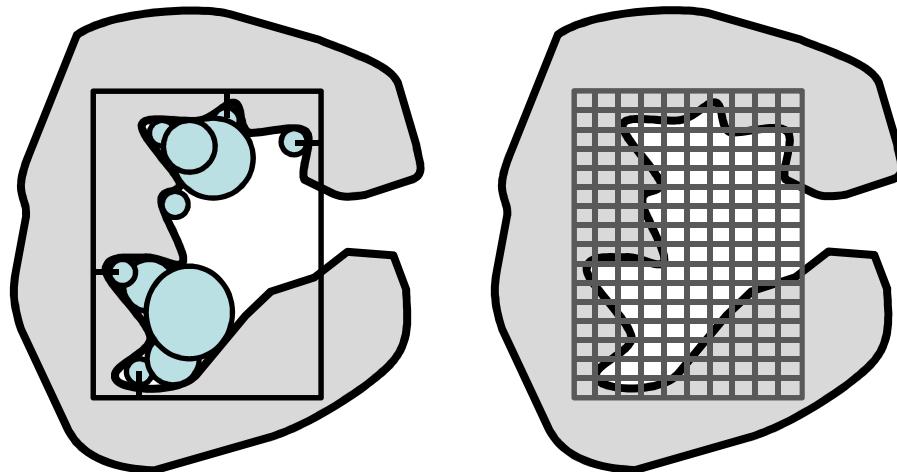
Grow segments in Layer 1

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 Balius, 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

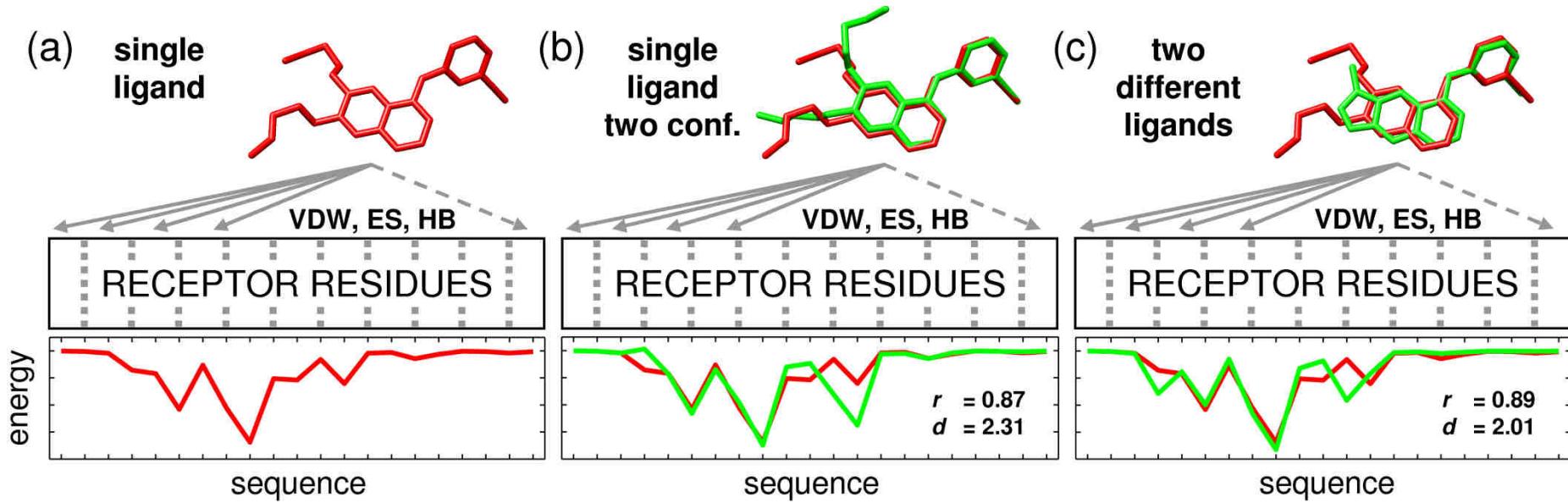


Table I. Examples of possible reference types to derive molecular footprints.

Reference Types

Known inhibitor
Natural substrate
Transition state
Modified structure
Text file footprint
Ensemble-weighted

Description

FDA-approved drug or experimental inhibitor validated to bind
Native peptide or cofactor
Predicted transition state geometry for a chemical reaction
Key functionality/substructure (side-chain mediating protein-protein interactions)
Modified entries to increase/decrease importance of select residues (resistance mutations)
Averaged footprints derived from MD/MC simulations

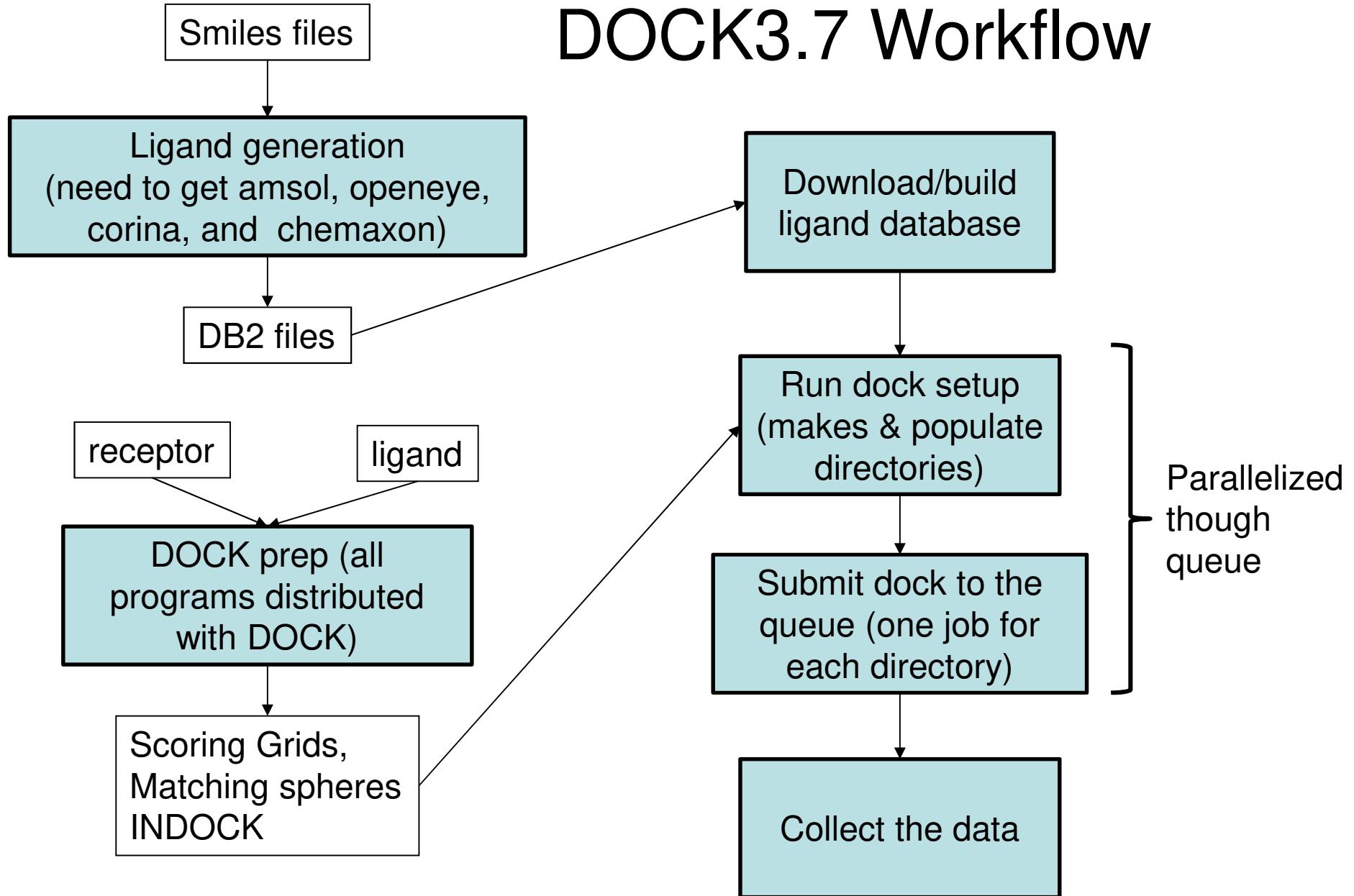
DOCK: Where it is and where its going

SBGrid Lunch Break

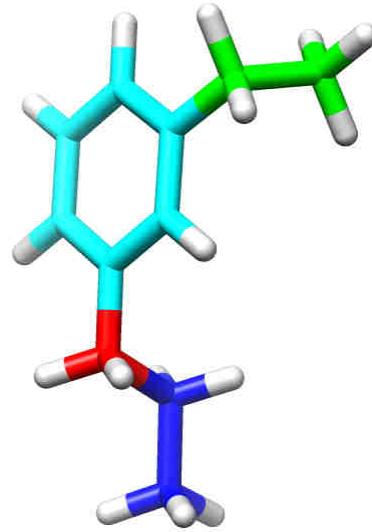
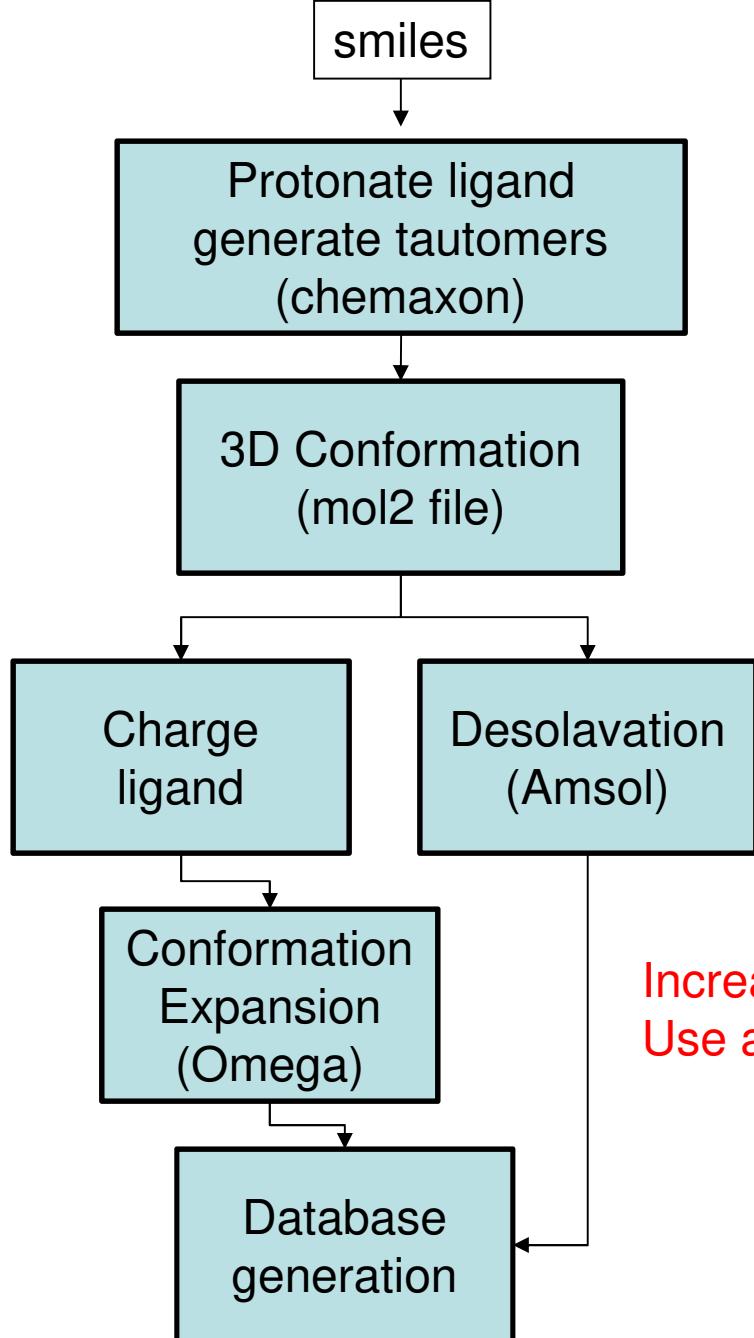
Trent Balius

June 20, 2017

DOCK3.7 Workflow



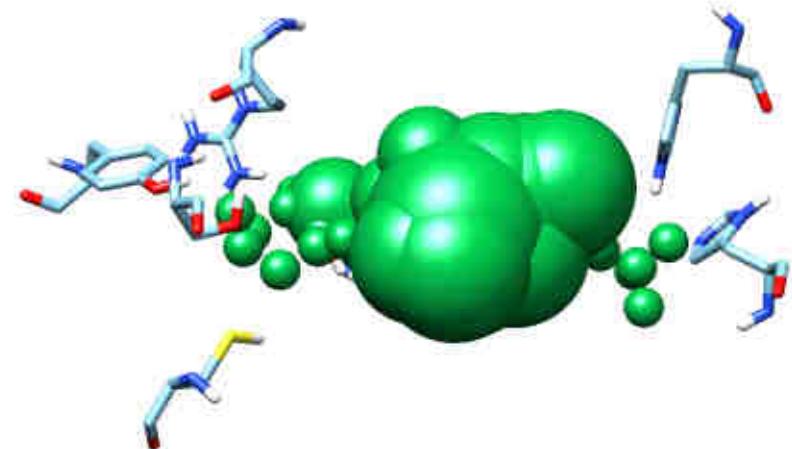
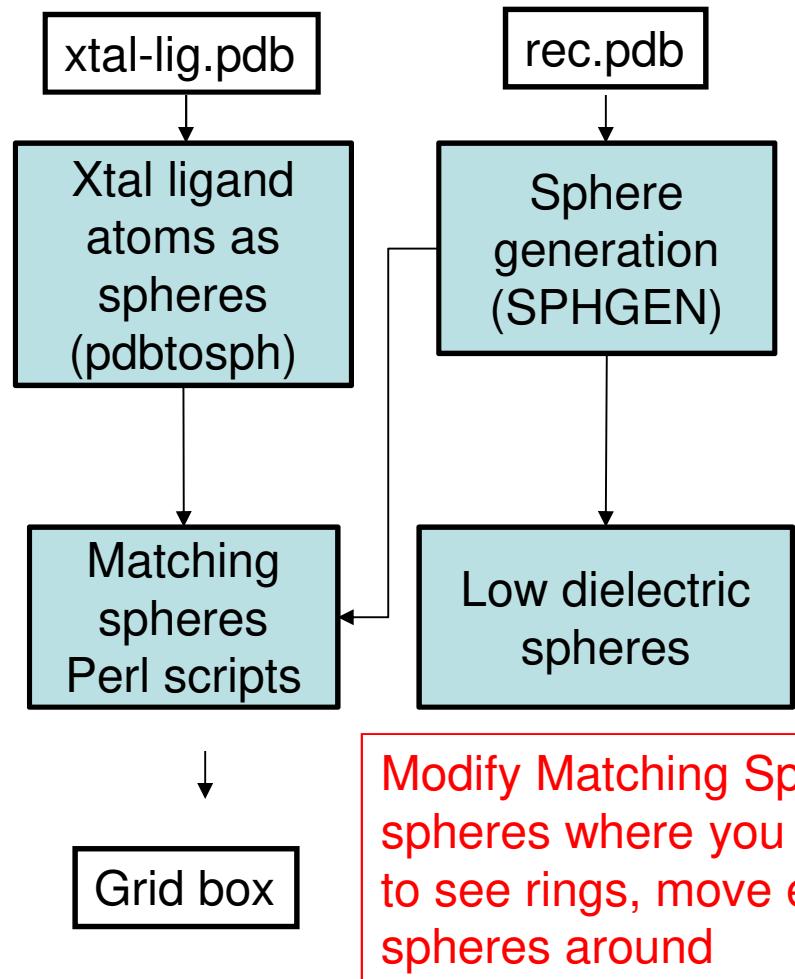
Ligand Preparation



DOCKable Ligand databases are available through **ZINC15**

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

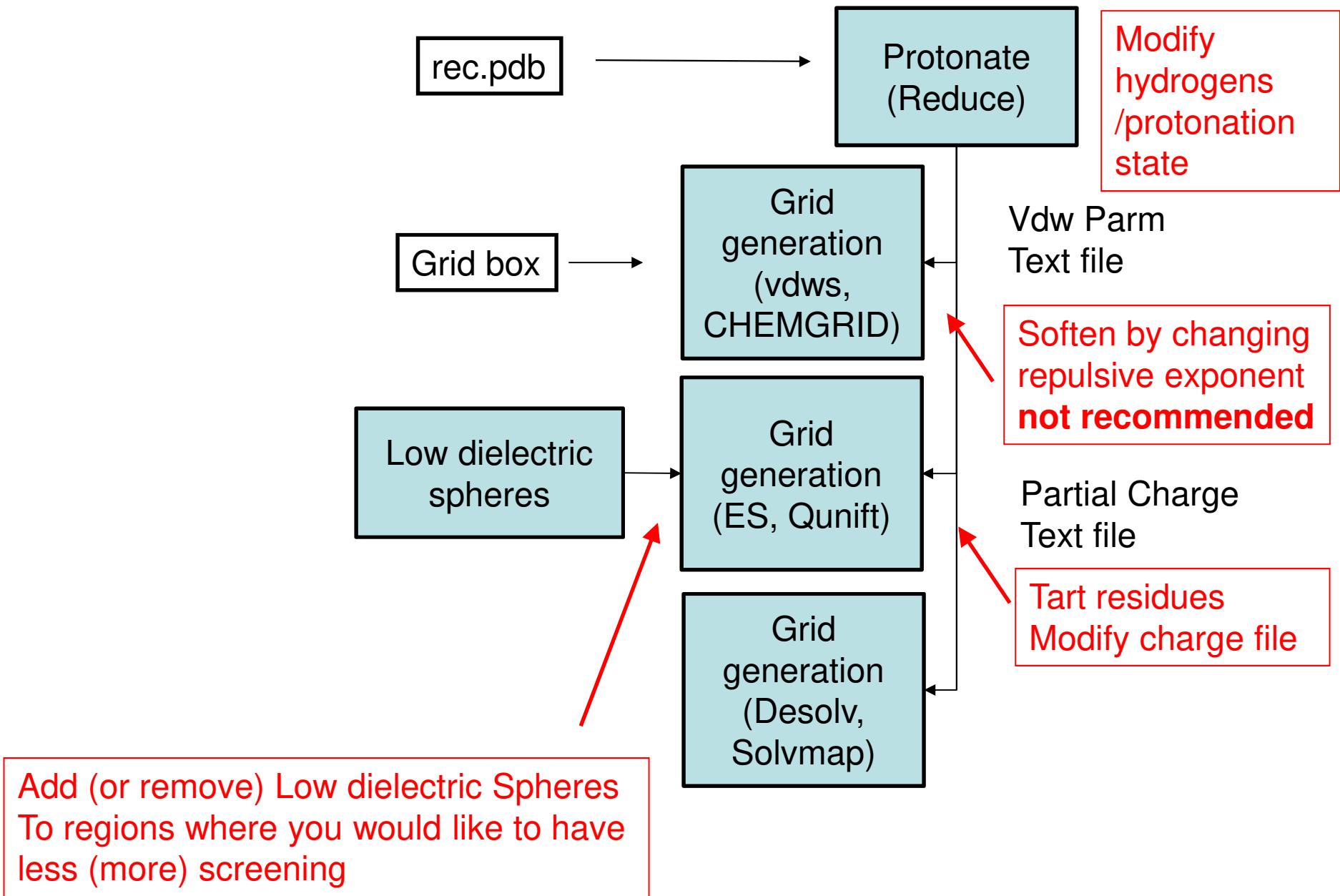
Sphere Generation: Receptor Prep.



