Accommodating

Protein Flexibility

in Computational Drug Design

Christopher N. J. Marai

Biochemistry and Structural Biology
Stony Brook University

Accommodating **Protein Flexibility** in Computational Drug Design;
http://molpharm.aspetjournals.org/cgi/content/full/57/2/213

Mol Pharmacol. 2000 Feb;57(2):213-8.Related Articles, Links

Accommodating protein flexibility in computational drug design.
Carlson HA, McCammon JA.
Department of Pharmacology, University of California, San Diego, La Jolla, California, USA. hcarlson@mccammon.ucsd.edu

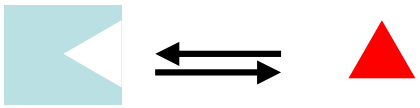
Note: For purposes of this presentation. (Receptor = Target = Protein) and (Ligand = Small Molecule = Drug) I interchange these frequently.

NOTES ATTACHED

- Use Notes Feature in PPT to read short summary of what was said during the presentation.

Accommodating Protein Flexibility in Computational Drug Design

- Given a receptor and an arbitrary small molecule, predict whether the small molecule will bind to the receptor and, if so, predict the geometry and affinity of the binding
- In other words, if the small molecule is a ligand of the receptor, predict the relative orientation of the two molecules in their bound state



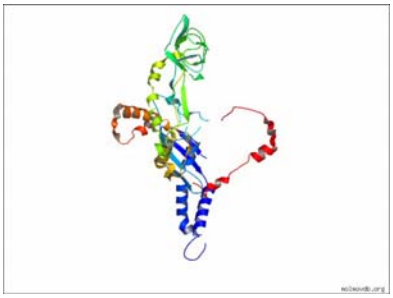
Receptor **Ligand**

The major problem to be solved in drug discovery.

Note that the interaction between the receptor and ligand is dynamic.

Protein Flexibility and Its Influence on Ligand Binding

- the **dynamic** behavior of receptors and ligands is a complicating factor and a reality
- typically receptor structures are obtained from X-ray and NMR methods and are consequently **static** and usually single structures



10/11/2004 Christopher N. J. Marai 4

Dynamic behavior of receptors and ligands is a complicating factor in drug design.

Typically receptor structures are obtained from x-ray and NMR methods and are consequently static. Imagine you have one snap shot from a home video.

This doesn't give you all the details on the events in the video just one small fraction of them. This is also true for NMR and XRAY structures. Proteins are typically in some form of motion.

(ie: The conformer which is crystallized is the one whose conformation is favorable for binding under these conditions, with the equilibrium shifting in its favor.)

When accessing a particular receptor a single structure is chosen most often or speed.

Figure: Note that this figure should be a movie of the protein moving. www.molmovdb.org

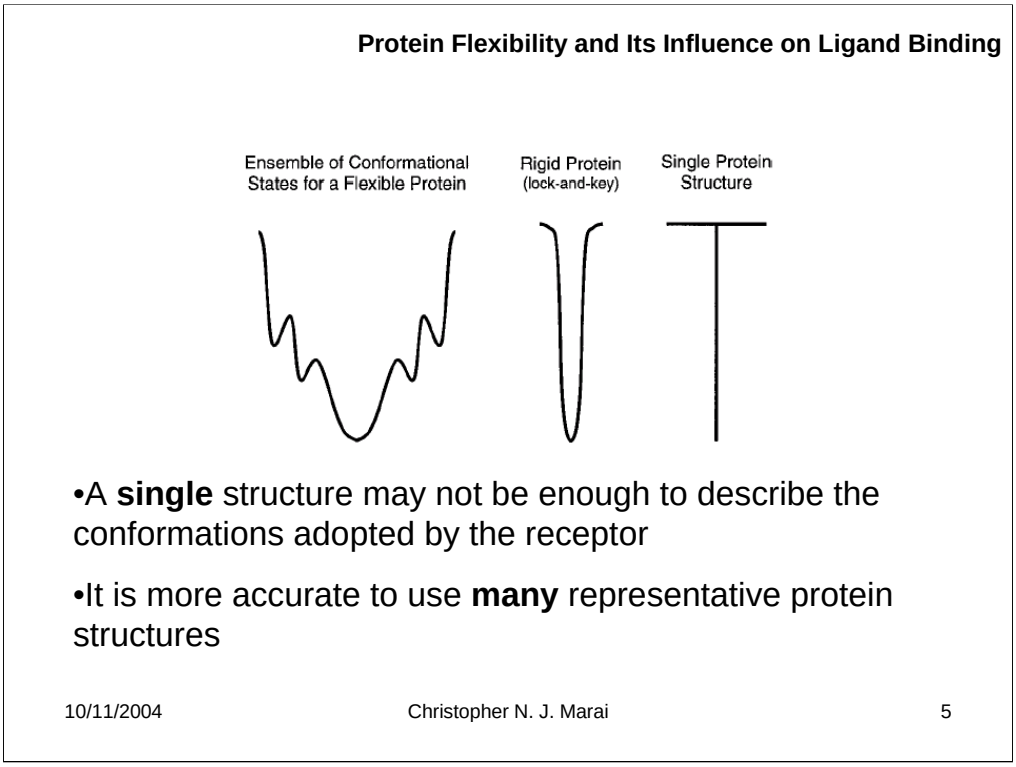
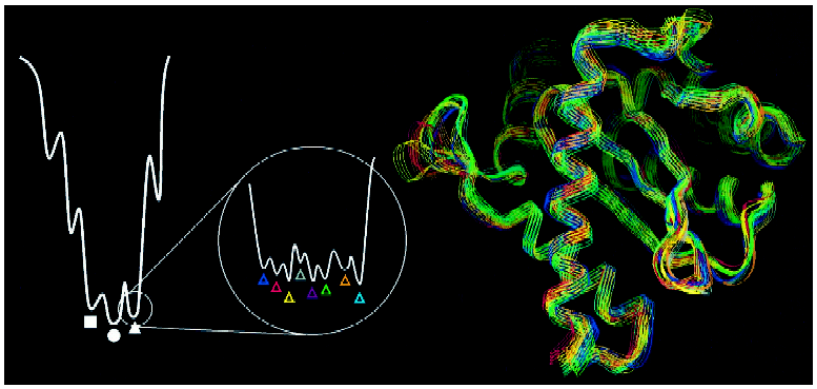


