

Recent Advances in Structure-Based Drug Design

Woody Sherman
Vice President, Applications Science

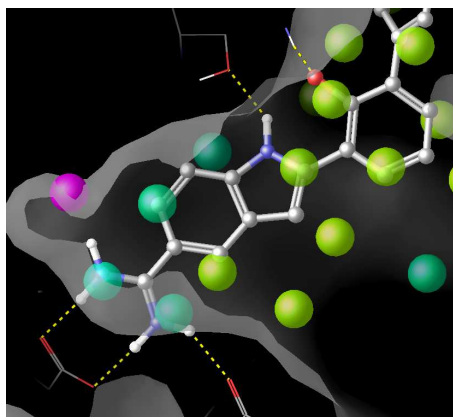


SCHRÖDINGER.

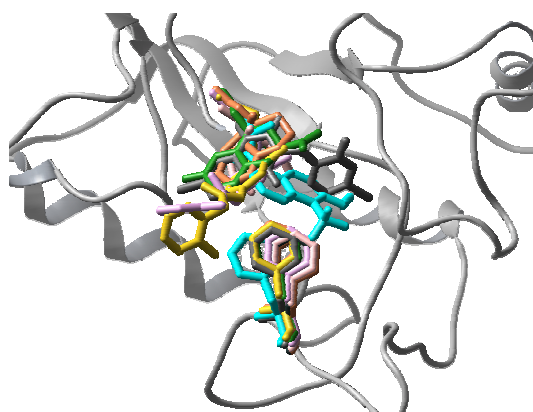
Overview

- Scope of the field
 - What we can and cannot do
 - What makes the hard things hard
- Examples of successes in SBDD
 - Docking and scoring
- Recent advances
 - Induced fit
 - Molecular dynamics
 - Structure-based ADME-Tox calculations
 - hERG
 - P450 site of metabolism predictions
 - Accounting for explicit waters
 - Free energy perturbation
 - Force field development

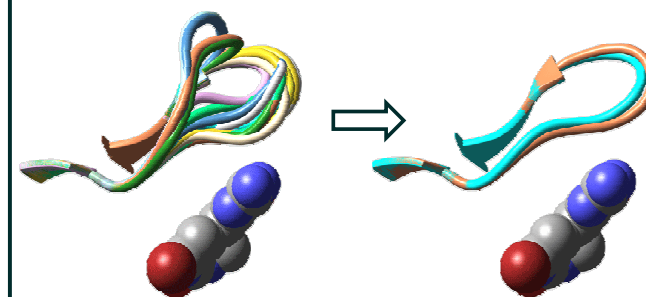
Visualization and Analysis



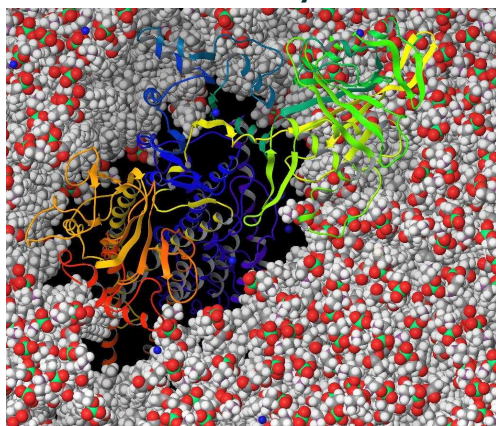
Docking



Protein Structure Prediction

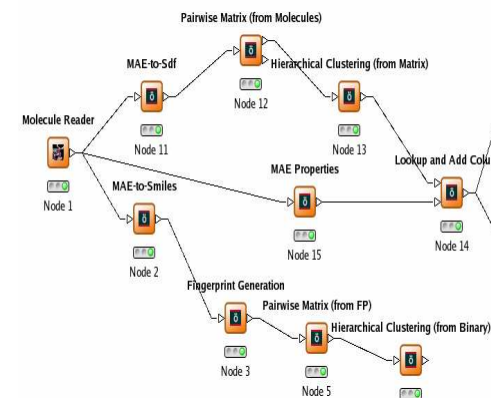


Molecular Dynamics

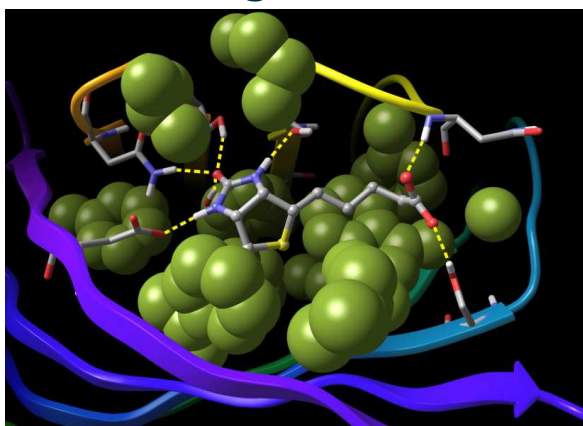


Core Technologies in Computer-Aided Drug Design

Workflows



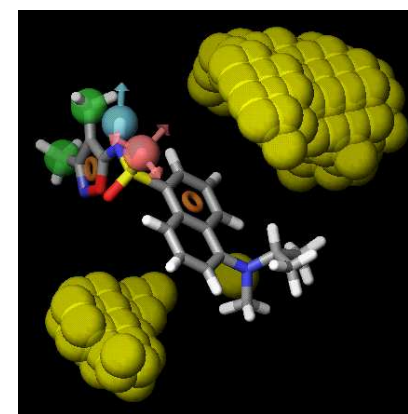
Scoring Functions



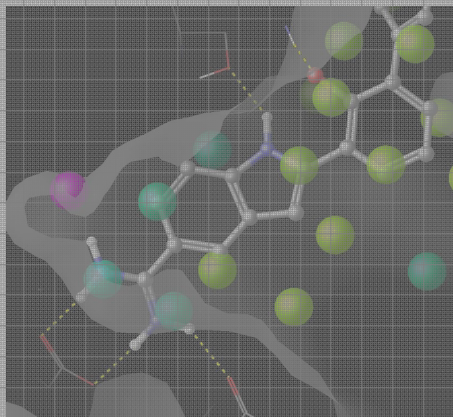
Cheminformatics

	2	4	7	8	1	3	5	6
4	0	0	0	0	1	1	1	1
6	0	0	0	0	1	1	1	1
9	0	0	0	0	1	1	1	1
5	0	0	1	0	0	1	1	1
3	0	0	0	0	1	1	0	0
7	1	1	1	1	0	0	0	0
8	1	1	1	1	0	0	0	0
2	1	1	1	1	0	0	1	0
10	1	1	1	1	1	0	1	0
1	1	1	1	0	0	0	0	1

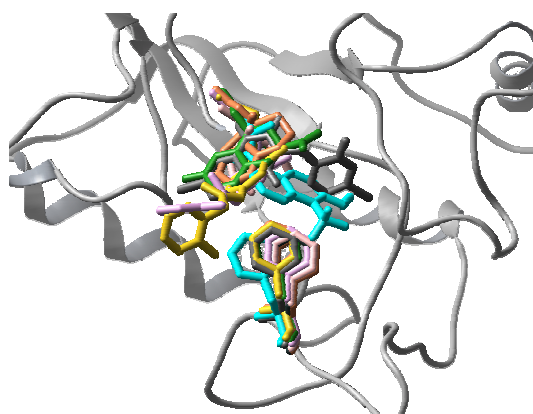
Pharmacophore Modeling



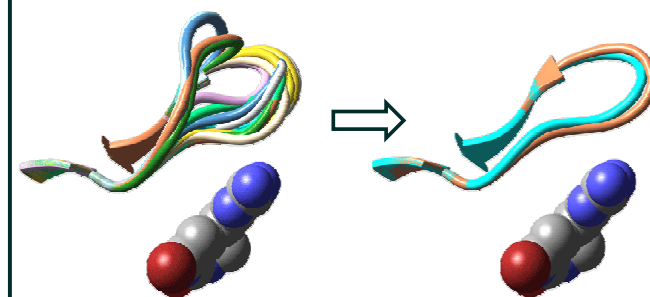
Visualization and Analysis



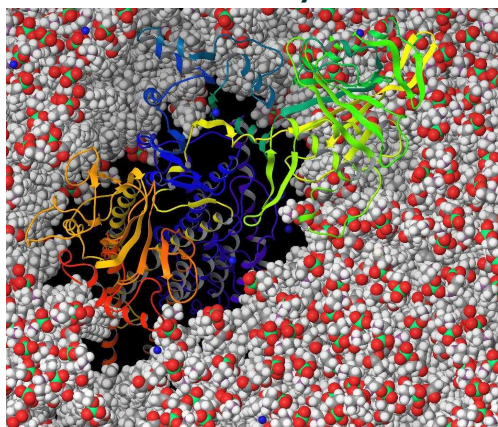
Docking



Protein Structure Prediction

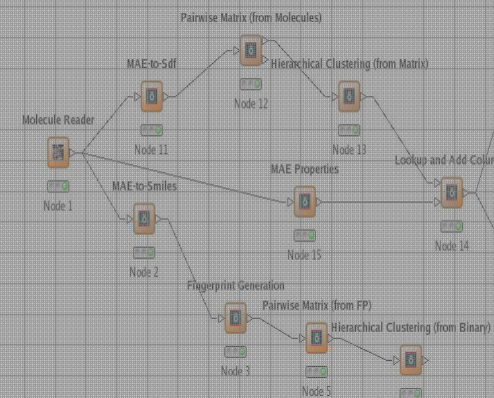


Molecular Dynamics

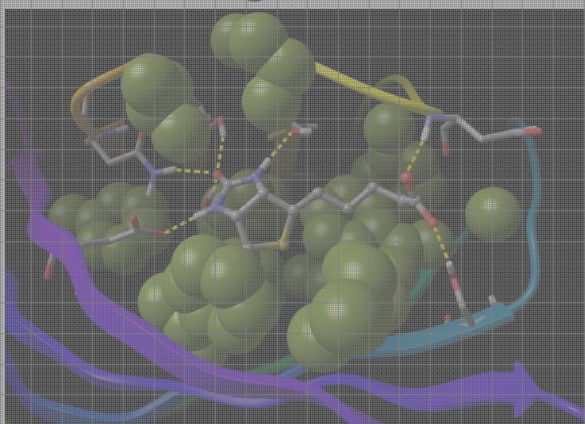


Core Technologies in Computer-Aided Drug Design

Workflows



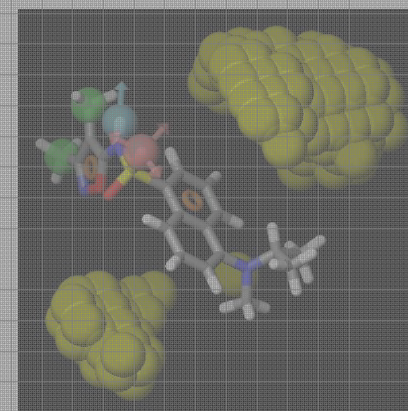
Scoring Functions



Cheminformatics

	2	4	7	8	1	3	5	6
4	0	0	0	0	1	1	1	1
6	0	0	0	0	1	1	1	1
9	0	0	0	0	1	1	1	1
5	0	0	1	0	0	1	1	1
3	0	0	0	0	1	1	0	0
7	1	1	1	1	0	0	0	0
8	1	1	1	1	0	0	0	0
2	1	1	1	1	0	0	1	0
10	1	1	1	1	1	0	1	0
1	1	1	1	0	0	0	0	1