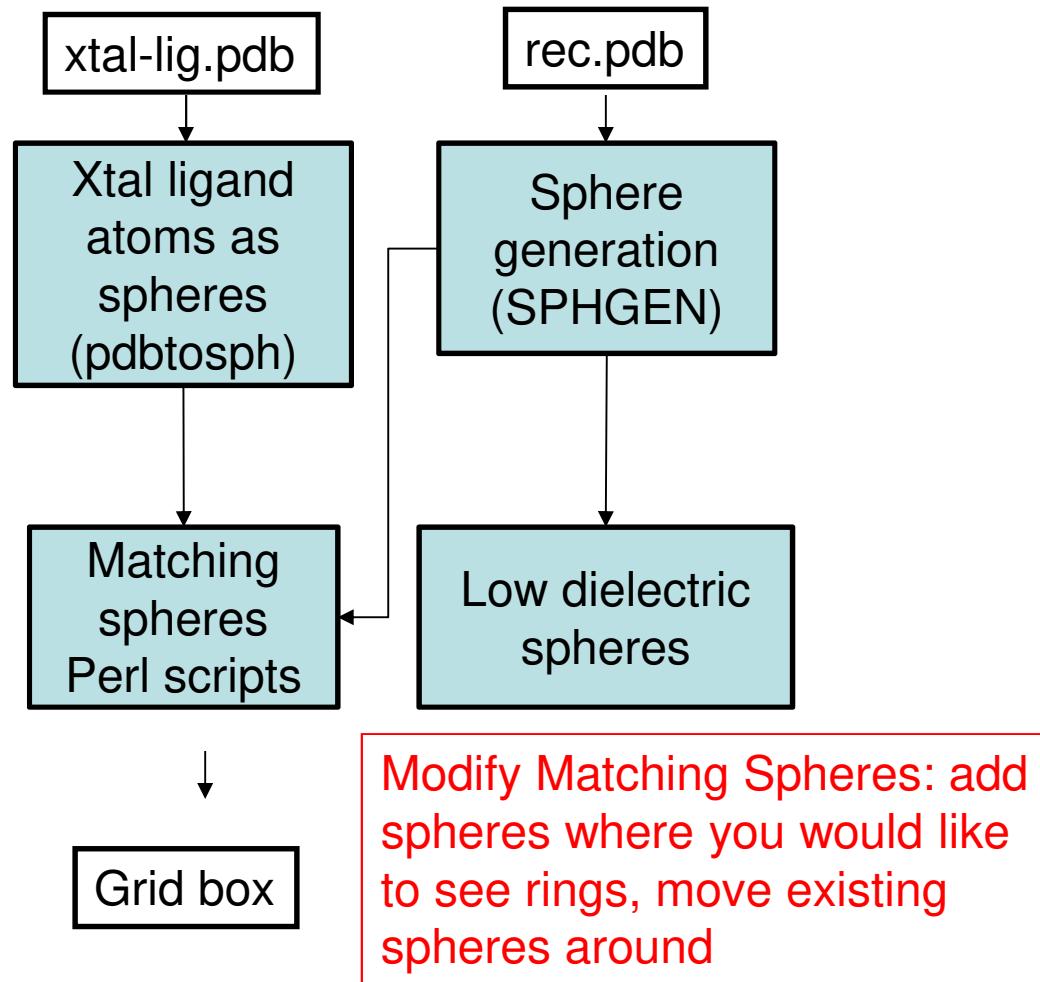
Add (or remove) Low dielectric Spheres
To regions where you would like to have
less (more) screening

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

Grid Generation: Receptor preparation



Sphere Generation

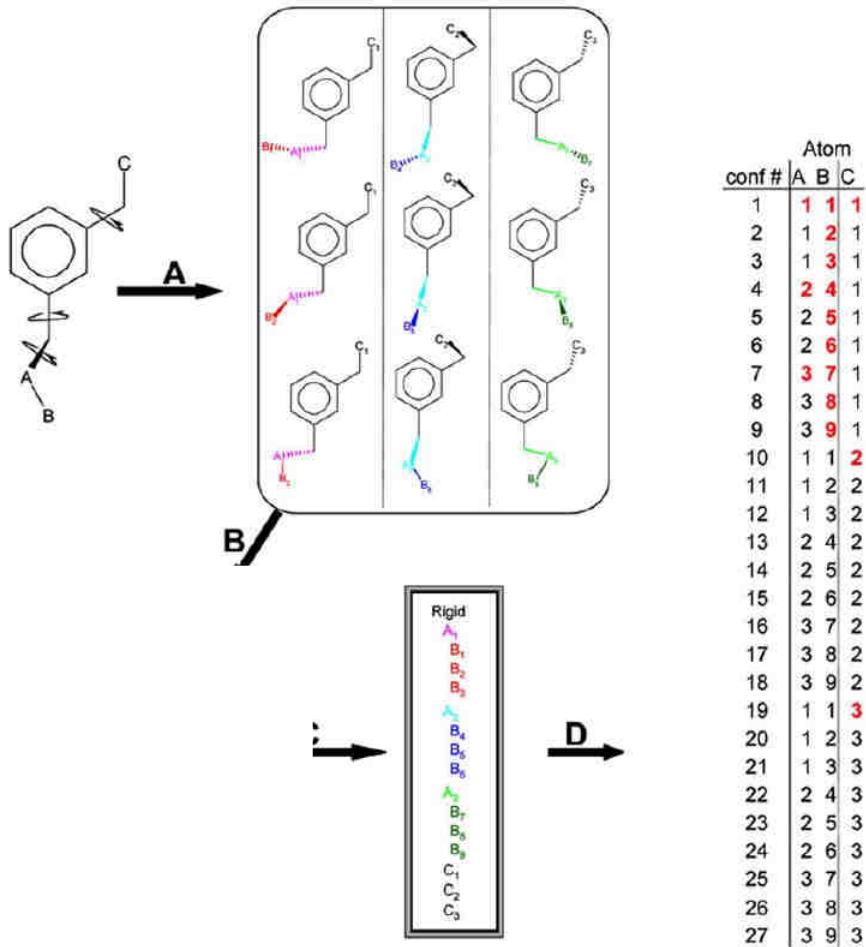


Add (or remove) Low dielectric Spheres
To regions where you would like to have
less (more) screening

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

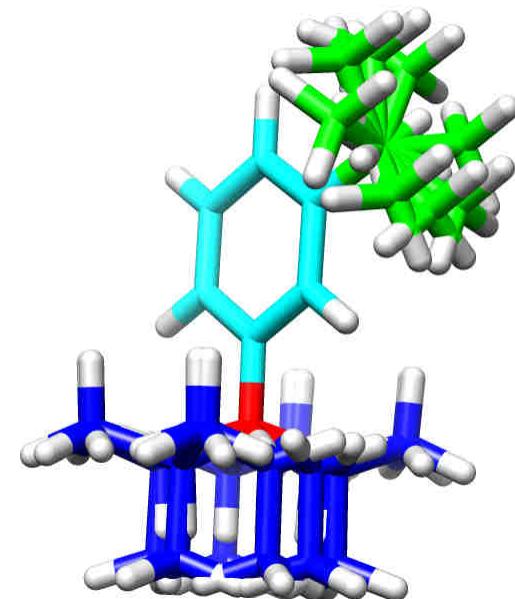
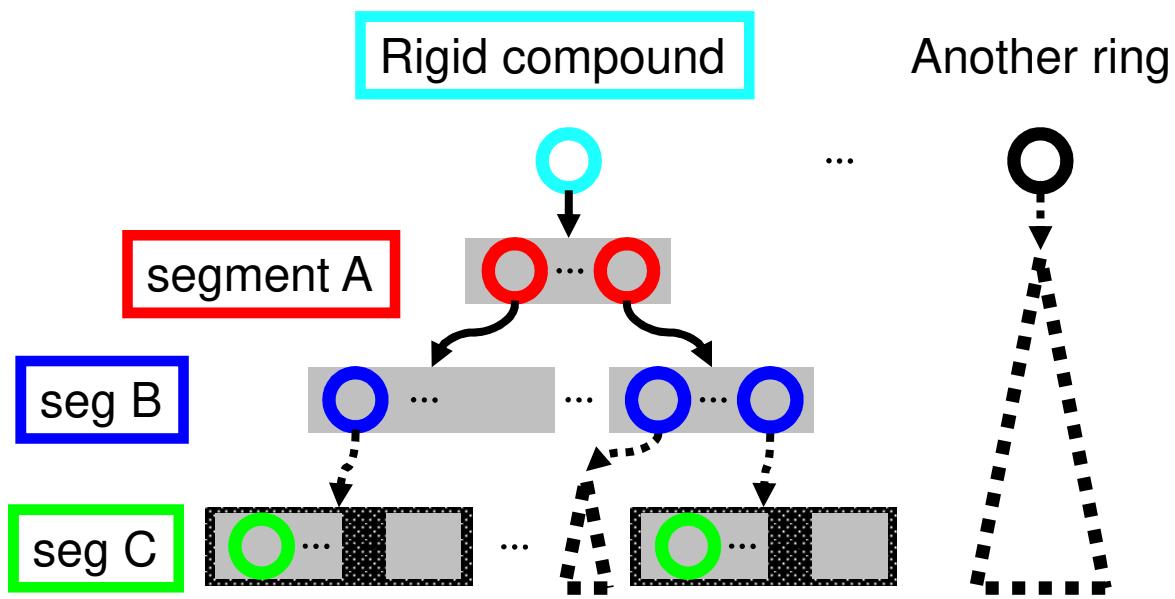
How Sampling Works (Internal)

Hierarchical docking of databases



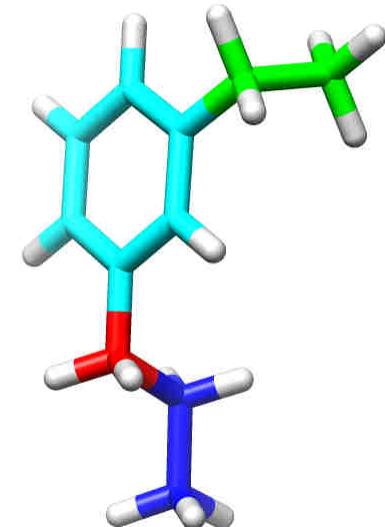
Lorber DM and Shoichet BK., Curr Top Med Chem. 2005;5(8):739-49.

Ligand Sampling – Database Construction



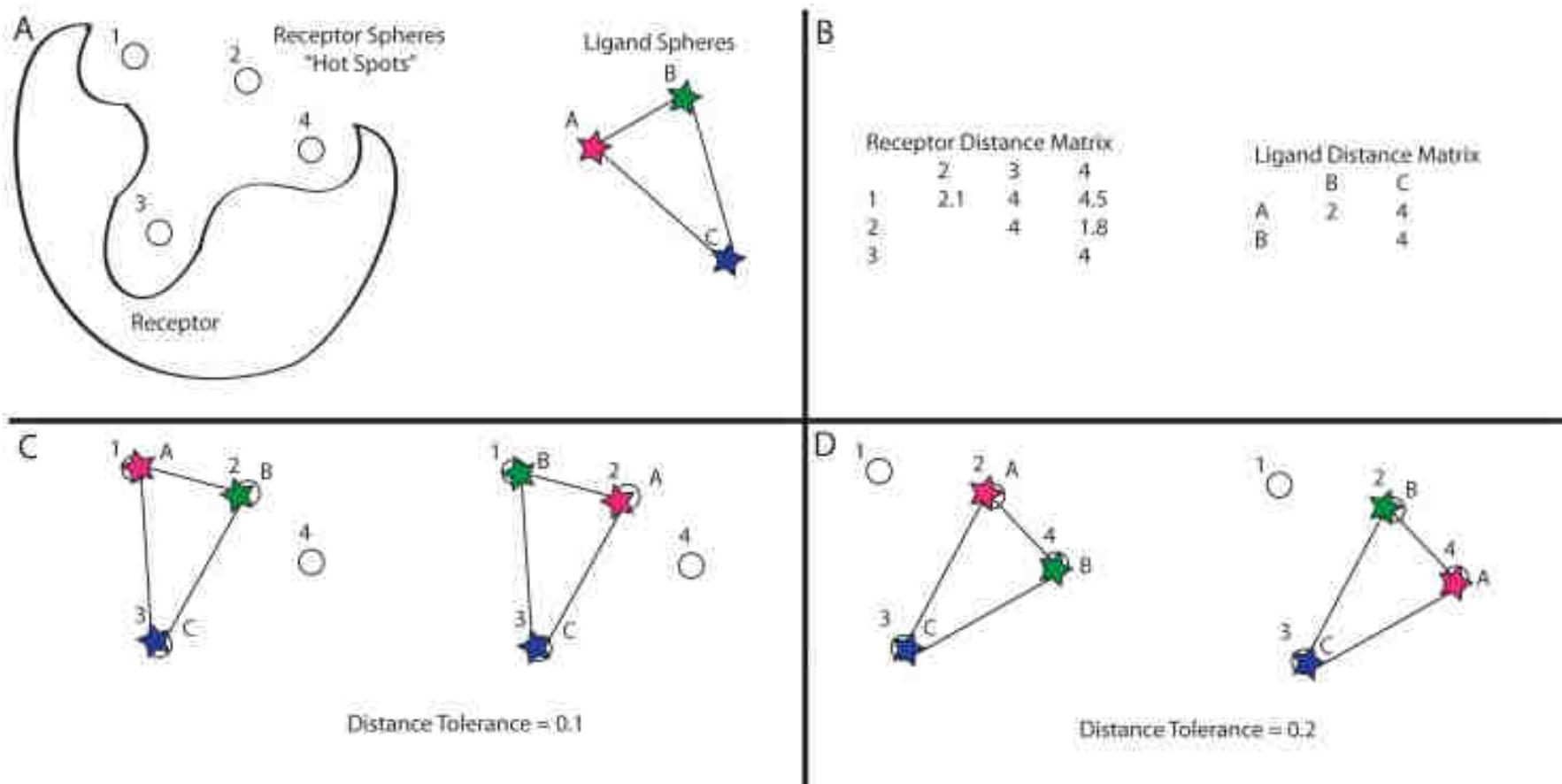
Fully Grown Conformers

Branch
(conformer)

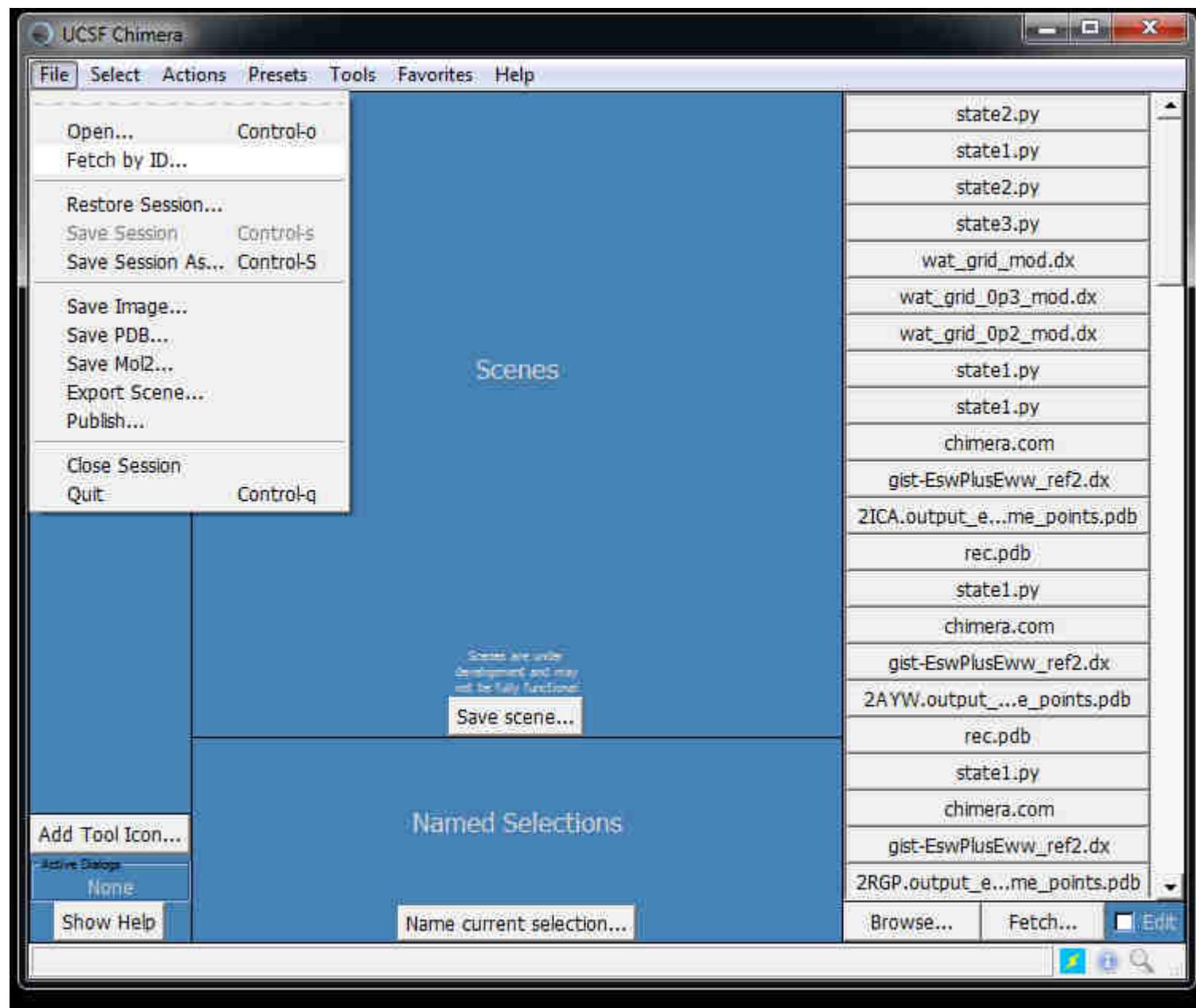


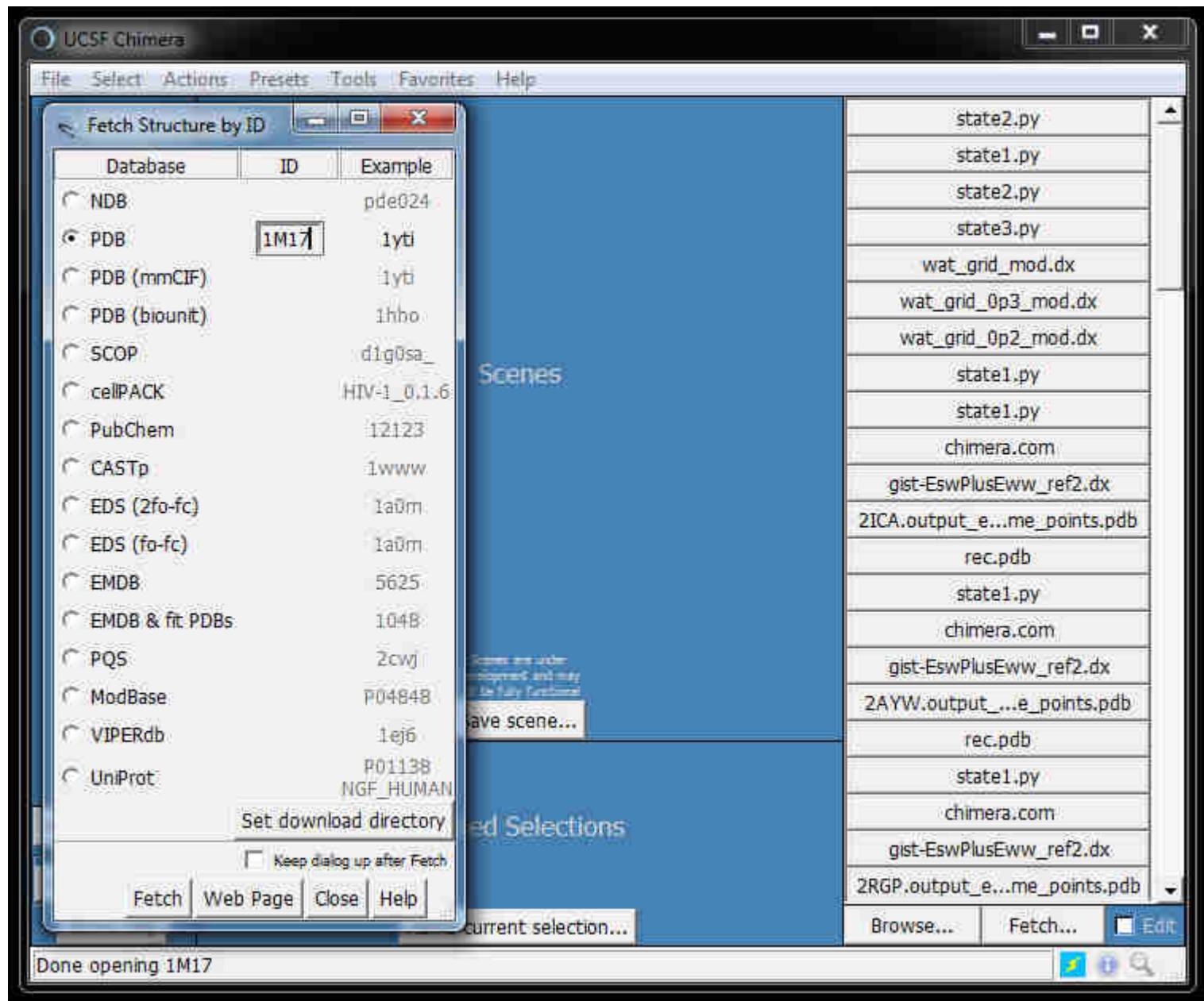
How Sampling Works (Oriental)

