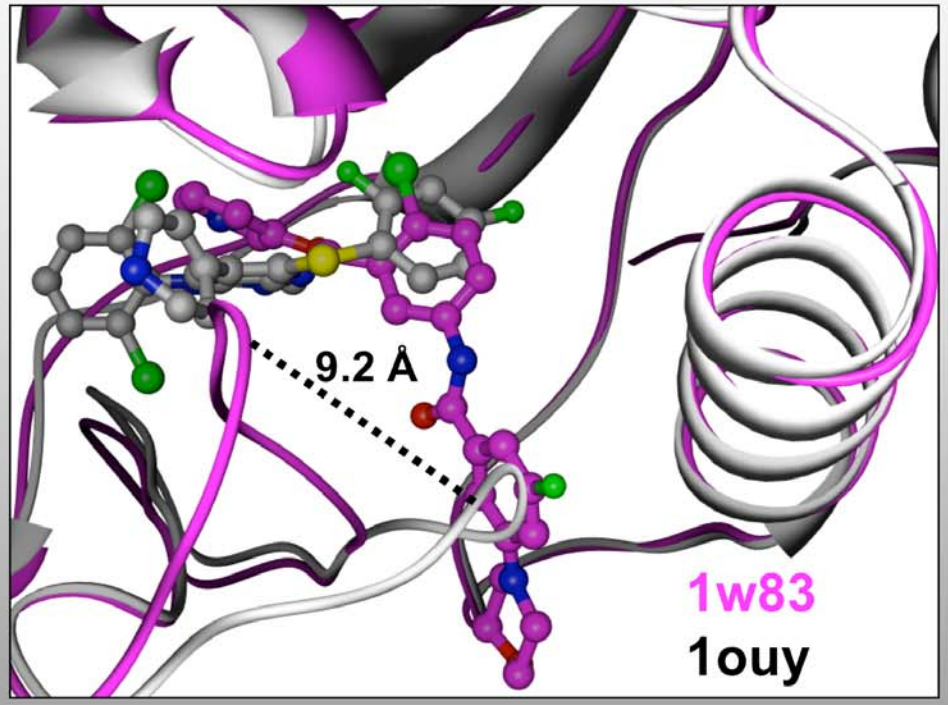
Computational Multi-Scale Modeling In Protein-Ligand Docking



THE SCRIPPS RESEARCH INSTITUTE

MICHIGAN

Roger S. Armen, Michela Taufer and Charles L. Brooks III

Motivation: Why Protein-Ligand Docking?

What are the goals of docking ?

1. **Accurate prediction of binding geometry**
   (the docking problem)

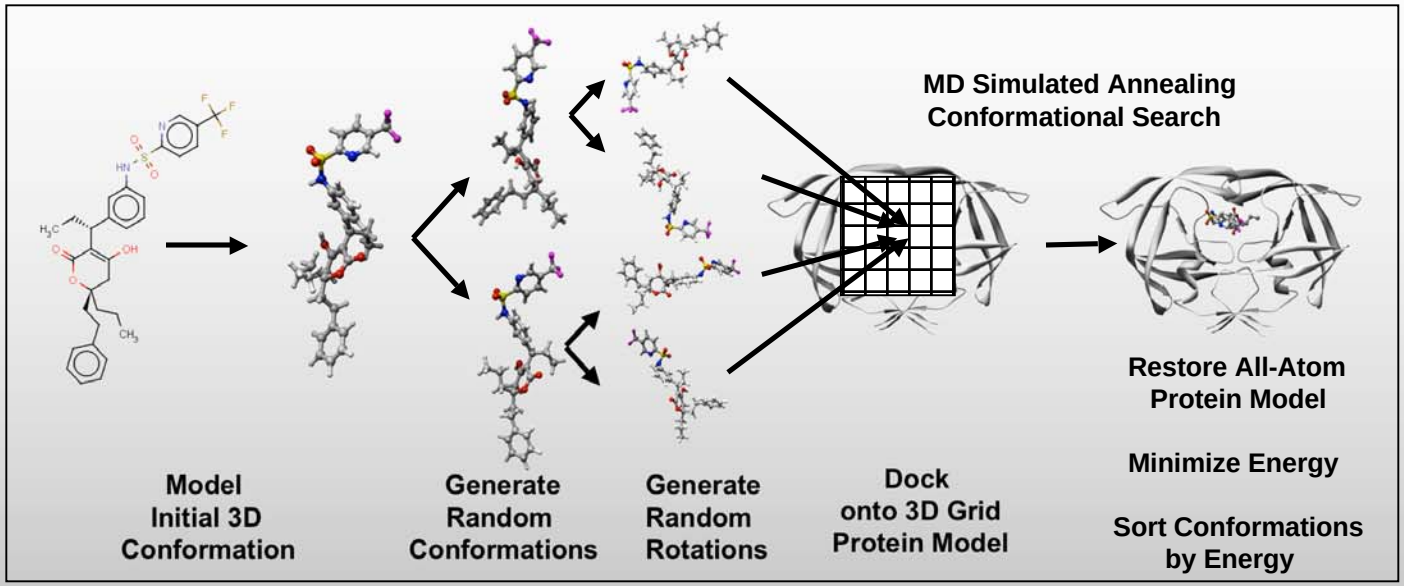
2. **Accurate prediction of binding free energy**
   (the scoring problem)

Can we augment the experimental approach ?

1. **Virtual screening**
   (discover novel chemotypes)

2. **Investigation of known lead compounds**
   (structure-guided-design of novel
   derivative compounds)

Overview of (Flexible Ligand / Rigid Receptor) Docking



1. Ongoing development of CHARMM-based molecular docking methods (CDOCKER)
2. Linear Interaction energy (LIE) free energy methods for scoring
3. Incorporation of protein flexibility into molecular docking
4. Use docking to predict the selectivity of kinase inhibitors
5. Use docking to design specific protein-ligand interactions
6. Use docking to peruse a fragment-based approach to inhibitor design

Computational Multi-Scale Modeling In Protein-Ligand Docking

The term multi-scale modeling usually refers to approaches to solving problems with important features on multiple spatial/temporal scales.

This definition can be extended to include non-orthogonal descriptive scales that allow a hierarchical approach to problem solving.