Pharmacophore Modeling



What We Can and Cannot Do

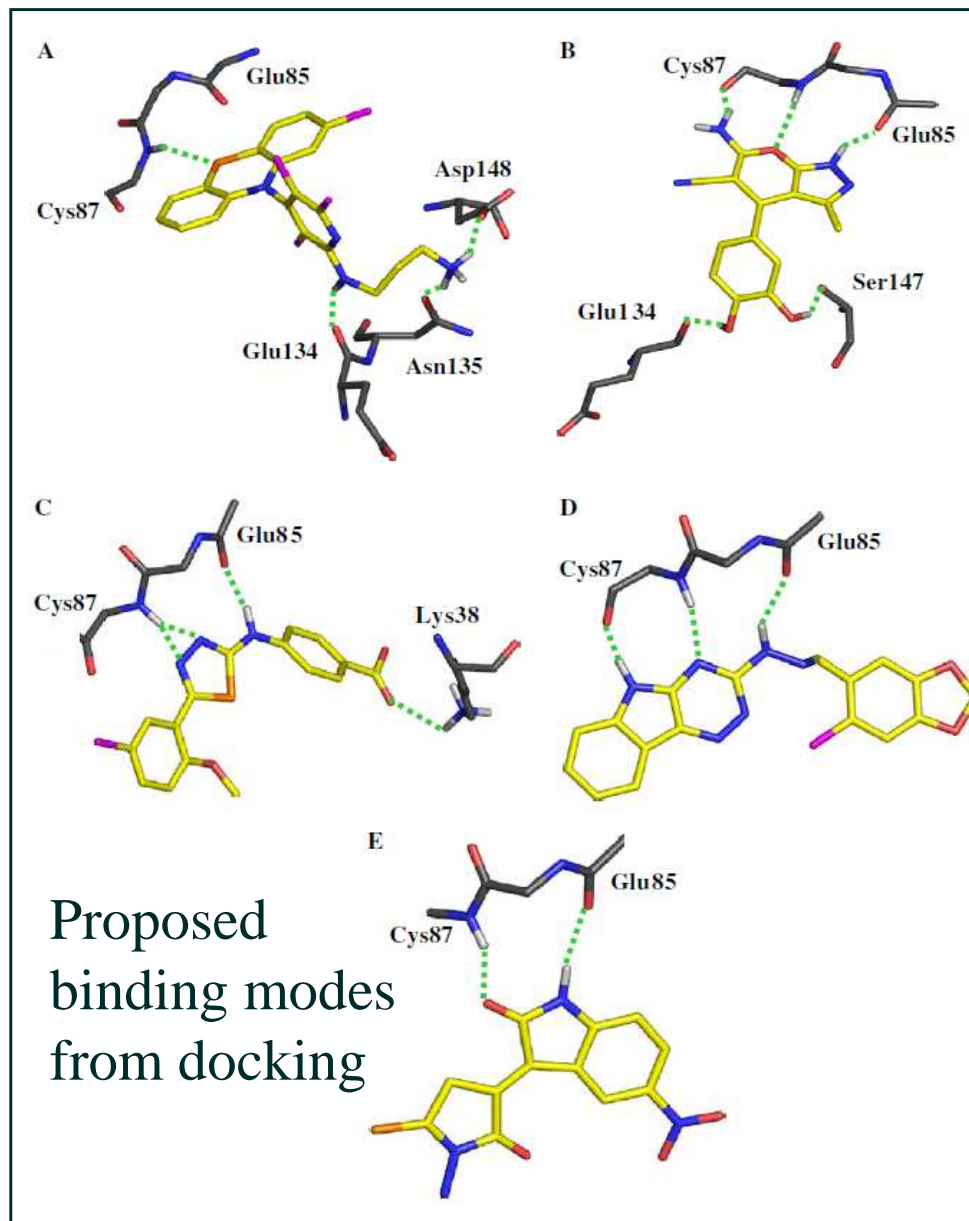
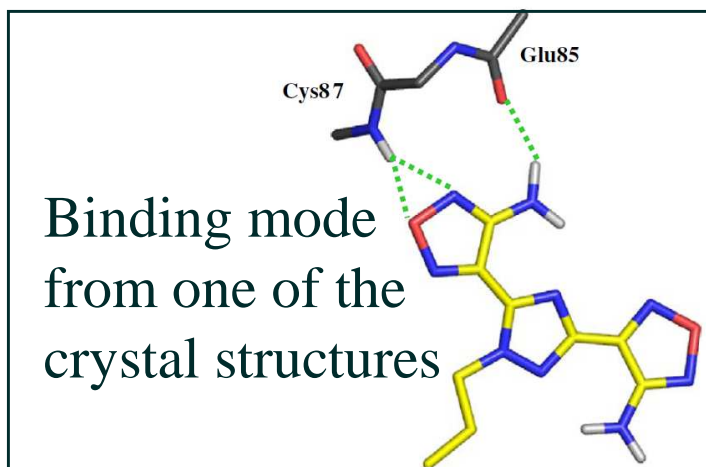
- Routine
 - Small molecule conformation generation and energy profiling
 - Visualizing crystal structures
 - Binding site characterization
 - Virtual screening to enrich databases for actives
 - Cheminformatics, ligand-based, and structure-based
 - Predict binding modes when receptor can be treated rigidly
- Difficult
 - Separating highly from weakly active compounds
 - Predicting side chain rearrangements and backbone relaxation
- Very Challenging
 - Predicting binding free energies
 - Predicting large scale protein movements
 - Mapping free energy surfaces
 - Understanding off-target effects
 - Other ADME-Tox

What Makes the Difficult Things Difficult?

- Force fields are approximate
 - Quantum mechanics would be better, but is too computationally expensive for most tasks
- Conformational sampling can be limiting
 - Typical drug like molecules can have many thousands of local minima that must be evaluated
 - Proteins have a significantly larger accessible conformational space
- The solution
 - Focus on specific problems
 - Know the limits of your method
 - Keep up with current methods
 - Methods are always improving
 - New resources can make old problems accessible
 - Cloud Computing
 - GPGPU

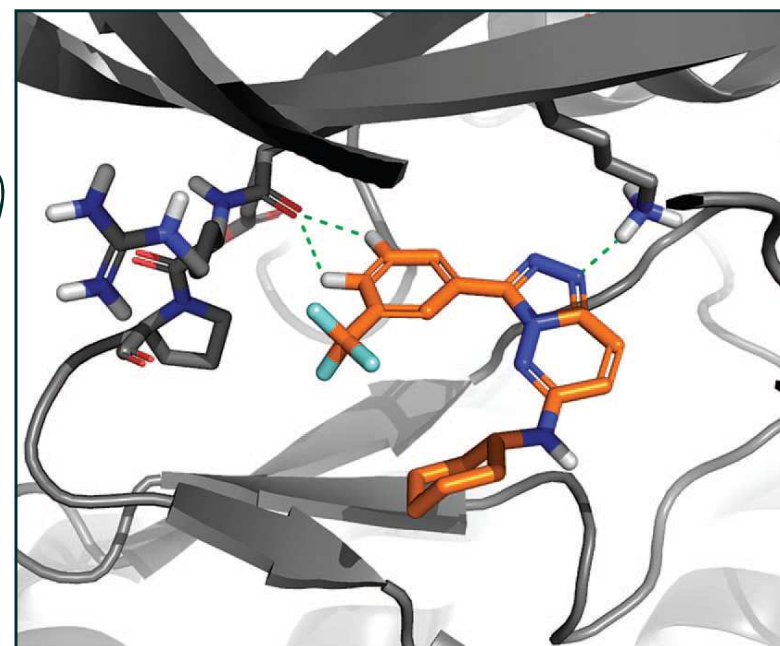
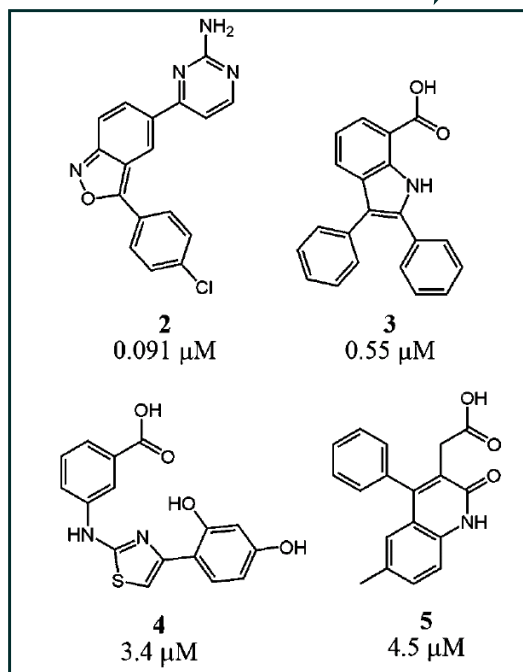
Structure-based Virtual Screening Example 1

- Researchers at Vernalis used docking to screen commercially available compounds; found 10 novel inhibitors to Chk1 kinase
- Novel hinge interaction motifs were discovered
- Crystal structures were obtained for 4 inhibitors
 - The others were docked



Structure-based Virtual Screening Example 2

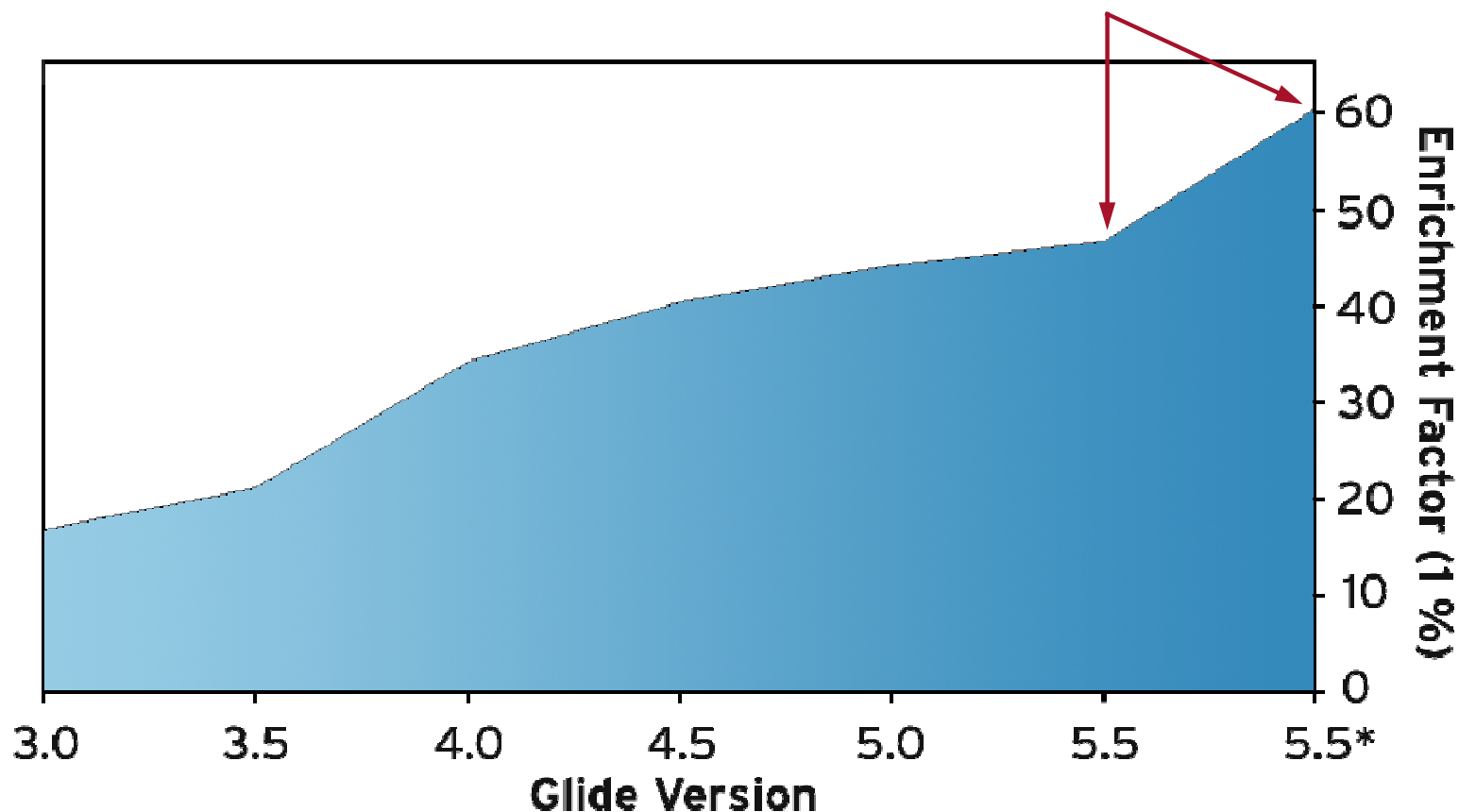
- Researchers at Vertex used docking to supplement experimental HTS and found 4 novel hits for Pim-1 kinase
- Used special aromatic CH••O hydrogen-bond constraint to the hinge
- Enrichment of actives 14x over HTS



Crystal structure used for docking
(PDB code 3BGQ)

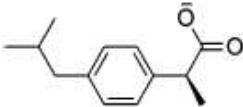
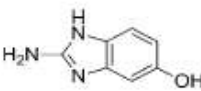
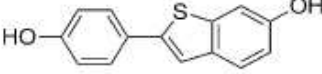
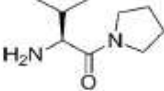
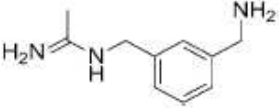
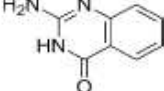
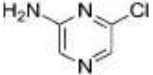
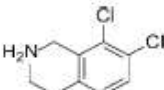
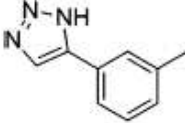
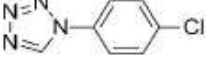
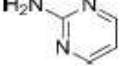
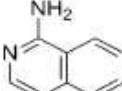
Glide Enrichments – Including Epik State Penalty

- Simply including many states degrades enrichments
- Need energetic penalty of the ionization/tautomer states
- Using the state penalty improves enrichments

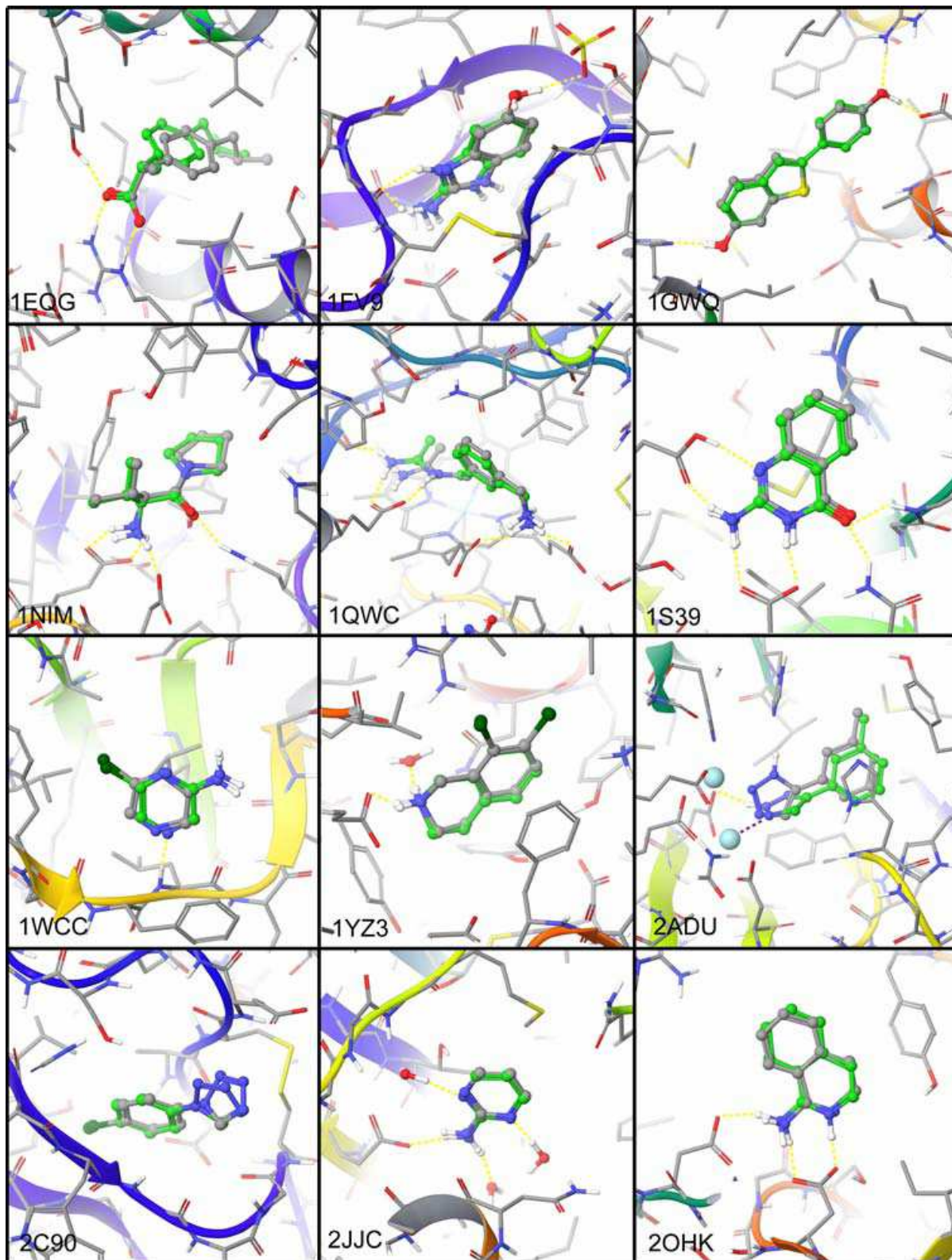


Fragment Docking

- Docking can generate accurate poses for fragments
- 12 cases^[1]
 - Maximum RMSD 1.3 Å
 - Most cases less than 0.5 Å RMSD
 - Accounting for tautomer/ionization state energies is key
- Loving K, et al., *J Comput Aided Mol Des* **2009** 23:541–554
- Cross docking can be considerably more challenging due to induced-fit, but we are making progress
- More data is needed