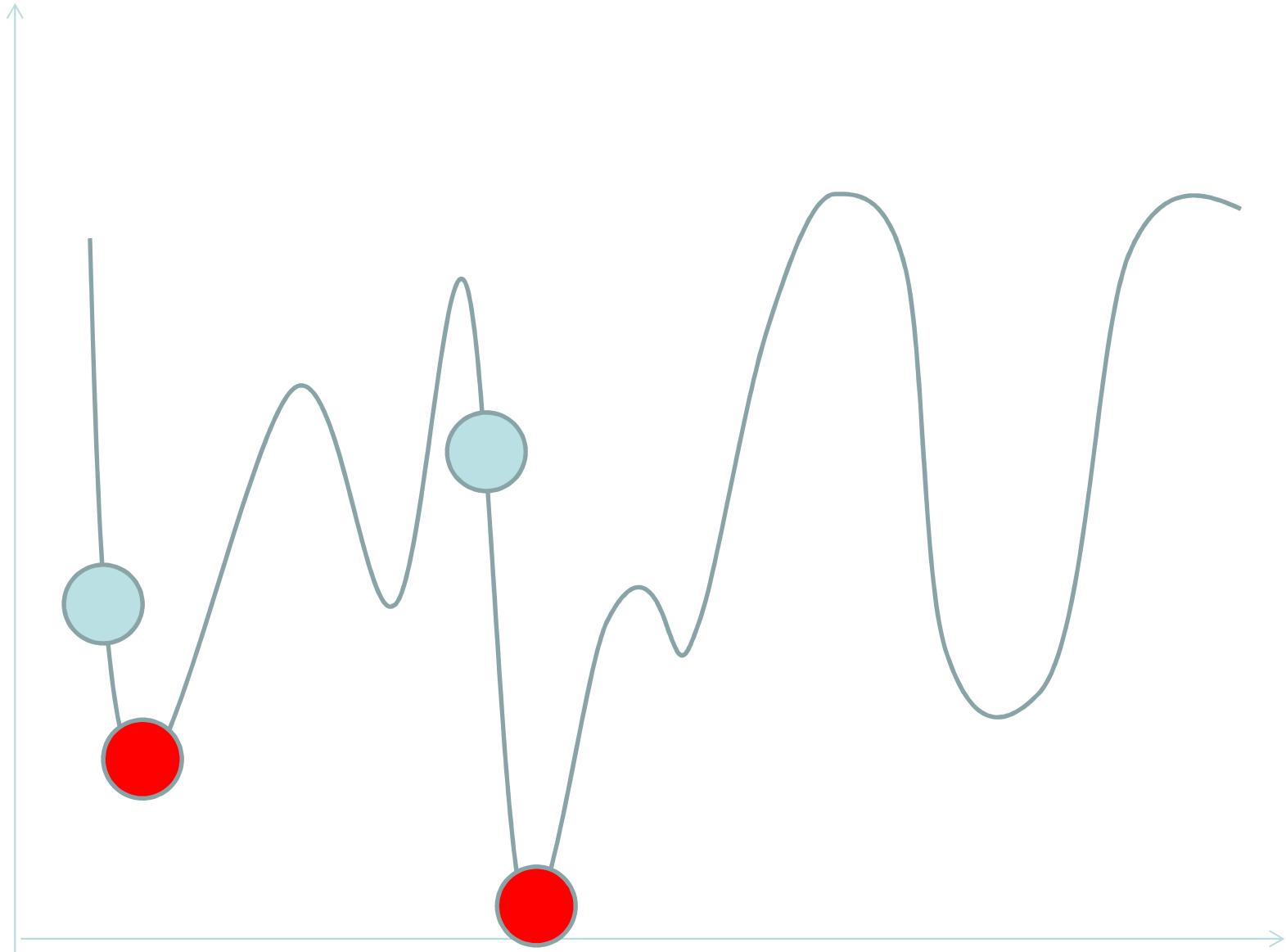
A toy example illustrating the matching sphere orientational matching algorithm



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

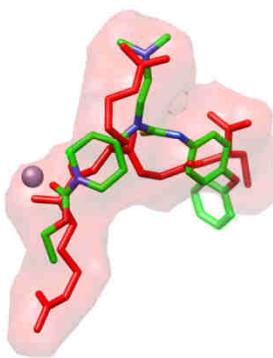




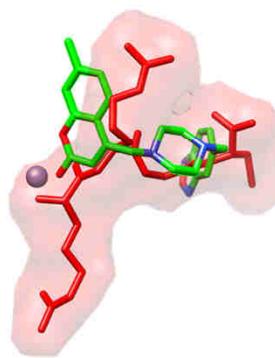


Lab Projects Employing FPS: Bot NT/A

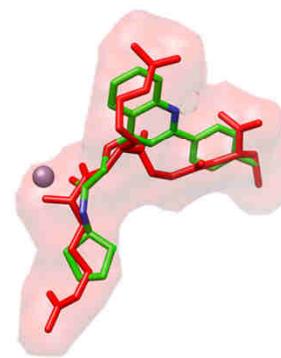
~900,000 screened → 99 purchased → 5 active (picked using FPS)



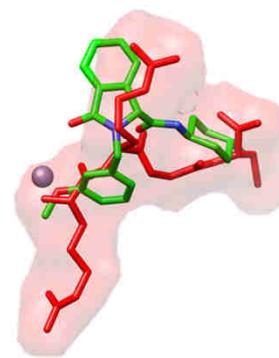
$IC_{50} = 6 \text{ } (\mu\text{M})$
DCE = -73.78
FPS = 0.89



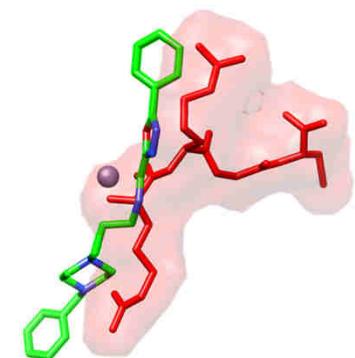
$IC_{50} = 16 \text{ } (\mu\text{M})$
DCE = -60.57
FPS = 0.89



$IC_{50} = 18 \text{ } (\mu\text{M})$
DCE = -60.98
FPS = 0.92



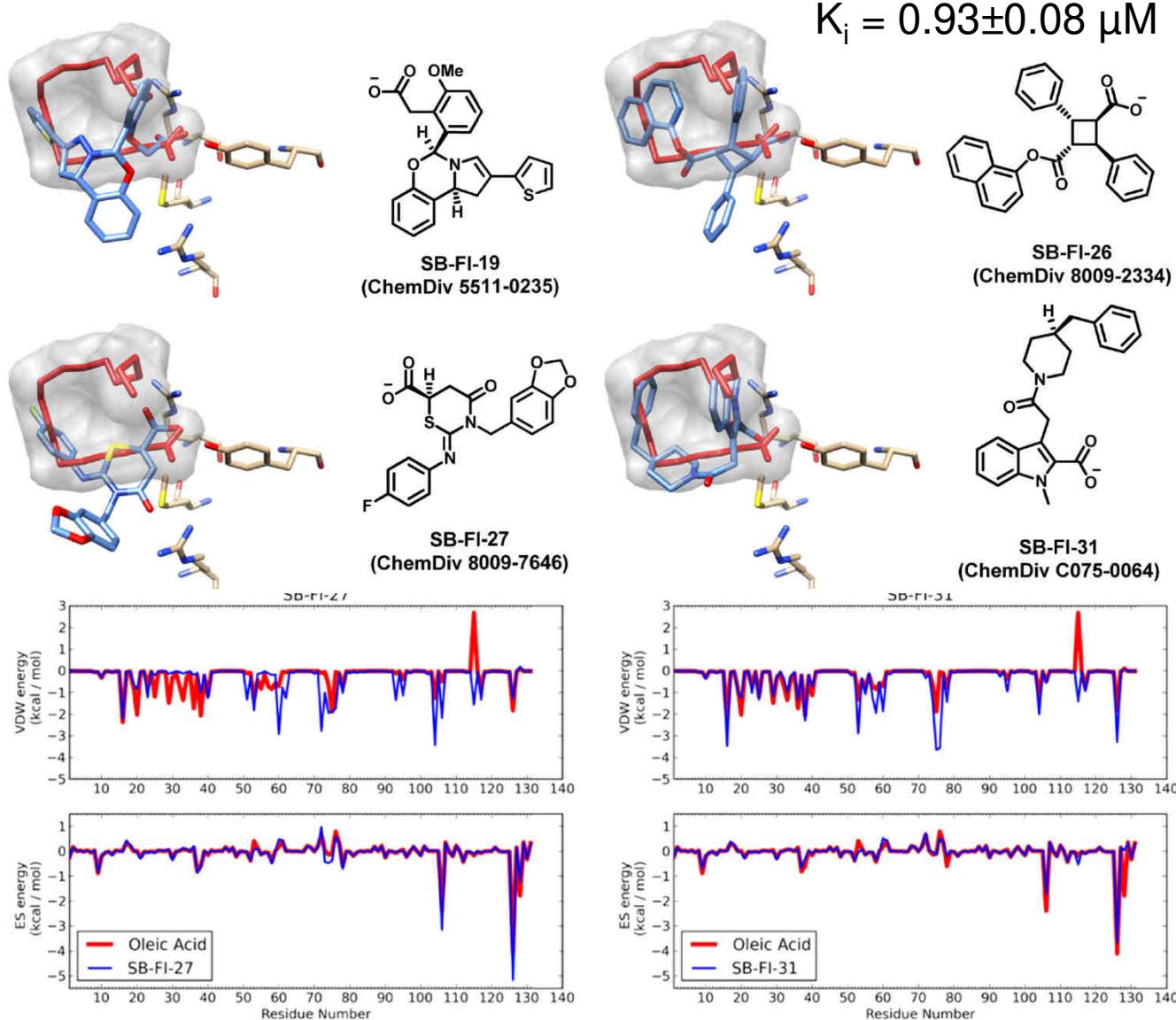
$IC_{50} = 18 \text{ } (\mu\text{M})$
DCE = -62.53
FPS = 0.86



$IC_{50} = 18 \text{ } (\mu\text{M})$
DCE = -67.97
FPS = 1.63

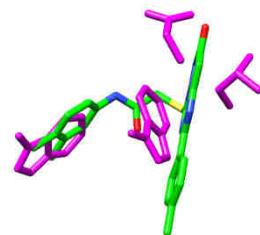
Preliminary results, work in progress: Bill Berger, Trent Balias, Eduard Melief, Peter Tonge, Subramanyam Swaminathan, Iwao Ojima, and Robert Rizzo

Lab Projects Employing FPS: FABP

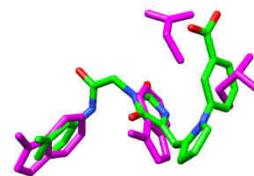


Lab Projects Employing FPS: HIVgp41

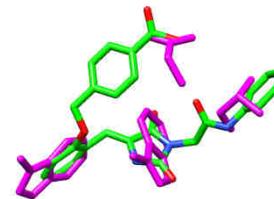
~500,000 screened → 115 purchased → 7 active (3 / 7 picked using FPS)



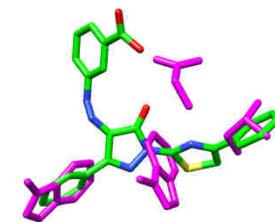
sum_fp
 $K_i = 0.46 \text{ } (\mu\text{M})$



es_fp
 $K_i = 4.45 \text{ } (\mu\text{M})$



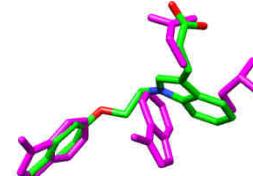
tots
 $K_i = 7.1 \text{ } (\mu\text{M})$



tots
 $K_i = 0.20 \text{ } (\mu\text{M})$



tots
 $K_i = 1.6 \text{ } (\mu\text{M})$



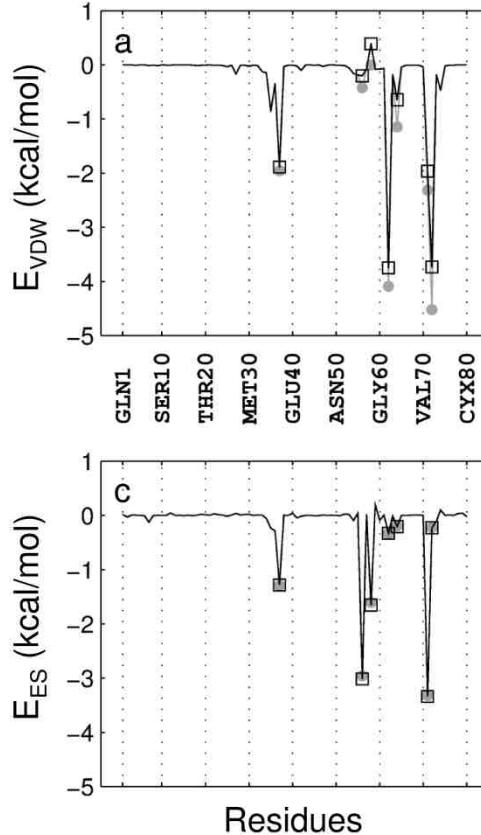
es_fp
 $K_i = 13 \text{ } (\mu\text{M})$



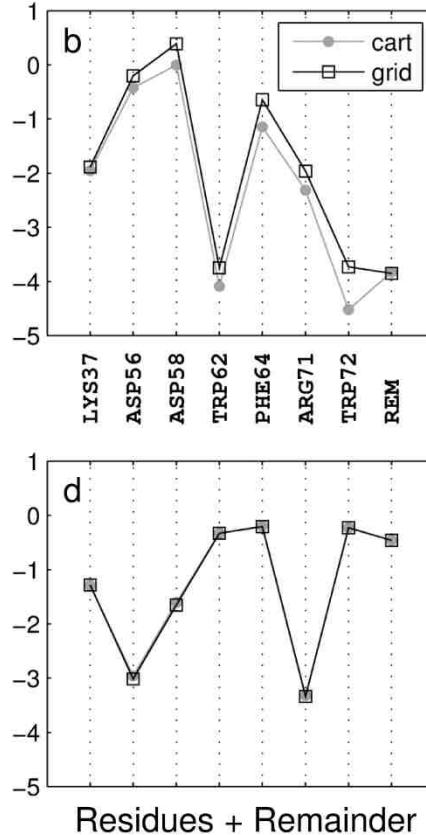
tots
 $K_i = 2.55 \text{ } (\mu\text{M})$

Grid-Based Footprints

All Residues



Threshold-based
Subset



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

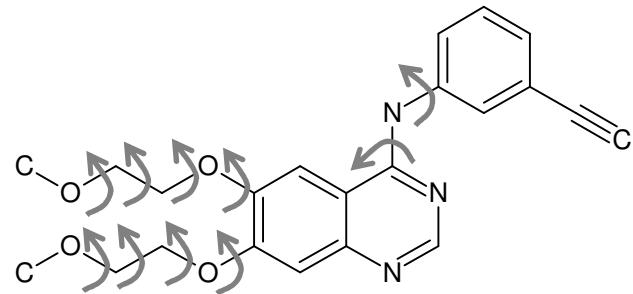
All 80 shown on the left, Subset shown on right

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

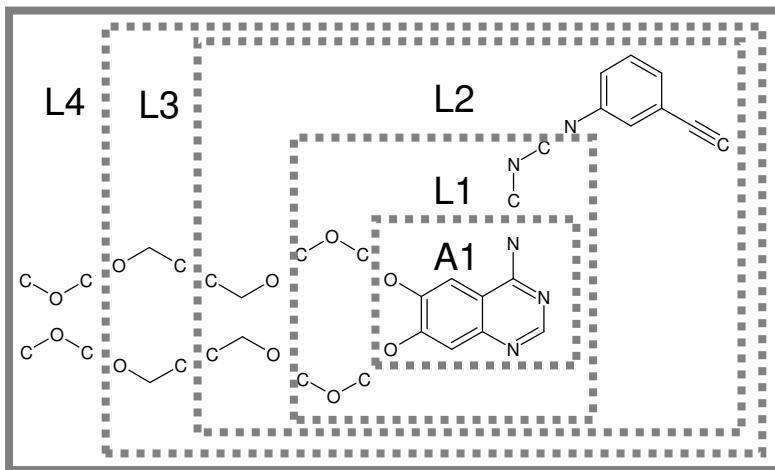
Arrange segments by layer about anchor center

