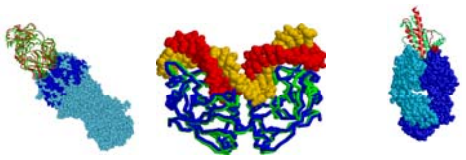
Rigid Protein-Protein Docking



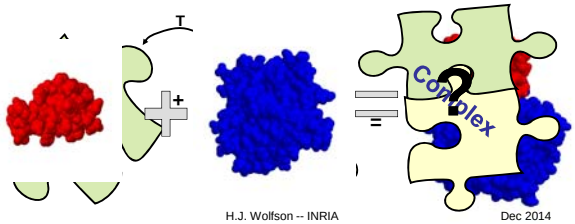
Haim J. Wolfson
School of Computer Science
Tel Aviv University

The Docking Problem

- Input : A pair of molecules represented by their 3D structures.
- Tasks :
  - Decide whether the molecules will form a complex (interact/bind).
  - Determine the binding affinity.
  - **Predict the 3D structure of the complex.**
  - Deduce function.

The (Pairwise)Docking Problem

Given 2 input molecules in their native conformation, the goal is to find the "native" 3D structure of their complex.



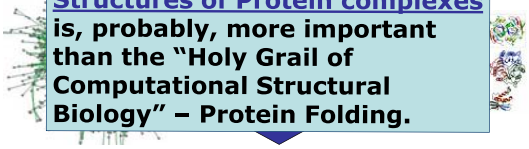
Biological Motivation

- **Proteins act by interaction – assembly and disassembly of multimolecular complexes.**
- **Drug development:**
  - **Disruption of multi-molecular interactions.**
  - **Design of protein-drug complexes.**
- **Structural Elucidation of the Large Molecular Machines of the Cell – Ribosome, Proteasome etc.**

Bioinformatics Motivation

- Large amounts of data with putati… the intera… w, due t… in the
- The number of protein-…

Given the current state of the art of Experimental structure elucidation methods, Computational Prediction of the Structures of Protein complexes is, probably, more important than the "Holy Grail of Computational Structural Biology" – Protein Folding.



Computational docking methods that predict the structures of protein complexes, are becoming indispensable tools.

H.J. Wolfson -- INRIA Dec 2014

Forces governing biomolecular recognition

Depend on the molecules involved and the solvent.

- Van der Waals.
- Electrostatics.
- Hydrophobic contacts.
- Hydrogen bonds
- Salt bridges .. etc.

All interactions act at short ranges.

Implies that a necessary condition for tight binding is molecular surface complementarity.

H.J. Wolfson -- INRIA Dec 2014

Geometric Docking Algorithms

- Based on the assumption of shape complementarity between the participating molecules.
- Molecular surface complementarity - protein-protein, protein-ligand, (protein - drug).
- Hydrogen donor/acceptor complementarity - protein-drug.

Remark : usually "protein" here can be replaced by "DNA" or "RNA" as well.

H.J. Wolfson -- INRIA Dec 2014

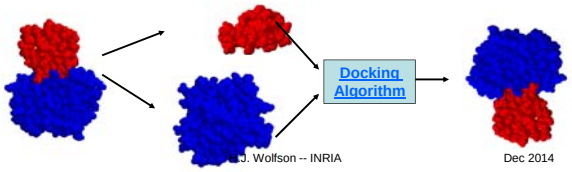
Issues to be examined when evaluating docking methods

- **Rigid docking vs Flexible docking :**
  - **If the method allows flexibility:**
    - **Is flexibility allowed for ligand only, receptor only or both ?**
    - **No. of flexible bonds allowed and the cost of adding additional flexibility.**
- **Does the method require prior knowledge of the active site ?**
- **Performance in "unbound" docking experiments.**
- **Speed - ability to explore large libraries.**