Fig. 1. A single structure of a protein implies an all-or-nothing folding funnel, perhaps best described with the mathematic singularity shown on the right. The folding funnels on the left and in the center demonstrate the conformational flexibility of a “standard” flexible protein and a rigid protein, respectively. Although the rigid system may be described adequatelyby a single structure (depending on the degree of rigidity as reflected by the narrowness of the funnel), a typical system is not.

Protein Flexibility and Its Influence on Ligand Binding



•this conformational funnel demonstrates multiple states for a protein (represented by their weighted-average structures ■, ●, ▲; collection of colored triangles represent subensemble structures)

10/11/2004 Christopher N. J. Marai 6

Fig. 2. This conformational funnel demonstrates multiple states for a protein (represented by their weighted-average structures: Square, Circle, and Triangle). The

flexibility inherent in a folded state (Triangle) is described by a subensemble of conformations (shown here as a collection of colored Triangle). We have used a

collection of structures from an MD simulation of HIV-1 integrase (Lins et al., 1999) to demonstrate a subensemble, showing a range of flexibility with

modest sampling of the backbone and wide sampling of a small flexible loop on the left. Expanding the minimum of a single state in the subensemble

(any of the colored Triangle) would reveal an additional series of subminima that arise from variations in the orientations of the side chains.