Grow each segment one at a time

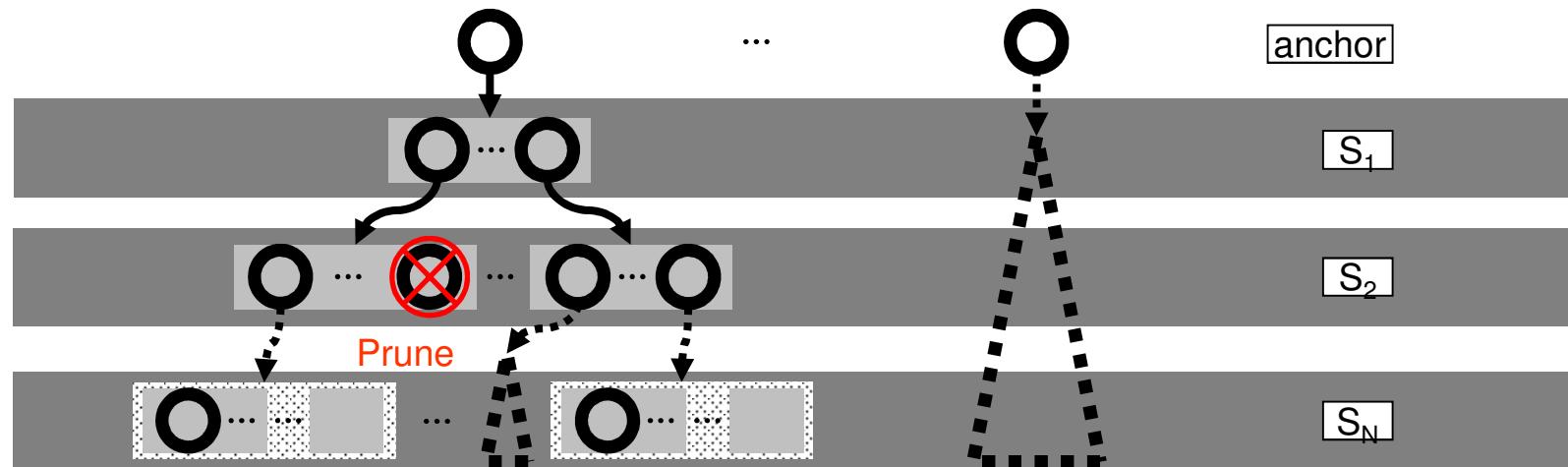
Grow segments in Layer 1

Grow segments in Layer 2 and so on

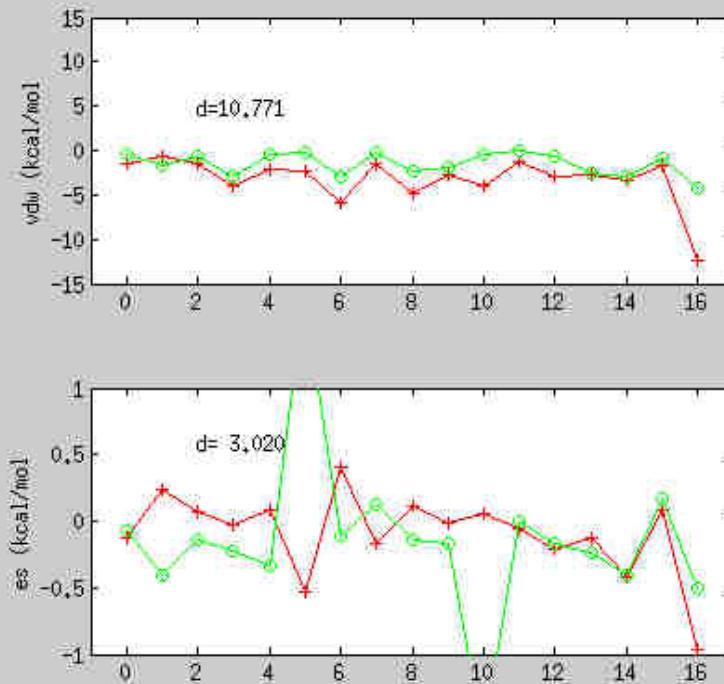
b



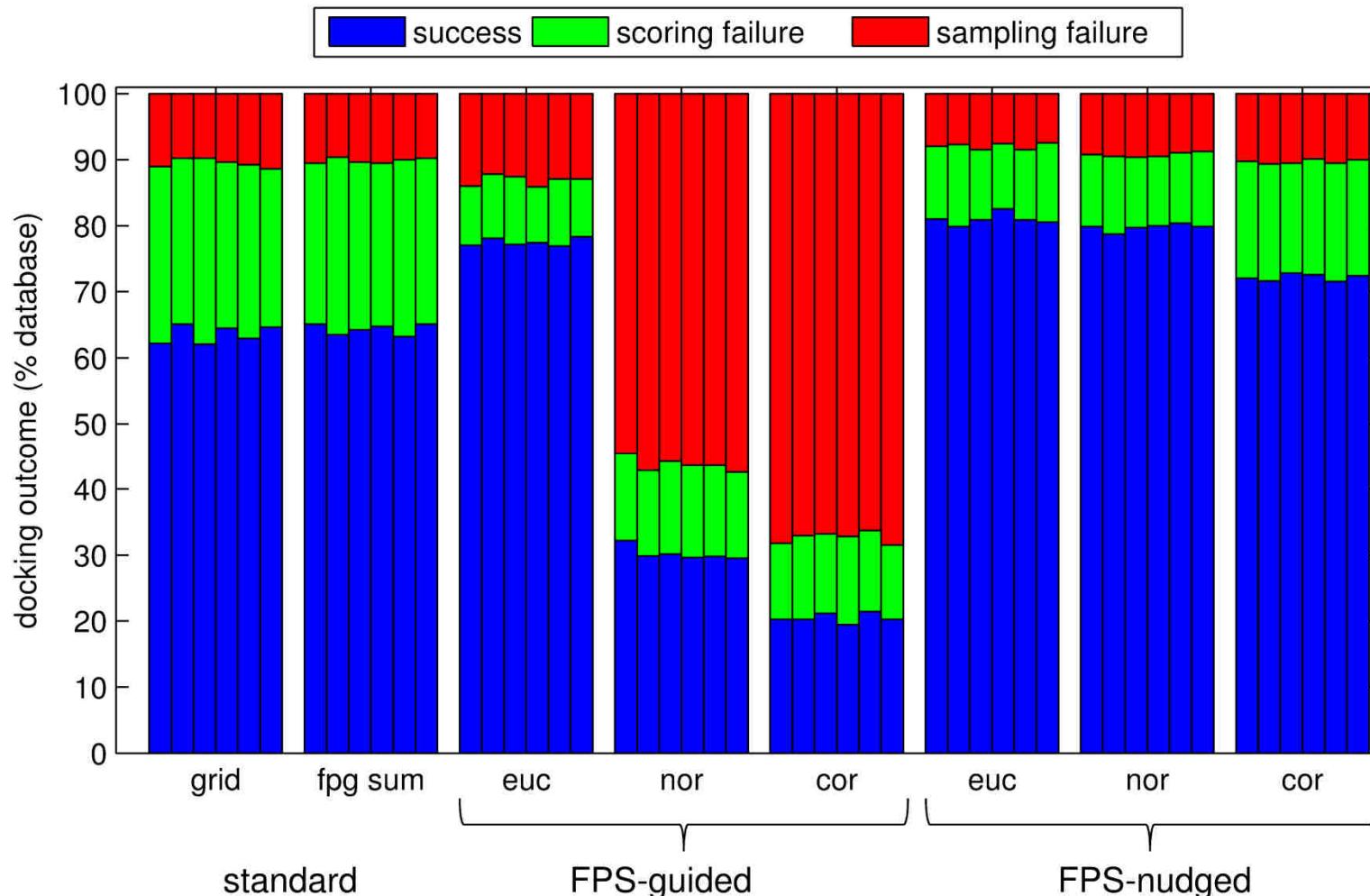
c



Footprint-Guided Growth



Grid-Based Footprints

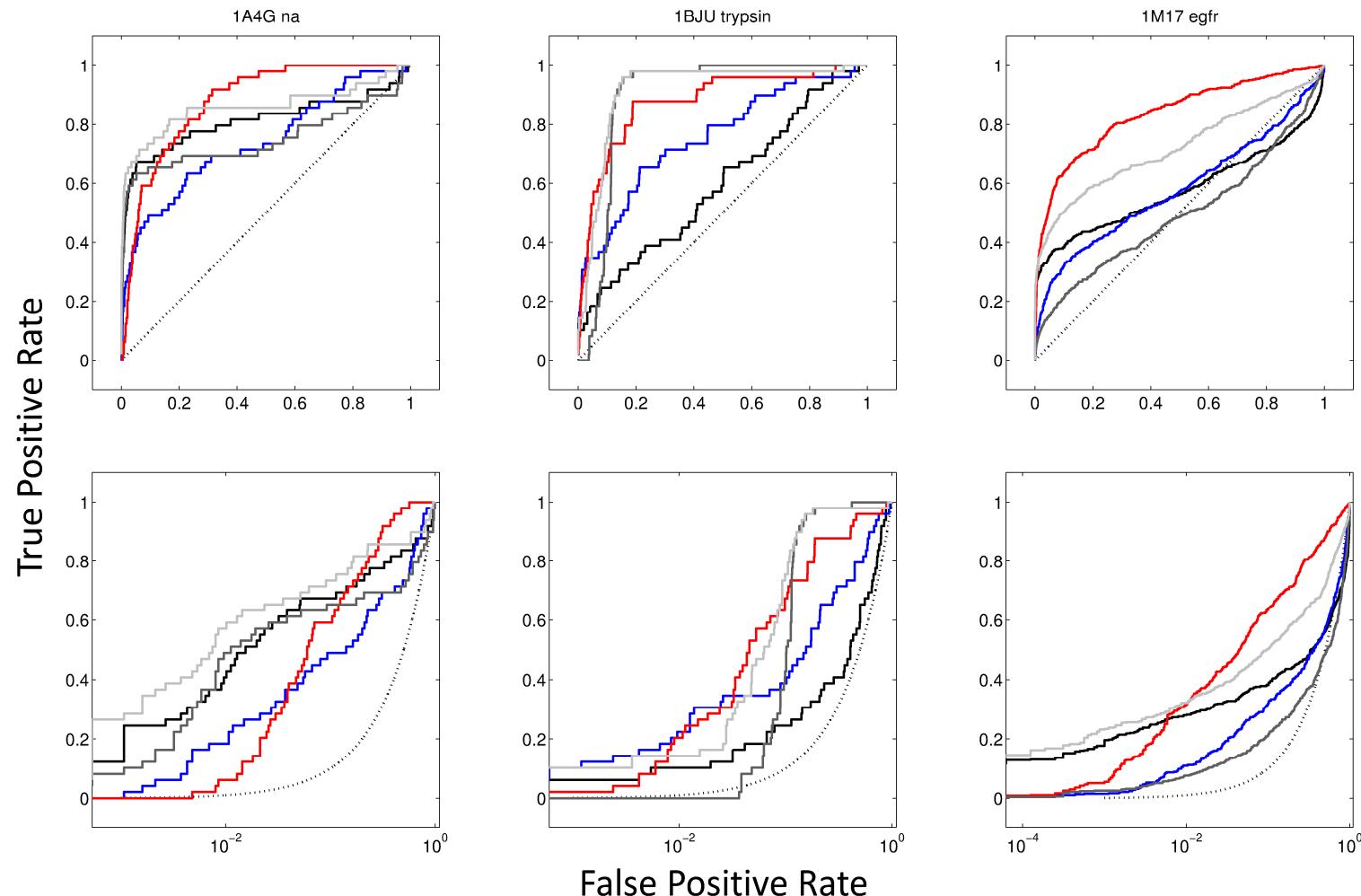


Grid is $O(N)$, Multi-grid is $O(N*M)$, Cartesian $O(N*K)$ $M \ll K$
Multi-grid is about 5 times slower

Effect of scoring function on sampling

d = driven; s = stared; 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      Delphi_nsize is electrostatic grid size
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
```

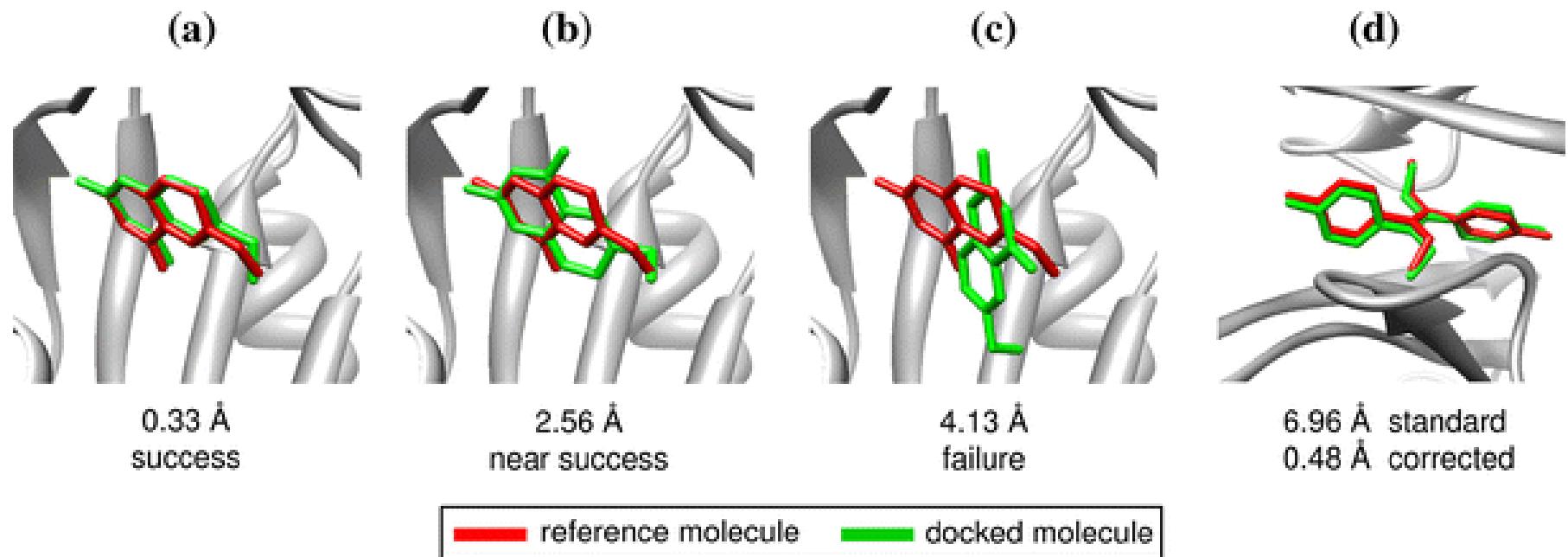
Things are structured differently in Github