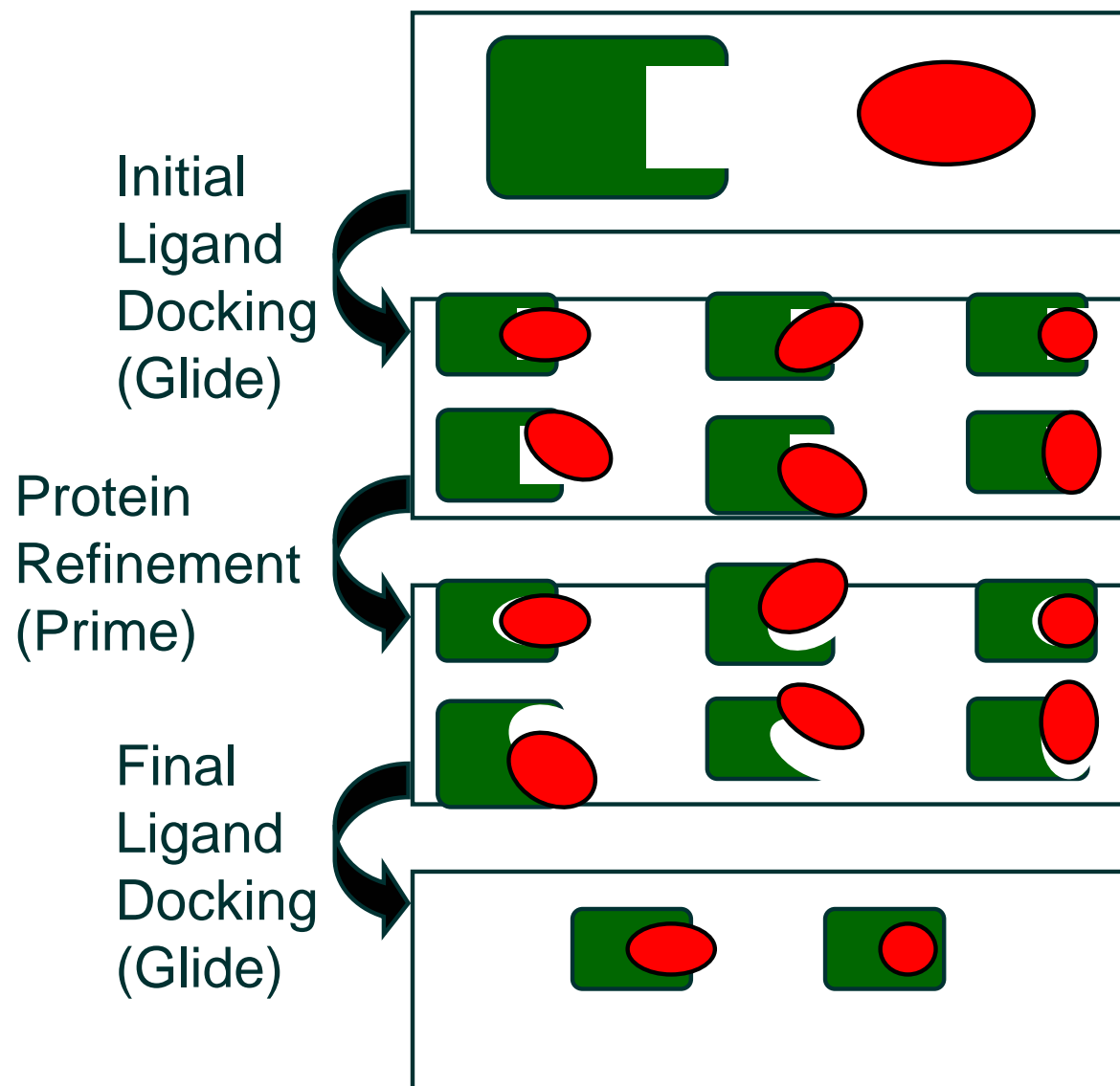
Fragment	PDB	Target	Chain	RMSD, default protocol
	1EQG	Cox-1	A B	0.67 1.09
	1FV9	Urokinase	A	0.54
	1GWQ	Estrogen receptor	A B	0.36 0.31
	1N1M	Dipeptidyl peptidase	A B	0.60 0.51
	1QWC	NO synthase oxygenase	A	0.47
	1S39	TGT	A	0.31
	1WCC	CDK2	A	0.37
	1YZ3	N-methyl-transferase	A B	0.26 0.43
	2ADU	Amino peptidase	A	0.98
	2C90	Thrombin	A	0.50
	2JJC	HSP90	A	0.20
	2OHK	Beta-secretase	A	0.31

[1] Congreve M, et al. *J Med Chem* **2008** 51:3661



Loving K, et al., *J Comput Aided Mol Des* **2009** 23:541–554

Induced Fit Docking



Sherman, "Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects", *J. Med. Chem.* 49, 2006 534-553.

Induced Fit Docking: Performance

Target	Receptor	Ligand From:	Ligand RMSD (Å)	
			Rigid Receptor Docking	Induced Fit Docking
Aldose Reductase	2acr:_	1ah3	6.5	0.9
Antibody DB3	1dba:H	1dbb	7.6	0.3
CDK2	1dm2:A	1aq1	6.2	0.8
CDK2	1aq1:_	1dm2	0.6	0.8
CDK2	1buh:A	1dm2	6.4	1.1
COX-2	3pgh:A	1cx2	11.1	1.0
COX-2	1cx2:A	3pgh	6.6	1.0 (0.5 ¹)
Estrogen Receptor	3ert:A	1err	2.3	1.4 (1.0 ¹)
Estrogen Receptor	1err:A	3ert	5.3	1.0
Factor Xa	1ksn:A	1xka	9.3	1.5
Factor Xa ¹	1xka:C	1ksn	5.3	1.5
HIV-RT	1rth:A	1c1c	2.5	1.3
HIV-RT	1c1c:A	1rth	12.0	2.5
Neuraminidase	1nsc:A	1a4q	3.9	0.8
Neuraminidase	1a4q:A	1nsc	1.0	1.7
PPAR- γ	2prg:A	1fm9	9.8	3.0 (1.5 ²)
PPAR- γ	1fm9:D	2prg	9.1	1.8 (0.4 ³)
Thermolysin	1kr6:A	1kjo	1.1	1.3
Thermolysin	1kjo:A	1kr6	3.5	3.2 (1.6 ⁴)
Thymidine Kinase	1kim:A	1ki4	4.7	0.4
Thymidine Kinase	1ki4:A	1kim	0.5	1.2

- Average ligand RMSD for docking to a flexible receptor for the 21 pairs is 1.4 Å
- RMSD ≤ 1.8 Å for 18 cases
- For the 3 cases with RMSD > 1.8 Å, the core of the ligand is properly docked and all key protein/ligand interactions are captured
- Still, a substantially larger validation set is needed

2nd most cited J Med Chem publication from 2006 (>150 citations, meaning this is working in the real world)

¹RMSD of 2nd ranked IFD structure that has nearly identical composite score as top ranked structure

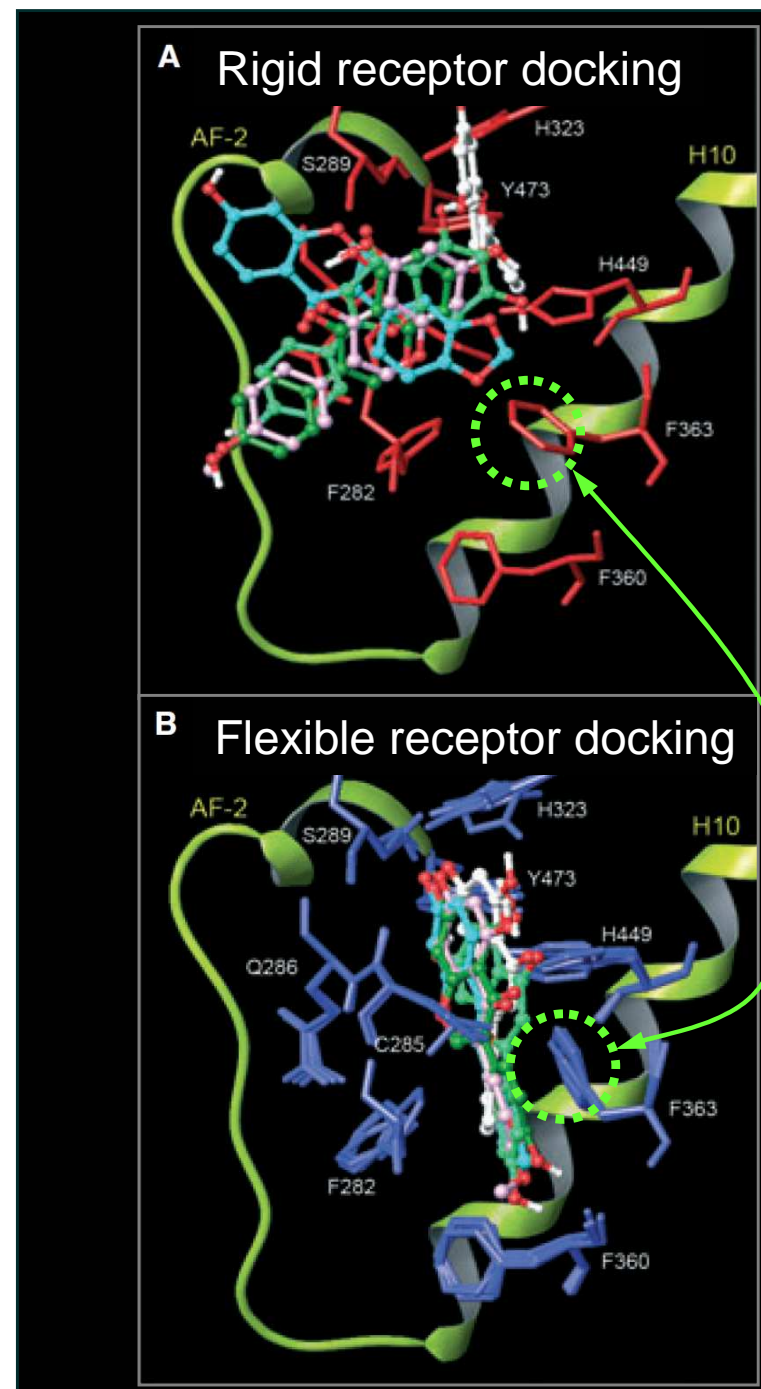
²RMSD excluding 13 atoms in solvent exposed methylphenyloxazole tail of the ligand

³RMSD excluding 10 atoms in solvent exposed methyl-2-pyridinylamino tail of the ligand

⁴RMSD excluding 6 atoms in the quasi-symmetric di-carboxylate that are flipped 180°