H.J. Wolfson -- INRIA Dec 2014

Bound Docking

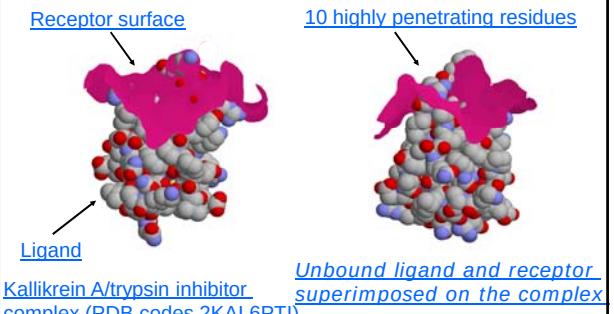
- In the bound docking we are given a complex of 2 molecules.
- After artificial separation the goal is to reconstruct the native complex.
- No conformational changes are involved.
- Used as a first test of the validity of the algorithm.



Unbound Docking

- In the unbound docking we are given 2 molecules in their native conformation.
- The goal is to find the correct association.
- **Problems:** conformational changes (side-chain and backbone movements), experimental errors in the structures.

Bound vs. Unbound

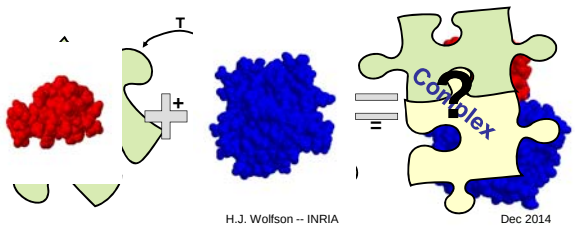


Recommended Literature – survey papers (see references of major methods therein)

- I. Halperin, B. Ma, H. Wolfson & R. Nussinov, Principles of Docking: An overview of Search Algorithms and a Guide to Scoring Functions, *PROTEINS*, 47, 409—443, (2002).
- A.M.J.J. Bonvin, Flexible protein-protein docking, *Curr. Opin. Struct. Bio.*, 16, 194—200, (2006).
- N. Andrusier, E. Mashiach, R. Nussinov, H.J.Wolfson, Principles of flexible protein-protein docking, *PROTEINS*, 73, 271—289, (2008).

The (Pairwise)Docking Problem

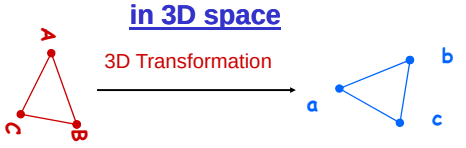
Given 2 input molecules in their native conformation, the goal is to find the "native" 3D structure of their complex.



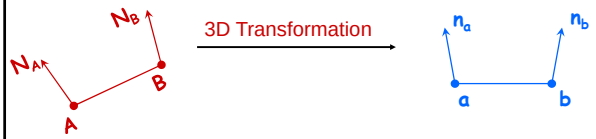
PatchDock - Motivation

Detect a 3D rigid transformation of one of the molecules that docks it to the other maximal interface and negligent shape penetration.

Geometry of complementarity detection in 3D space



• Three non-collinear points correspondence is necessary in order to compute a rigid transformation in 3D.



• Two points are enough if the normals are given.

PatchDock- *algorithm outline*:

Rigid Geometric Docking

PDB files

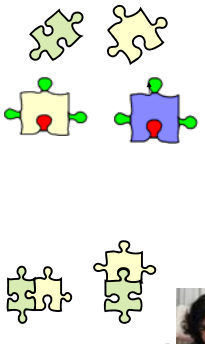
Surface Representation

Feature Extraction

Shape complementarity detection by Geometric Hashing

Scoring & Filtering

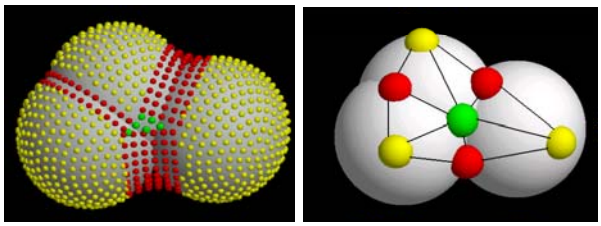
Candidate complexes



Schneidman-Duhovny et al. Proteins 2003
Duhovny (Schneidman), D., Nussinov, R. Wolfson,HJ. WABI 2002

Surface Representation (sampling)

- Dense MS surface (Connolly)
- Sparse surface (Shuo Lin et al.)