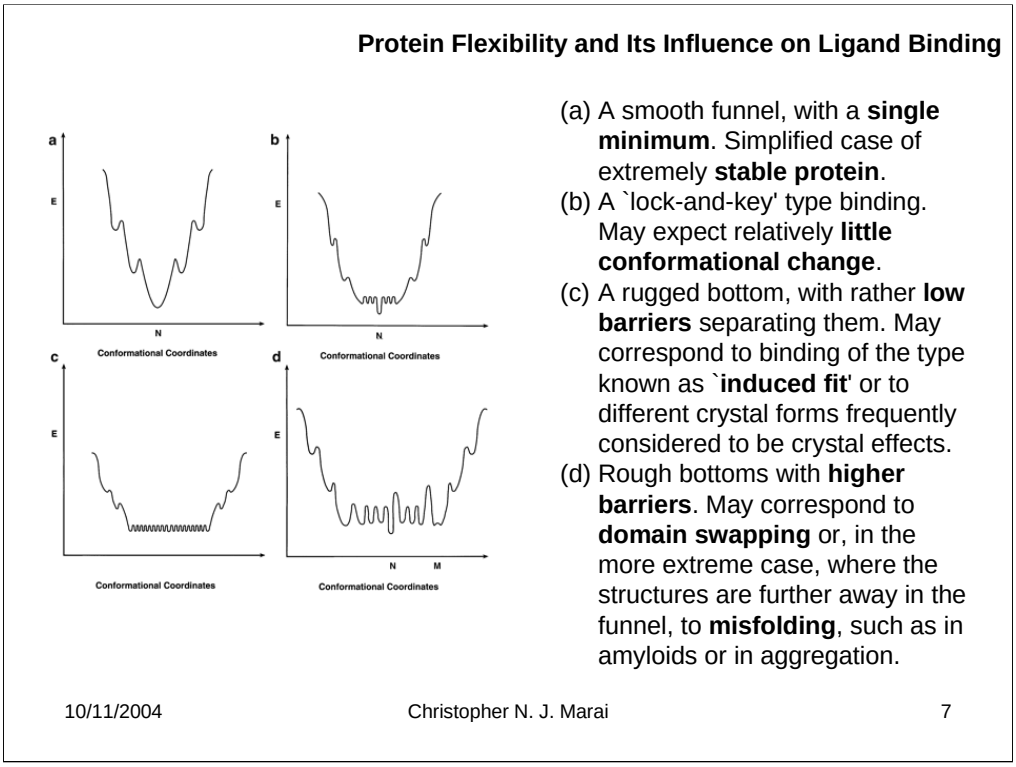


Fig. 2. Schematic depiction of different landscapes around the bottom of the funnels and their correlation with binding mechanisms. In general, the ruggedness correlates with flexibility. (**a**) A smooth funnel, with a single minimum. This may be a simplified case of an extremely stable protein. (**b**) A `lock-and-key' type binding. Here we may expect relatively little conformational change. The landscape is shown schematically as having few minima, with the conformations nearby on the energy landscape very similar geometrically. (**c**) A rugged bottom, with rather low barriers separating them. This may correspond to binding of the type known as `induced fit' or to different crystal forms frequently considered to be crystal effects. Depending on the range of minima and on the barrier heights, we may have a case like that corresponding to non-specific binding. (**d**) Rough bottoms with higher barriers. These types of cases may correspond to domain swapping or, in the more extreme case, where the structures are further away in the funnel, to misfolding, such as in amyloids or in aggregation.

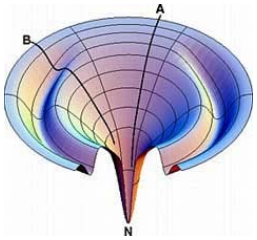
From: Protein Eng. 1999 Sep;12(9):713-20.

Folding funnels and binding mechanisms.

Ma B, Kumar S, Tsai CJ, Nussinov R.
Laboratory of Experimental and Computational Biology and Intramural Research Support Program-SAIC, Laboratory of Experimental and Computational Biology, NCI-FCRDC, Frederick, MD 21702, USA

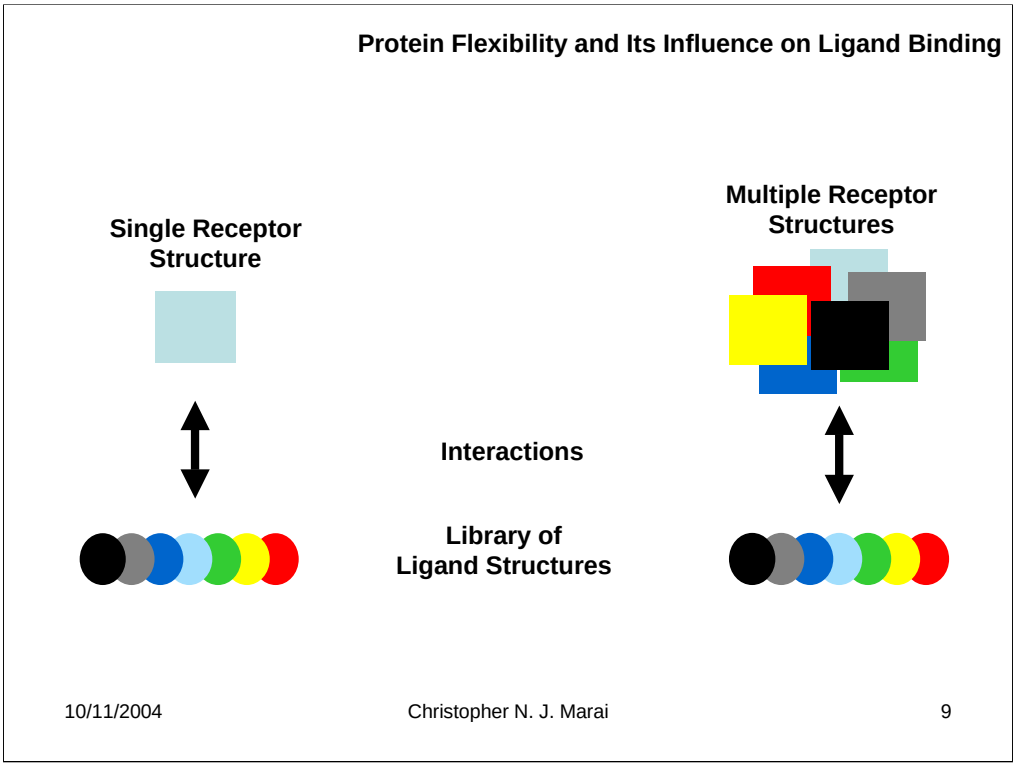
Protein Flexibility and Its Influence on Ligand Binding

- The conformational funnel and the states it represents are dependent upon the environment
- Conditions include ionic strength, pH, temperature, etc.
- Altering environments can reveal additional possible conformations
- Environment includes the **ligand**



Thus the ligand alters the conformation of the receptor.

