

**GSK, we love them even if they
don't buy our stuff**

And they have a good docking test set



Top 10 reasons to go to GSK

- 10) Offer to return Snake Warren for human sacrifice in exchange for buying Fred
- 9) Discuss **CENSORED**, religion and fashion
- 8) Be reminded how slow SGI's really are
- 7) Avoid Boston traffic, by sitting in Philly traffic
- 6) Listen to Marti complain about Anthony



Top 10 reasons to go to GSK

- 5) Get to go out to expensive steak dinner on the company
- 4) Steal donuts from birthday celebration of GSK employee we don't know
- 3) Fix bugs in Fred
- 2) Remember to be thankful OpenEye doesn't have an IT group
- 1) Run GSK test case



What is the GSK docking test?

- 8 Targets
 - Chk1, fxa, gyrb, hcvp, mrs, pdfc, pdfs, ppard
 - Up to 180 actives per target
- Data provided for test
 - One protein structure per target
 - One bound ligand per target
 - Residues making up the active site for each target
 - Database of ~1300 ligands to dock
 - Database made up of actives for all 8 targets

Didn't Marti already give this presentation at CUP 4?

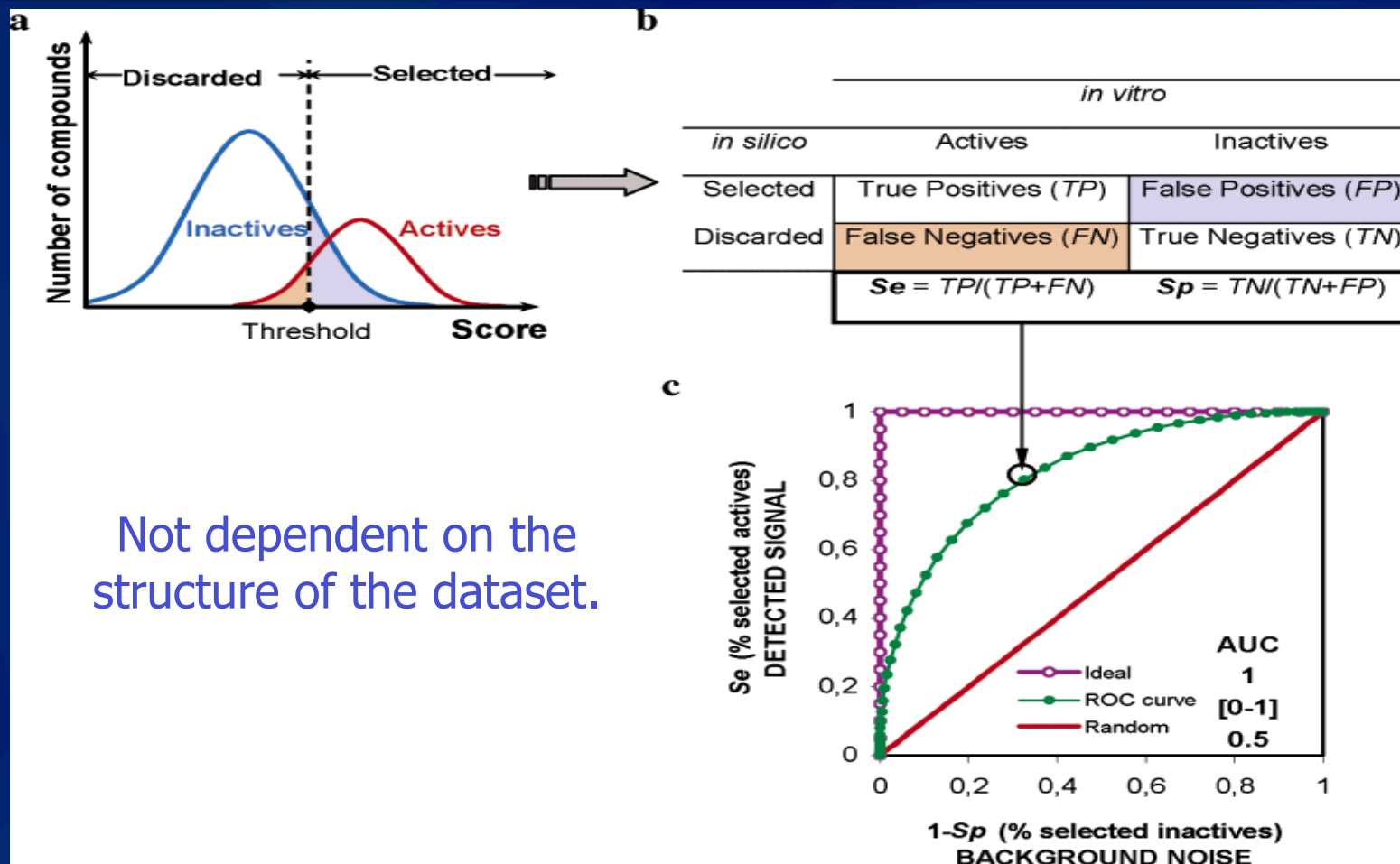
- Results were from Fred 1.2.10
 - Fred 1.2.10 was outperformed by more other docking programs
 - Smashed MOE-Dock!
 - When a single constraint was introduced results improved significantly
- Fred 2.2 is three generations improved

Key Improvements since original test

- Graphical setup of receptor site
 - Dedicated site setup program
 - Easy to specify constraints
- Chemgauss3 scoring function
 - Approximate Desolvation* => less size bias
- Exhaustive Scoring
 - No longer restricted to shape alone
 - Better negative image

*not real physics

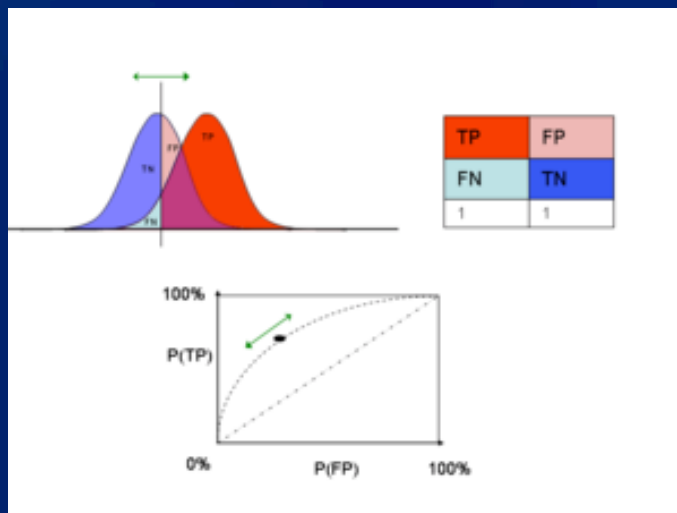
Receiver Operator Characteristic



Not dependent on the structure of the dataset.