Connolly's MS algorithm

- A 'water' probe ball (1.4-1.8 A radius) is rolled over the van der Waals surface.
- Smoothes the surface and bridges narrow 'inaccessible' crevices.

Connolly's MS algorithm - cont.

- Convex, concave and saddle patches according to the no. of contact points between the surface atoms and the probe ball.
- Outputs points+normals according to the required sampling density (e.g. 10 pts/A^2).

Critical points based on Connolly rep. (Lin, Wolfson, Nussinov, Proteins 1994)

- Define a single point+normal for each patch.
- Convex-caps, concave-pits, saddle - belt.

Active Site Focusing (optional)

There are major differences in the interactions of different types of molecules (protease-inhibitor, antibody-antigen, protein drug). Studies have shown the presence of energetic ***hot spots*** in the active sites of the molecules.

Protease/inhibitor – select patches with high enrichment of hot spot residues (Ser,Gly,Asp and His for protease; and Arg,Lys,Leu,Cys and Pro for protease inhibitor).

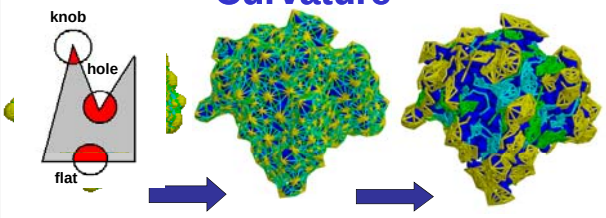
Antibody/antigen – 1.detect CDRs of the antibody.
2. select hot spot patches (Tyr,Asp,Asn,Glu,Ser and Trp for antibody; and Arg,Lys,Asn and Asp for antigen)

Protein/drug – select largest protein cavity (highest value of average shape function for the patch)

Local Feature Extraction

- Connolly points + normals - dense.
- Lin et al. points - sparser.
- Knobs - holes (Connolly; Norel-Nusinov-Wolfson) – sparse crude curvature evaluation.

Segmentation by Crude Curvature



- The surface is segmented into connected components (patches), which preserve the essence of the shape (convex, concave, flat - almost equal area patches).
- The shape complementarity step is initially computed between patches (convex-concave/flat etc.)

Local Patch Detection

Goal: ***divide the surface into connected, non-intersecting, equal sized patches of critical points with similar curvature.***

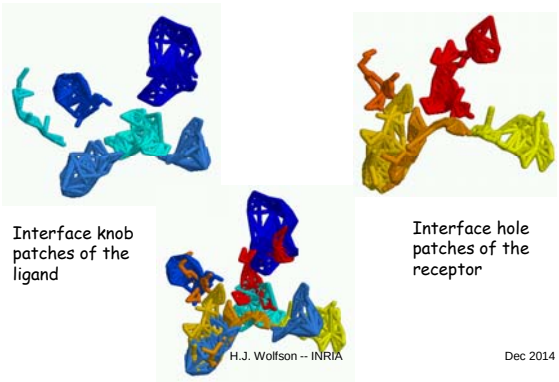
- **connected** – the points of the patch correspond to a connected sub-graph of critical point.
- **similar curvature** – all the points of the patch correspond to only one type: knobs, flats or holes.
- **equal sized** – to assure better matching we want shape features of almost the same size.

Patch Detection by Segmentation

- Construct a sub-graph for each type of points: knobs, holes, flats. For example G_{knob} will include all surface points that are knobs and an edge exists between two 'knobs' if they belong to the same atom.
- Compute **connected components** of every sub-graph.
- Problem: the sizes of the connected components can vary.
- Solution: apply '**split**' and '**merge**' routines.