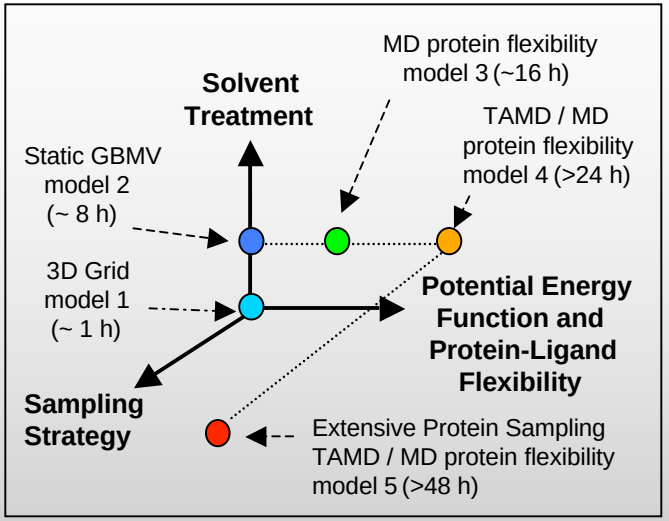
Our scales describe:

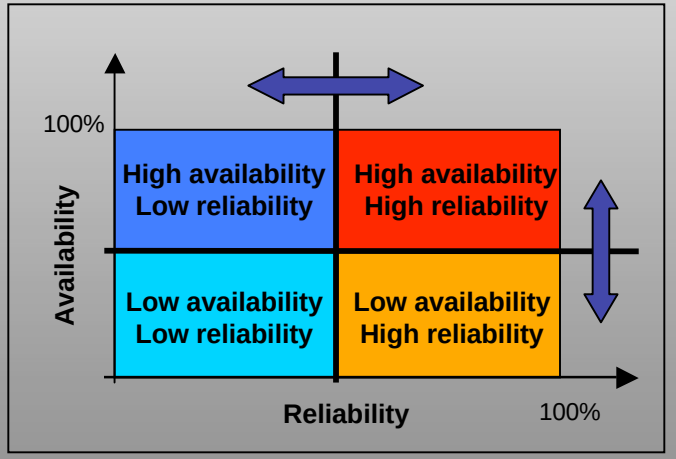
Computational complexity of "docking models"

As docking tasks are serial (not parallel), docking is a good computational problem for:

Volunteer distributed computing

Docking job runtime scales with the computational complexity of the "docking model".





Docking@Home

Docking@Home Project

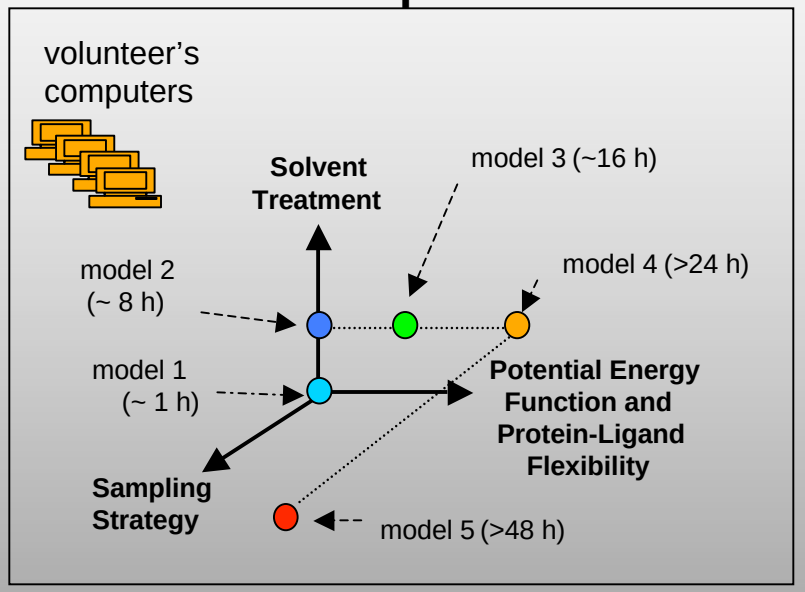
http://docking.gcl.cis.udel.edu

Volunteer distributed computing for high-throughput protein-ligand docking simulations:

BOINC (Berkeley Open Infrastructure for Network Computing)

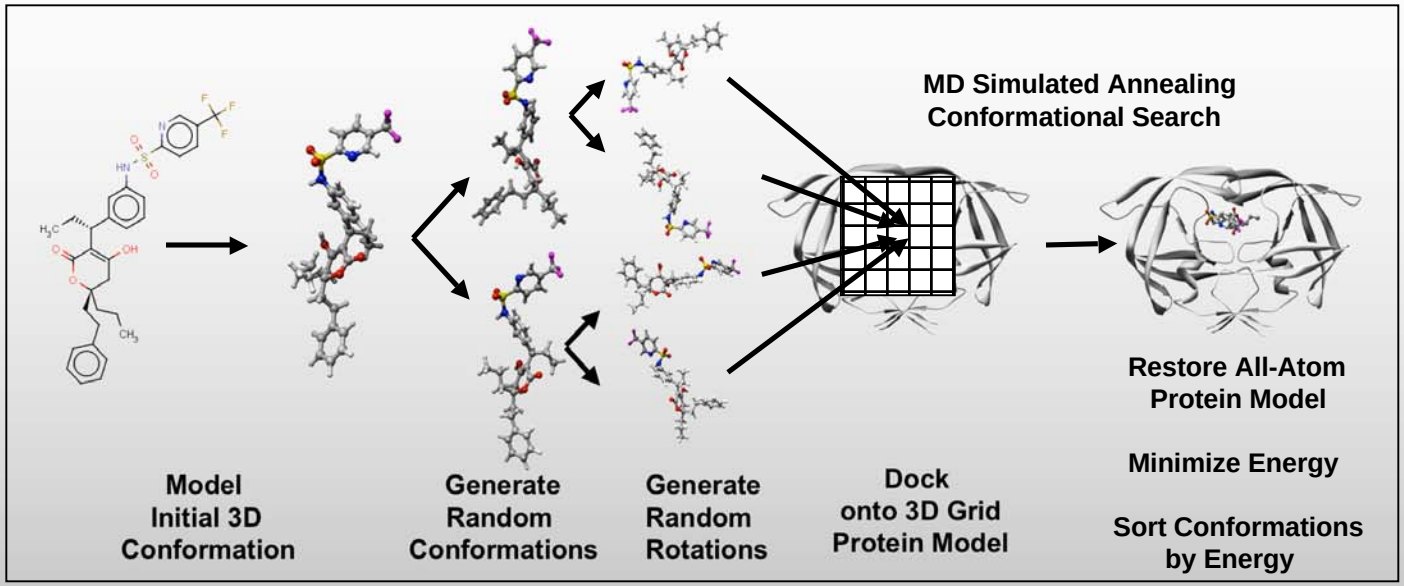
Initial scientific goals are to validate our existing docking methods on broader test sets of protein-ligand complexes, and to develop and validate new methods.

Multi-Scale Computational Models



UNIVERSITY OF DELAWARE 1743

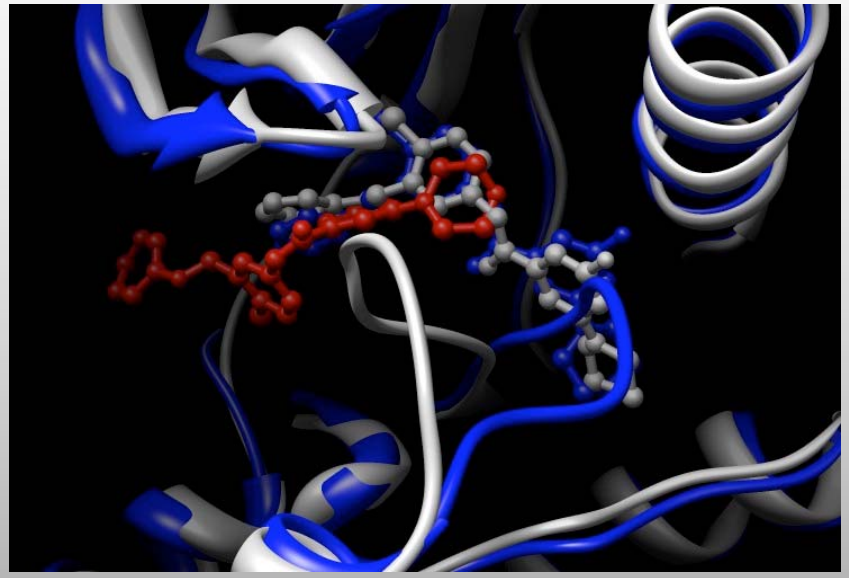
Overview of (Flexible Ligand / Rigid Receptor) Docking