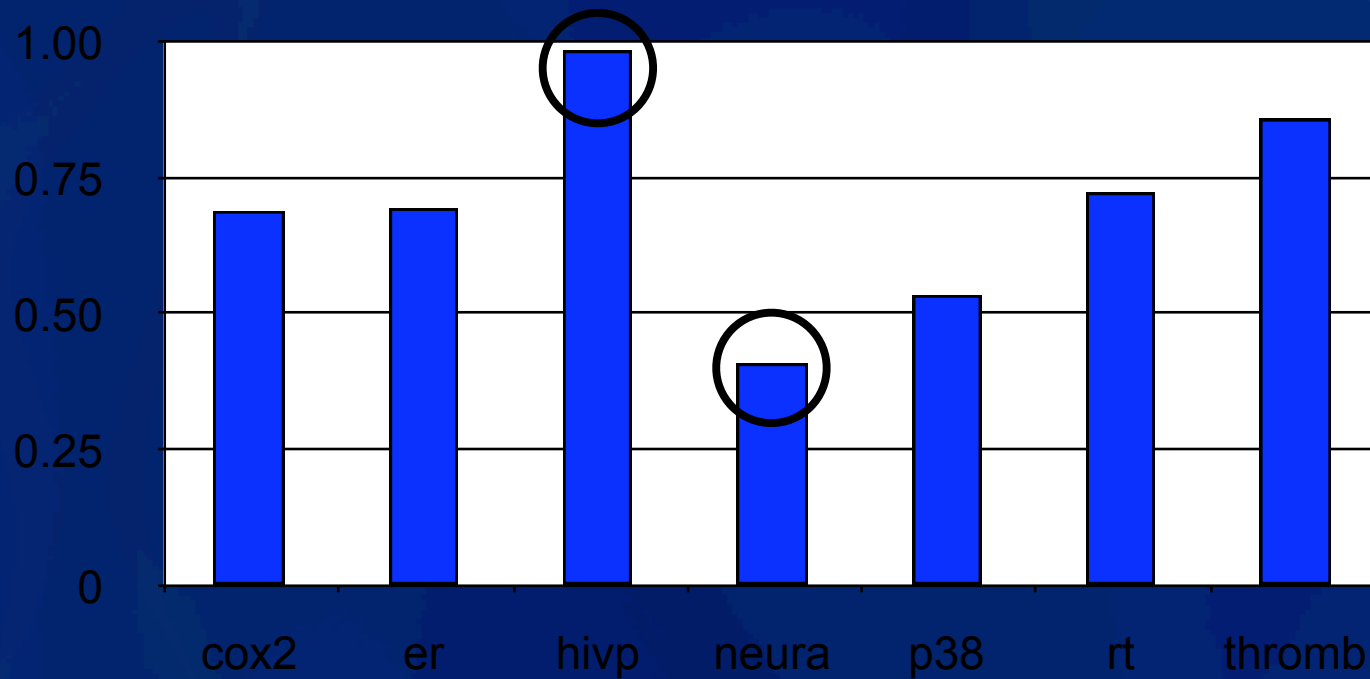
Triballeau et al., J. Med. Chem., **48**, 2534 (2005)

Discrimination



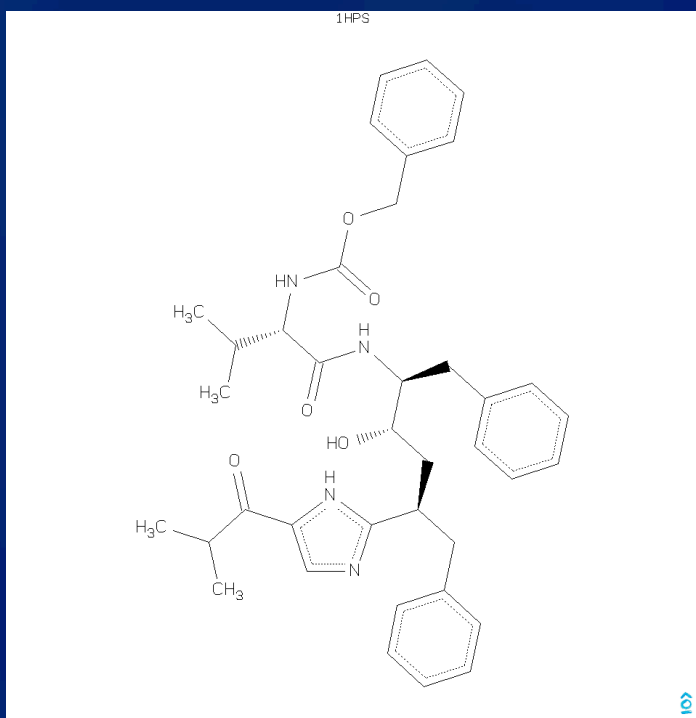
- AUC measures discrimination
 - Select one TP and one TN randomly
 - Discrimination = %tage of pairs that TP is given higher score than TN
 - R, randomness (Baber et al., J. Chem. Inf. Model., **2006**, 46, 277)