H.J. Wolfson -- INRIA Dec 2014

Complementarity of the Patches:

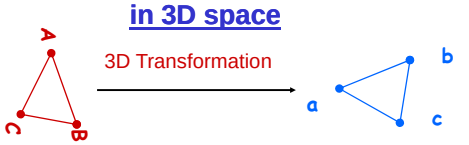


Interface knob patches of the ligand

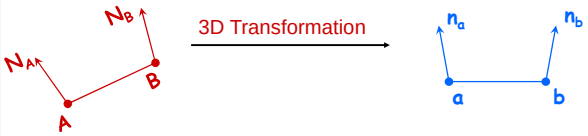
Interface hole patches of the receptor

H.J. Wolfson -- INRIA Dec 2014

Geometry of complementarity detection in 3D space



• Three non-collinear points correspondence is necessary in order to compute a rigid transformation in 3D.



• Two points are enough if the normals are given.

H.J. Wolfson -- INRIA Dec 2014

Single Patch Matching

Receptor hole patch

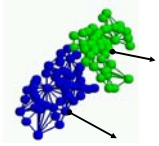
Ligand knob patch

- **Basis:** a pair of critical points with their normals from one patch.
- Match every basis from a receptor patch with all the bases from complementary ligand patches.
- Compute the transformation for each pair of matched bases.

H.J. Wolfson -- INRIA Dec 2014

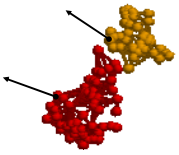
Patch-Pair Matching

Receptor patches



Transformation

Ligand patches



Basis: 1 critical point with its normal from one patch and
1 critical point with its normal from a neighboring patch.

- Match every basis from the receptor patches with all the bases from complementary ligand patches.
- Compute the transformation for each pair of matched bases.

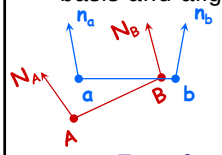
Geometric Hashing

Preprocessing:

Insert all ligand bases into a hash table,
Using a transformation invariant hash-key.

Recognition:

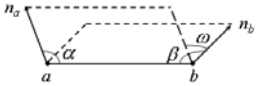
Access the hash table with each receptor basis and align matching ligand bases.



Transformation

Hash Table Key is Invariant to the Rigid (Euclidean)Transformation

- Euclidean and geodesic distances between the points: **dE, dG**
- The angles ***α, β*** between the [a,b] segment and the normals
- The torsion angle ***ω*** between the planes



5D key:
dE, dG, α, β, ω

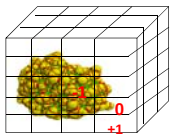
Two bases are matching if their keys are similar up to a pre-defined threshold vector

Pose Clustering, Clash Detection & Scoring Stage

- Since local features are matched, we usually have multiple instances of "almost" the same transformation.
- Some transformations may induce steric clashes.
- Pose clustering, steric clash filtering and scoring are applied to the transformation list.

Steric Clash (Penetrations) Filtering

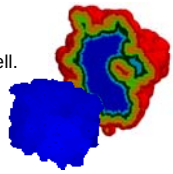
- Define the Distance Transform Grid, which stores for each voxel its distance from the surface of the molecule. The distance is negative inside the molecule and positive outside.
- Steric clashes are detected by accessing the **receptor** grid with **ligand surface** points.



H.J. Wolfson -- INRIA Dec 2014

Scoring

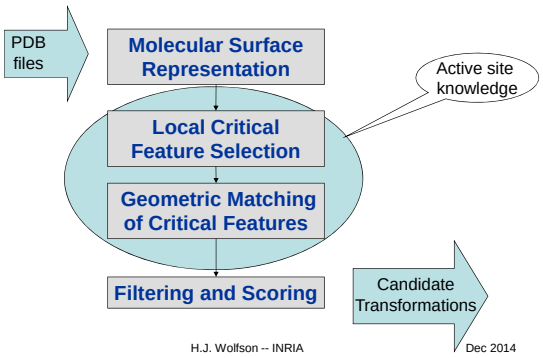
- The surface of the **receptor** is divided into five shells according to the distance function: **S1-S5**
- The number of **ligand** surface points in every shell is counted.
- The geometric score is a weighted sum of the number of ligand surface points inside every shell.