1. Ongoing development of CHARMM-based molecular docking methods (CDOCKER)
2. Linear Interaction energy (LIE) free energy methods for scoring
3. Incorporation of protein flexibility into molecular docking
4. Use docking to predict the selectivity of kinase inhibitors
5. Use docking to design specific protein-ligand interactions
6. Use docking to peruse a fragment-based approach to inhibitor design

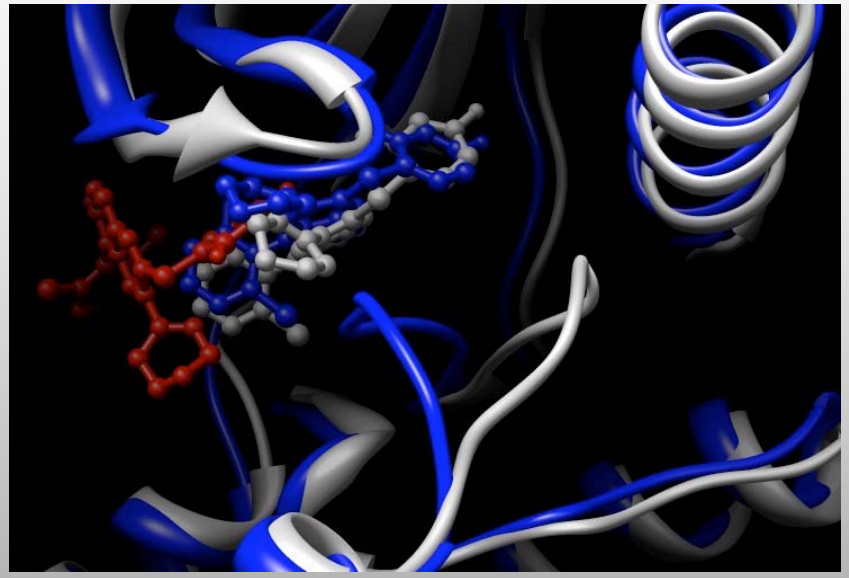
Why Do We Need Protein Flexibility In Docking?

Dock ligand 1w83 into receptor 1ouy



Reference structure (Gray)
Docking Success (Blue) RMSD (1.8 Å)
Rigid Receptor (Red) RMSD (8.1 Å)

Dock ligand 1ouy into receptor 1w83



Reference structure (Gray)
Docking Success (Blue) RMSD (1.5 Å)
Rigid Receptor (Red) RMSD (8.7 Å)

Approaches to Incorporate Protein Flexibility

If multiple crystal structures of the target protein exist:

Use multiple rigid protein conformations

Different approaches for protein conformational search:

1. **(rigid backbone) flexible side chain**
2. **(flexible backbone) flexible side chains**
3. **Entire protein flexible**

MD (Molecular Dynamics in Cartesian Space)
TAMD (Torsion Angle Molecular Dynamics in Internal Coordinates)

Small side chain Movements

Large backbone Movements

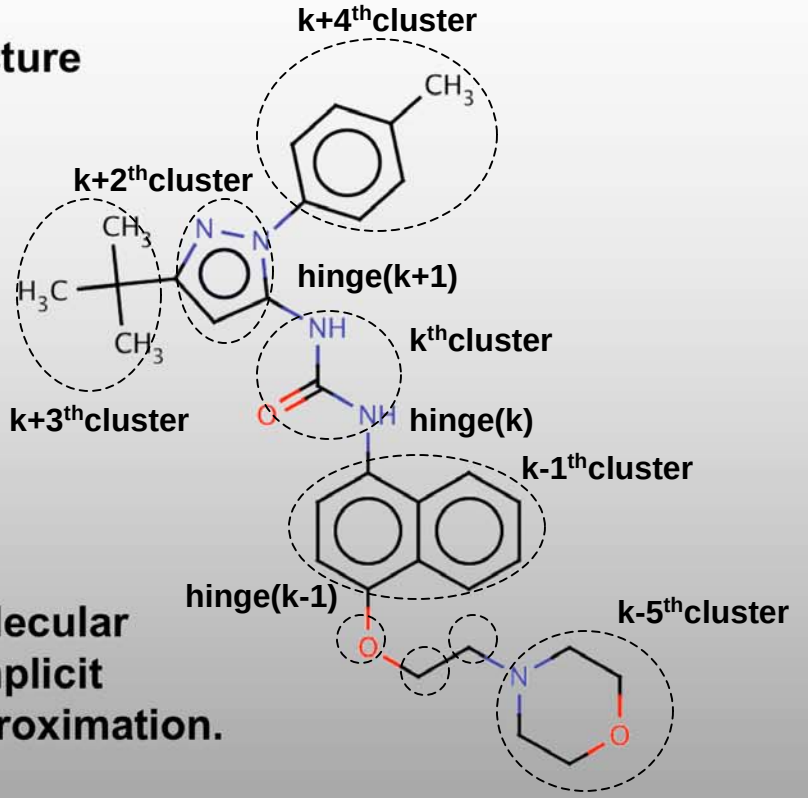
Torsion Angle Molecular Dynamics (TAMD)

Molecules are represented with a branched tree structure of rigid bodies (clusters) connected by hinges.

An efficient Newton-Euler inverse mass operator (NEIMO) algorithm is used for solving the equations of motion in internal coordinates (IC).

Torsional cross-terms are constructed from local molecular fragments, using a soft-core potential to introduce implicit bond and angle flexibility into the rigid geometry approximation.

(This effectively removes high barriers on the IC potential energy surface)



MMTSB

Chen, J., Im, W., & Brooks, C.L. III *J. Comput. Chem.* (2005) 26:1565-1578

New Flexible Receptor Docking Algorithm Using TAMD

Step 1	**Generate a diverse ensemble of flexible ligand conformations using MD. (N= 200)**
Step 2	***TAMD sampling In the absence of the ligand (apo TAMD):*** **Generate a diverse ensemble of flexible receptor conformations using TAMD simulated annealing. (N= 200)**
Step 3	**Using all-atom models and a soft-core potential, for each new receptor-ligand pair:** **Perform 1000 ligand rotations to identify the optimal rotation for each new ligand conf. in a given flexible receptor conf.**
Step 4	***TAMD sampling with the optimal ligand rotation (holo TAMD):*** **Refine the structure of the receptor-ligand complex using TAMD simulated annealing. The protein and ligand are flexible simultaneously.**
Step 5	**Calculate ΔG binding for the ensemble of refined receptor-ligand complexes, and select the "top 5" conformations.**
Step 6	**For the top 5 conformations, perform additional MD conformational sampling using the GBMV implicit solvent model. Use these conformations for an improved calculation of ΔG binding.**