Secret Origin of Chemgauss3

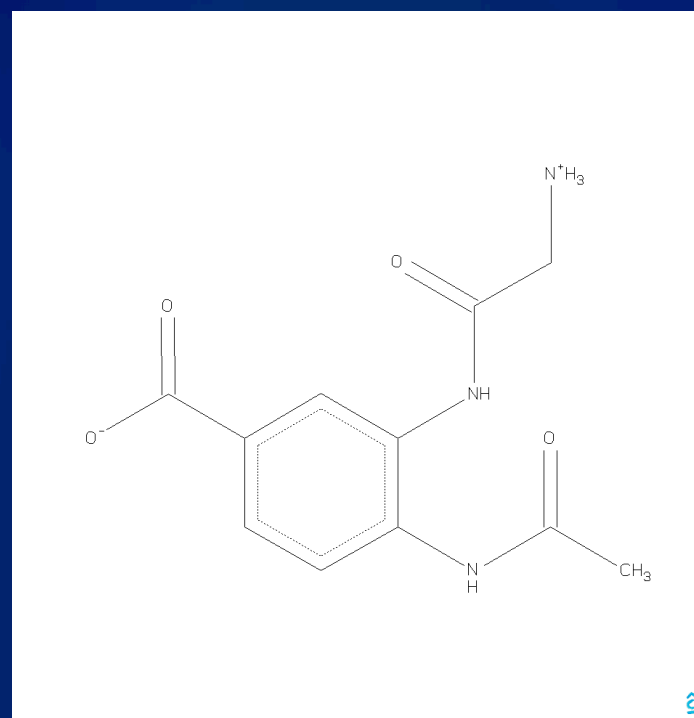


Size Bias

An HIV protease active



A neuraminidase active



Average MW

HIV actives

580

Neura. actives

290

Decoys

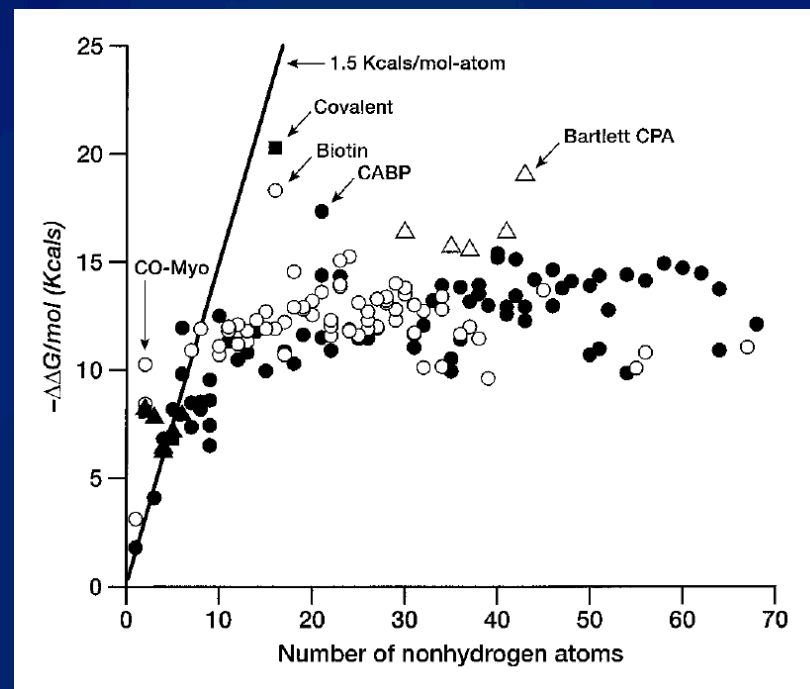
350



Does size matter?

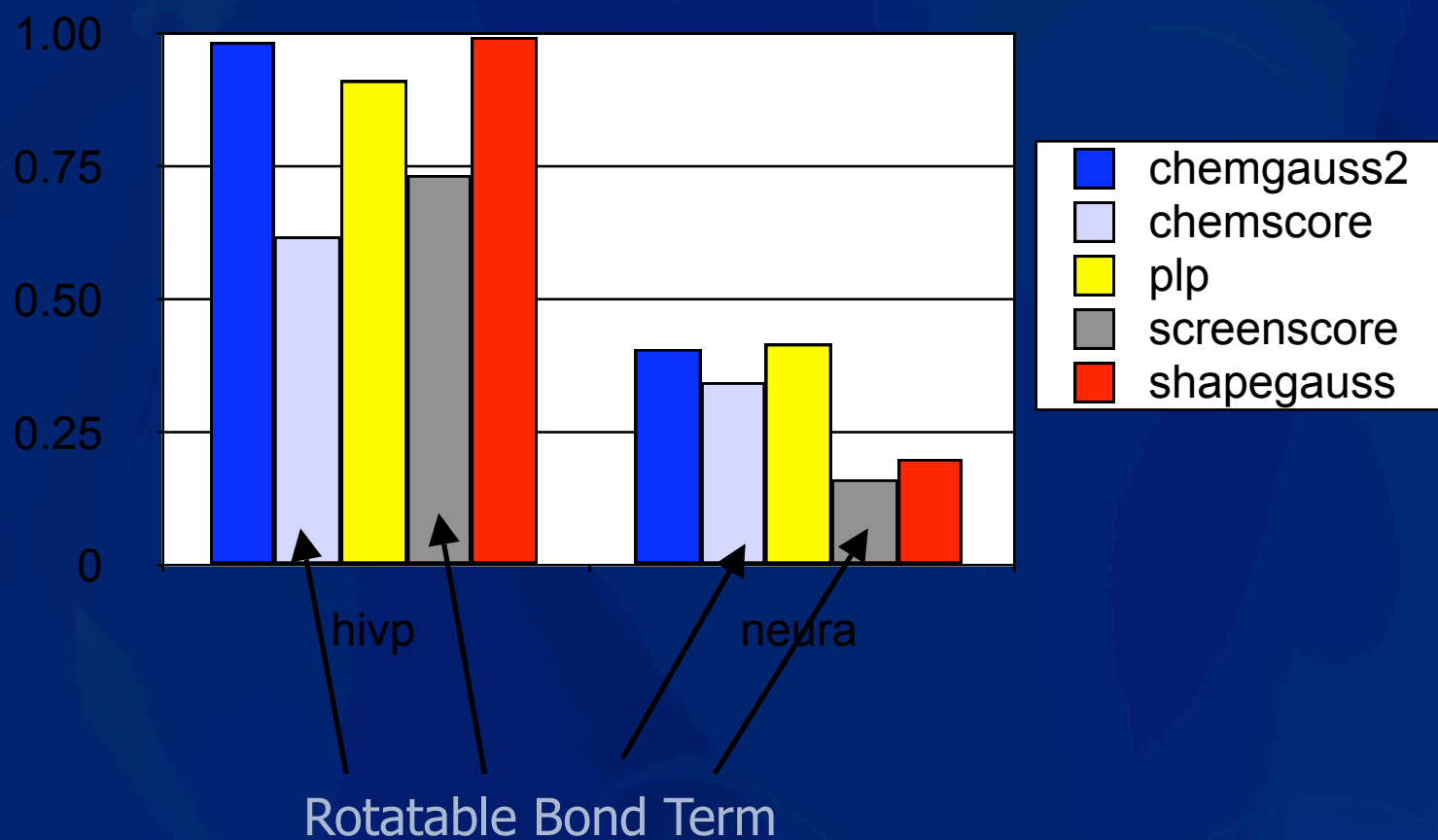
- Physical size of the site
- Entropy
- Desolvation

Size **should** matter, but scoring functions with only positive interactions likely overestimate the importance of size



“Maximal Affinity of Ligands”, Kuntz *et al.*, *Proc. Natl. Acad. Sci.* pg 10001, v 96, 1999

Are all Fred 2.1. scoring functions sized biased?