```
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
```

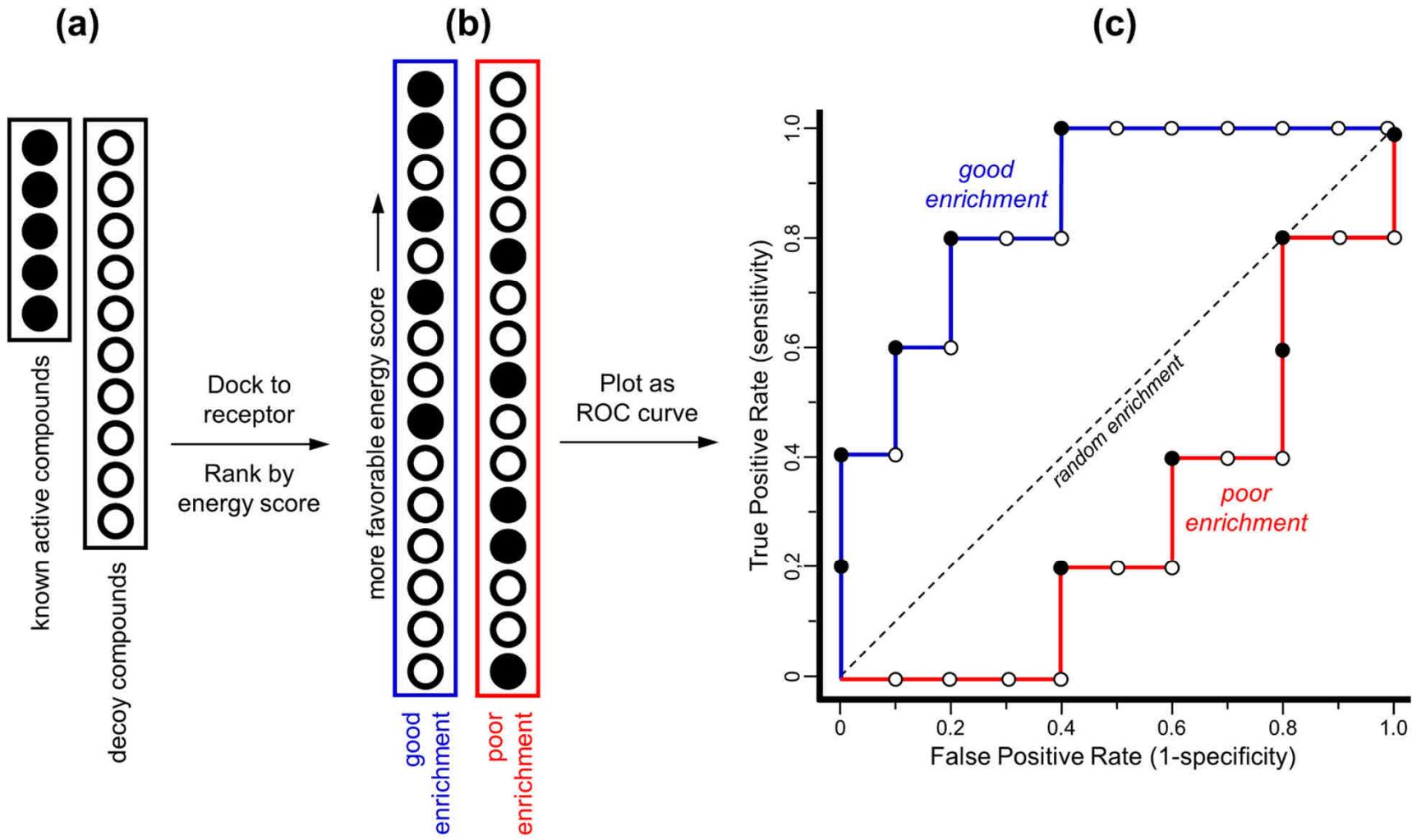
work done by Teague and Ryan.

Evaluation Methods

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

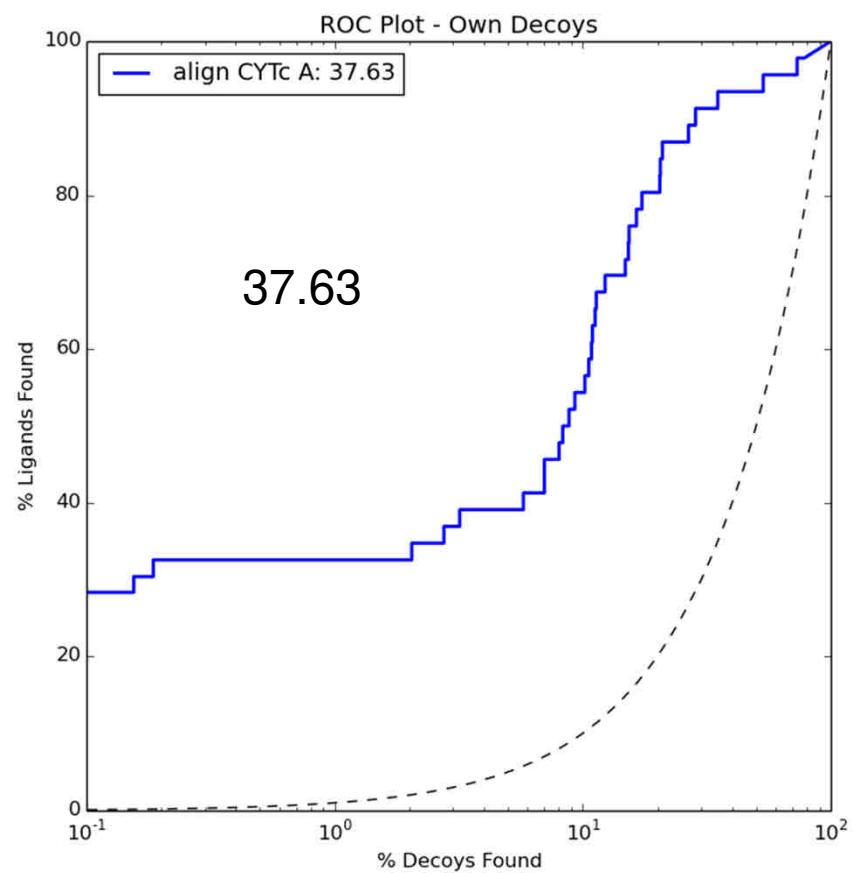
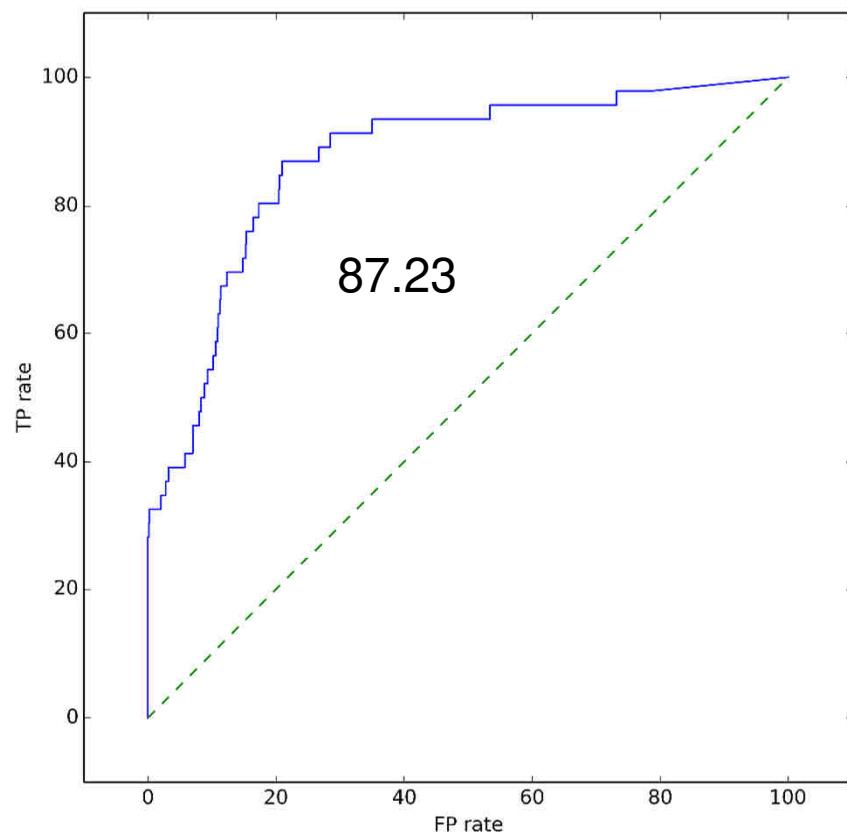


Enrichments: ROC Curves



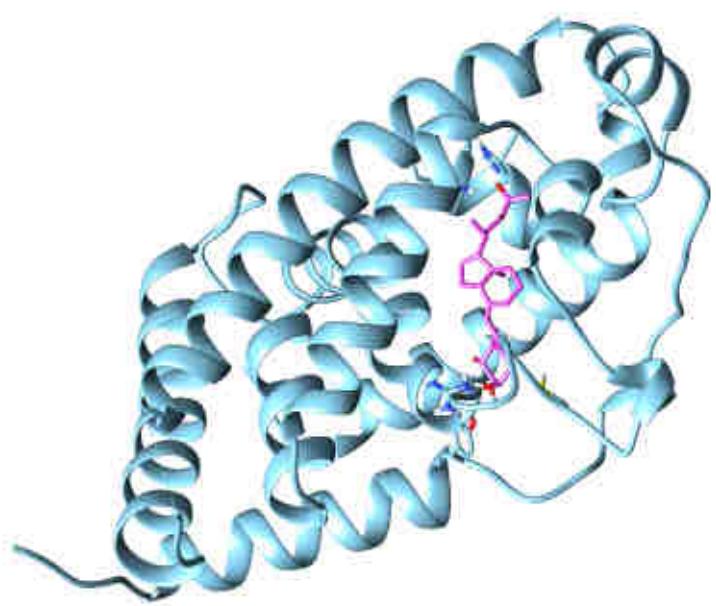
Log-adjusted AUC: Early Enrichment Weighted More

CcP conformation A
ROC curve Log adjusted ROC curve

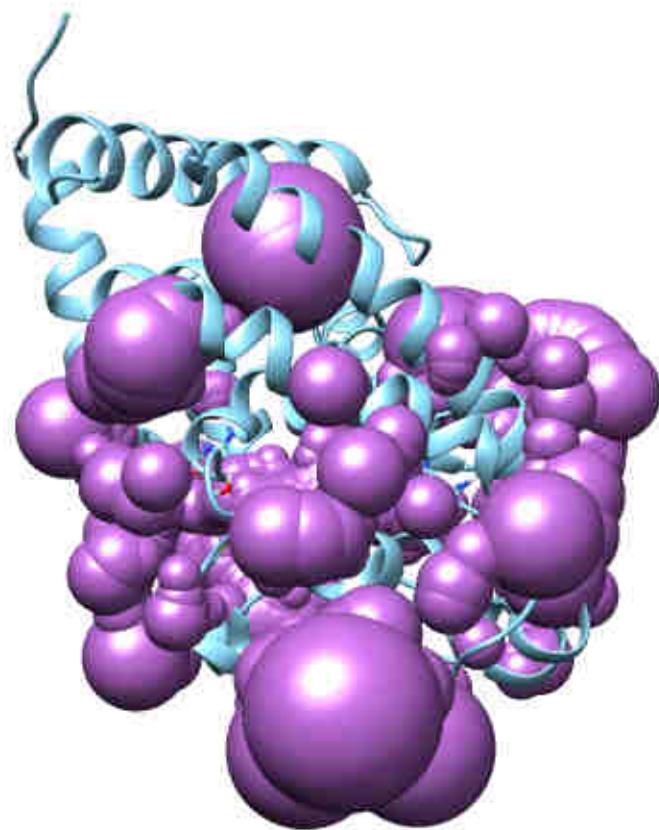


Tweaking your set up – Spheres

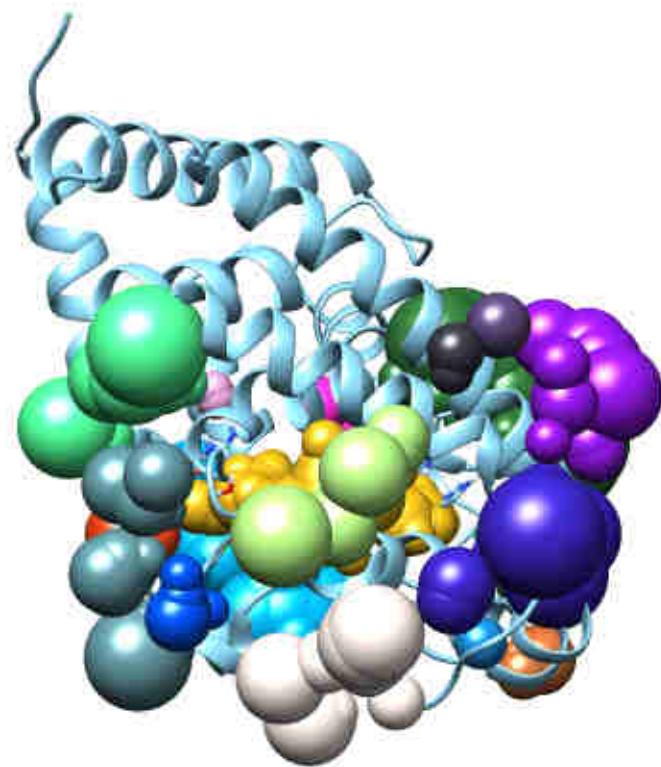
- Place spheres where you want your rings to go
- Use previous screens for ideas



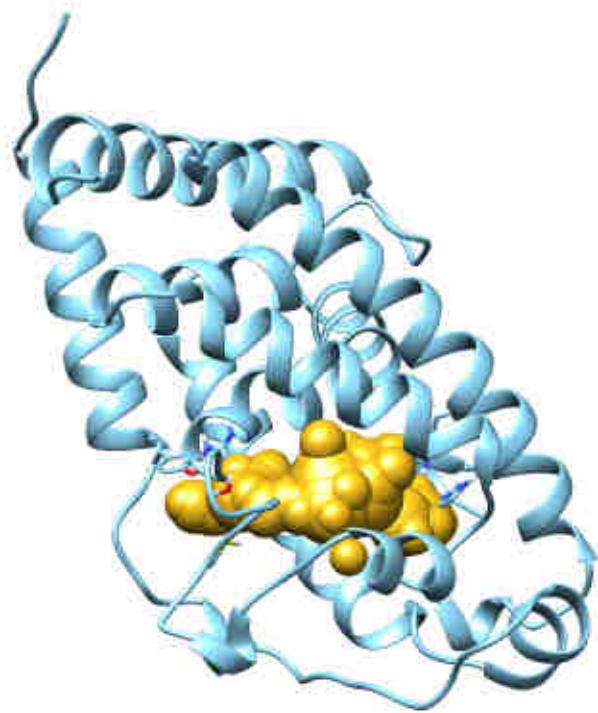
SPHGEN Spheres



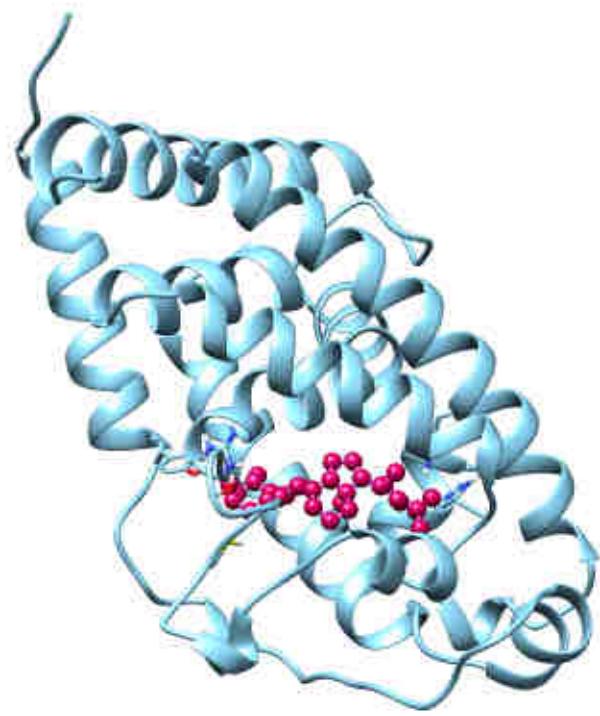
SPHGEN Sphere Clusters



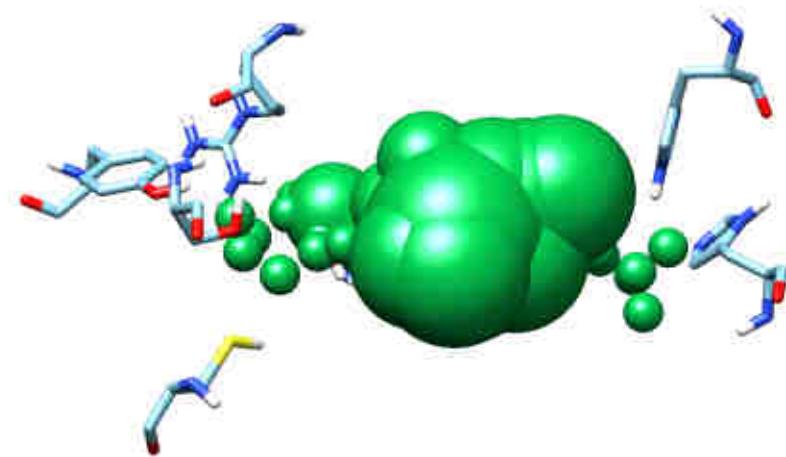
SPHGEN Cluster1 Spheres



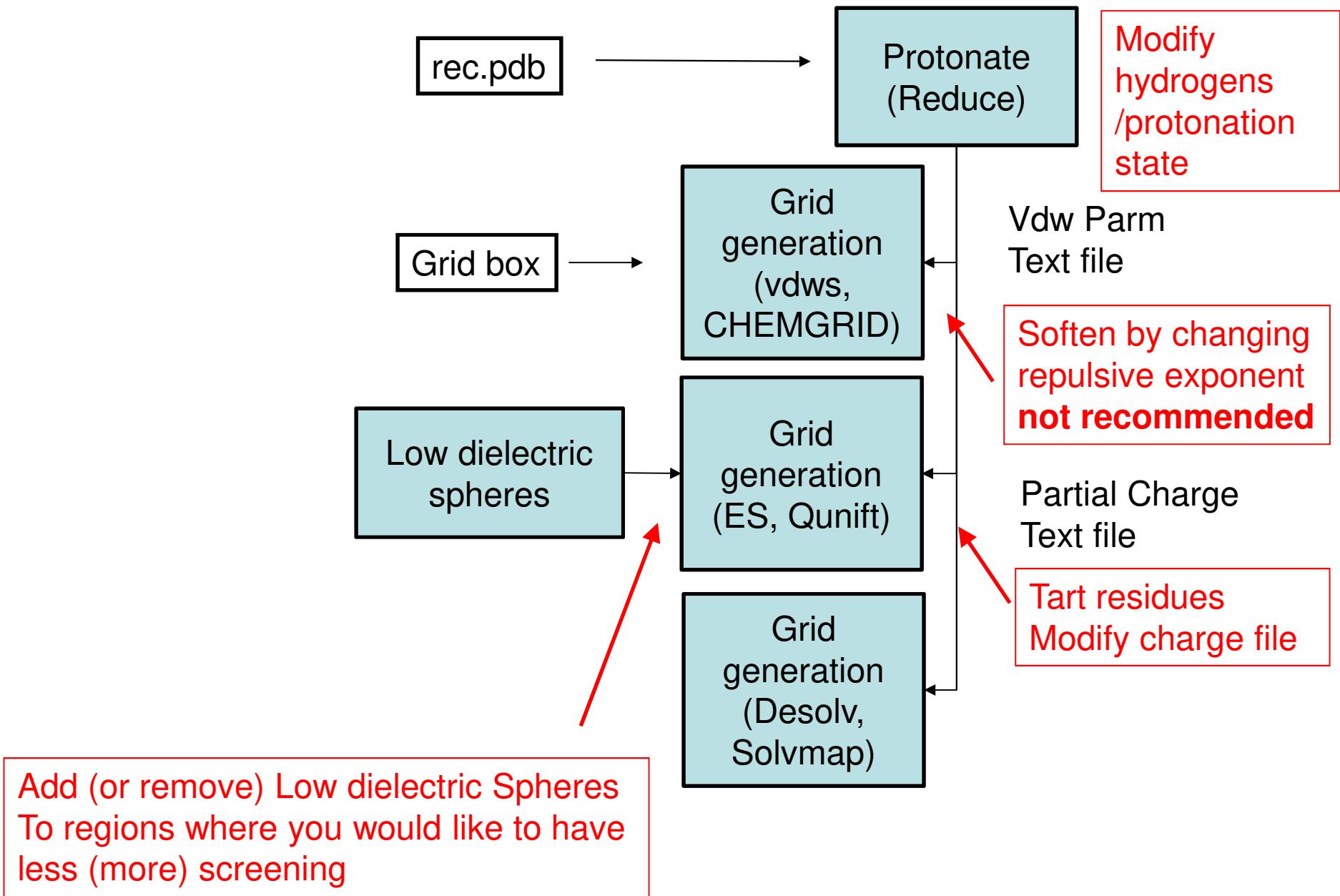
Ligand Atoms Spheres



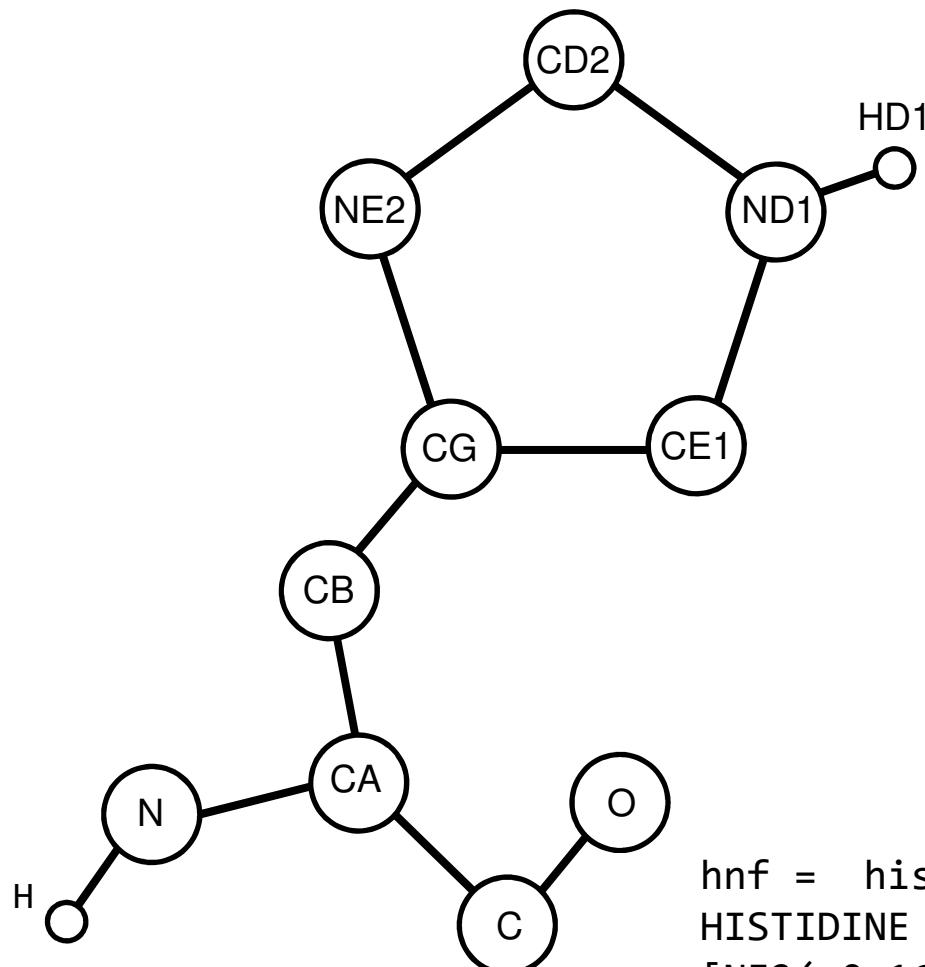
Matching Spheres



Receptor preparation



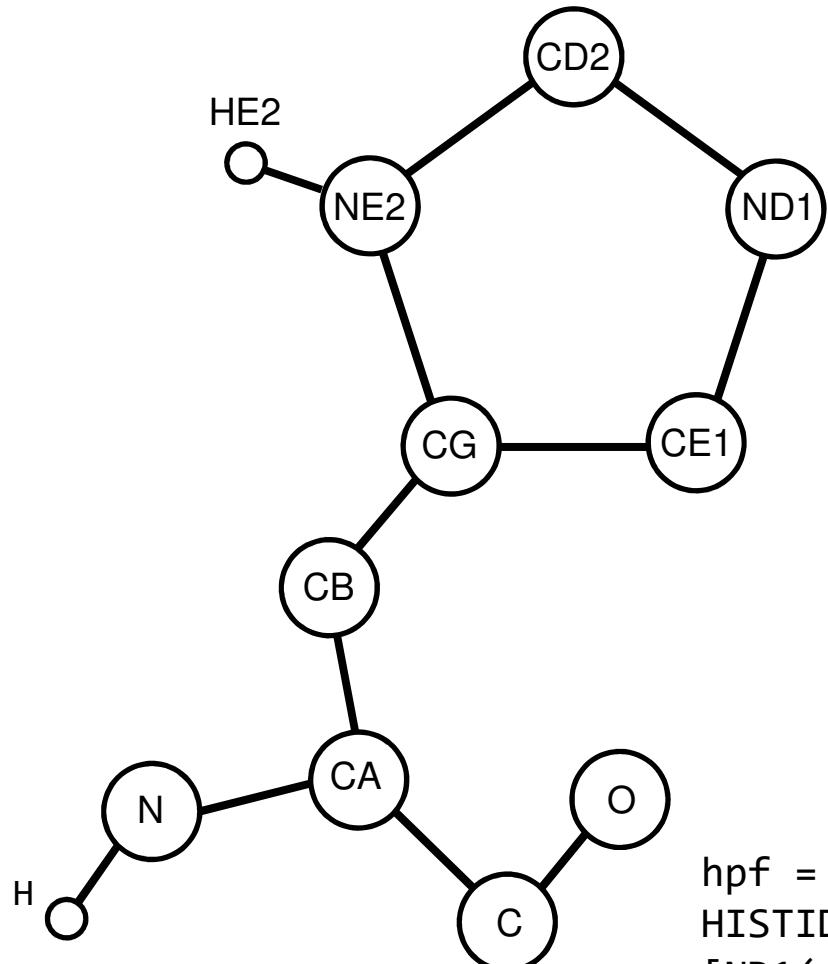
HISTIDINE neutral delta proton [tart2]



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
NE2	-0.527	NE2	-0.687
H	0.248	H	0.248
HD1	0.320	HD1	0.400

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

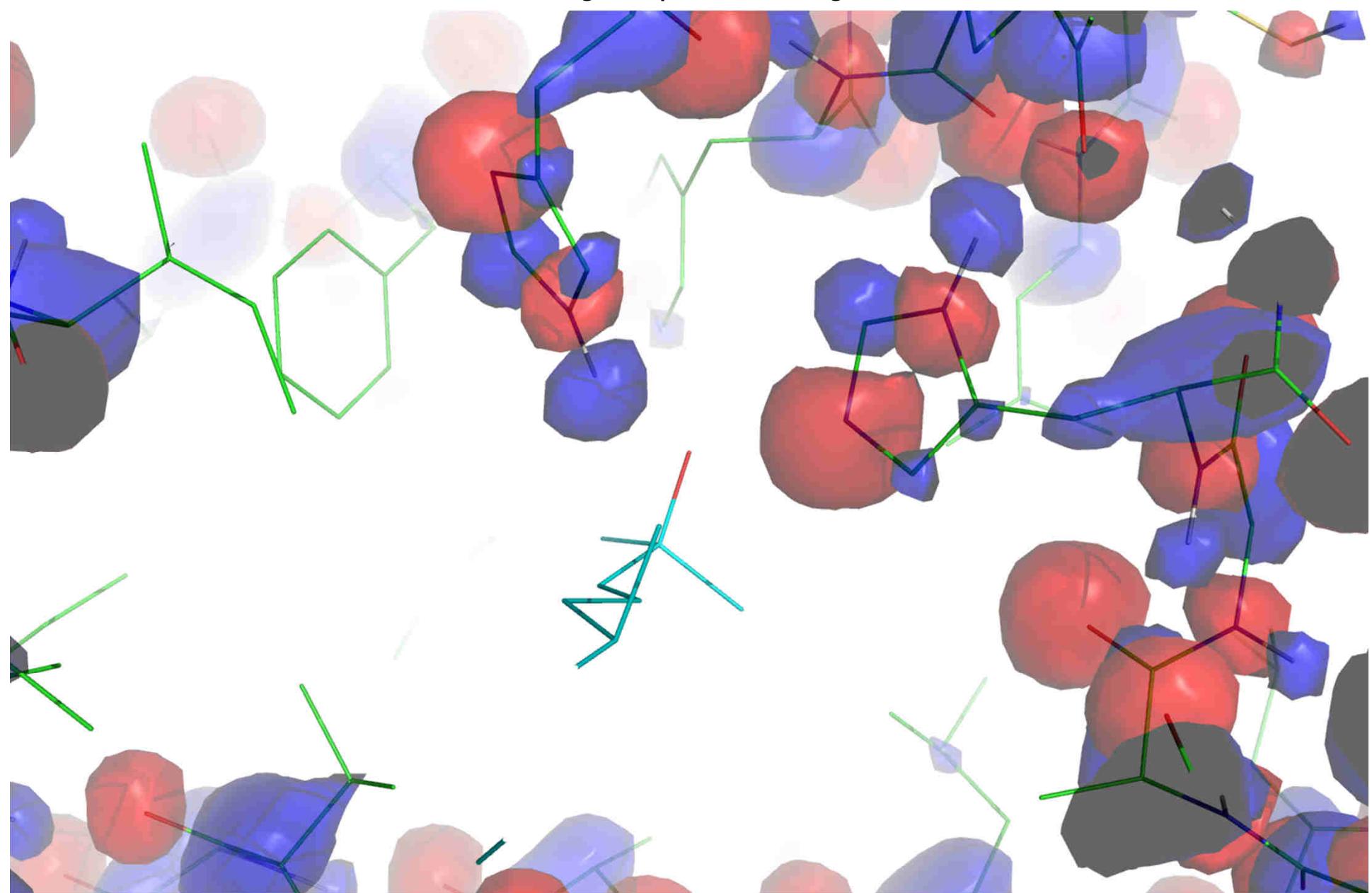
HISTIDINE neutral epsilon proton [tart2]



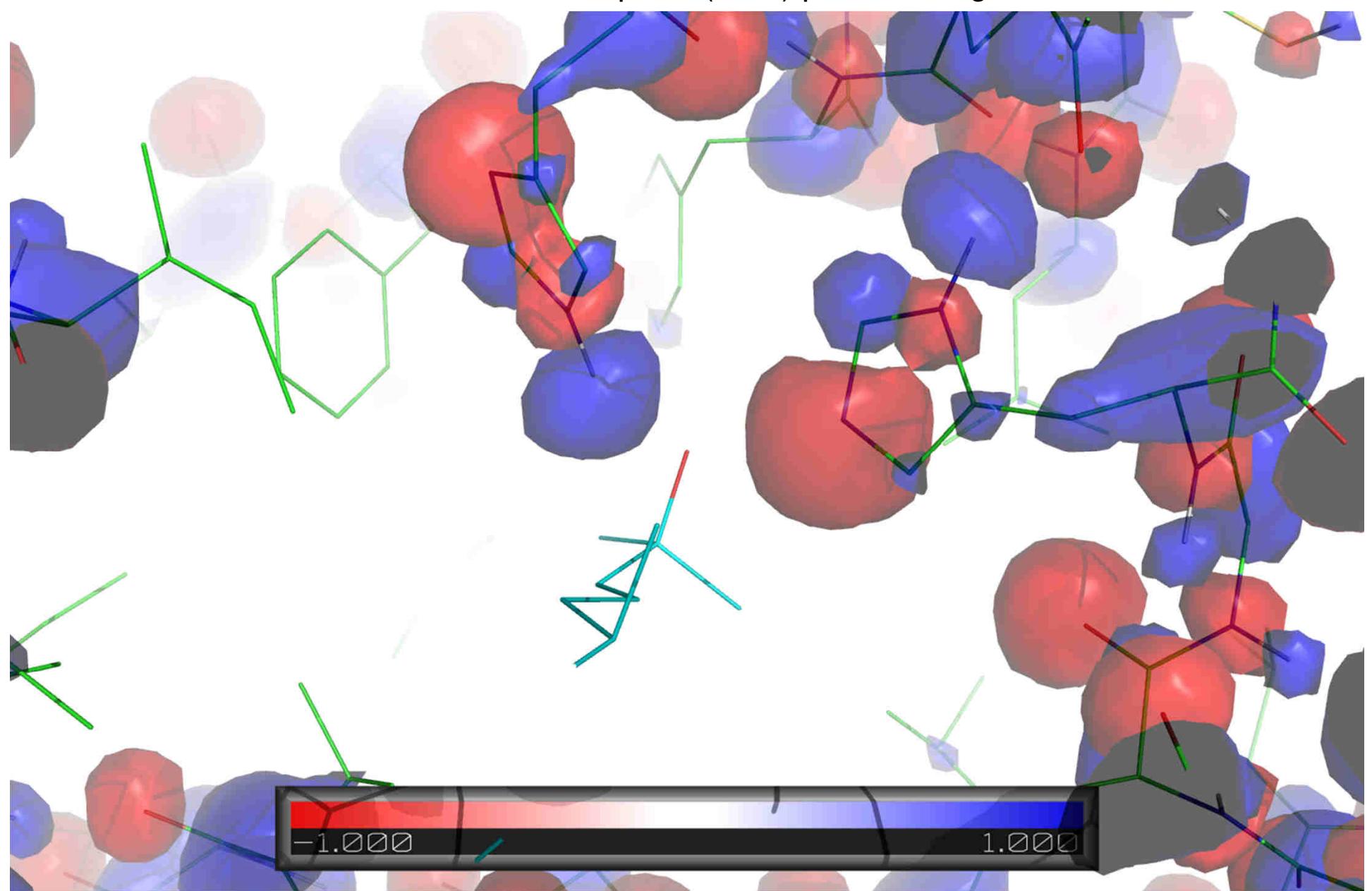
N	-0.520	N	-0.520