Taufer, M., Armen, R. S., Chen, J., Teller, P.J., & Brooks, C.L. III *IEEE Eng. Med. & Biol.* (2008) *in press*

“Cross-Docking” to Validate Flexible Protein Docking

What is Cross-Docking?

	prot 1a9u	prot 1bl6	prot 1bl7
lig 1a9u	self-dock	cross-dock	cross-dock
lig 1bl6	cross-dock	self-dock	cross-dock
lig 1bl7	cross-dock	cross-dock	self-dock

Cross-Docking can determine the sensitivity of docking results to changes in protein conformation:

1. **Consider results from a row:**
   Each individual ligand is docked into an ensemble of experimentally determined receptor conformations

2. **Consider results from a column:**
   The entire ligand series is docked into each individual receptor conformation

“Cross-Docking” to Validate Flexible Protein Docking

	prot 1a9u	prot 1bl6	prot 1bl7
lig 1a9u	self-dock	cross-dock	cross-dock
lig 1bl6	cross-dock	self-dock	cross-dock
lig 1bl7	cross-dock	cross-dock	self-dock

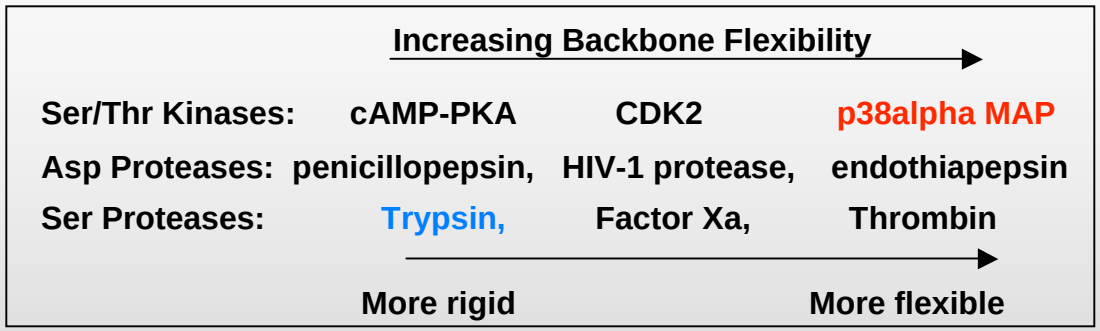
We aim to validate the flexible docking algorithm by comparing docking into a rigid receptor to flexible receptor TAMD docking.

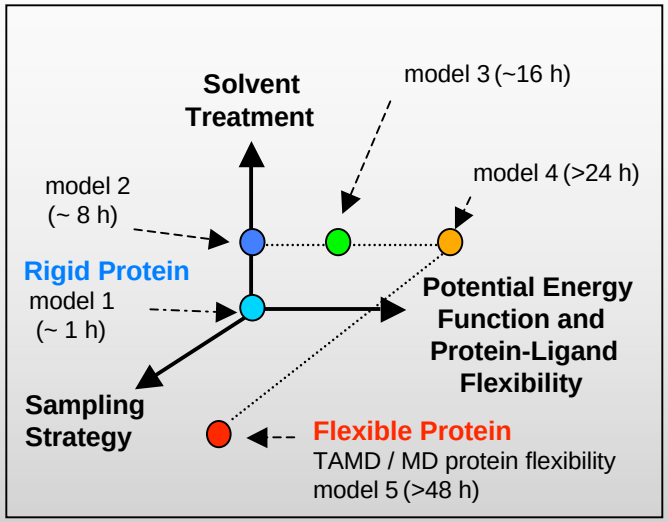
We successfully demonstrate:

1. **that reasonable receptor conformations are sampled regardless of the initial receptor conformation**

2. **the correct “native-like” receptor-ligand conformation can be selected from an ensemble of fully flexible complexes.**

Cross-Docking Test Set for Validation of TAMD Method





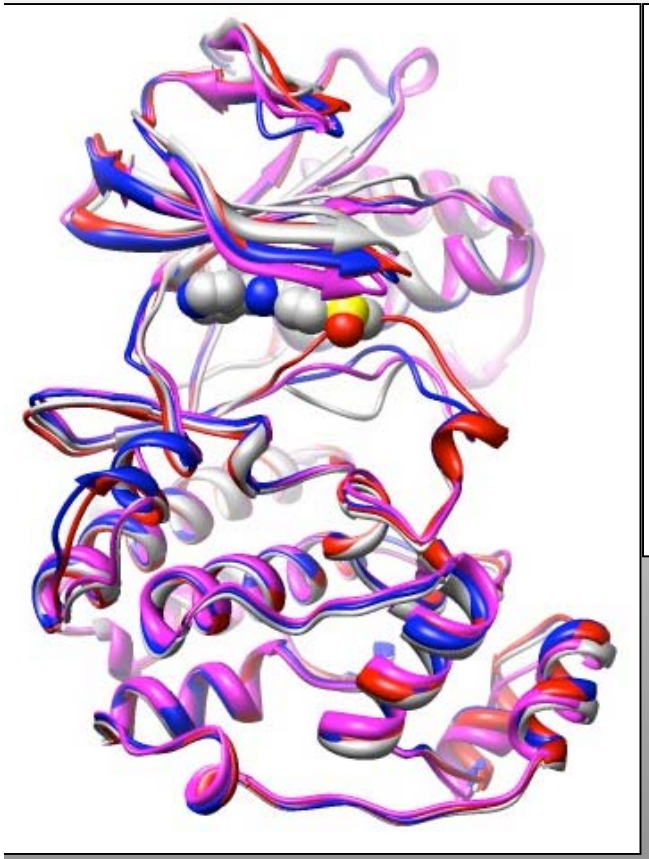
1. **Less challenging cases (rigid proteins):**
   Some ligands only require a very slight relaxation of the binding pocket to find the correct conformation.

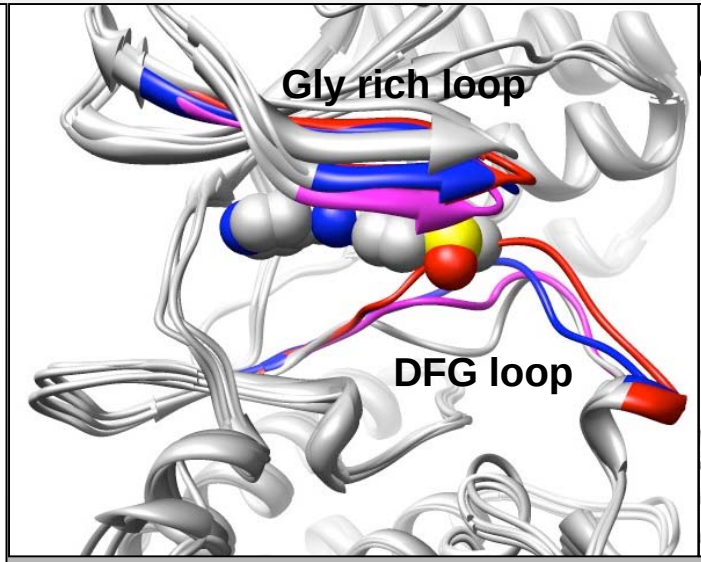
2. **Very difficult cases (flexible proteins):**
   Some ligands require significant rearrangement of the binding pocket. These receptors will need to exhibit enough flexibility to successfully sample "native-like" conformations (the cross-dock reference structure)