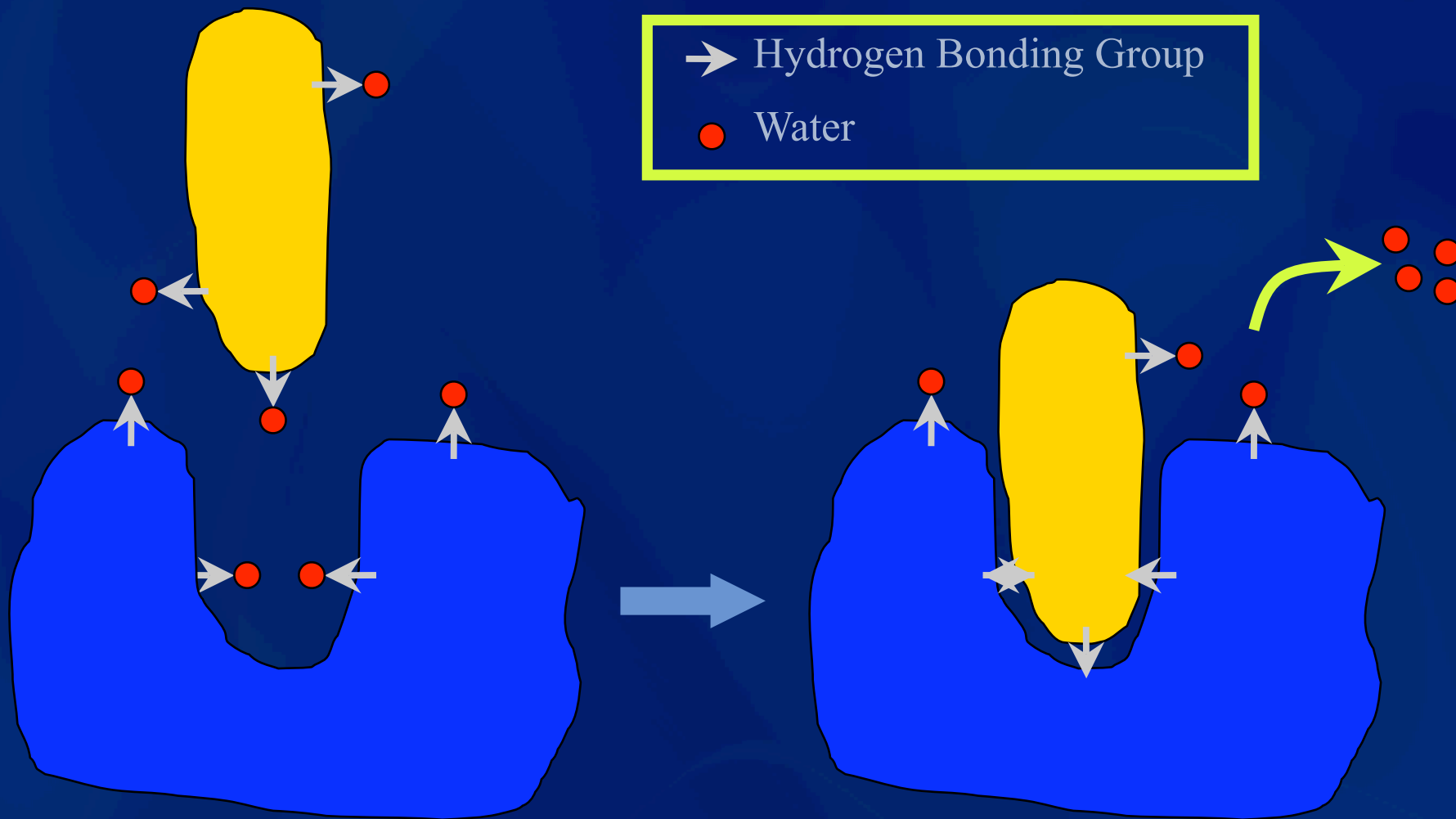


Partial Desolvation : PB

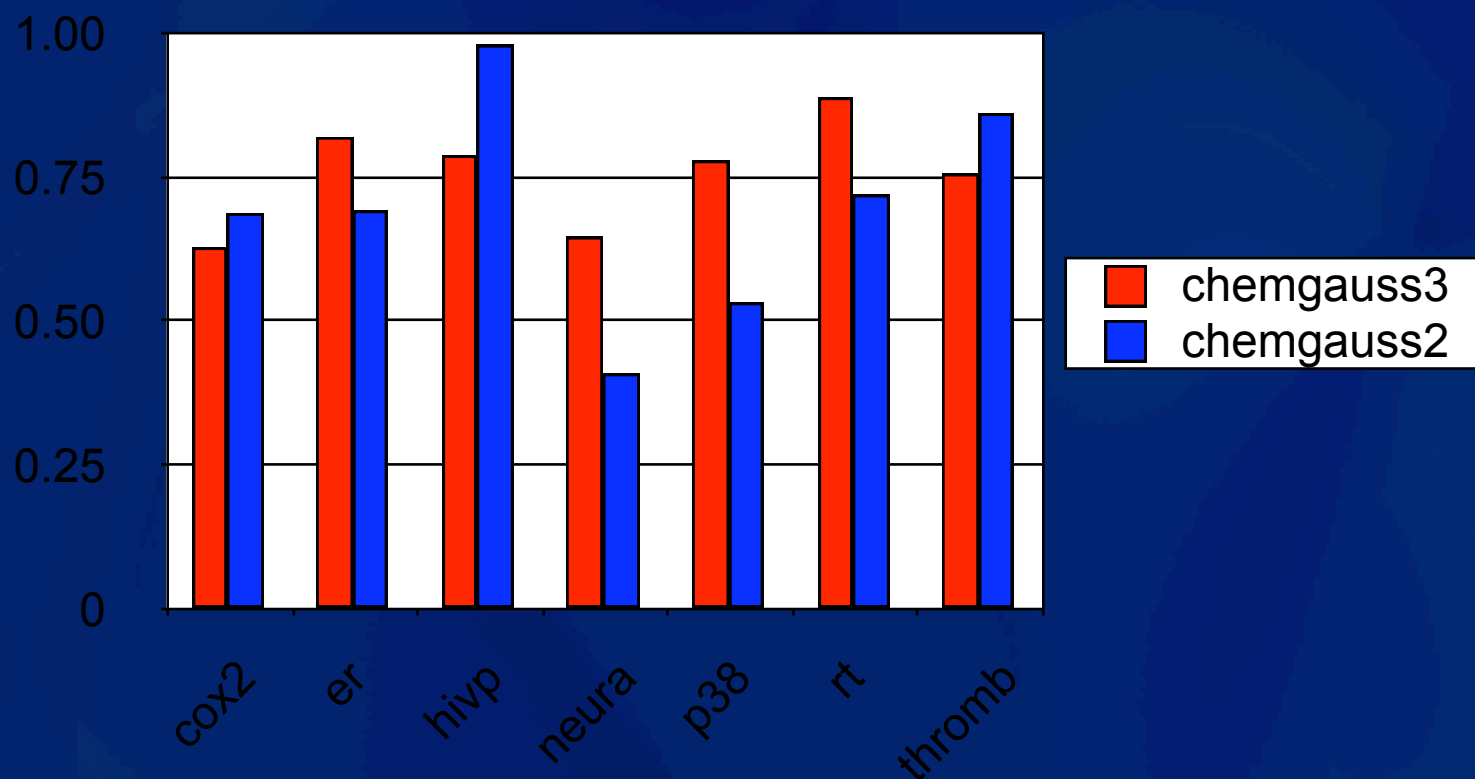
$$\nabla \varepsilon \cdot \nabla \phi + \kappa \phi = -4\pi q$$

Approximating Desolvation : S.A.D.T.T.A.W.M.V.C.I.R.P.