- **Multi-resolution surface** data structure was developed to speed up this stage.

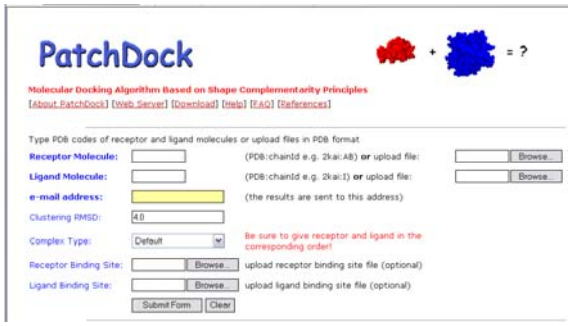
H.J. Wolfson -- INRIA Dec 2014

Geometric Docking Algorithms



H.J. Wolfson -- INRIA Dec 2014

WWW server: http://bioinfo3d.cs.tau.ac.il/PatchDock

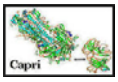


Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. NAR 2005

H.J. Wolfson -- INRIA Dec 2014

CAPRI

Critical **A**ssessment of **PR**ediction of **I**nteractions

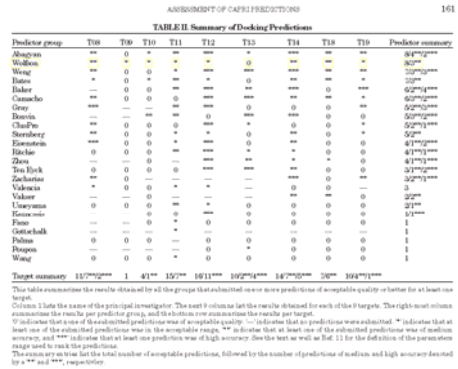


CAPRI community-wide experiment on the co ein

Orga lak

Ranking	Conditions based on Capri computed parameters
High	$f_{nat} \geq 0.5$ AND (L_rms ≤ 1.0 OR I_rms ≤ 1.0)
Medium	($f_{nat} \geq 0.3$ AND $f_{nat} < 0.5$) AND (L_rms ≤ 5.0 OR I_rms ≤ 2.0) OR $f_{nat} \geq 0.5$ AND L_rms > 1.0 AND I_rms > 1.0
Acceptable	($f_{nat} \geq 0.1$ AND $f_{nat} < 0.3$) AND (L_rms ≤ 10.0 OR I_rms ≤ 4.0) OR $f_{nat} \geq 0.3$ AND L_rms > 5.0 AND I_rms > 2.0
Incorrect	$f_{nat} < 0.1$ OR (L_rms > 10.0 AND I_rms > 4.0)

10 predictions can be submitted by each group



ASSESSMENT OF CAPRI PREDICTIONS 161

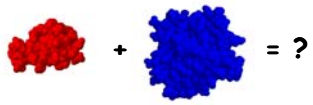
TABLE II. Summary of Docking Predictions