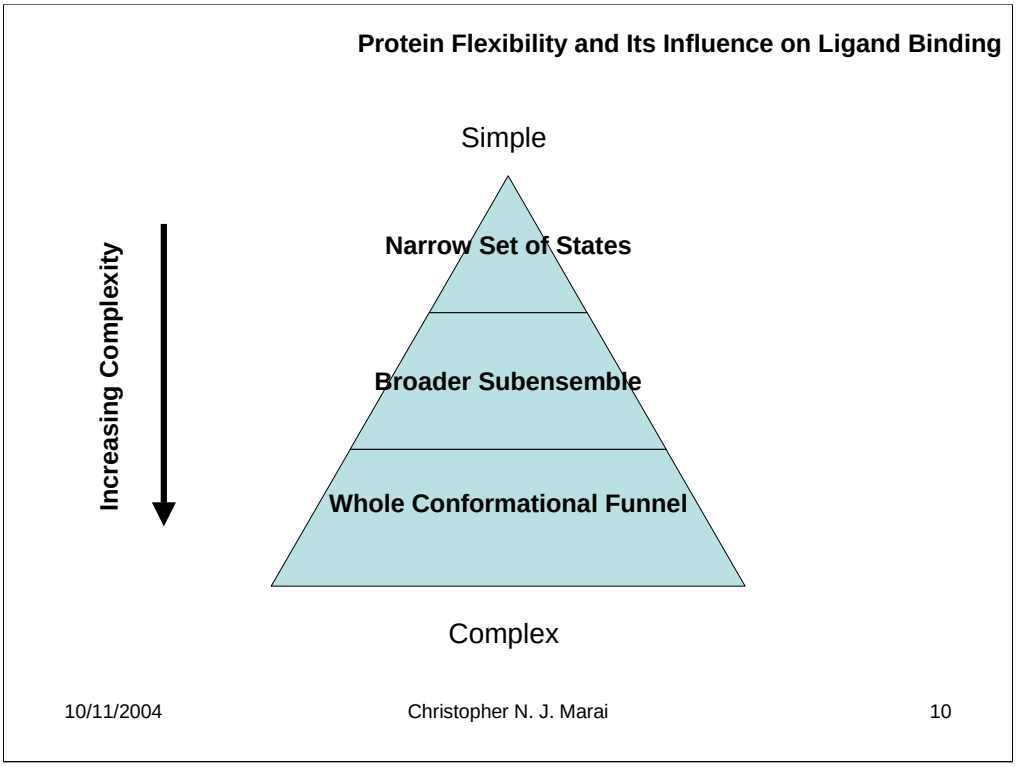
Image: Levinthal to pathways to funnels, *Nature Structural Biology*, Volume 4, No. 1, January 1997



RIGHT Current (most common) and most simple situation we try to dock many ligands with one single receptor structure to identify a possible ligand that could be used as a drug. Typically libraries are huge 100,000's of possible ligands and many of there possible conformations.

This is analogous to high throughput screening processes of experimental drug discovery.

LEFT Ideally we represent the single receptor as several different conformations (to represent its dynamic behavior) then we would interact each of the library structures with each of the receptor structures. This problem become large fast with many possible combinations. Ahead are some variations attempting to include multiple structures and thus the dynamic behavior of the receptor.



Some ways to introduce the dynamic receptor into models. Complexity equates to a lot of computational time typically.

Soft-Docking Methods

- Soft functions allow for some **overlap** between the ligand and protein, this gives a small **estimate** of the **plasticity** of the receptor

- The protein is held **fixed** but a 'soft-scoring' function *(smaller energetic penalty for overlaps and close contacts)* is used to evaluate the fit of the ligand to the receptor

- Scoring functions are typically derived from **force fields** from molecular mechanics

Soft-Docking Models are one way to account for dynamic behavior of the receptor binding site. See picture next slide.