C	0.526	C	0.526
O	-0.500	O	-0.500
CA	0.219	CA	0.219
CB	0.060	CB	0.060
CG	0.112	CG	0.112
CD2	0.122	CD2	0.122
CE1	0.384	CE1	0.384
ND1	-0.527	ND1	-0.607 -0.08
NE2	-0.444	NE2	-0.524 -0.08
H	0.248	H	0.248
HE2	0.320	HE2	0.480 +0.16

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

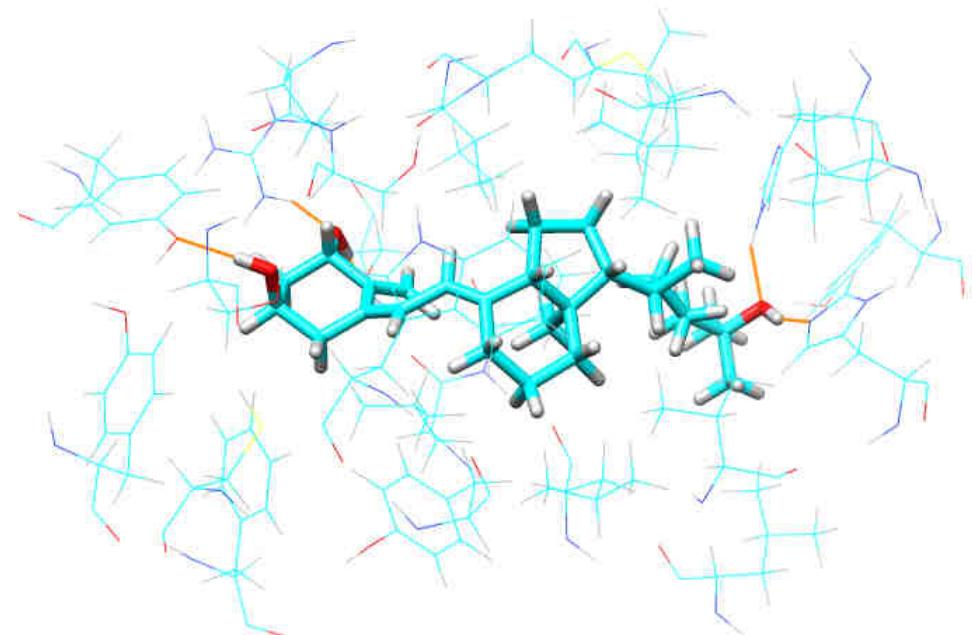
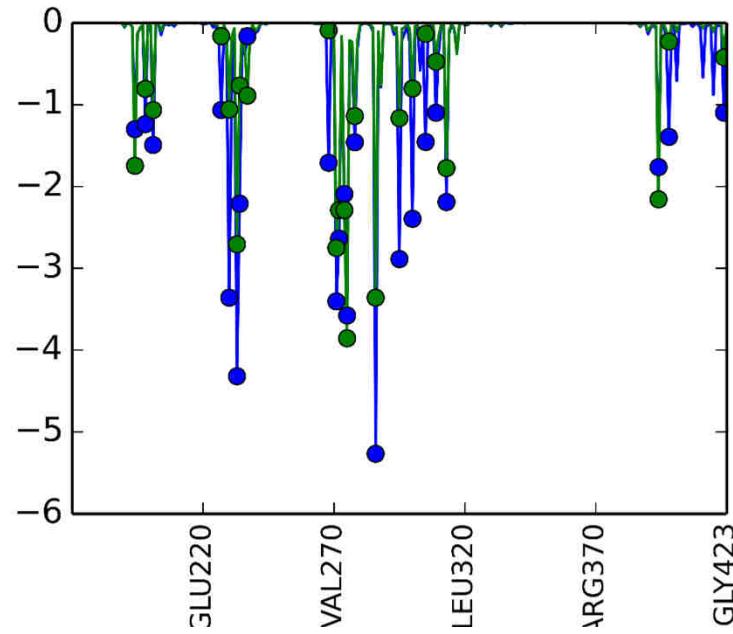
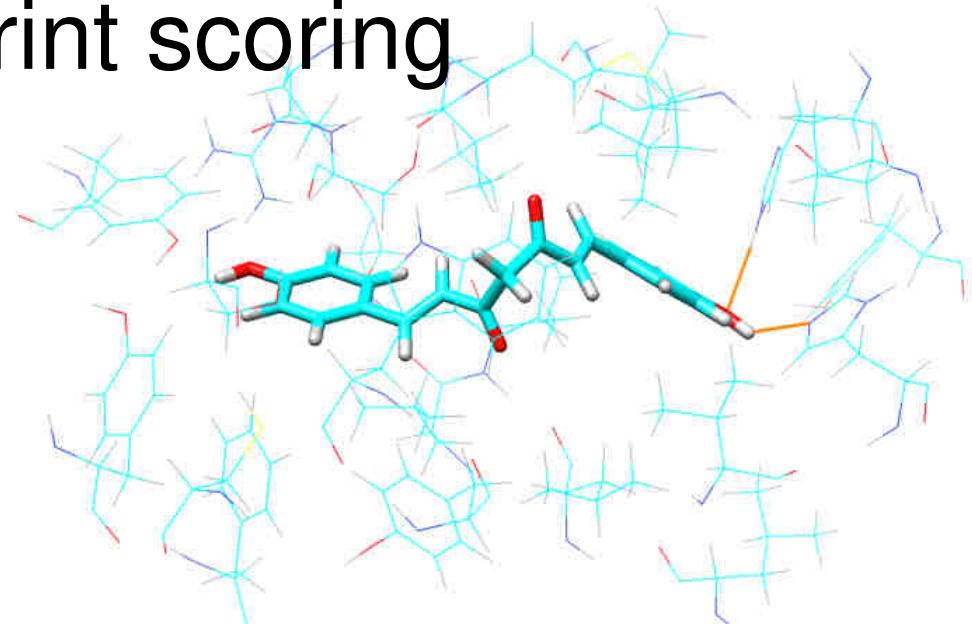
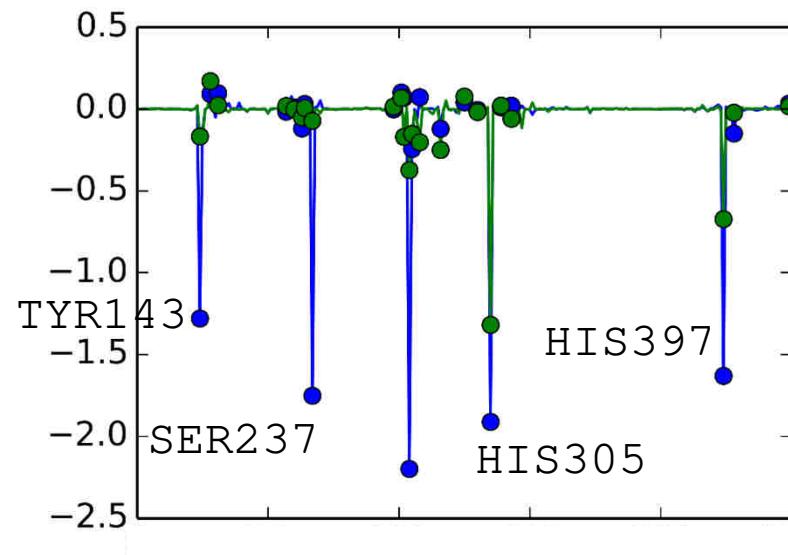
Original partial charges



More polar (tart2) partial charges



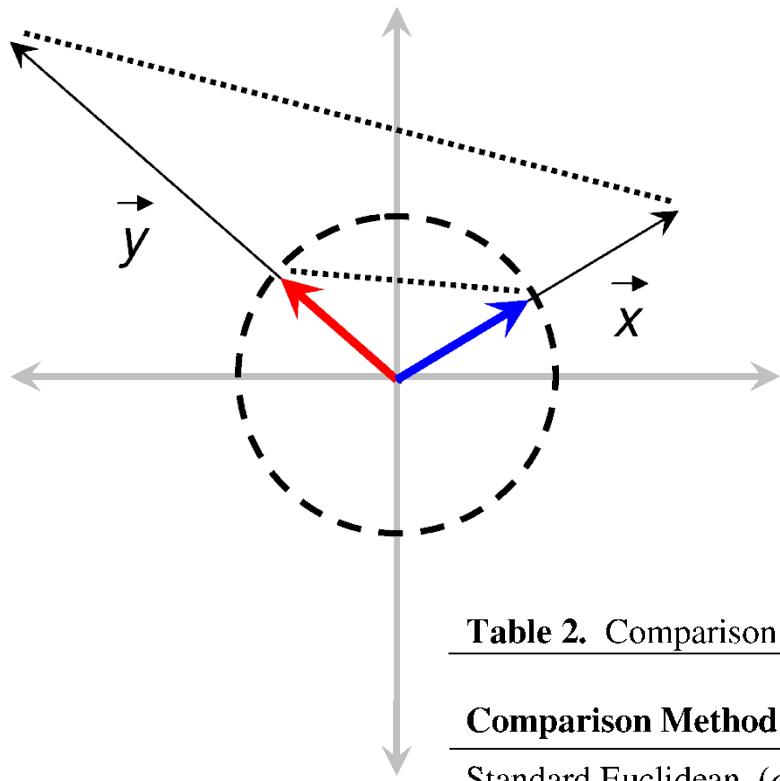
Footprint scoring



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
- $\text{FPS}_{\text{VDW+ES}} = \text{FPS}_{\text{VDW}} + \text{FPS}_{\text{ES}}$ (consensus score)

Methods for Quantifying Similarity



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

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

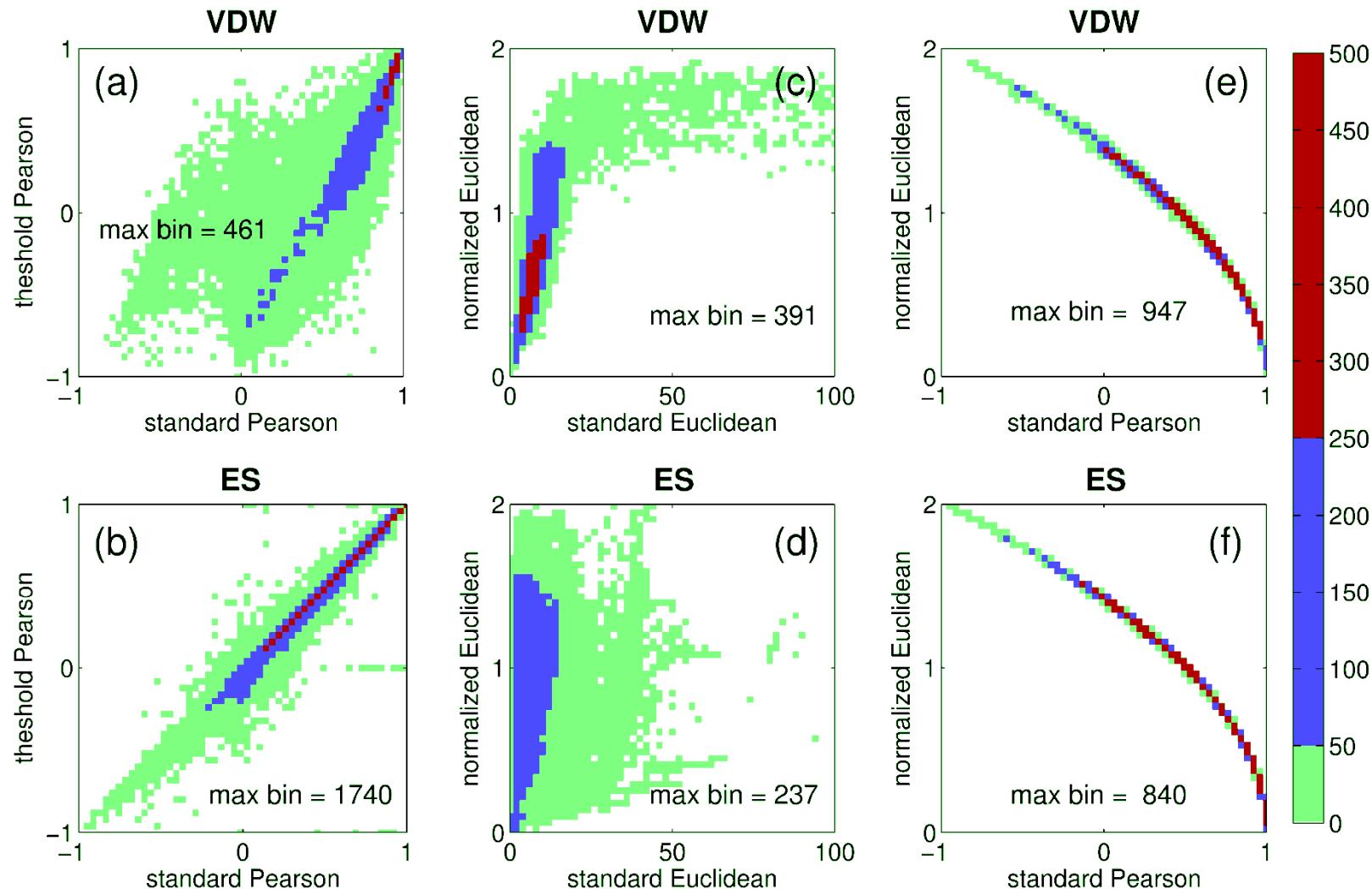
$$r = \frac{\text{cov}(\vec{x}, \vec{y})}{\sqrt{\text{var}(\vec{x})} \sqrt{\text{var}(\vec{y})}}$$

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

Comparison Method	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



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

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,

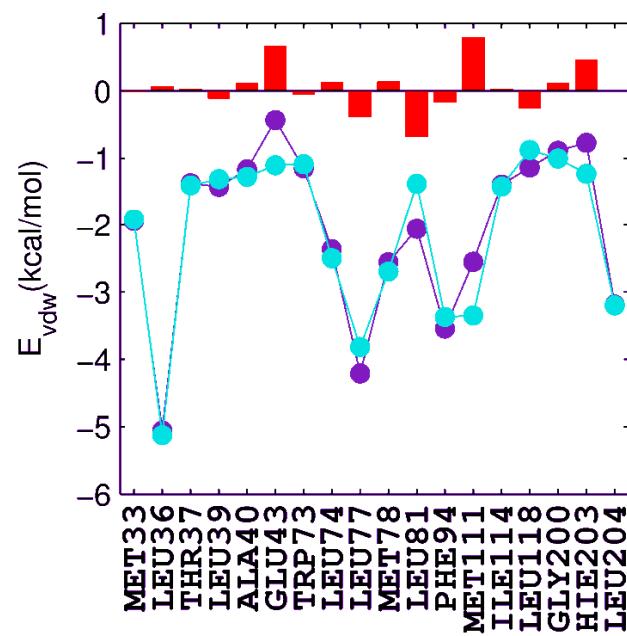
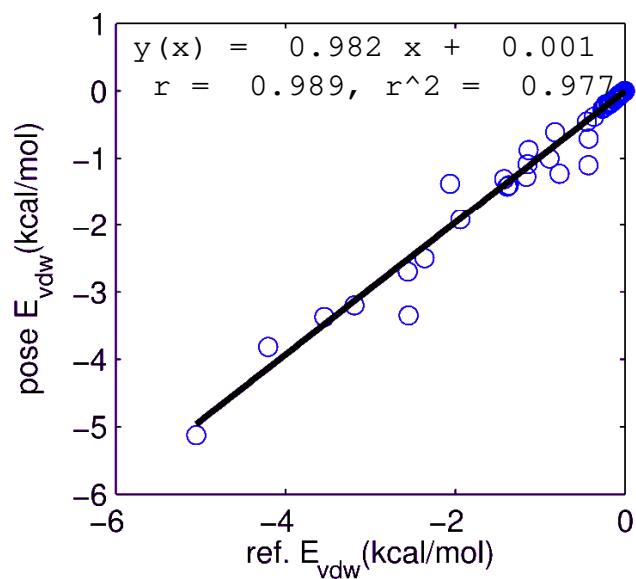
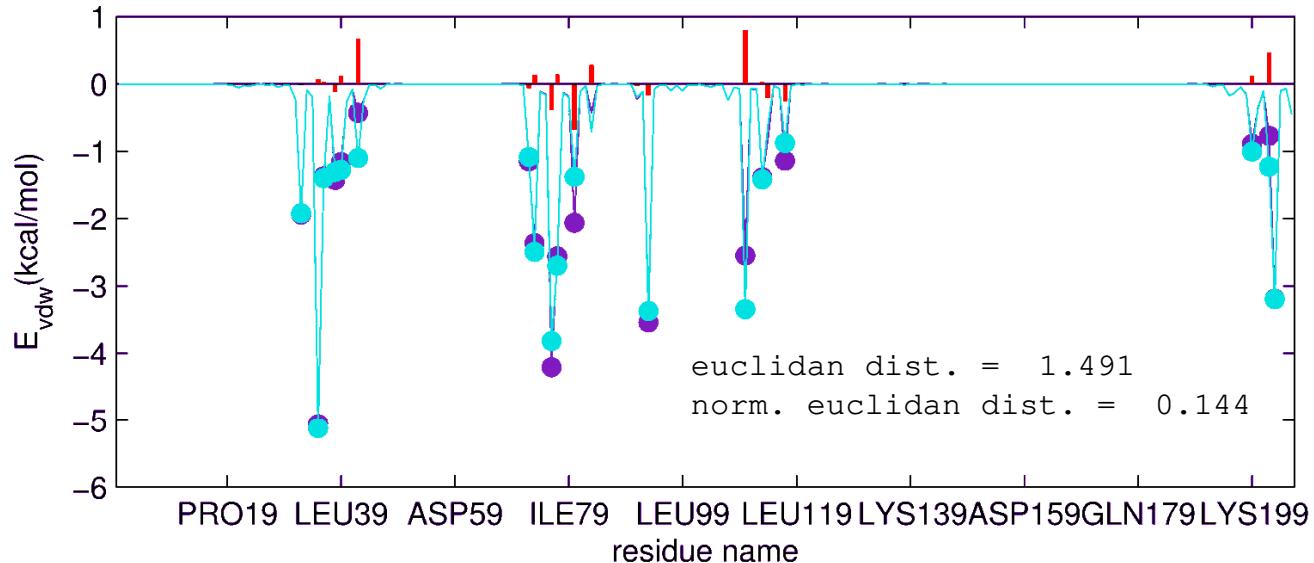