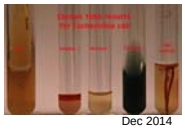
Predictor group	T08	T09	T10	T11	T12	T13	T14	T18	T19	Predictor summary
Abagyan	**	0	*	**	***	*	***	**	**	8/4***/***
Wodak	**	*	*	*	*	0	**	**	*	8/2**
Weng	**	0	0	*	***	***	***	**	**	7/2**/***
Bates	*	0	*	**	*	0	**	**	*	7/2**
Baker	—	0	0	**	***	**	***	0	***	6/2**/***
Camacho	**	0	0	0	***	***	**	**	*	6/2**/***
Gray	***	—	—	**	***	0	0	0	**	5/2**/***
Bonvin	—	—	**	**	0	***	***	0	0	5/2**/***
ChaPro	**	0	0	0	***	*	0	0	*	5/2**/***
Sternberg	**	0	0	*	*	0	**	0	*	5/2**
Eisenstein	***	0	0	*	***	0	**	0	0	4/1**/***
Ritchie	0	0	0	**	***	*	*	0	0	4/1**/***
Zhou	—	—	0	—	***	**	*	*	0	4/1**/***
Ten Eyck	0	0	0	0	***	***	**	0	0	3/1**/***
Zacharias	**	0	—	—	—	—	***	0	**	3/2**/***
Valencia	*	0	0	*	*	—	0	0	—	3
Vakser	—	—	0	—	—	—	**	**	0	2/2**
Umeyama	0	0	0	**	*	0	0	0	0	2/1**
Kozakov	—	—	0	0	***	0	0	0	0	1/1***
Fiser	—	—	0	*	0	0	0	0	0	1
Gottschalk	—	—	—	*	—	—	—	—	—	1
Palma	0	0	0	—	0	0	0	0	0	1
Poupon	—	—	—	—	0	*	0	0	0	1
Wang	0	0	0	*	0	0	0	0	0	1
Target summary	11/7***/***	1	4/1**	15/7**	16/11***	10/2**/***	14/7***/***	7/0**	10/4**/***	

This table summarizes the results obtained by all the groups that submitted one or more predictions of acceptable quality or better for at least one target.
Column 1 lists the name of the principal investigator. The next 9 columns list the results obtained for each of the 9 targets. The right-most column summarizes the results per predictor group, and the bottom row summarizes the results per target.
'0' indicates that none of the submitted predictions was of acceptable quality. '—' indicates that no predictions were submitted. '*' indicates that at least one of the submitted predictions was in the acceptable range, '**' indicates that at least one of the submitted predictions was of medium accuracy, and '***' indicates that at least one prediction was of high accuracy. See the text as well as Ref. 11 for the definition of the parameters range used to rank the predictions.
The summary section lists the total number of acceptable predictions, followed by the number of predictions of medium and high accuracy denoted by a '**' and '***', respectively.

From: Mendez, Leplae, Lensink, Wodak, Assesment of CAPRI Predictions in Rounds 3-5 Shows Progress in Docking Procedures, PROTEINS 60: 150-169 (2005).

The Real Challenge: Can we help biologists?

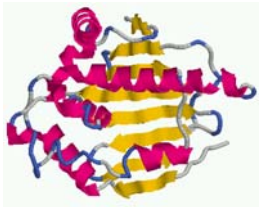


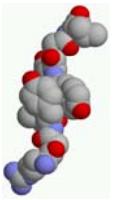


Identification of the N-terminal peptide binding site of GRP94

GRP94 - Glucose regulated protein 94

VSV8 peptide - derived from vesicular stomatitis virus





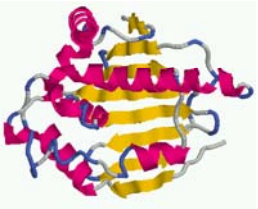
Gidalevitz T, Biswas C, Ding H, Schneidman-Duhovny D, Wolfson HJ, Stevens F, Radford S, Argon Y. J Biol Chem. 2004

Biological motivation

- The complex between the two molecules strongly stimulates the response of the T-cells of the immune system.
- The grp94 protein alone does not have this property. The activity that stimulates the immune response is due to the ability of grp94 to bind different peptides.
- Characterization of peptide binding site is highly important.