Induced Fit Docking Application

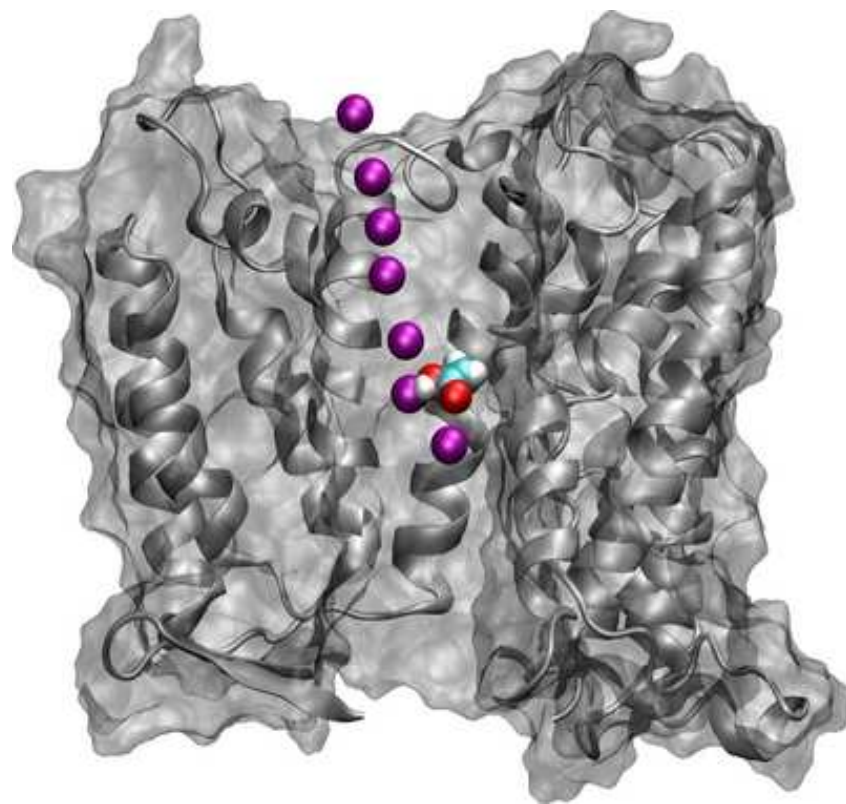
- PPAR- γ is a highly flexible target
 - See superposition of PDB structures
 - Most nuclear receptors are flexible
- Researchers at the University of Sydney identified novel PPAR- γ agonists from a natural product library
 - Flexible ligand docking to a rigid receptor of known active compounds produced inconsistent poses (see top right)
 - Receptor flexibility was required to get good and consistent poses (see bottom right)
- IFD has been used in this project to find new PPAR- γ inhibitors and novel IP



Salam, N.K., et al. Novel PPAR-gamma agonists identified from a natural product library: A virtual screening, induced-fit docking and biological assay study. *Chem Biol Drug Des* **2008** (71) 51-70

Molecular Dynamics

- Probing protein flexibility
- Generation of structural ensembles
- Visualization molecular processes
- Estimation binding energies
 - Solvation free energies
 - Binding free energies
 - Conformational free energies



G-protein coupled receptors

- Largest gene family in the human genome
- Represent the target for >30% of drugs
- Structural data has historically been scarce

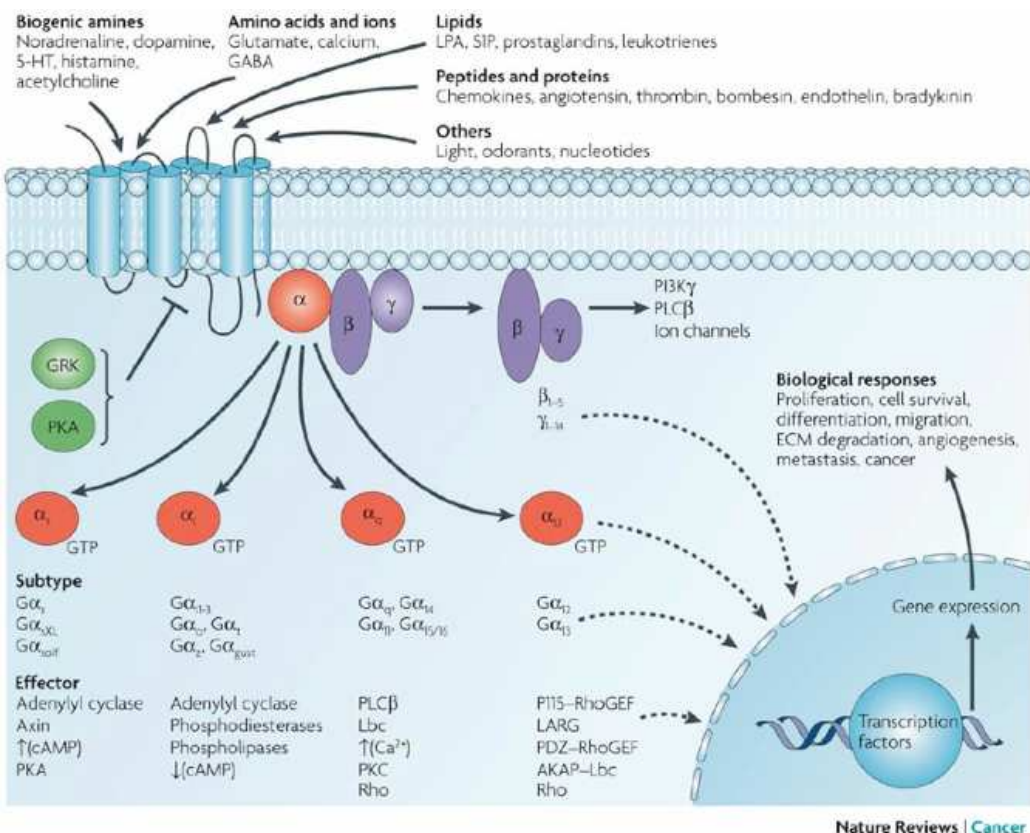


Table 2. Best-selling therapeutics targeted at GPCRs in 2000 [2]

Product	Receptor(s)	Indication	Company
Zyprexa®	Serotonin 5-HT ₂ and dopamine	Schizophrenia or antipsychotic	Eli Lilly, Indianapolis, IN, USA
Claritin®	Histamine H ₁	Rhinitis or allergy	Schering-Plough, Kenilworth, NJ, USA
Risperdal™	Serotonin 5-HT ₂	Schizophrenia	Johnson & Johnson, Titusville, NJ, USA
Imigran™	Serotonin 5-HT _{1B/1D}	Migraine	GlaxoSmithKline, Harlow, UK
Cozaar®	Angiotensin AT ₂	Hypertension	Merck and Co., Whitehouse Station, NJ, USA
Serevent®	β ₂ -adrenoceptor	Asthma	GlaxoSmithKline
Singulair®	BLT ₁	Asthma	Merck and Co.
Gastridin™	Histamine H ₂	Peptic ulcer	Merck and Co.
Zantac/Tagamet™	Histamine H ₂	Peptic ulcer	GlaxoSmithKline
Zirtec™	Histamine H ₁	Rhinitis or allergy	Pfizer, Sandwich, UK
BuSpar®	Serotonin 5-HT _{1a}	Anti-depressant	BMS, New York, NY, USA
Gaster®	Histamine H ₂	Peptic ulcer	Yamanouchi, Tokyo, Japan

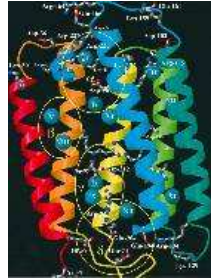
G-protein-coupled receptors and cancer. Robert T. Dorsam and J. Silvio Gutkind. Nature Reviews Cancer **2007** 7, 79-94

Target validation of G-protein coupled receptors. Wise A, Gearing K, Rees S. Drug Discov Today. **2002** Feb 15;7(4):235-46.

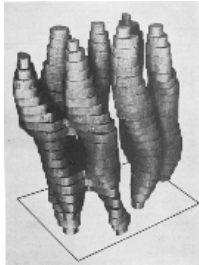
Milestones in structure determination



Henderson et al., Model for the structure of **bacteriorhodopsin** based on high-resolution **electron cryo-microscopy** (3.5Å). J Mol Biol. 1990



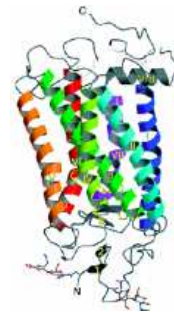
Grigorieff et al., Electron-crystallographic refinement of the structure of **bacteriorhodopsin**. J Mol Biol. 1996



Henderson et al., . Three-dimensional model of **purple membrane** obtained by **electron microscopy**. Nature. 1975.



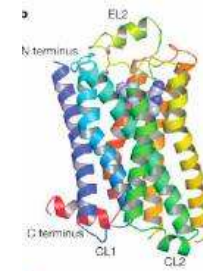
Unger et al., Arrangement of **rhodopsin** transmembrane alpha-helices. Nature. 1997



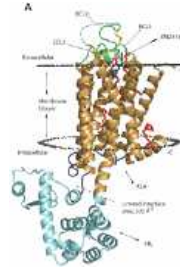
Palczewski et al., **Crystal structure of rhodopsin**: A G protein-coupled receptor. Science. 2000



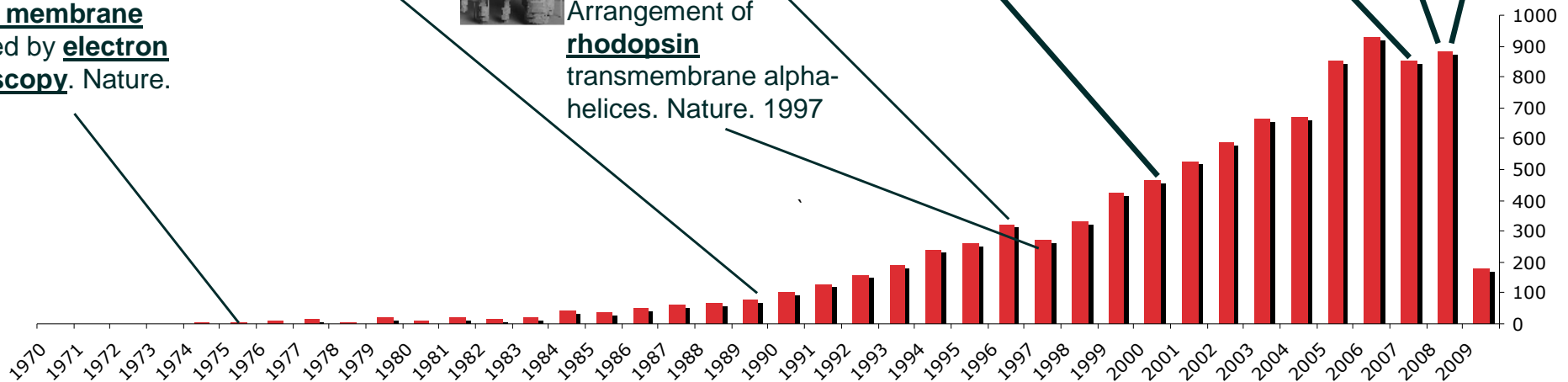
Cherezov et al., High-resolution **crystal structure** of an engineered human **beta2-adrenergic G protein-coupled receptor**. Science. 2007



Warne et al., Structure of a **beta1-adrenergic G-protein-coupled receptor**. Nature. 2008



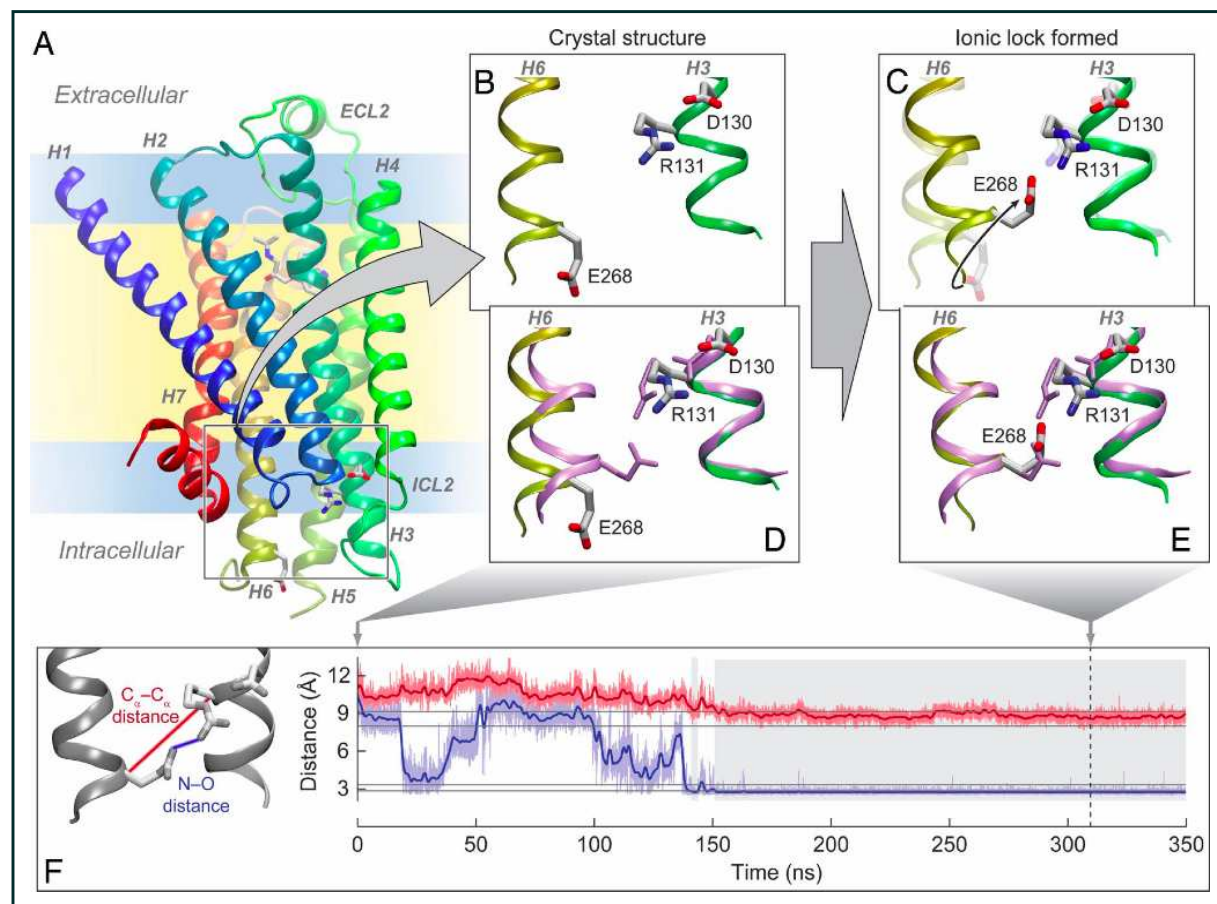
Jaakola et al., The 2.6 angstrom **crystal structure** of a human **A2A adenosine receptor** bound to an antagonist. Science. 2008