→ Hydrogen Bonding Group
● Water



Effects of Hydrogen Bond Desolvation



Median AUC Chemgauss3 : **0.79** (SADTTAWMVCIRP)

Median AUC Chemgauss2 : **0.70** (no desolvation)

Goals of test

1. Run Fred on a test set it was not trained on
Not truly prospective, but very close
2. Compare Fred to other programs
Rocs
Glide
3. Evaluate effectiveness of constraints
4. Compare site preparation techniques
Expert setup with GUI
Automatic setup by Fred



Setting up the site

1. Automatic

Setup by Fred using protein and box

2. Programmer/Engineer –Mark

Modified #1 in GUI

No prior knowledge of actives or site

3. Expert - Greg

Setup completely in GUI

Used knowledge of site



Running Fred

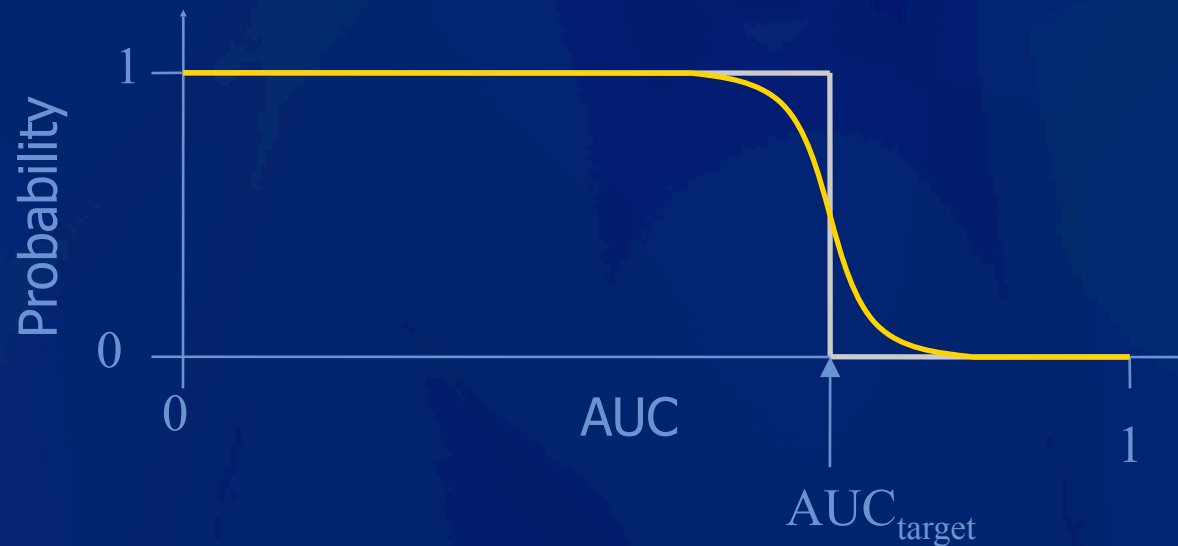
Default settings were used for Fred

1. Exhaustive Search
Chemgauss3 scoring
100 Poses per molecule returned
2. Optimization
All 100 poses optimized vs. Chemgauss3
3. Consensus structure
Single best pose selected by consensus of PLP,
Chemgauss2, Chemgauss3 and OEChemscore
4. Scoring
Chemgauss3

Headaches

- **Missing results**
 - Gyrb : Mark and Greg's constraint runs
 - Hcv : Automatic constraint runs
 - Fxa : Mark's setup. Rerun successfully
- **Metals**
 - Pdfe and Pdfs both have metals (Ni)
 - Nickel interactions half strength
 - Not all valid geometries recognized