$$\cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|} = \vec{u} \cdot \vec{v} \quad (\because \|\vec{u}\| = 1)$$

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

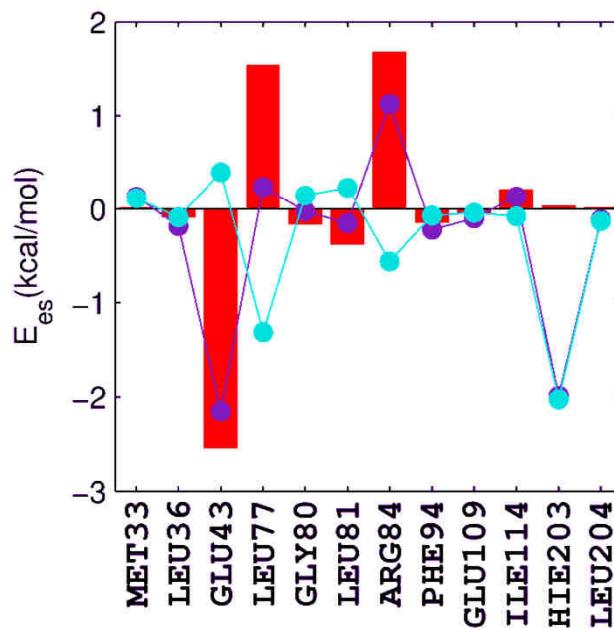
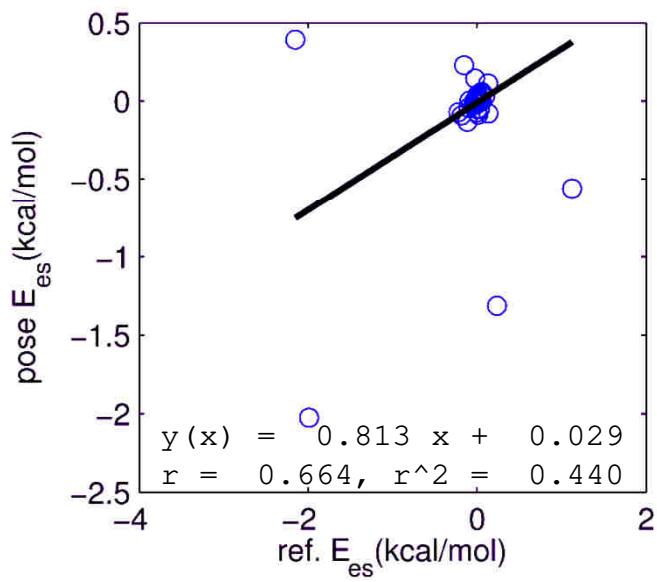
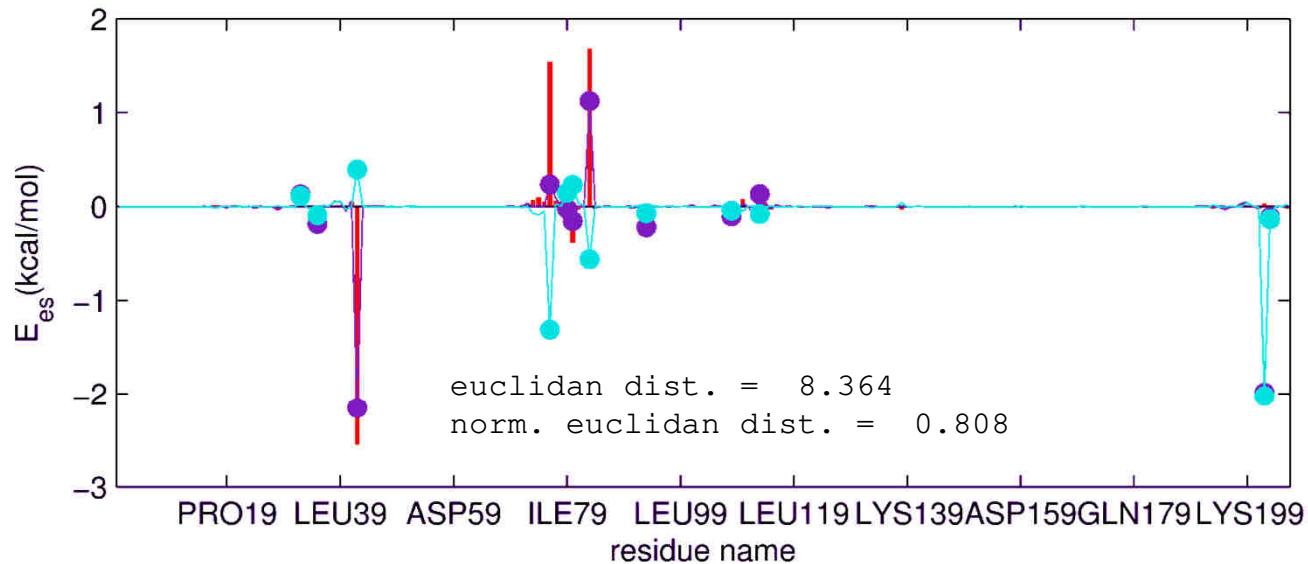
Therefore,

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

2QE4 van der Waals Footprint (Good Comparison)



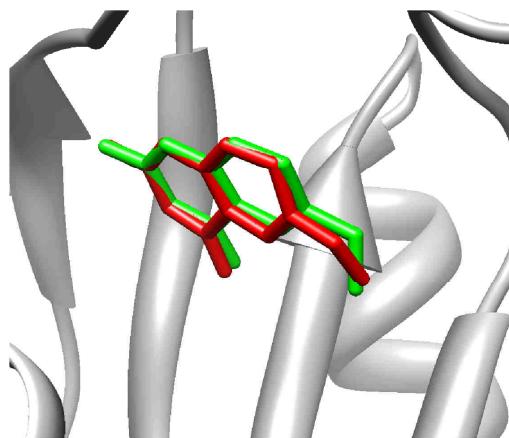
2QE4 Electrostatics Footprint (Bad Comparison)



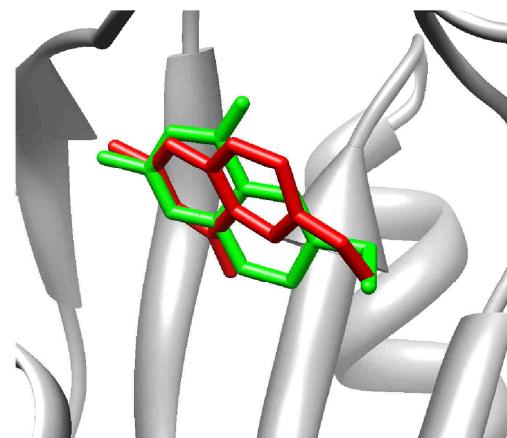
Structural Comparisons using rmsd

— reference molecule — docked molecule

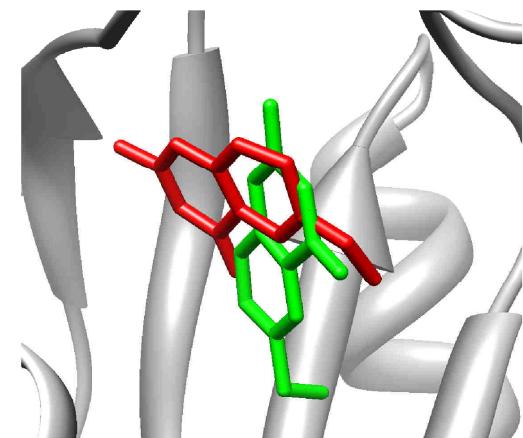
(a)



(b)



(c)



0.33 Å

success

2.56 Å

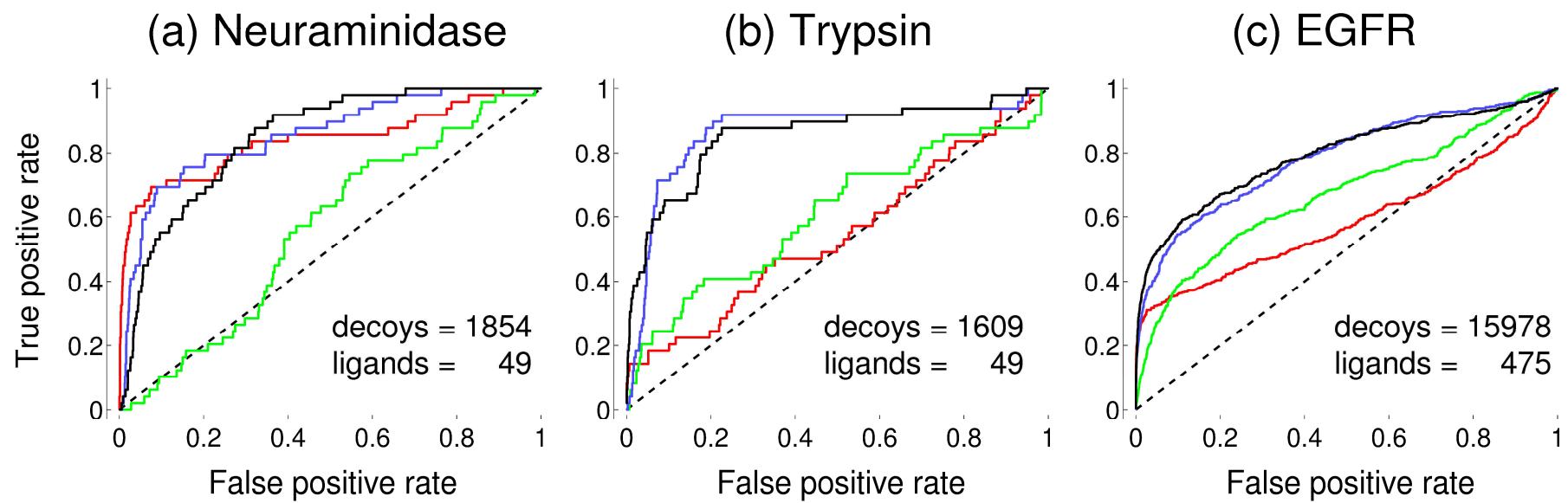
near success

4.13 Å

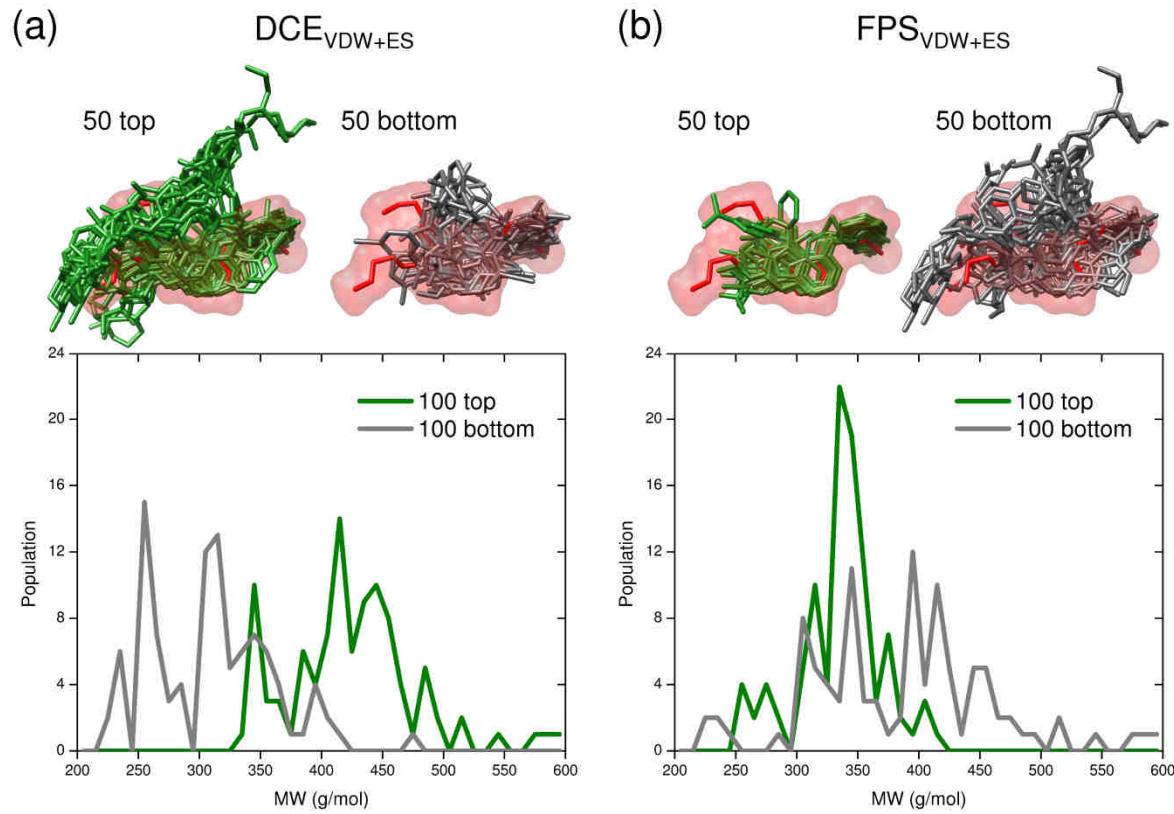
failure

Enrichment (ROC Curves)

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

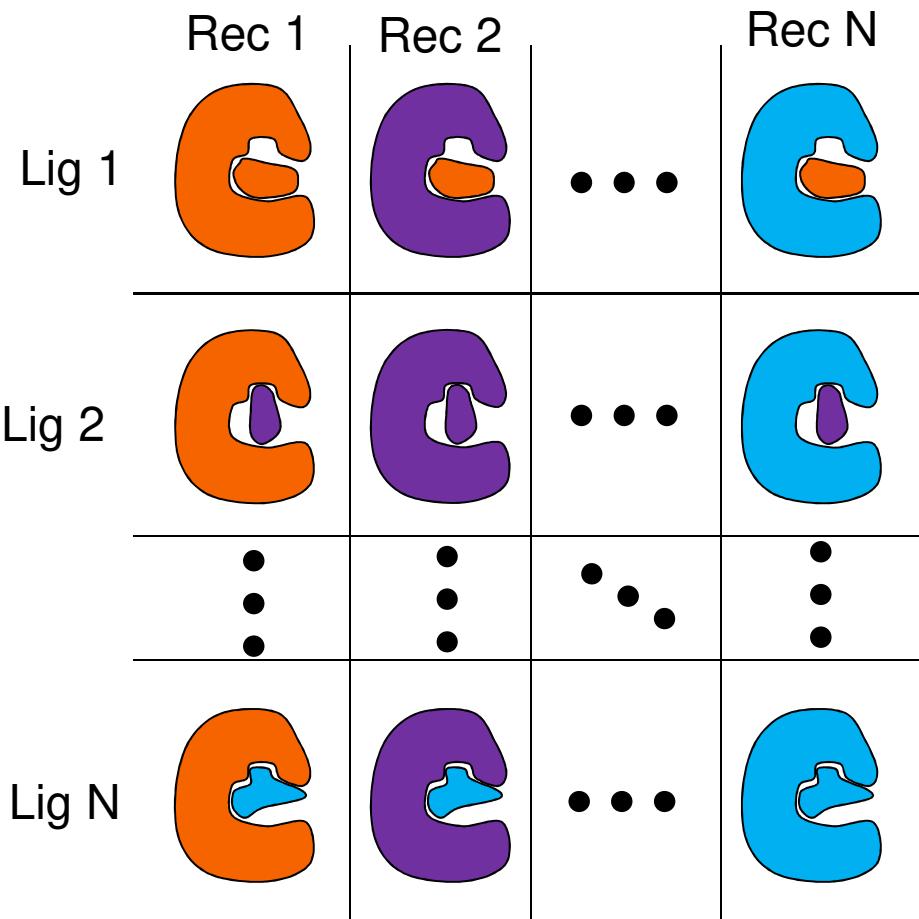


Enrichment (Molecular Weight Bias)



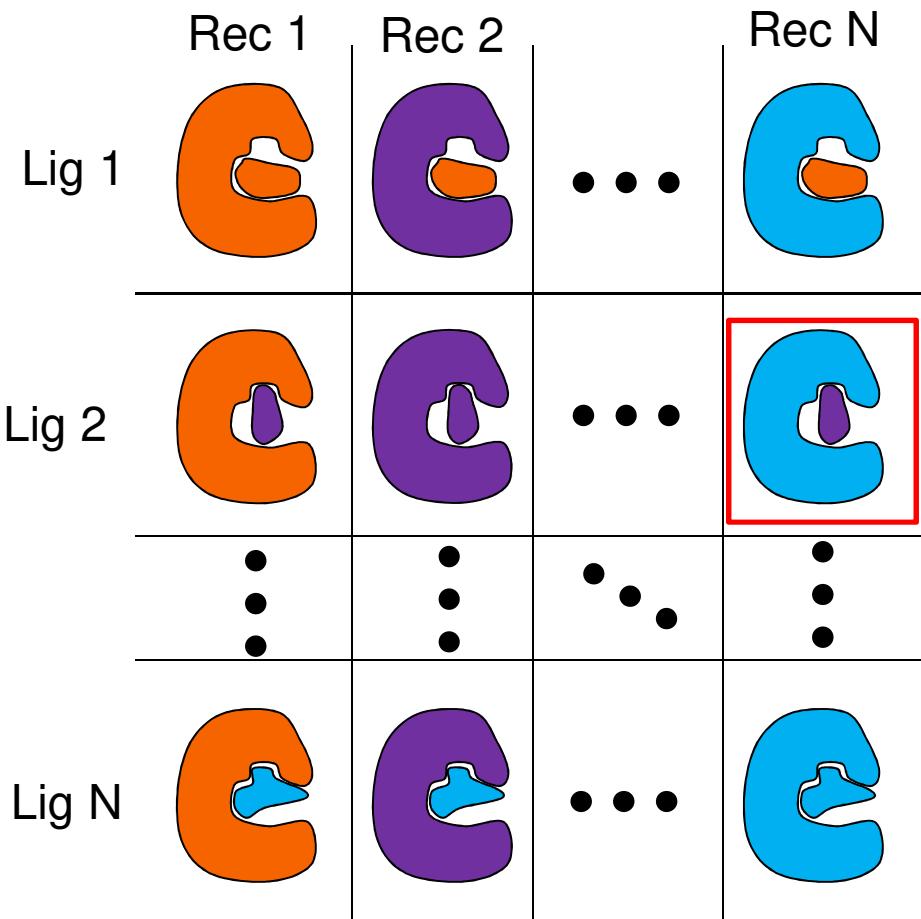
- DCE prefers larger ligands
- FPS prefers molecular weight close to that of erlotinib (393.44 g/mol)

Cross-Docking



Dock every ligand into every receptor

Cross-Docking

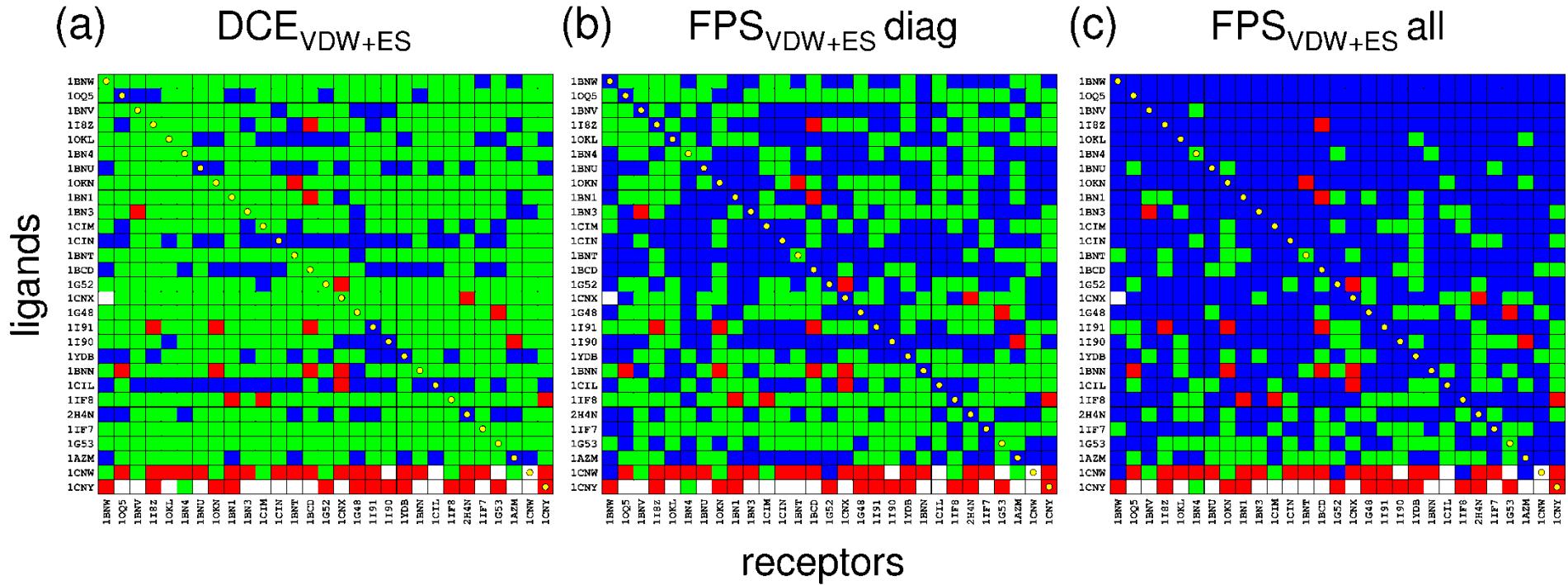


Dock every ligand into every receptor

Docking ligand 2 in to receptor N

[dock lig i in rec j]

Carbonic Anhydrase Crosdocking



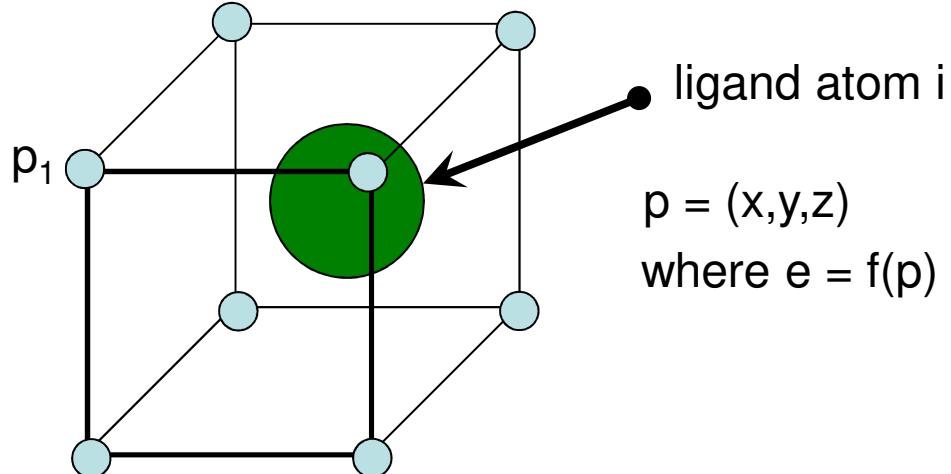