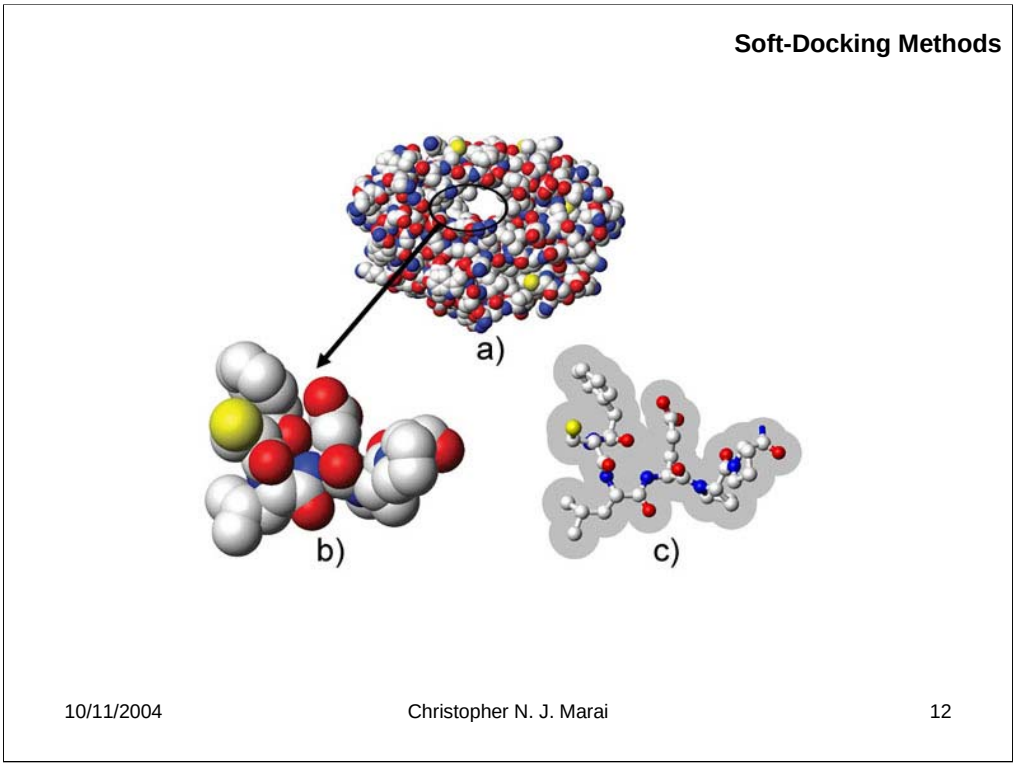


Figure: Three dimensional van der Walls representation of a target receptor.

For the purposes of rigid protein docking, the receptor is commonly described by the union of the volumes occupied by its atoms. The steric collision of any atom of the candidate ligand with the atoms of the receptor will result in a high energetic penalty.

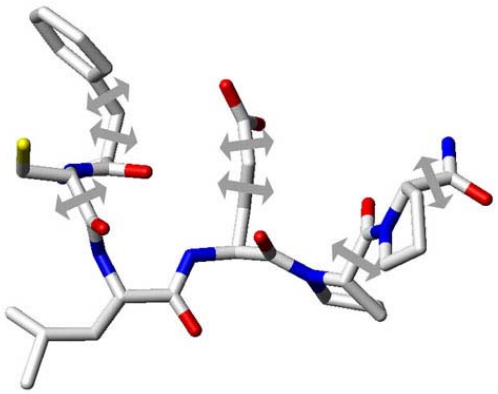
Same section of the binding site as shown in b) but with reduced radii for the atoms in the receptor. This type of soft representation allows ligand atoms to enter the shaded area without incurring a high energetic penalty.

http://cnx.rice.edu/content/m11456/latest/

Conformational Sampling of Receptor Side Chains

•Soft docking still involves a **static single conformation** description of the receptor

•However, **conformational sampling** of receptor side chains can provide a better (dynamics) representation of the binding site



10/11/2004 Christopher N. J. Marai 13

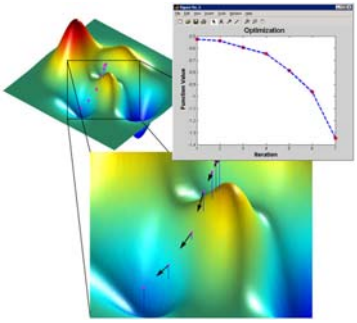
Figure: Stick representation of the same binding site section as shown in Figure 1. In order to approximate the flexibility of the receptor it is possible to carefully select a few degrees of freedom. These are usually select torsional angles of sidechains in the binding site that have been determined to be critical in the induced fit effect for a specific receptor. In this example the selected torsional angles are represented by arrows.
http://cnx.rice.edu/content/m11456/latest/

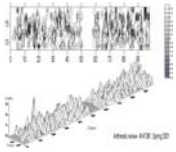
Conformational Sampling of Receptor Side Chains

- Conformational sampling accounts for rotational freedom of the receptor
- May allow for just H rotations
(allowed for more favorable h-bonding interaction between ligand and receptor)
- Has been improved to sample side chain dihedral changes, while restricting angle and bond distance changes

Conformational Sampling of Receptor Side Chains

- During ligand docking the conformations can be minimized through **internal coordinate optimization** or random thermodynamic sampling methods such as **metropolis**





Conformational Sampling of Receptor Side Chains

- Minimization routines typically have problems with conformational barriers and can be computationally expensive
- To overcome this **Rotamer Libraries** can be employed to sample conformational space of side chains
- Provides lowest energy conformers, thus provides a first estimate of available conformational space
- Fast way to search through receptor conformations
- Reduces problem of conformational barriers that minimization routines typically have

10/11/2004 Christopher N. J. Marai 16

So make a library to sample to the conformational space Instead of minimizing through the space which may result in an non-minimum energy structure (local minima).

Jones G, Willet P and Glen RC (1995) Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol* **245:** 43-53

Totrov M and Abagyan R (1997) Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* **1 (Suppl):** 215-220.