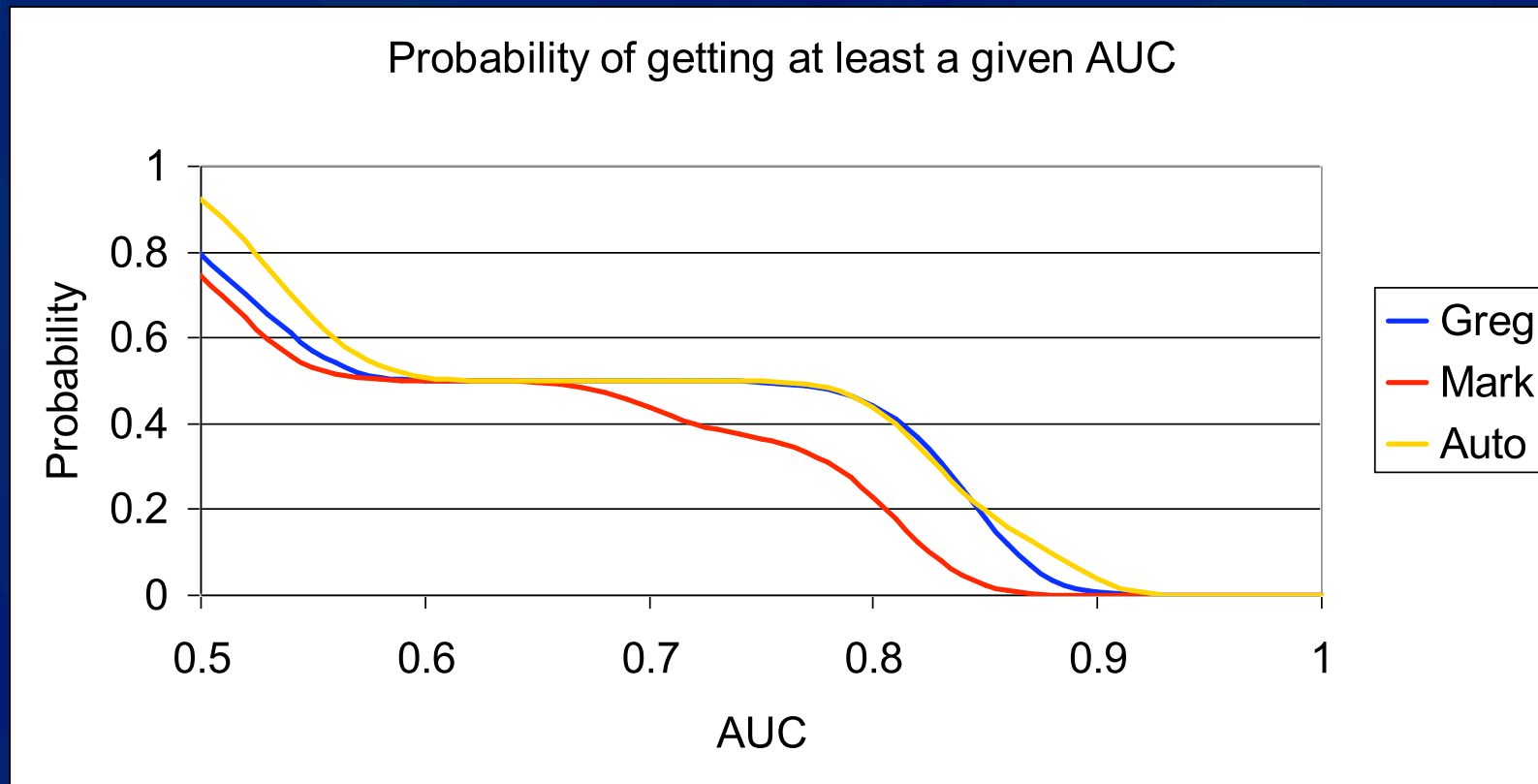
Probability of success



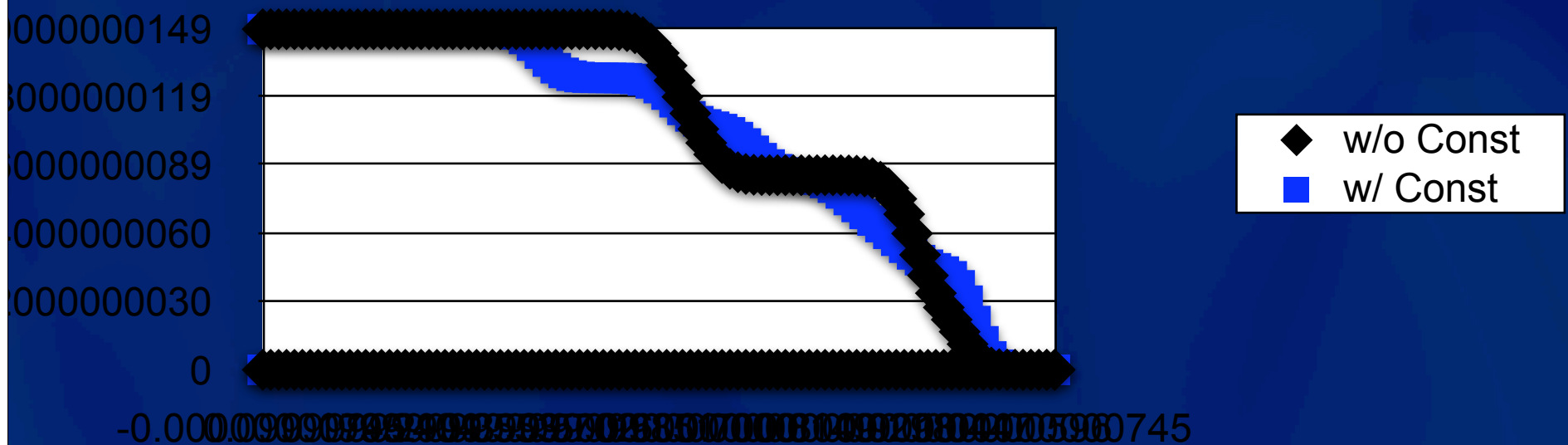
$$P(x) = \begin{cases} 1, & x < AUC_{\text{target}} \\ 0, & x \geq AUC_{\text{target}} \end{cases}$$

$$P(x) = 1 - \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - AUC_{\text{target}}}{\sigma \sqrt{2}} \right) \right]$$

Site Setup Comparison

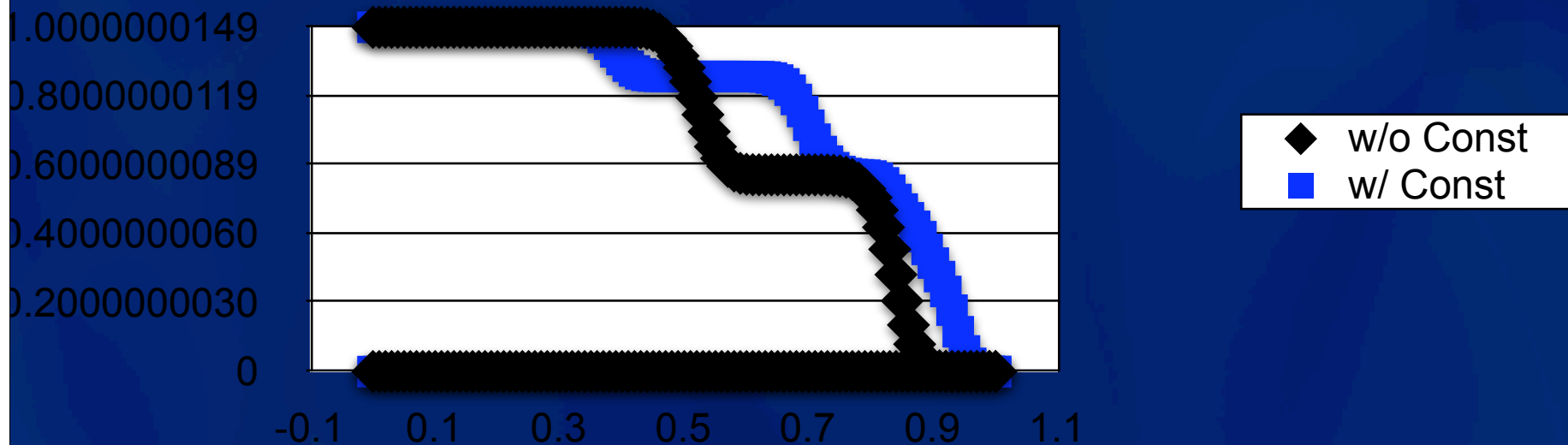


Automatic Setup w/Constraints



Constraints are a gamble, they can make results really good for some systems at the price of degrading overall performance