GRP94 molecule

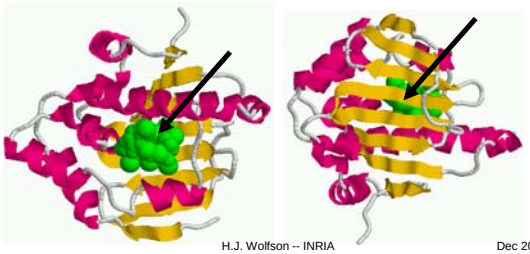
- There was no structure of grp94 protein. Homology modeling was used to predict a structure using another protein with 52% identity.



- Recently the structure of grp94 was published. The RMSD between the crystal structure and the model is 1.3A.

GRP94 molecule

- There is a binding site for inhibitors between the helices.
- There is another cavity produced by a β-sheet on the opposite side.



Docking

- PatchDock was applied to dock the two molecules, without any binding site constraints.
- Interestingly, the better scoring docking results were clustered in the two cavities:

Experimental Verification

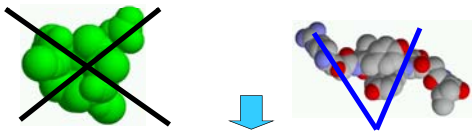
Goals:

- Try to eliminate one of the binding site hypotheses.
- Find the positions of the amino acids which are important for peptide bindi

H.J. Wolfson -- INRIA Dec 2014

Experimental Verification 1

- Experimental data shows that the inhibitor and the peptide can bind simultaneously.
- Two key residues in the inhibitor binding site were mutated.
- The mutant did not bind the inhibitor, however it could still bind the peptide.

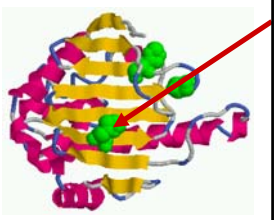


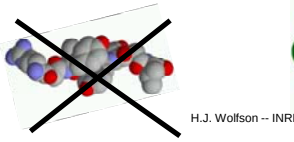
- The binding sites of the inhibitor and the peptide are probably distinct.

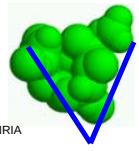
H.J. Wolfson -- INRIA Dec 2014

Experimental Verification 2

- The peptide binding was pH sensitive. Therefore involvement of a His residue was suspected.
- **His125** was mutated to Asp and Tyr. The first mutated protein did not bind the peptide at all and the second had only partial activity.
- Both mutants were soluble and could bind the inhibitor.

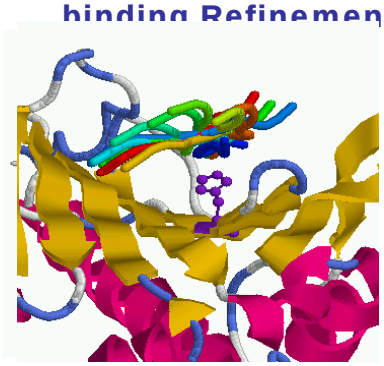






H.J. Wolfson -- INRIA Dec 2014

Computational Verification and binding Refinement



H.J. Wolfson -- INRIA Dec 2014

What did we learn?

- Rigid Docking by shape complementarity works efficiently, when there are no significant conformational changes. Will work also for non-protein molecules.
- Docking can help in guiding wet lab experiments.
- Experimental results can guide docking.
- Scoring and ranking of the resulting hypotheses is a major challenge.

Some PatchDock Publications

- **D. Duhovny, R. Nussinov, H.J. Wolfson**, *Efficient Unbound Docking of Rigid Molecules,* 2'nd Workshop on Algorithms in Bioinformatics (WABI'02), Sept. 2002, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag.
- **D. Schneidman-Duhovny, et al.**, *Taking Geometry to its Edge: Fast Unbound Rigid (and Hinge-bent) Docking*, Proteins, 52, 107—112, (2003).
- **D. Schneidman-Duhovny, Y. Inbar, R. Nussinov and H. J. Wolfson,** *PatchDock and SymmDock: servers for rigid and symmetric docking,* Nuc. Acids Res., 33 (NAR, web server issue), W363—W367, (2005).