Schnecke V, Swanson CA, Getzoff ED, Tainer JA and Kuhn LA (1998) Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* **33:** 74-87

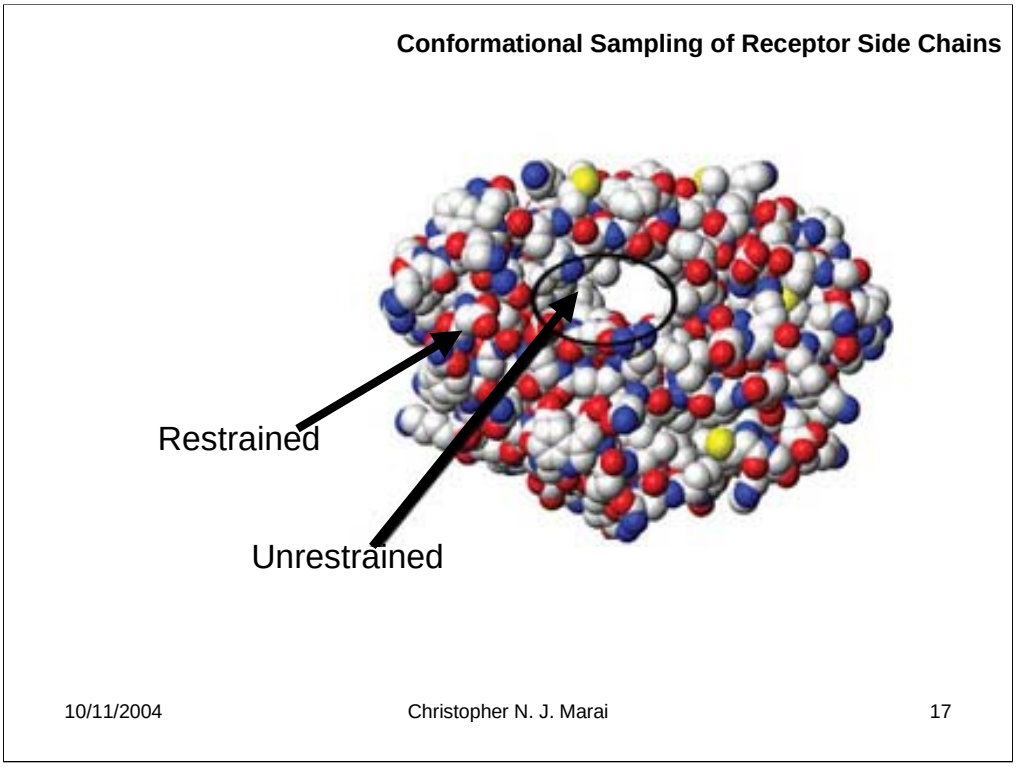


Figure: Area treated computational includes a smaller region, ie the binding site than the entire protein. Thus intensive computation is focused on area of interest – hopefully with accurate description of the whole system.

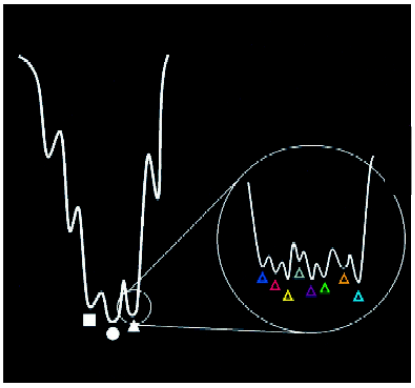
Improving upon this approach library is used then structures are minimized.

Rotamer library is used to generate likely conformations of side chains and then the docked system undergoes unrestricted minimization (optimization) in the area of the active site

Thee opt facilitating strongest contacts between ligand/receptor and all possible conformations between them .

Ensemble Methods

- A **subensemble** is a collection of structurally similar and nearly energetically equivalent conformations of a protein making up the folded state



Different colored triangles represent different conformations in the subensemble

Depending upon the scale you are investigating proteins involve many LEVELS of structure.