PubMed search: molecular model G-protein coupled receptor

Probing Flexibility in β_2 Adrenergic Receptor (B2AR) GPCR

- Scientists at D. E. Shaw Research ran a series of long simulations to understand the ionic lock between helices 3 & 6 in B2AR
- While these simulations were relatively long, it is possible to gain insights using shorter simulations or by taking advantage of enhanced sampling techniques
- Note: Simulations can be tricky and require some degree of user expertise

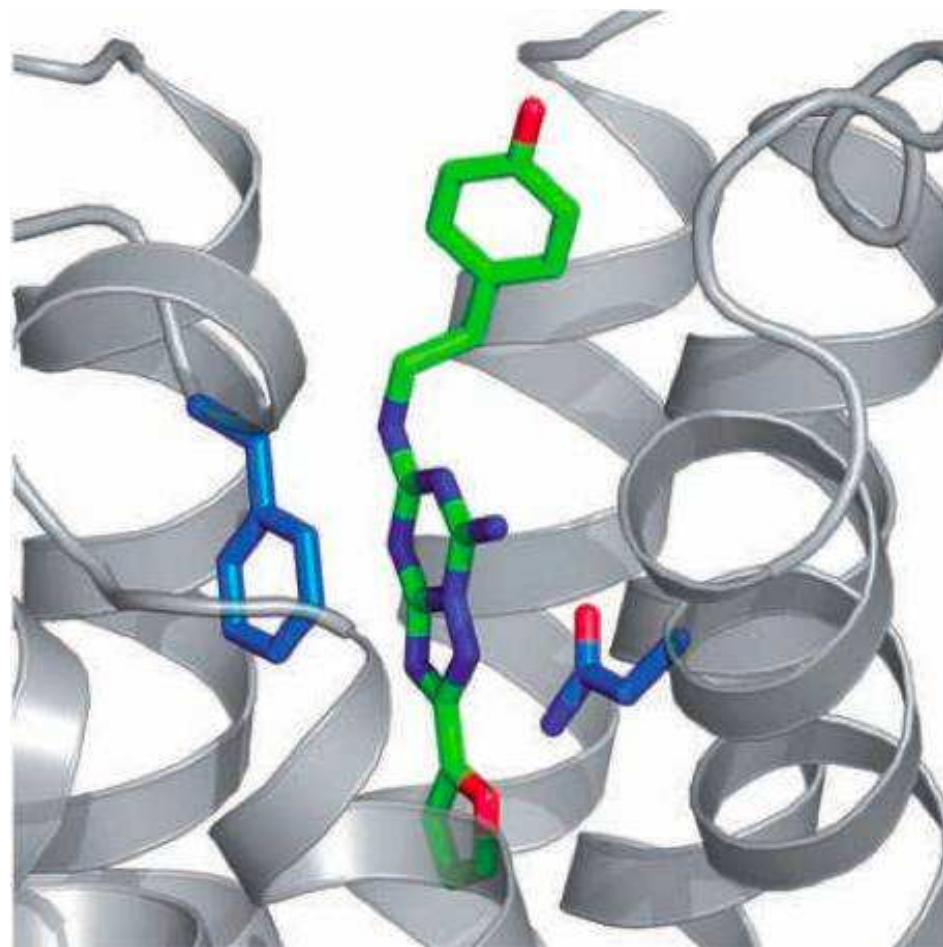


Dror, R.O., et al. Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proc Natl Acad Sci USA*, **2009** (106) 4689-4694

Induced Fit Docking in GPCR Dock 2008

Application to the A2A receptor

“With the aim of evaluating the current status of GPCR structure prediction and ligand docking, a community-wide, blind prediction assessment — GPCR Dock 2008 — was conducted in coordination with the publication of the crystal structure of the human adenosine A2A receptor bound to the ligand ZM241385. Twenty-nine groups submitted 206 structural models before the release of the experimental structure, which were evaluated for the accuracy of the ligand binding mode and the overall receptor model compared with the crystal structure.”



Michino et al., Nat. Rev. Drug Disc., 2009

Software used: Autodock, DOCKER, Fred, Glide, Gold, ICM, IFD, Membstruck, Modeller, Moe, Q-dock, Rosetta, Tasser

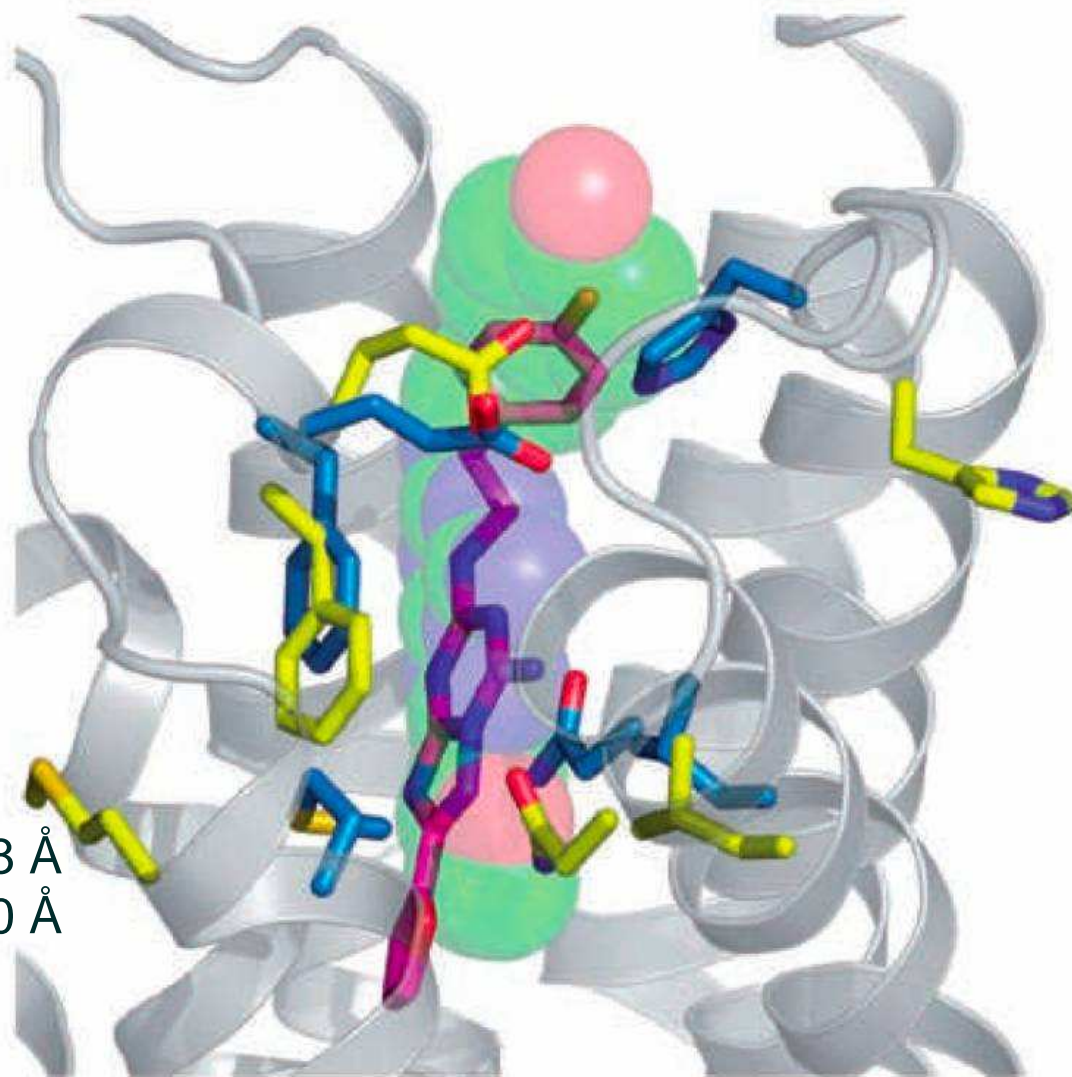
SCHRÖDINGER.

GPCR Dock 2008: Results

Best model submitted by Stefano Costanzi (NIH)

“The molecular docking of ZM241385 at the homology model of the A2A receptor has been conducted with the Schrödinger package. The ligand was sketched and subsequently docked at the receptor by means of the *“Induced Fit”* docking procedure, which combines automatic docking of flexible ligands to a rigid representation of the receptor (*Glide*) with reconstructions of the receptor binding site around the docked ligands (*Prime*).”

Ligand RMSD 2.8 Å
Protein RMSD 3.0 Å



CPK: xray ligand, stick: ligand model
Blue: xray receptor, yellow: receptor model

SCHRÖDINGER.

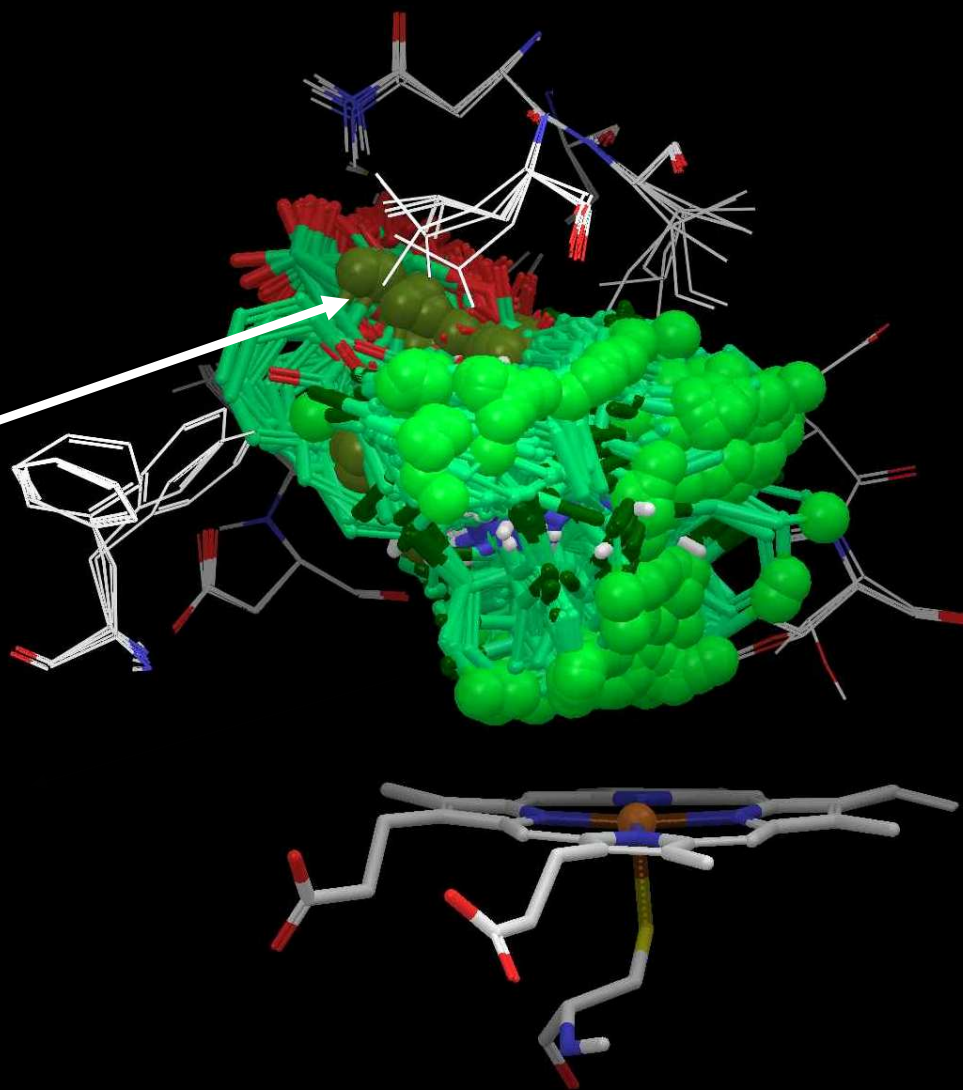
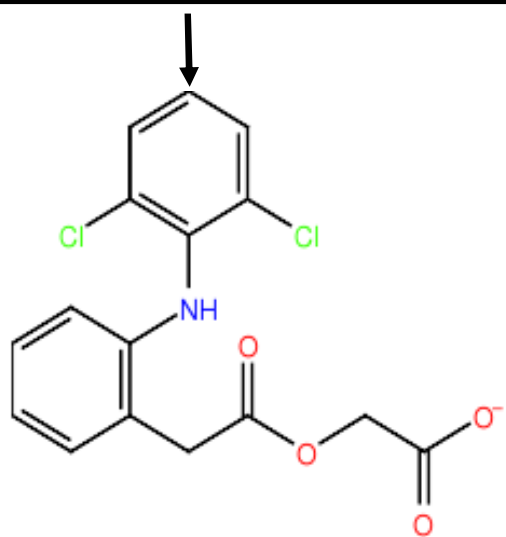
Note: GPCR Dock 2010 was recently announced. Submission deadline mid-July).
New Targets: D3 and CXCR4

Predicting P450 Sites of Metabolism

- hERG predictions were previously successful
 - Farid, R. et al. "New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies", *Bioorg. & Med. Chem.*, 2006, 14, 3160-3173
 - Induced fit structure available for download
- Goal:
 - Method to accurately predict all sites of metabolism
 - Also get poses, to help with design
 - Can account for difficult cases, like stereoisomers
- Methods:
 - Structure-based combined with intrinsic reactivity
 - If an atom has high intrinsic reactivity and can get an atom close to the reactive iron, then it is considered to be a site of metabolism
 - More sophisticated free energy-based scoring is being investigated
- Next steps
 - Reactivity rates
 - Inhibition