- success, blue; scoring failure, green; sampling failure, red; incomplete growth, white
- (a) standard DCE_{VDW+ES} ,
- (b) FPS_{VDW+ES} in which cognate ligands (diagonals) were used as the footprint-reference corresponding to each receptor
- (c) FPS_{VDW+ES} in which footprint-references were derived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

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

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

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

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



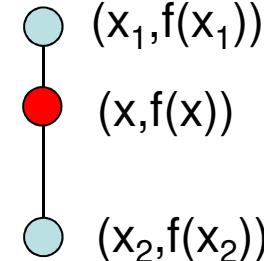
http://dock.compbio.ucsf.edu/DOCK_6/dock6_manual.htm#Grid

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

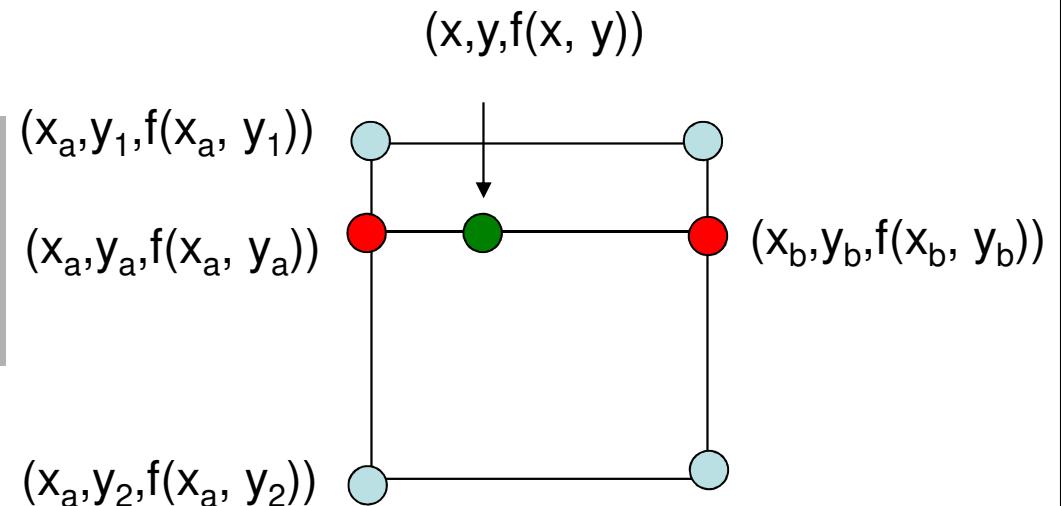
Interpolation

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**)

Decomposition for the grid

grid point

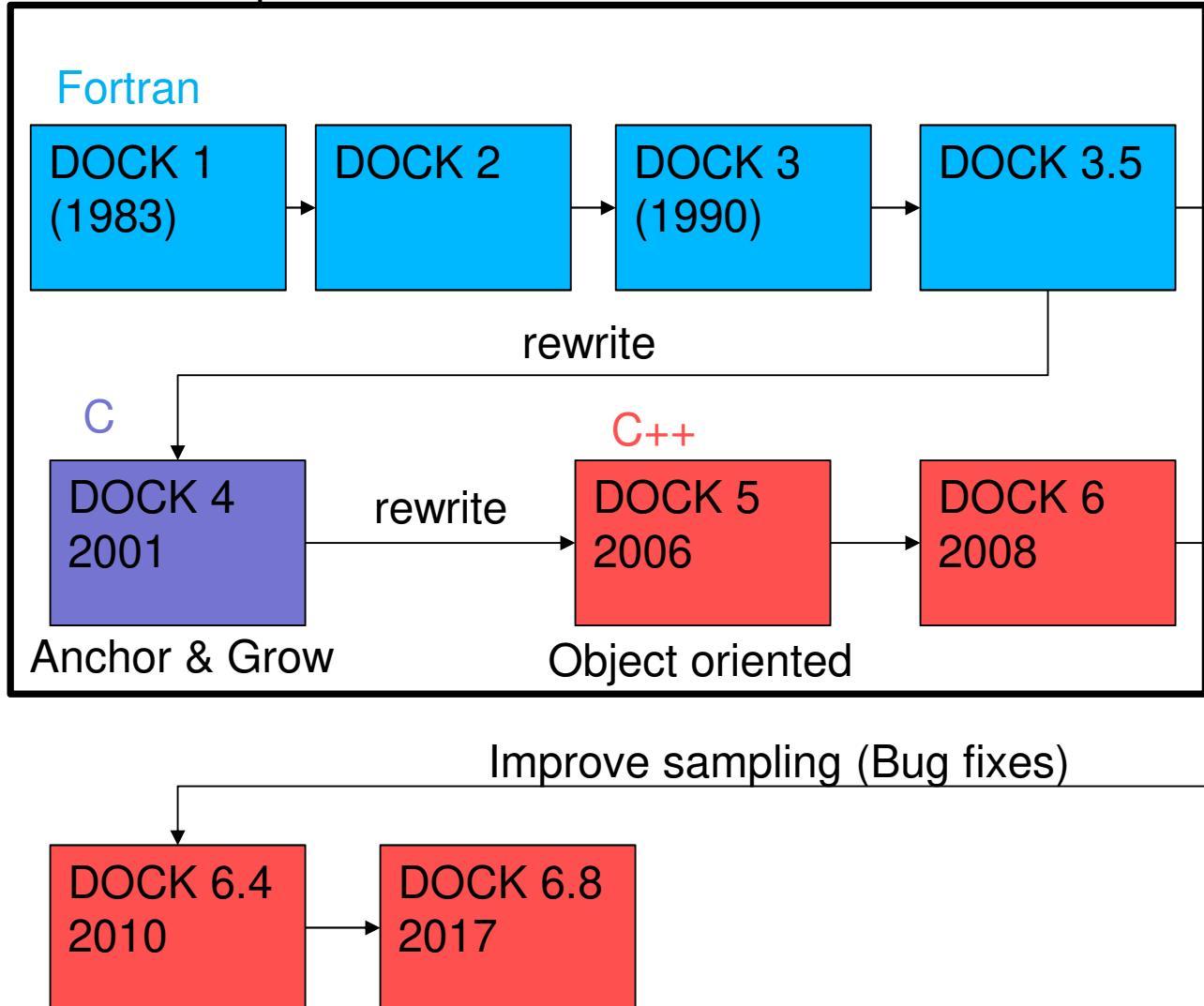
$$\begin{aligned} G_X(p) &\stackrel{\downarrow}{=} \sum_{l \in R} E_X(p, l) \\ &= \sum_{l \in SC_1} E_X(p, l) + \sum_{l \in SC_2} E_X(p, l) + \cdots + \sum_{l \in SC_M} E_X(p, l) \\ &= \sum_{k \in [1, M]} \sum_{l \in SC_k} E_X(p, l) & R = [1, N] \\ &= \sum_{k \in [1, M]} G_{X, SC_k}(p) & \text{set of receptor atoms} \\ & & \text{receptor has } N \text{ atoms} \\ & & \text{receptor has } M \text{ residues} \end{aligned}$$

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

Kuntz Group



Rizzo Group