Ensemble Methods

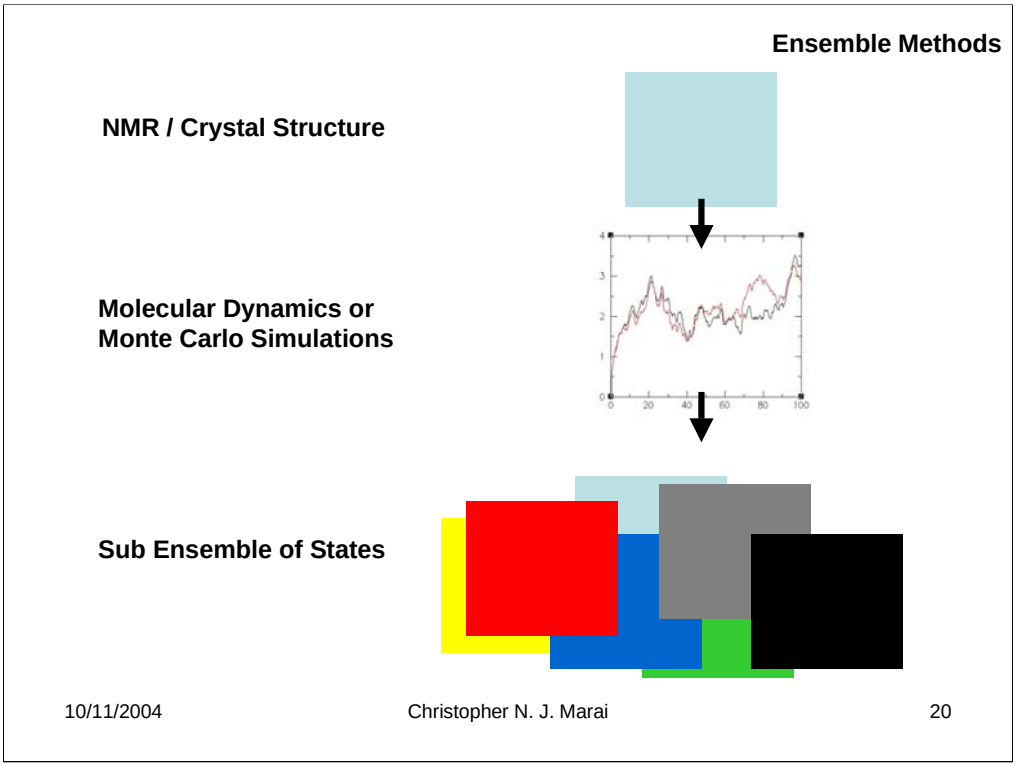
- MC or MD simulations are rigorous ways to obtain ensembles from crystal structures, (*typically with in 1kcal/mol exp.*)
- Computationally expensive
- Extract many structures then use individual conformers from the simulation to test a ligand with *(compares favorably to many NMR snap shots)*
- NOTE: It's not ligand binding by MD

10/11/2004 Christopher N. J. Marai 19

Since we have x-ray and NMR structures available to us but they are snap shots we could use MD or MC simulations to generate more of the possible lowest energy states in which the protein exists. This approach gives a good indication of the states.

IE: Take a picture of a party, based upon what we know about what is in the snap shot, generate more snap shots which hopefully represent what went on at the party!

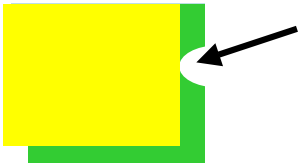
Lamb ML and Jorgensen WL (1997) Computational approaches to molecular recognition. *Curr Opin Chem Biol* **1:** 449-457.



Pictorial view of the creation of several snap shot states of the initial experimentally derived (NMR XRAY) structure.

Overlay Model

- Incorporation of multiple crystal structures *(overlay 9 inhibitor complexes of a receptor then used to create a composite binding site)*
- A map of the surface area that encompasses all the inhibitors used to describe the shape available in the receptor

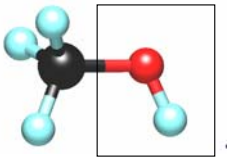


- Small molecules then docked to one receptor using conjugate gradient minimization of the ligands and all residues within 5A of ligands
- The docked structure compared to the surface area made by the 9 structures

So your ligand is docked to a single structure then it is compared to the composite binding site to determine if that single docked picture was accurate in light of the more dynamic representation of the protein.

Pharmacophore Model

- Methanol molecules were docked with 11 receptor conformations to create a model
- Determines the complementary positions for H-bonding groups (*important for ligand binding*)
- Overlay structures to identify complementary binding regions, thus describing the 'dynamic receptor'



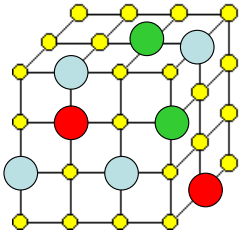
Methanol

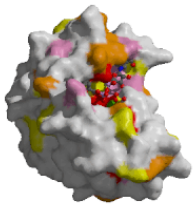
Methanol being a hydrophobic/hydrophilic molecule can interact with such sites on the receptor. It acts like a model drug to scope out important details of the binding site. Since it is small docking is quick. Docking with 11 structures gives some idea of the dynamic nature of the receptor.

Masukawa KM, Carlson HA and McCammon JA (2000) Technique for developing a pharmacophore model that accommodates inherent protein flexibility: An application to HIV-1 integrase, in *Pharmacophore Perception, Development, and Use in Drug Design* (Güner OF ed)in press, International University Line, La Jolla, CA.

Grid Method

- A grid contains the shape and specificity of the receptor
- Grid contains hydrophobic regions, hydrogen bond donors/acceptors, charged regions etc
- MD often used to make grid, weighted averages of grids for individual conformations are considered





This is essentially coarse modeling of the receptor site.

Knegtel RMA, Kuntz ID and Oshiro CM (1997) Molecular docking to ensembles of protein structures. *J Mol Biol* **266:** 424-440.