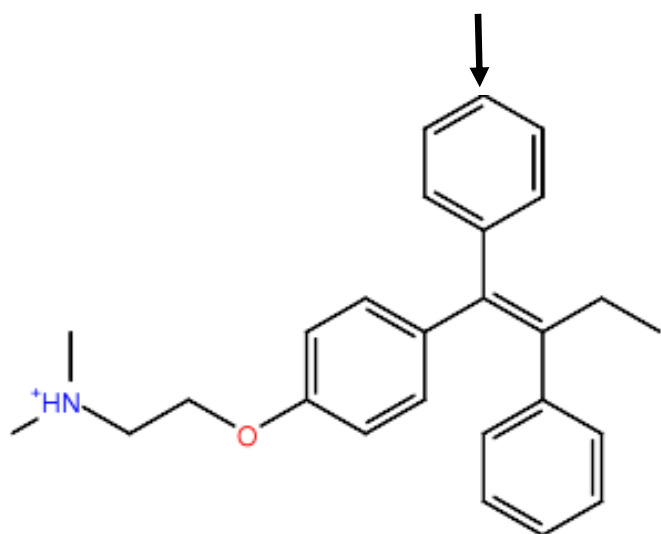
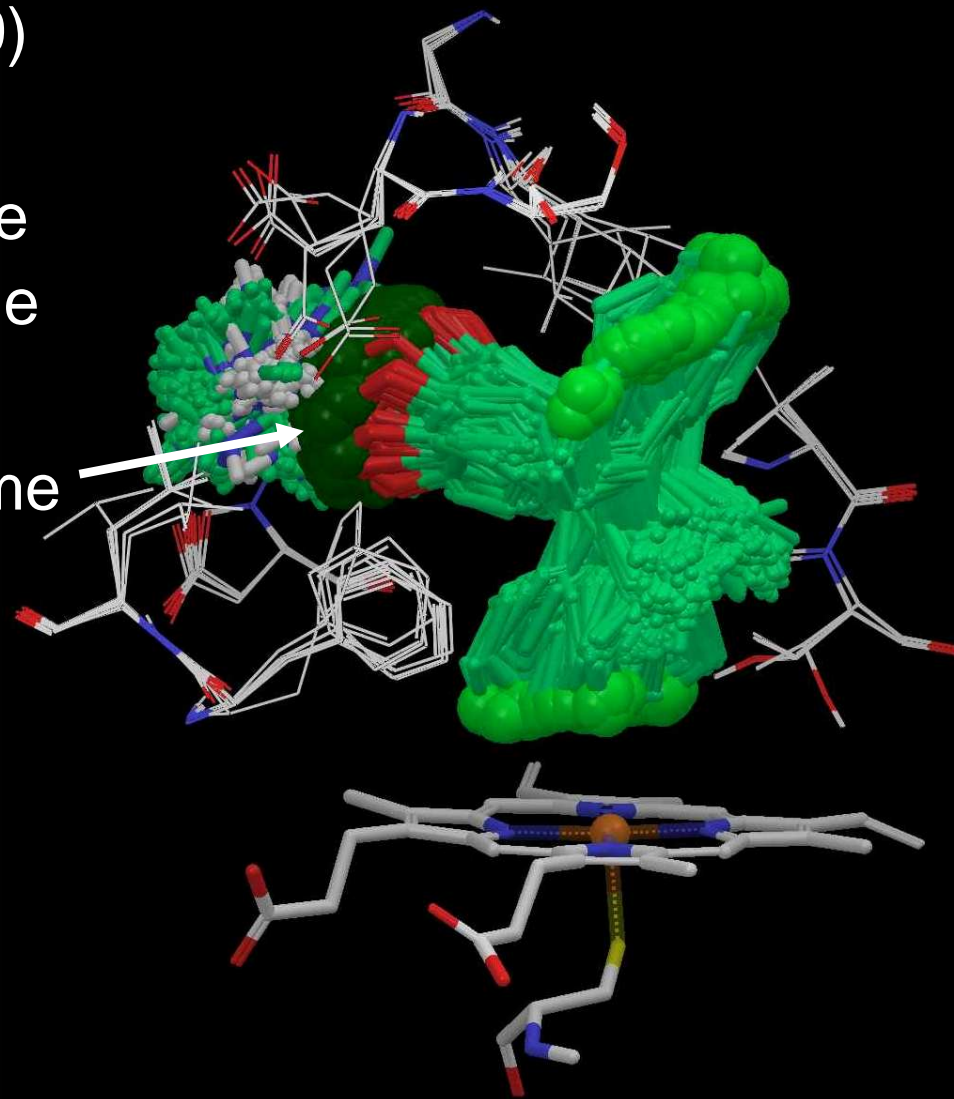
Aceclofenac SOM in CYP2C9 – Value of IFD

- Highlighted atom in light green is SOM, but has poor reactivity (2.0) vs non-SOM in dark green (-3.6)
- Many IFD derived poses have the SOM close to the heme. It has the best overall score
- The non-SOM is far from the heme in all poses



Tamoxifen SOM in CYP2D6 – Value of IFD

- Highlighted atom in light green is SOM, but has poor reactivity (0.9) vs non-SOM in dark green (-5.5)
- Many IFD derived poses have the SOM close to the heme. It has the 2nd best overall score
- The non-SOM is far from the heme in all poses

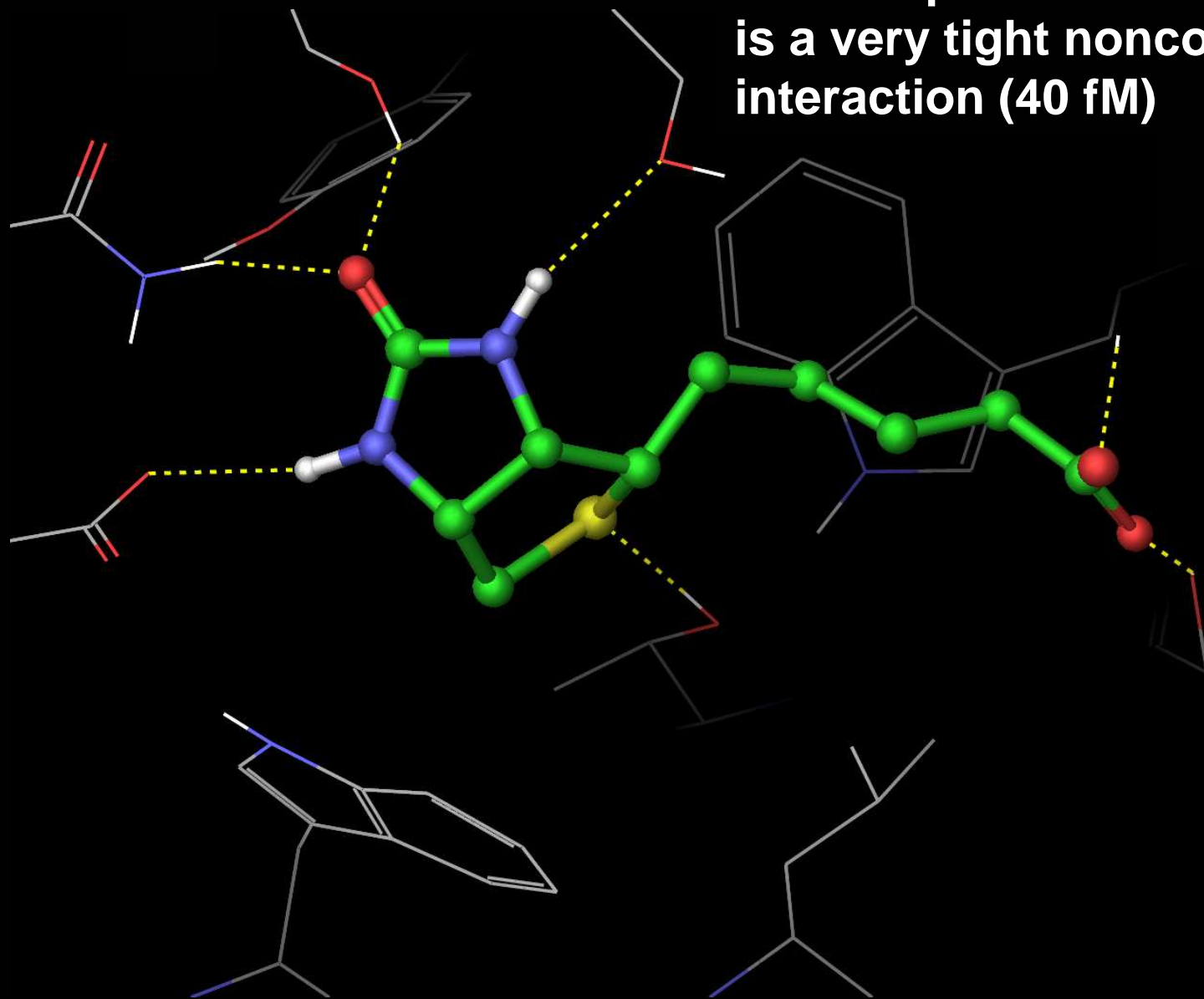


The Role of Waters in Drug Discovery

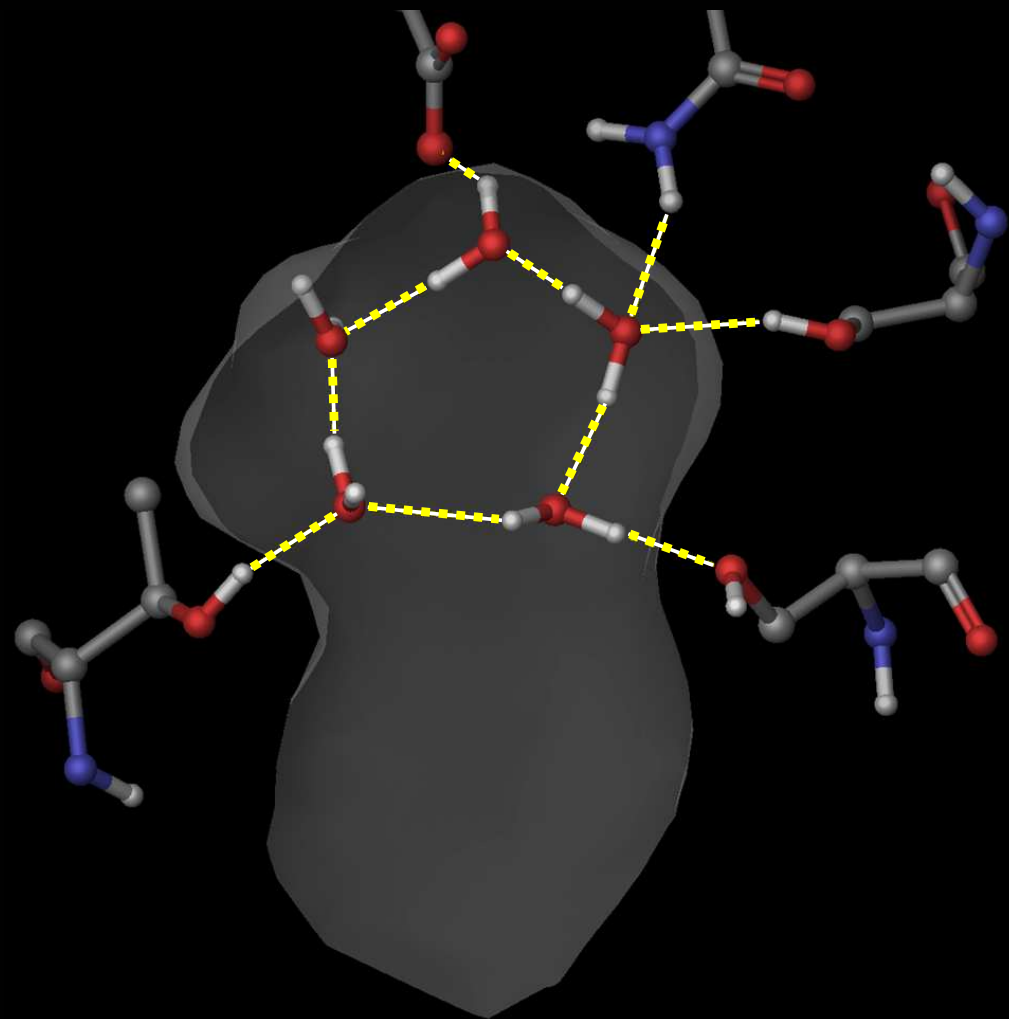
- “Hydrophobic effect” is driven by release of water molecules
- Explicit nature of water is essential to describe water behavior
- Water molecules have a great deal of mobility, making them ideal chemical probes with real physical implications
 - Water molecules are competing with ligands
 - We can learn from water locations and interactions
- Recent advances in methods and algorithms have allowed for the quantification of water molecule thermodynamics
 - Lazaridis, T. (1998) *J Phys Chem. B* 102:3531–3541
 - Inhomogeneous fluid approach to solvation thermodynamics
 - Young, T., et al. (2007) *PNAS* 104:808-813
 - Initial validation to describe hydrophobic enclosure motif
 - Abel, R., et al. (2008) *J Am Chem Soc* 130:2817-2831
 - Predicting affinity for congeneric pairs of factor Xa inhibitors
 - Beuming, T., et al. (2009) *Prot Sci* 18:1609-1619
 - Affinity and selectivity insights for PDZ domains
 - Robinson, D., et al. (2010) *ChemMedChem* 5:618-627
 - Kinase selectivity
 - Guimaraes, C., et al. (2010) *J Chem Inf Model* 50:547–559
 - Scoring of factor Xa and CDK2 ligands
 - Higgs, C., et al. (2010) *Med Chem Lett* (accepted)
 - Affinity predictions for a series of A2A GPCR inhibitors

What Can we Learn from Waters?