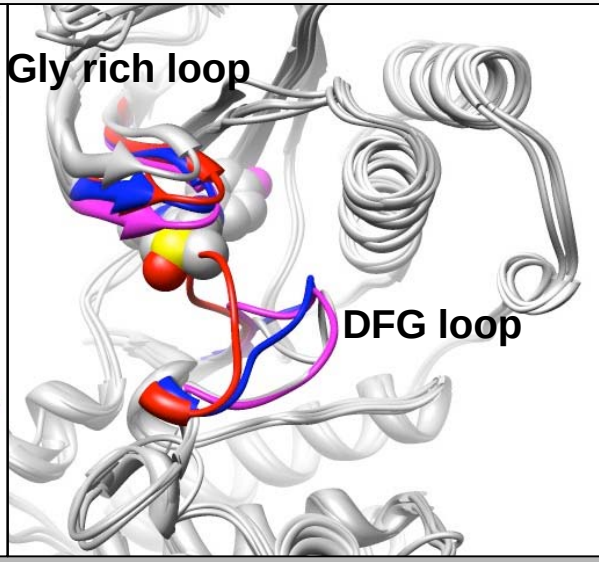
   However, if our proteins exhibit too much flexibility:
   (a) we will unlikely be able to find the correct conformation
   (b) selecting a "native-like" conf. from an ensemble will be more difficult
   (c) "less challenging cases" will become difficult

p38alpha MAP Kinase: Very Difficult Cross-Docking Test

Shown are 4 diverse protein-ligand complex crystal structures:





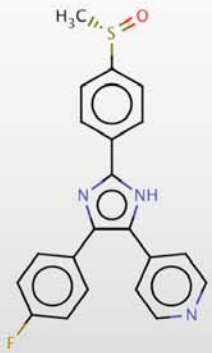


Two flexible loops define the binding site:

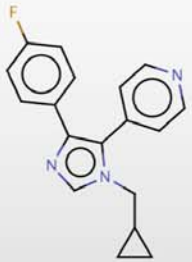
1. **The DFG Loop (res:165-177)**
2. **The Gly rich loop (res:30-41)**

p38alpha MAP Kinase: A Diverse Set of Ligands

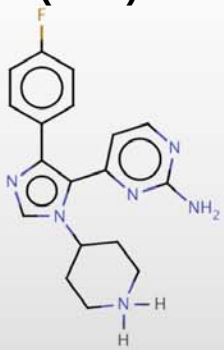
1 (1a9u)



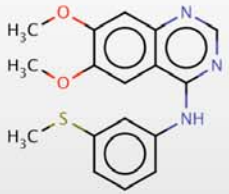
2 (1bl6)



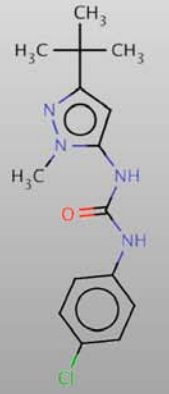
3 (1bl7)



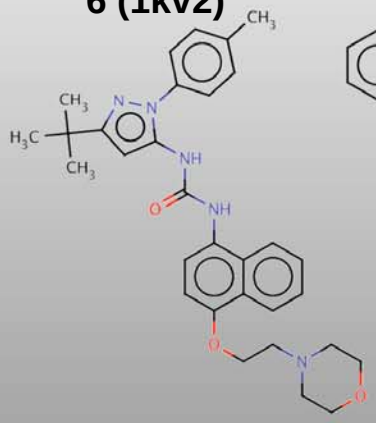
4 (1di9)



5 (1kv1)



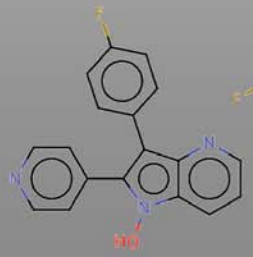
6 (1kv2)



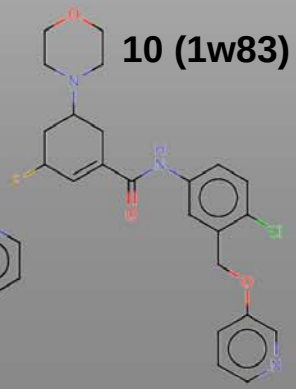
7 (1ouk)

8 (1ouy)

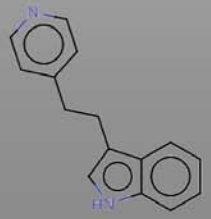
9 (1oz1)



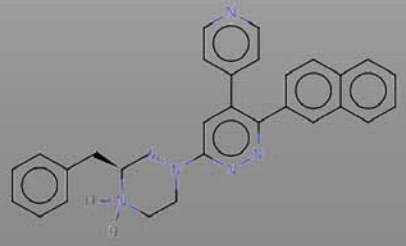
10 (1w83)



11 (1w84)



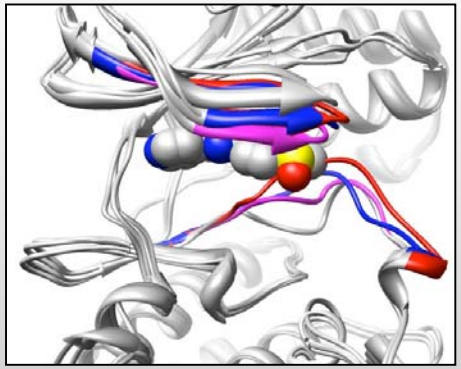
12 (1yqj)

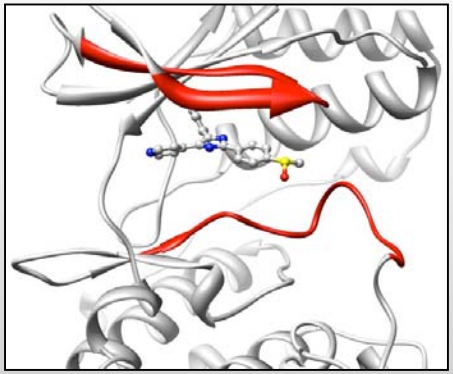


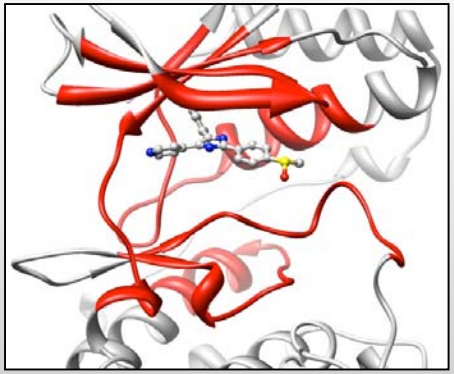
Models of Protein Flexibility & Degrees of Freedom

TAMD model 1
(2 flexible loops)

TAMD model 2
(6 flexible segments)







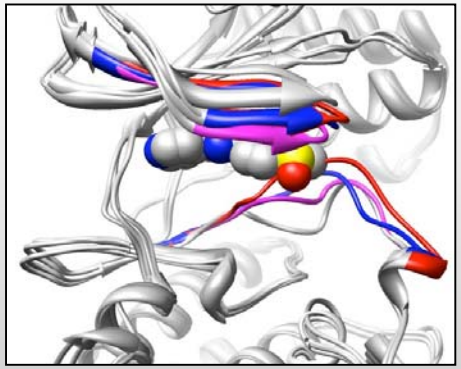
The flexibility of the protein can be defined by any combination of fully flexible backbone segments and flexible side chains.