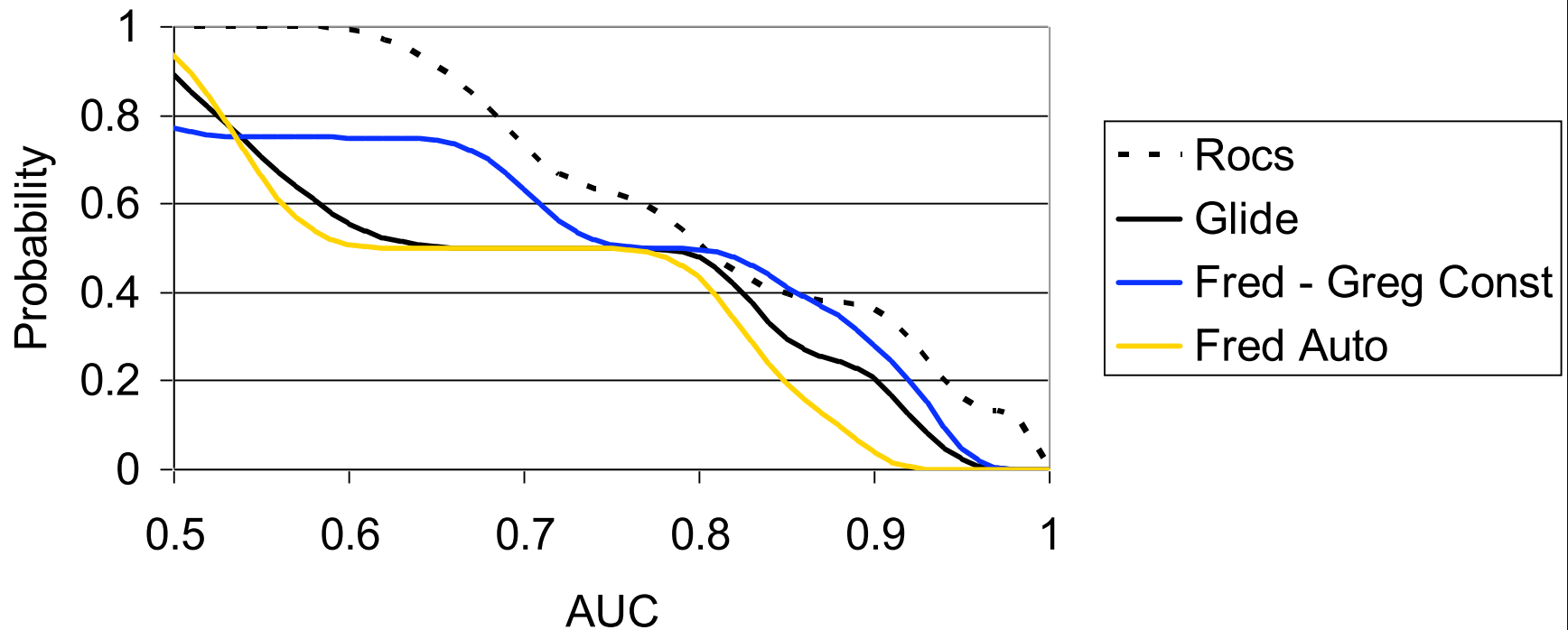
Expert Setup w/Constraints



When setup by an expert constraints are beneficial

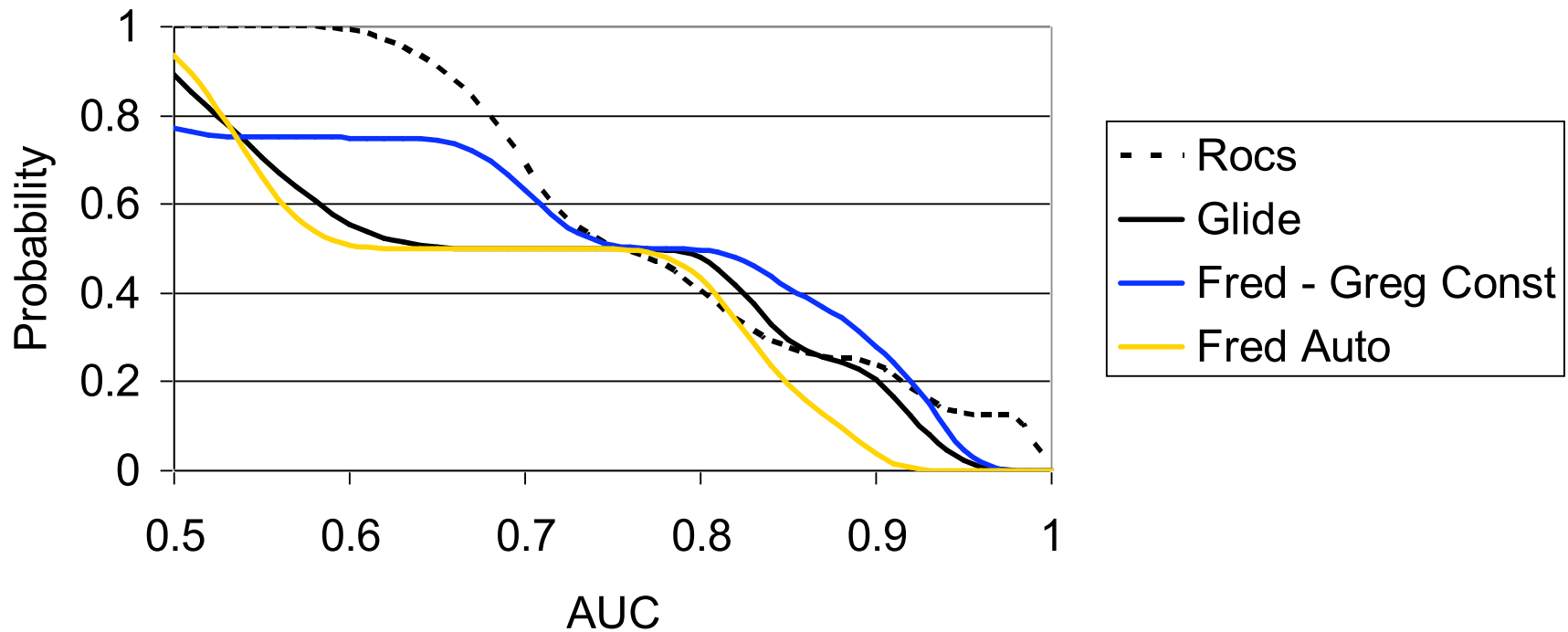
Glide, Rocs and Fred

Probability of getting a given AUC

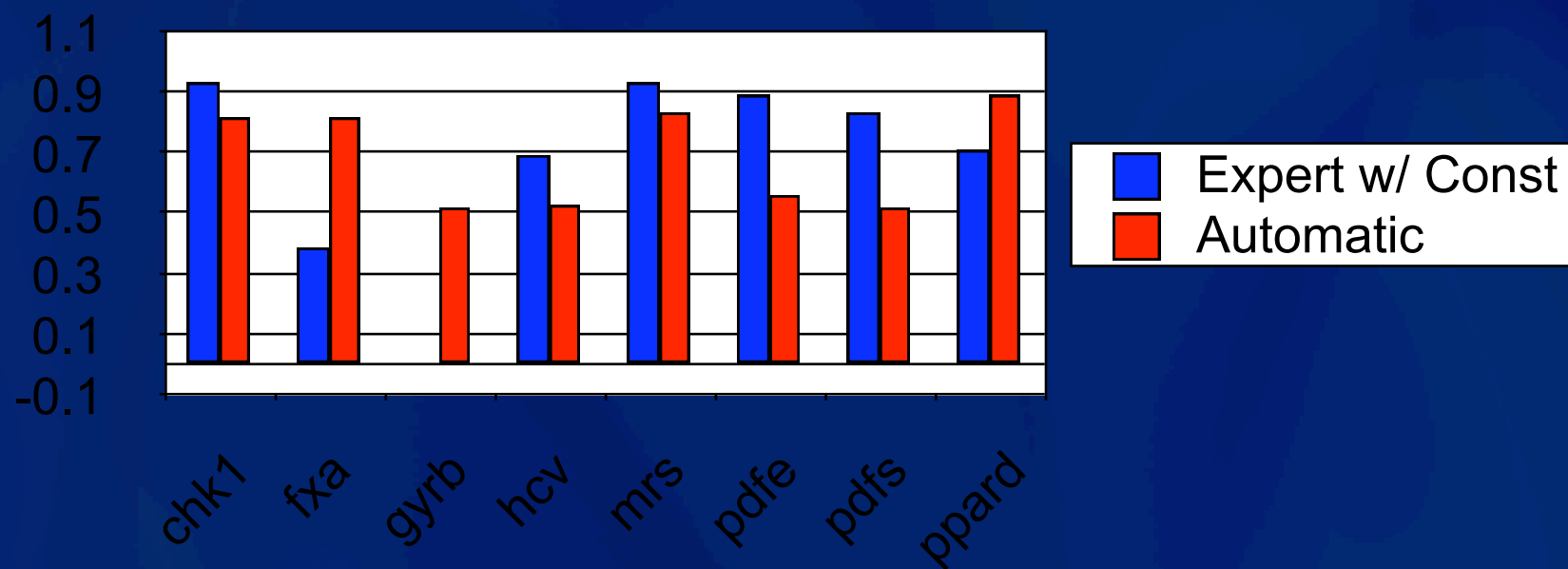


Glide, Rocs (average) and Fred

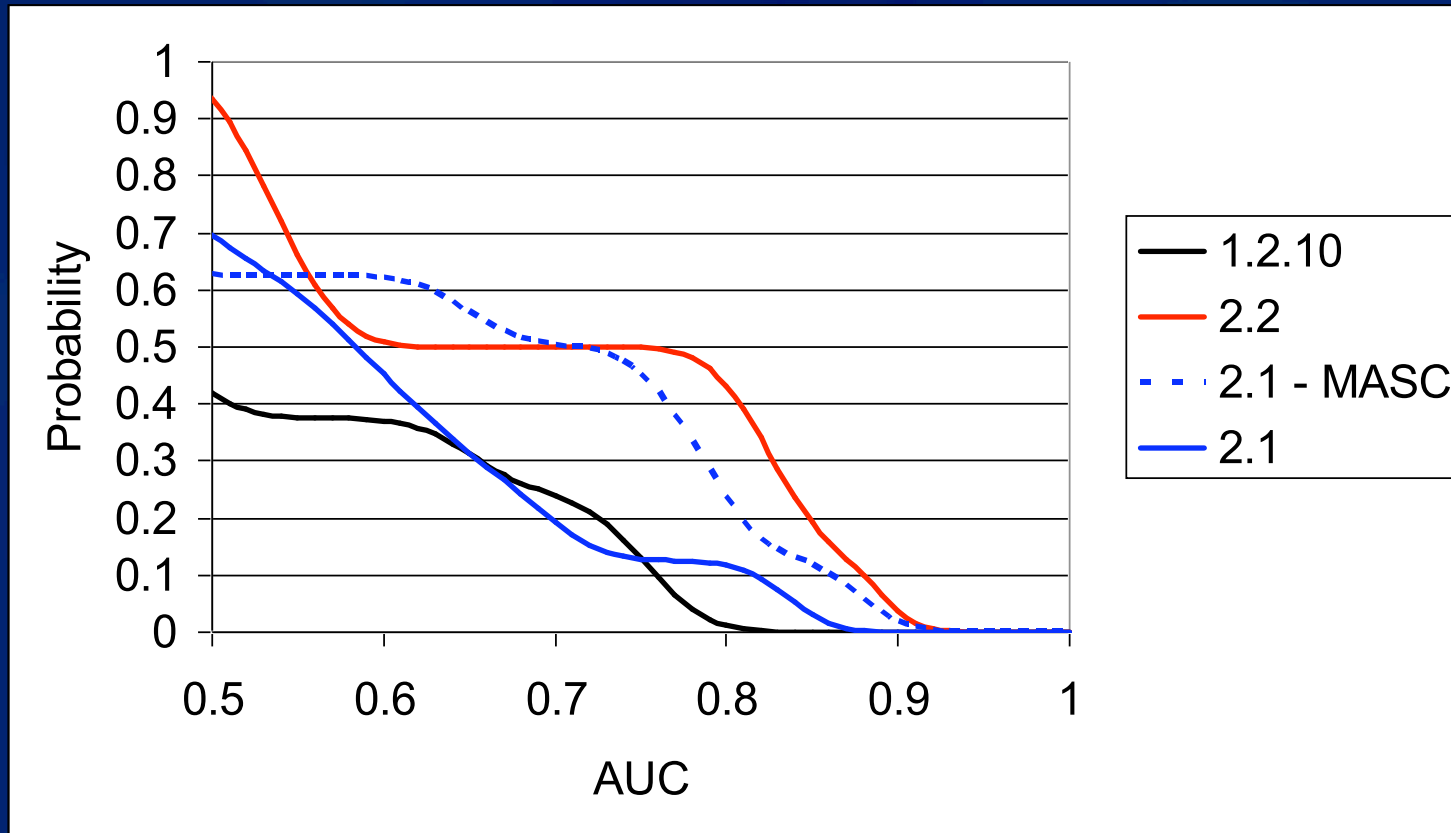
Probability of getting at least a given AUC



Fred AUC's by Target



Fred a historical perspective



Conclusions

- Without constraints expert and automatic setup gave similar results
 - Don't let the developer setup your site
- Constraints setup by experts are worthwhile
- Blindly using all automatic constraints is a gamble
- Glide and Fred performed similarly
- Rocs outperformed both Fred and Glide
- Fred 2.2 performed better than previous version



Acknowledgments

Marti Head, GSK

- Indulged my somewhat large personality

Kaushik Raha, GSK

- Did some very helpful post-processing

Greg Warren, OpenEye

- Went to Ruth's Chris, despite being a vegetarian

Geoff Skillman, OpenEye

- For believing in docking

Anthony Nichols, OpenEye

- For letting me work on Chemgauss3 even though it's not real physics

Conclusions

- Receptor Setup
 - Fred automatic setup with GSK site definition did as well as expert setup
 - Don't let the developer setup your site
- Constraints
 - Always worthwhile with expert knowledge
 - Blindly using all auto constraints is a gamble
- Fred 2.2 performed better than previous versions