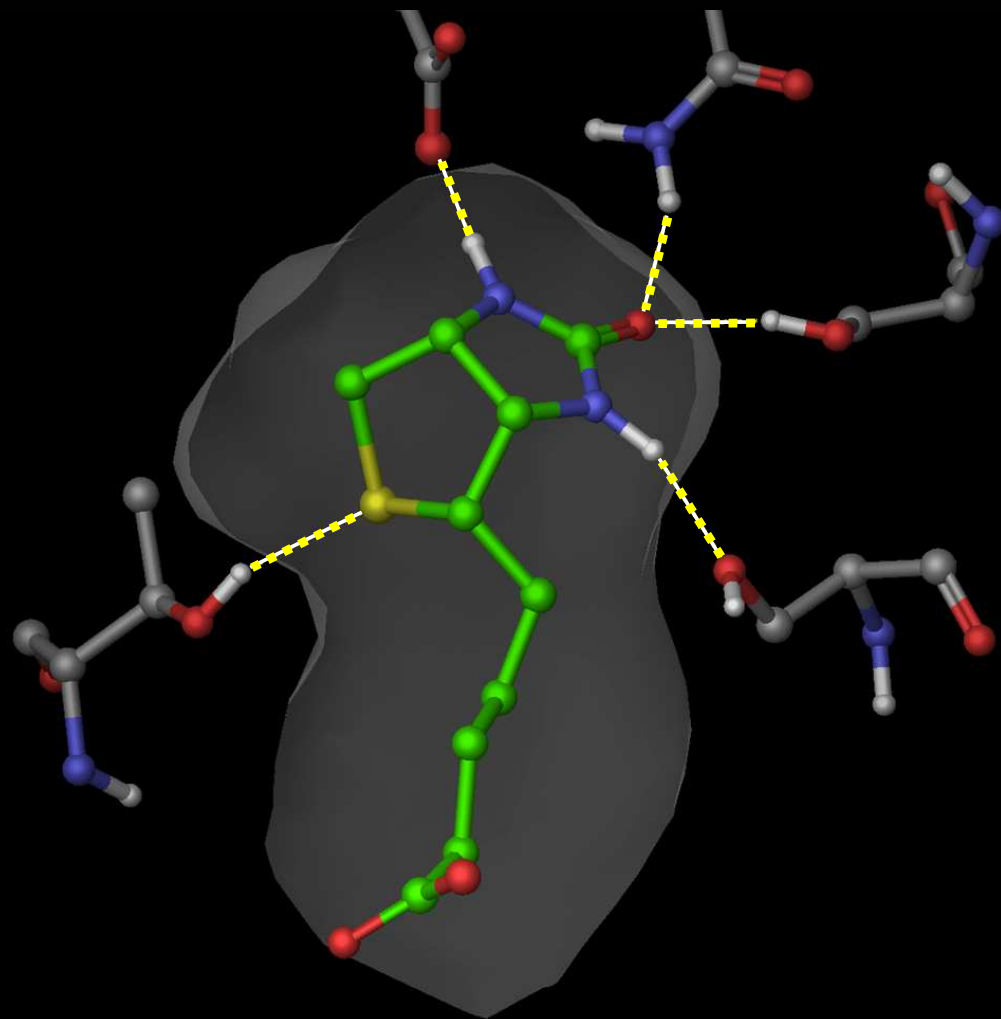
The streptavidin/biotin complex is a very tight noncovalent interaction (40 fM)



Streptavidin/biotin complex



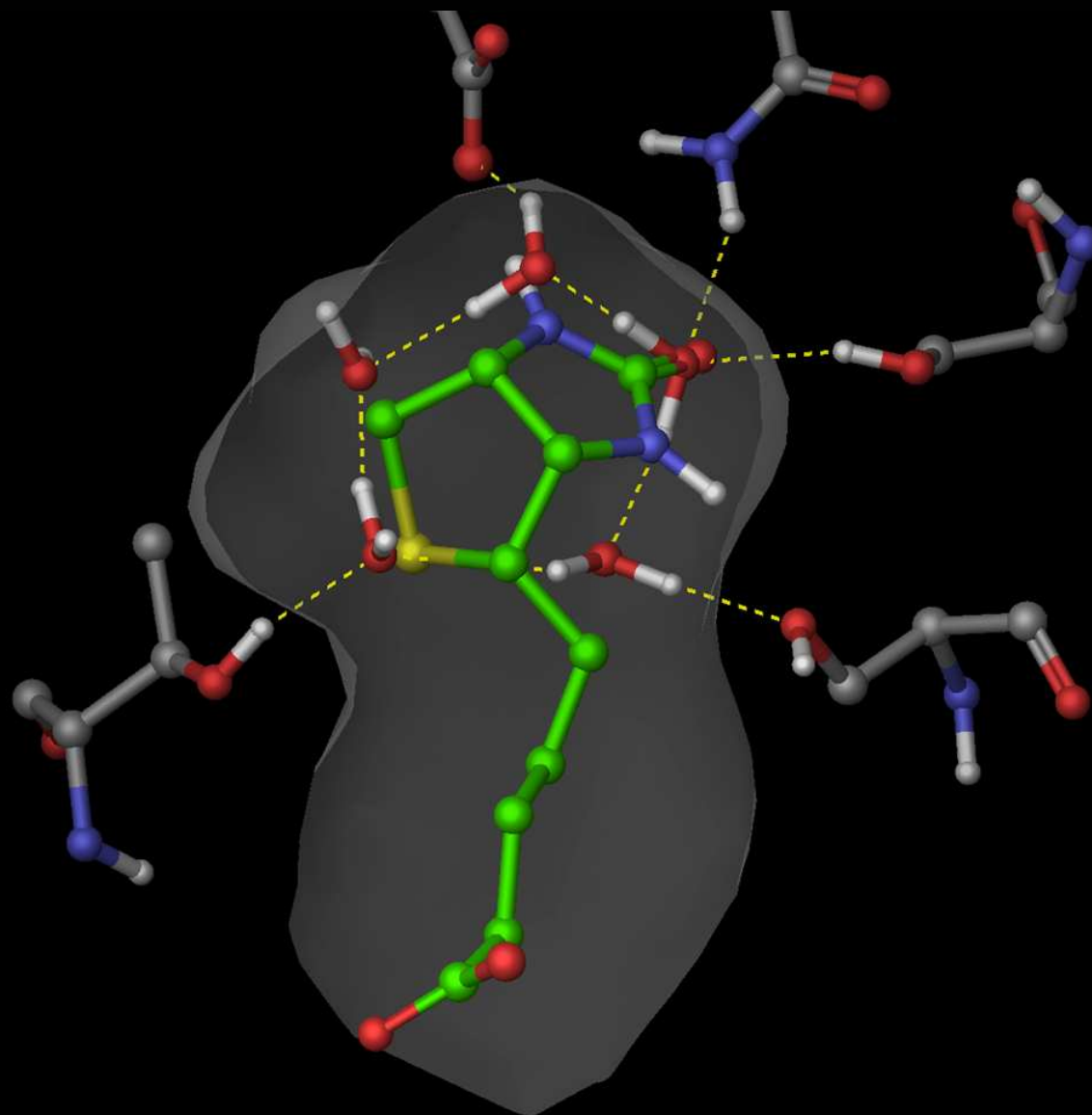
Clathrate ice-like water structure persists through an MD simulation



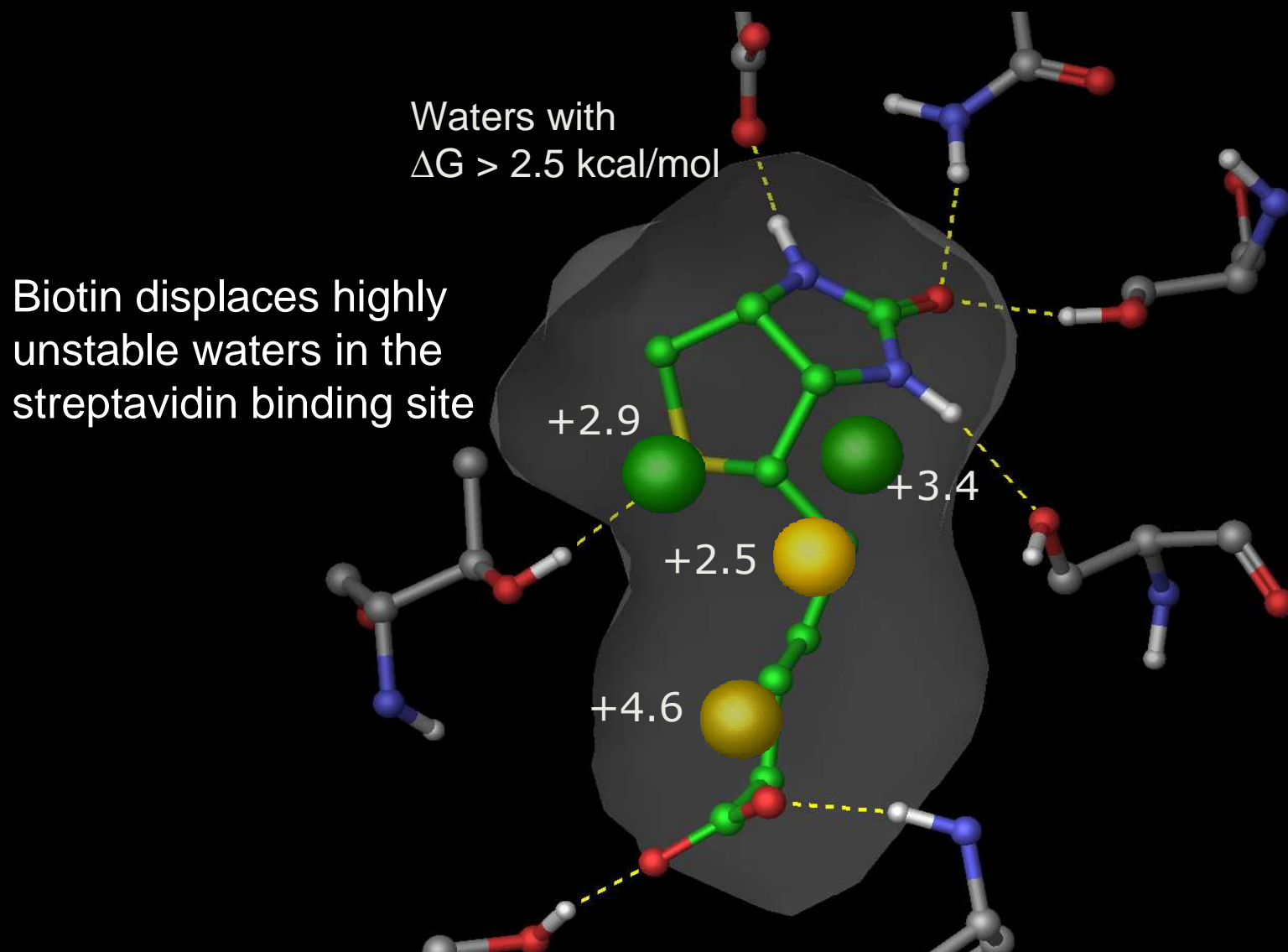
Biotin displaces these five waters and makes back the key H-bond interactions

SCHRÖDINGER.

Streptavidin/biotin complex

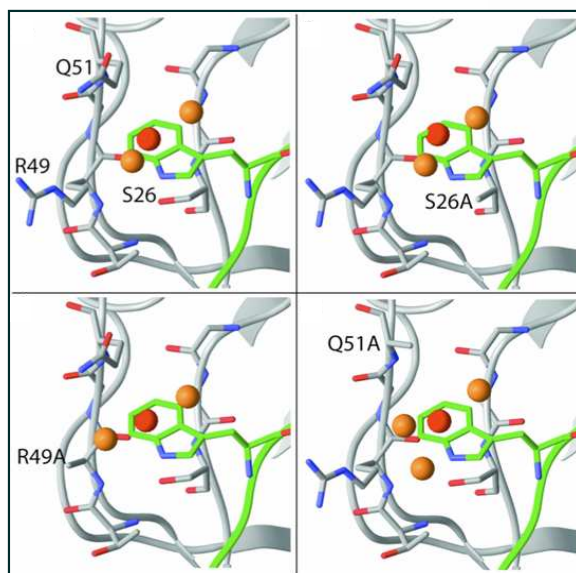
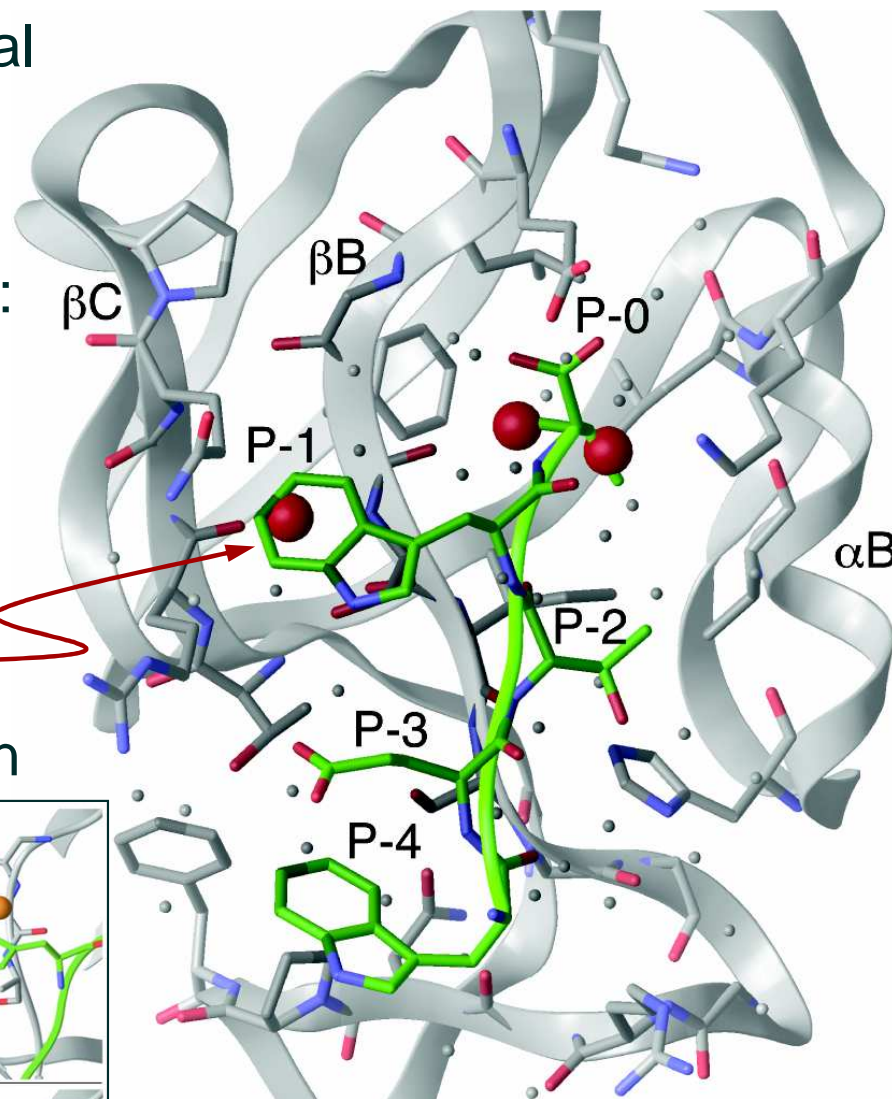


Streptavidin/biotin complex



High-energy Water Determines P-1 Affinity/Specificity in PDZ Domains

- Phage display shows WETWV is the optimal Erbin binder (Skelton et al., JBC, 2003)
- However, mutagenesis of the PDZ domain cannot explain the crucial role of Trp at P-1:
 - “curiously, alanine substitutions of residues that contact Trp-1 did not decrease peptide binding” (Skelton et al., JBC, 2003)
- High-energy water at P-1 position explains preference for Trp
- Alanine mutations in P-1 have little effect on activity



Beuming, T., et al. High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Prot Sci* **2009** 18:1609-1619

Free Energy Methods

- A long history of academic literature suggests we can accurately predict binding free energies
 - We are not really there
- FEP (and related methods) feel like they are parameter free
 - There are many parameters associated with the protocol
- We have started a large-scale relative binding free energy FEP validation study
 - 10+ people
 - 20+ targets
 - Pharmaceutically relevant perturbations
- Goals
 - Define scope of applicability and limitations
 - Understand what can be expected from FEP
 - Identify areas for improvements (force fields, sampling time, λ , etc.)