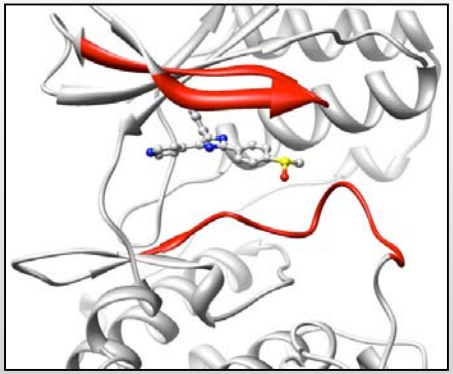
Flexible Protein Models	Degrees of Freedom	N Flexible Segments	N Flexible Side Chains	Description of Flexible Protein Models
TAMD model 1	328	2	50	2 flexible loops + flexible side chains
TAMD model 2	442	6	50	6 flexible segments
TAMD model 3	1651	1	351	entire protein flexible
Cartesian MD	14,133	1	351	entire protein flexible

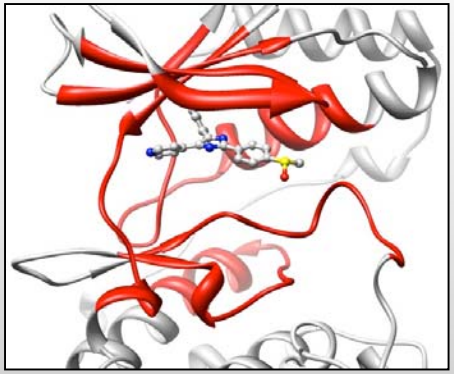
Cross-Docking Accuracy Validates the TAMD Approach

TAMD model 1
(2 flexible loops)

TAMD model 2
(6 flexible segments)







Cross-Docking Accuracy (% success over 12x12 = 144 docking simulations)

Control (Rigid Receptor)	best1 pose <= 2.0 A	best1 pose <= 3.0 A	any of top 5 <= 2.0 A	any of top 5 <= 3.0 A
Rigid Receptor Docking	15%	24%	38%	51%

Flexible Protein Models	Degrees of Freedom	best1 pose <= 2.0 A	best1 pose <= 3.0 A	any of top 5 <= 2.0 A	any of top 5 <= 3.0 A
TAMD model 1	328	34%	52%	83%	94%
TAMD model 2	442	31%	48%	59%	85%
TAMD model 3	1651	11%	28%	44%	78%

Cross-Docking Accuracy Validates the TAMD Approach

Yellow	RMSD <= 2.0 A
Orange	RMSD <= 3.0 A

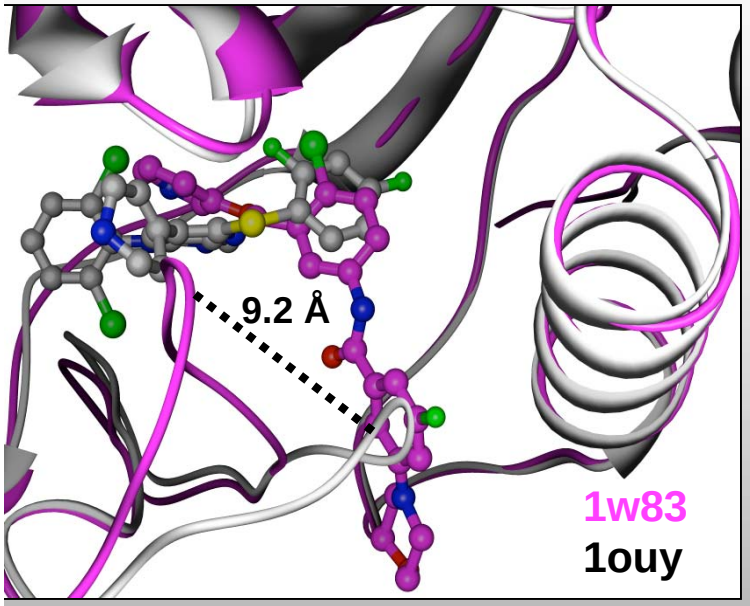
Control Rigid Receptor
(Lowest Ligand RMSD for any of the Top 5 Scoring Pose)

	1a9u	1bl6	1bl7	1di9	1kv1	1kv2	1ouk	1ouy	1oz1	1w83	1w84	1yqj	
1a9u	1.7	3.6	2.5	2.6	6.6	7.6	3.4	2.6	2.2	6.4	2.1	2.4	1a9u
1bl6	2.1	1.6	1.0	4.5	6.2	6.0	1.7	4.2	1.7	1.3	4.2	1.7	1bl6
1bl7	1.6	2.7	1.0	1.6	8.4	6.0	2.0	1.5	1.7	1.9	1.3	4.6	1bl7
1di9	4.4	0.9	0.9	0.7	2.1	1.9	2.8	2.8	0.7	6.6	5.1	2.8	1di9
1kv1	7.0	5.8	6.9	7.7	1.0	1.6	7.6	8.9	7.1	0.5	6.7	6.2	1kv1
1kv2	6.7	7.1	5.9	9.4	1.2	0.4	9.6	8.9	8.8	1.5	8.3	6.4	1kv2
1ouk	8.1	7.4	6.2	2.6		7.4	0.7	1.7	7.9	9.9	2.9	5.0	1ouk
1ouy	5.3	6.6	6.9	3.7	8.1	9.1	1.8	0.6	3.0	8.7	8.3	3.6	1ouy
1oz1	0.7	5.0	0.8	5.0	7.4	10.4	1.8	1.4	0.6	2.0	1.0	1.4	1oz1
1w83	4.3	5.7	5.9	7.4	1.4	1.4	7.7	8.1	9.6	0.9	6.0	6.0	1w83
1w84	2.5	1.5	1.6	3.0	1.8	2.7	2.0	1.7	1.8	0.8	0.6	1.9	1w84
1yqj	2.7	1.1	2.1	6.3	7.7	7.0	1.4	1.3	4.6	2.4	1.9	0.5	1yqj

TAMD Flexible Receptor Model 1
(Lowest Ligand RMSD for any of the Top 5 Scoring Pose)