Dock via MC to many

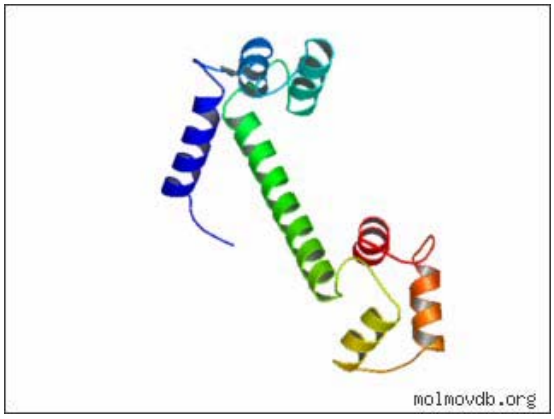
- Explicitly dock a ligand to numerous protein conformations
- Docked 2 different ligands to 10 different protein conformations
- MC simulated annealing minimization docked **flexible ligand** into **rigid protein** structures
- Soft-scoring is used for the receptor
- Results showed 2nd ligand bound almost equally to 4 conformations

More statistical based method.

Bouzida D, Rejto PA, Arthurs S, Colson AB, Freer ST, Gehlhaar DK, Larson V, Luty BA, Rose PW and Verkhivker GM (1999) Computer simulations of ligand-protein binding with ensembles of protein conformations: A Monte Carlo study of HIV-1 protease binding energy landscapes. *Int J Quantum Chem* **72:** 73-84.

Predicting Loop Flexibility and Domain Motions

- ligands in process of binding can also alter large scale receptor geometries (domain motions)
- ie: induced ligand binding



Calmodulin (CaM) - A Calcium Binding Protein. Upon Ca^{2+} binding large scale domain motions transpire

So imagine the binding site changes when the ligand binds.

The actual site of the ligand binding would be affected perhaps like a hand closing on the ligand. IE: induced binding.

In CaM Ca2+ binding at the end of an alpha helix disrupts favorable dipole interactions and results in destabilized helix which allows large domain of the protein to move flexibly.

Figure: Is actually a movie and should show the described a-helix motions www.molmovdb.org

3D Hinges

- a challenging task to drug design is predicting large domain motions
- can be done by identifying large stable regions of protein are modeled as being connected to **3D hinges** (could be a bond rotation etc.)
- hinges can be used for flexible points in the ligand also
- docking in such cases is scored by matching appropriate surface areas (which is more coarse, no side chain flex, or structural optimization)

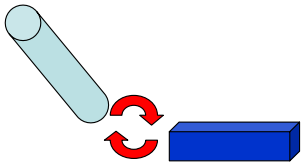


Figure: 2 large domains connected by a hinge. Hinge could be 3d also.

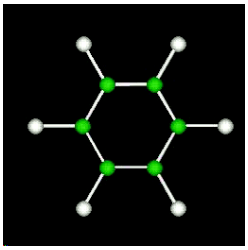
The domain surfaces are assessed to identify possible areas of interaction and reaction.

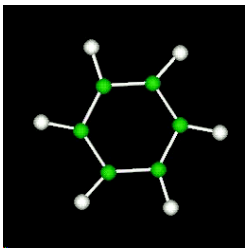
Sandak B, Wolfson HJ and Nussinov R (1998) Flexible docking allowing induced fit in proteins: Insights from an open to closed conformational isomers. *Proteins* **32:** 159-174.

Predicting Loop Flexibility and Domain Motions

Harmonic Modes

- Harmonic modes have been used to described such flexibility
- Mobility of large regions of proteins are characterized by low vibrational frequencies
- Have used relaxations of harmonic modes (since they are used to describe flexibility) to improve steric fit





Zacharias M and Sklenar H (1999) Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: Application to DNA minor groove ligand complex. *J Comput Chem* **20:** 287-300.

Figures: Note should show vibrational frequencies of the benzene rings in motion not displayed here.

Stability Model

- Map out stability of a protein on to individual residues (can be compared with H exchange protection factors from NMR)
- Compared stability for all residues when bound to 13 inhibitors
- Diff in inhibitors propagated into the structural and dynamic changes throughout the protein
- Also described the dual-character binding site (one part static, other plastic)

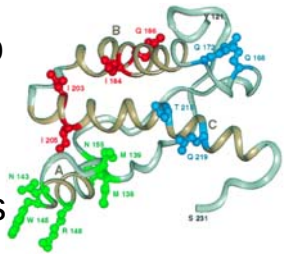


Figure: Indicates areas of interest on protein.

Bardi JS, Luque I and Freire E (1997) Structure-based thermodynamic analysis of HIV-1 protease inhibitors. *Biochemistry* **36:** 6588-6596.

Take Home Message

- **Dynamic behavior of receptors complicates drug design, the following attempt to account for this. . .**
- **Soft docking**
  - Allow for some overlap to sample flexibility
- **Conformational Sampling of Side Chains**
  - Use many conformations of side chains in active site area
- **Ensemble Methods**
  - Sample multiple conformations generated from one
- **Domain Motions**
  - Model large scale motions of the protein

VERY Brief GENERAL scope of what's going on.