Overview of Absolute Solvation Free Energy Work

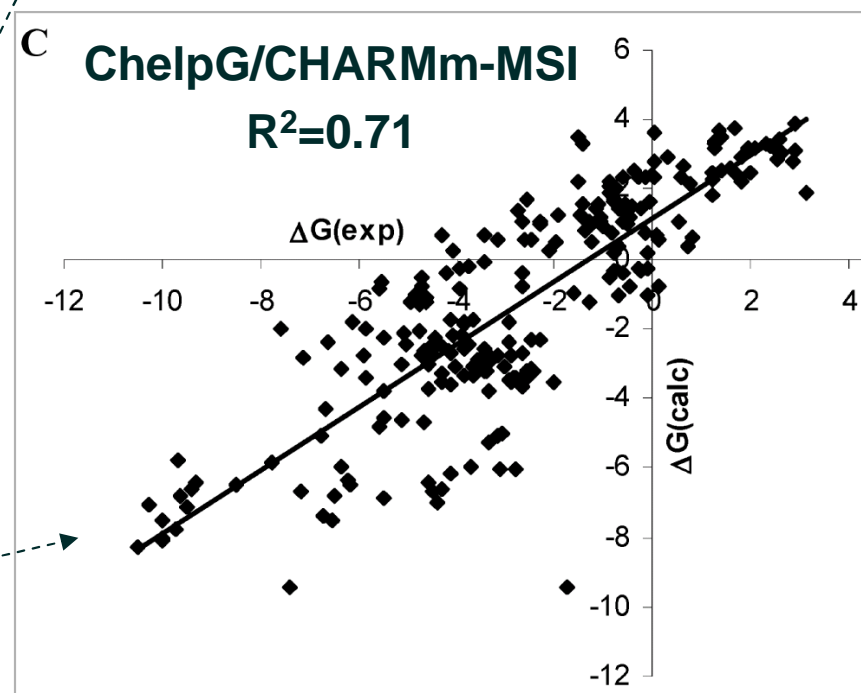
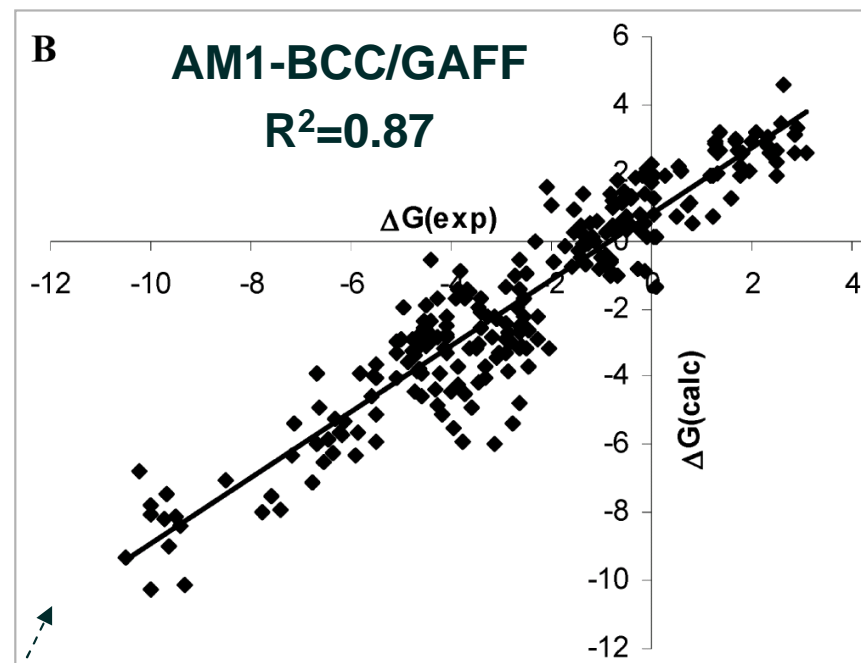
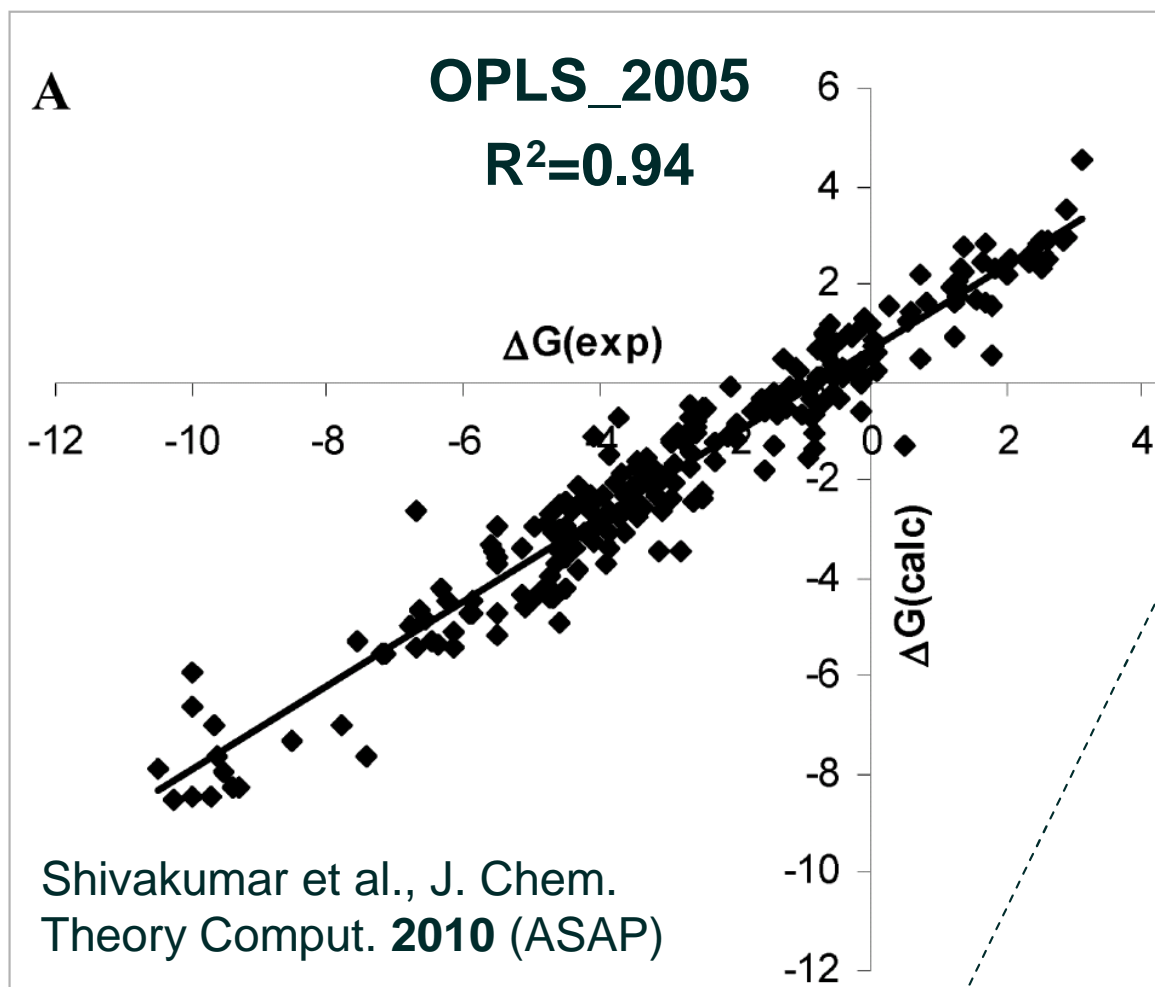
- Used existing dataset from Shivakumar et al.*
 - Allows for direct comparison with other programs and force fields
- First objective was to test Desmond FEP implementation
 - Overall results with default parameters worked well
 - OPLS_2005
 - SPC water model
 - 12 lambda windows
- Looked into sensitivity of various parameters
 - Force field
 - Solvent model
 - Charges
 - Lambda schedule
- Paper now available in JCTC
 - *J. Chem. Theory Comput.* **2010**, 6, 1509–1519
 - “Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field”

Absolute Solvation Free Energy Dataset

- Chemical functional groups and number of ligands from each

type	number	type	number
alkanes	8	esters	15
alkenes	10	ethers	11
alkynes	5	halogen, bromo	10
alcohols	17	halogen, chloroalkanes	11
aldehydes	6	halogen, chloroalkenes	5
aliphatic amines	16	halogen, chloroarenes	3
amides	5	halogen, fluoro	6
arenes	14	halogen, iodo	8
aromatic amines	14	thiols	4
bifunctional amine	3	ketones	12
bifunctional groups	5	multiple halogens	15
branched alkanes	7	nitriles	5
carboxylic acids	5	nitro	7
cycloalkanes	5	sulfides	5
disulfides	2	total	239

Overall ASFE Results and Comparison



Shivakumar et al., J. Chem. Theory Comput. 5; 919; 2009

What Makes Binding Free Energy Methods Difficult

- We really do not know
 - Not enough data has been generated
- We come back to same possible sources of errors
 - Sampling
 - We know this is a major problem
 - Force field
 - We cannot determine how much of a problem this is until we address the sampling problem
- At least we have over come the human time needed to setup FEP calculations
 - What used to take an expert days or weeks can now be done in minutes
- Enhanced sampling is a key step in moving forward
 - In addition to general methods, intelligent sampling can be done along a path if the endpoints are known

Relative Binding Free Energies – Datasets

- Abelson Tyrosine Kinase (4 ligands)
- Acetylcholinesterase (18 ligands)
- Aldose reductase (3 ligands)
- Aurora Kinase (3 ligands)
- cAMP Dependant Protein Kinase (9 ligands)
- CHK1 Checkpoint Homolog (6 ligands)
- Estrogen Receptor (2 ligands)
- Estrogen Related Receptor (2 ligands)
- Extra-Cellular Signal Regulated Kinase 2 (5 ligands)
- Factor VIIa (2 ligands)
- Factor Xa (4 + 24 + 24 ligands)
- HIV-1 Protease (9 ligand)
- HIV-1 Reverse Transcriptase (21 ligands)
- Ionotropic Glutamate Receptor (14 ligands)
- Lymphocyte-specific Tyrosine Kinase (20 ligands)
- P38 Map Kinase (16 ligands)
- Phosphodiesterase 4 (5 ligands)
- Pim-1 Kinase (4 ligands)
- Thrombin (2 + 8 + 5 ligands)
- Urokinase-type Plasminogen Activator (2 ligands)

20 targets

200+ ligands

Force Field Development

- Many limitations existing in current force fields
- Some are just issues of parameterization
- Others have to do with more fundamental methodological issues
- We have 5 people working over the past 5 years on the next generation OPLS force field
 - Over 1 million commercially available compounds searched for torsion types
 - Over 12,000 unique torsion types
 - Using CM1A-BCC charges
 - Special non-bond rules for water
 - Preliminary data suggests the new force field will help with free energy calculations

Conclusions

- Some tasks are relatively well validated
 - Pose prediction for rigid receptors
 - Virtual screening enrichment
 - Local protein structure rearrangements
- Significant work is still needed for certain tasks
 - Binding free energy predictions
 - Large-scale protein structure predictions
- Advances in method and increased compute resources are allowing us to address challenging problems
 - Prediction of water locations and thermodynamic properties
 - GPCR modeling
 - Structure-based predictions of ADME properties
 - hERG
 - P450

Final Thoughts

- The field needs more accurate lead optimization methods
 - Current empirical models have proven to be valuable for virtual screening, but are lacking for lead optimization
- Accurate scoring cannot be achieved without accurate prediction of the structure
- Confidence/error estimates are important
 - It is okay to only be right 50% of the time if you can be confident about when you will be right
- Small molecule energy predictions are only a small part of CADD
 - ADME/Tox
 - Selectivity/pathways
 - Kinetics (yes, biology is not happening at equilibrium)
 - Antibodies
 - Other biologics

Acknowledgements

Developers

Applications Scientists

Academic Collaborators

SCHRÖDINGER