Ligand Properties

| | Hiv protease actives | Decoys | Neuraminidase actives |
|----------------------|-------------------------|--------|--------------------------|
| Molecular Weight | 581 | 349 | 292 |
| Rotatable Bonds | 14.7 | 5.3 | 6.1 |
| Acceptor Strength | 4.0 | 1.2 | 3.9 |
| Donor Strength | 4.9 | 3.6 | 5.8 |

Deficiencies in Chemgauss3 metal

Pdfe and Pdfs both have metals (Ni)

- Nickel interactions half strength
- Not all valid geometries recognized
 - Chelators know positions around them when interacting metals can be placed
 - Chelators with a ring of positions were not completely sampled.

Ways to handle desolvation

- **Based on ligand properties alone**
 - Independent of the ligands pose in the site
 - E.g., M.A.S.C., Ligand solubility corrections
- **Partial desolvation**
 - Accounts for the waters that are displaced (from both the protein and ligand)
 - Depends on the pose of the ligand
 - More realistic than simply using ligand properties

Partial desolvation Options

- Calculate with PB
 - Uses real physics
 - Make the boss happy
 - Speed : a few seconds for one pose
 - Needs 'cleaner' poses than Fred normally makes
- Hydrogen Bond Desolvation
 - Expert system
 - Approximates how many hydrogen bonding groups can not interact with solvent due to the placement of the ligand in the site
 - Speed : thousands of poses per second

Decided to use Hydrogen Bond Desolvation and hired bodyguards for protection from the bosses hit squads