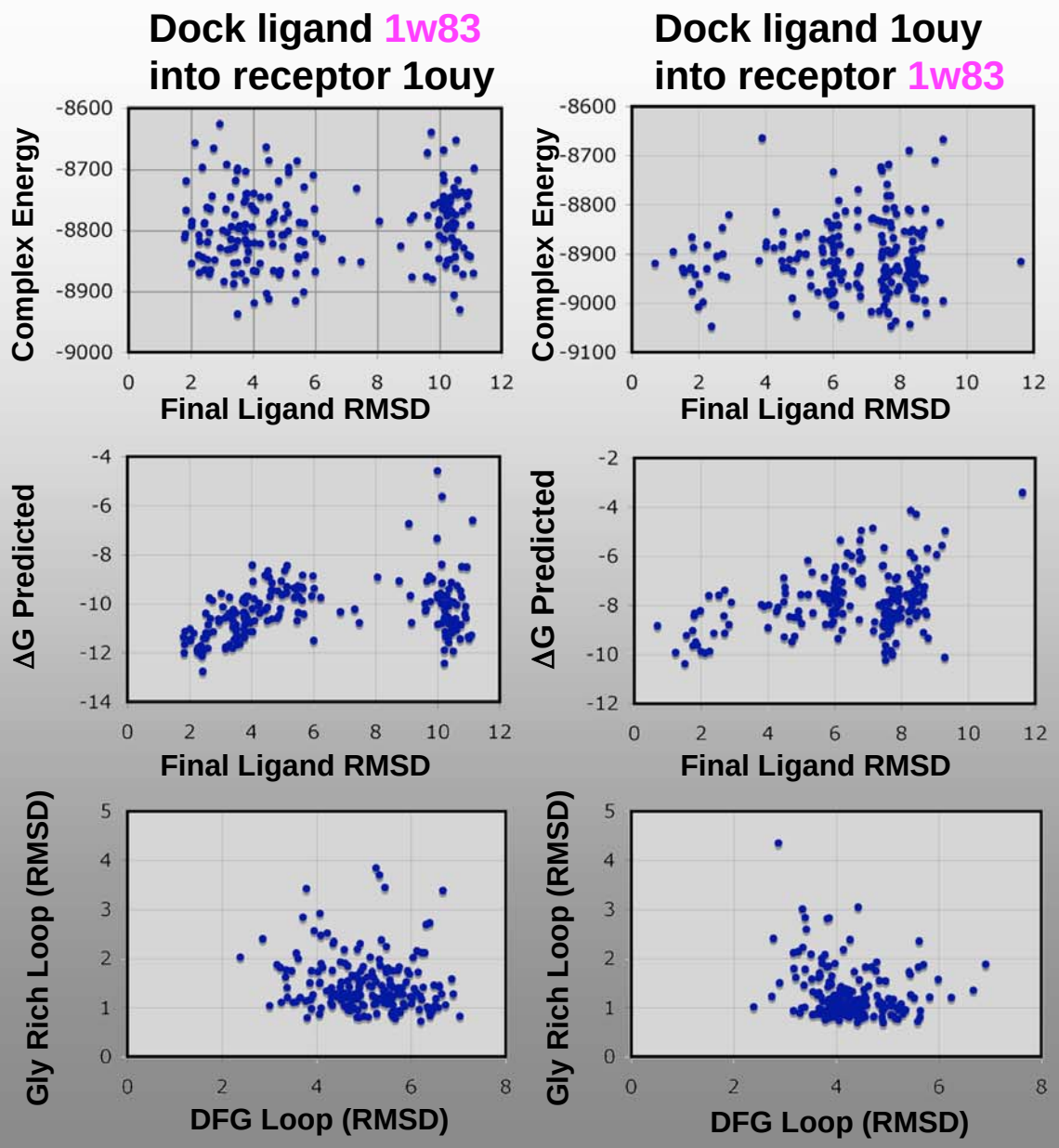
	1a9u	1bl6	1bl7	1di9	1kv1	1kv2	1ouk	1ouy	1oz1	1w83	1w84	1yqj	
1a9u	1.7	2.0	3.1	2.0	1.9	1.9	2.8	2.1	1.8	1.9	2.7	1.5	1a9u
1bl6	1.2	1.4	0.9	1.5	1.4	1.4	1.8	2.0	1.8	1.6	1.4	2.1	1bl6
1bl7	1.2	1.6	1.3	1.7	2.0	1.9	1.4	1.2	1.6	2.3	1.7	1.3	1bl7
1di9	1.8	1.1	1.0	1.9	2.8	2.1	2.3	1.8	1.1	1.4	3.0	1.9	1di9
1kv1	1.3	1.9	2.1	3.8	1.5	1.6	1.9	1.0	1.9	1.7	1.5	1.8	1kv1
1kv2	2.0	1.6	1.6	1.8	0.9	1.2	1.5	1.3	1.4	1.9	1.7	1.8	1kv2
1ouk	1.5	1.3	2.2	2.2	2.1	1.5	0.9	1.1	1.7	5.2	1.5	1.2	1ouk
1ouy	2.1	1.2	4.7	4.1	0.7	1.1	0.9	1.0	1.9	1.5	6.4	1.2	1ouy
1oz1	0.5	1.2	1.0	1.1	1.6	1.6	1.0	0.9	0.9	0.7	0.8	1.1	1oz1
1w83	2.0	1.7	2.1	2.3	1.8	1.3	1.8	1.8	1.4	0.9	1.6	1.8	1w83
1w84	0.8	0.6	0.9	0.9	1.1	1.4	0.8	0.9	0.7	0.9	0.4	0.7	1w84
1yqj	1.1	1.0	0.9	5.2	1.6	2.3	1.1	0.7	0.9	3.1	1.8	0.9	1yqj

ΔG (LIE) can Separate "Native-Like" Conf. From Ensemble



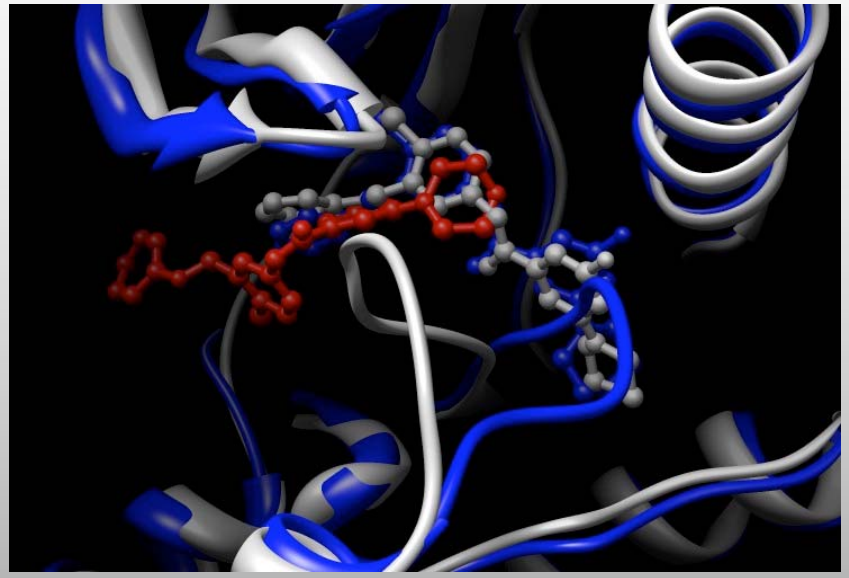
A Difficult Cross-Dock:

Both starting conformations have significant atom-clashes



Flexible Docking “Success” For a Difficult Test Case

Dock ligand 1w83 into receptor 1ouy

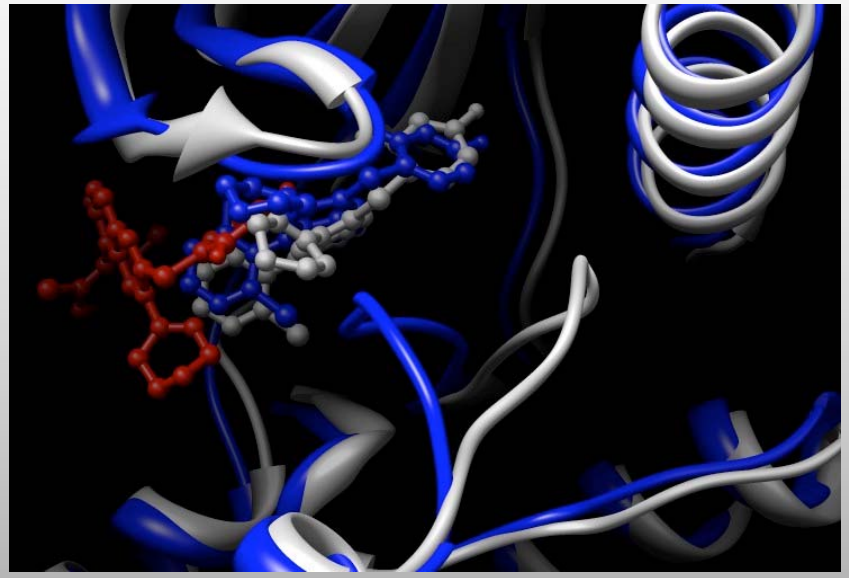


Reference structure (Gray)
Flexible Receptor (Blue) RMSD (1.8 Å)
Rigid Receptor (Red) RMSD (8.1 Å)

Flexible Receptor docking pose forms 70% of the native Prot-Lig contacts

Self-Docking pose forms up to 85%

Dock ligand 1ouy into receptor 1w83

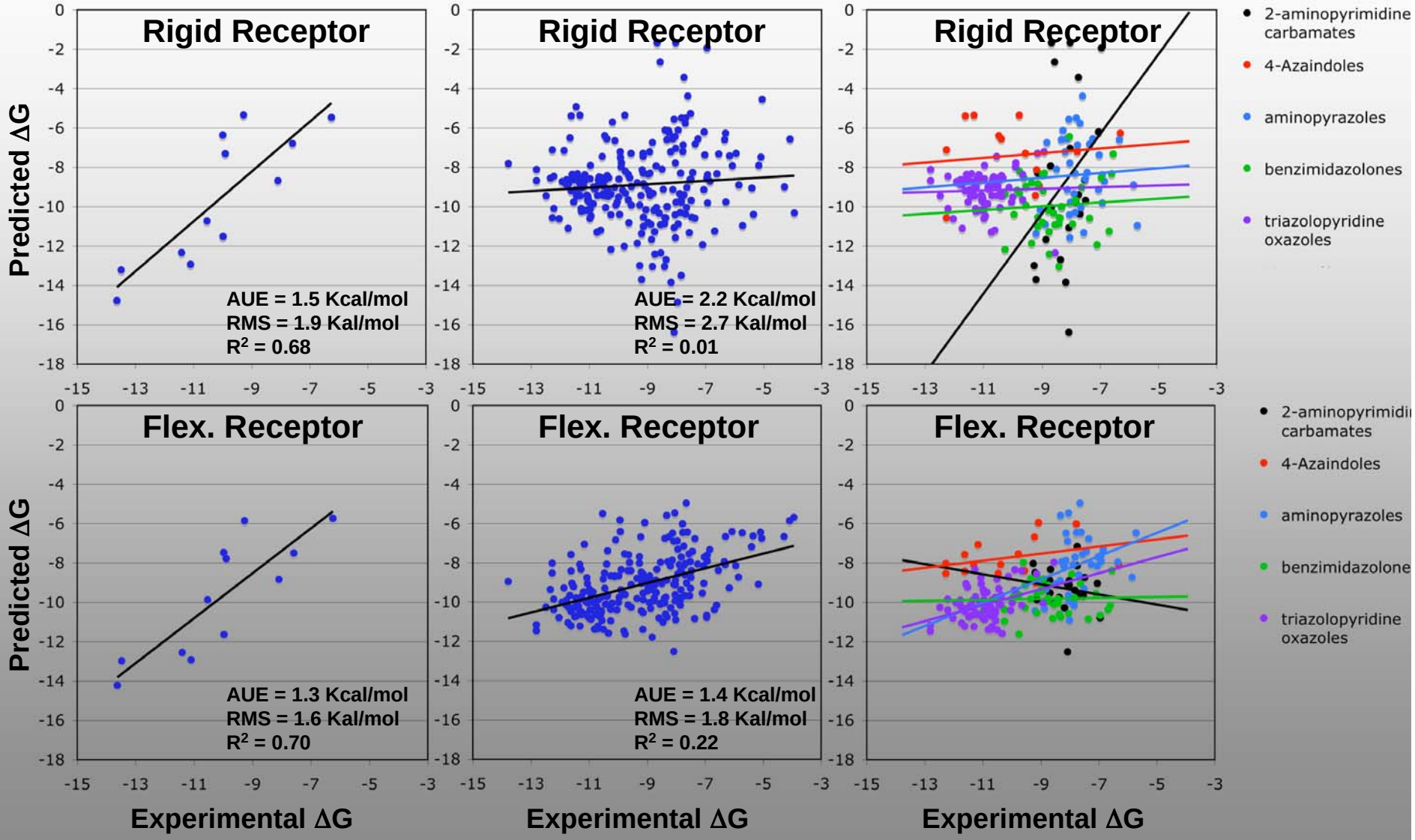


Reference structure (Gray)
Flexible Receptor (Blue) RMSD (1.5 Å)
Rigid Receptor (Red) RMSD (8.7 Å)

Flexible Receptor docking pose forms 68% of the native Prot-Lig contacts

Self-Docking pose forms up to 90%

Dock a series of 230 ligands (p38alpha inhibitors)



Conclusions & Future Directions

We have validated our flexible docking algorithm using cross-docking.

We successfully demonstrate:

1. **That reasonable receptor conformations are sampled regardless of the initial receptor conformation**

2. **The correct “native-like” receptor-ligand conformation can be selected from an ensemble of fully flexible complexes.**

3. **We can extend these results beyond simple studies of (N=12) to larger validation studies (N=230)**

Acknowledgements

NSF (OCI #0802650)
to Michela Taufer,
Charles L. Brooks III,
David P. Anderson,
Patricia J. Teller

Docking@Home:

http://docking.gcl.cis.udel.edu

NIH NRSA Postdoctoral Fellowship
to Roger S. Armen (GM076836)

NIH Research Resource Center
for the Development of
Multiscale Modeling Tools
for Structural Biology (RR12255)

MMTSB

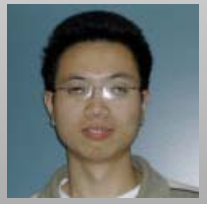
http://mmtsb.org

Charles L. Brooks III
University of Michigan

Michela Taufer
University of Delaware

David P. Anderson
University of Cal. Berkeley

Patricia J. Teller
University of Texas El Paso



Jianhan Chen
Kansas State University

Dan Price
Glaxo Smith Kline

Mike Crowley
National Renewable
Energy Lab (NREL)

Trilce Estrada
University of Delaware