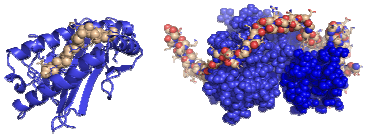
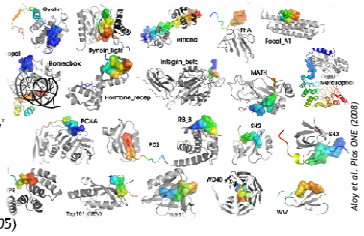
Rosetta FlexPepDock – High-Resolution Refinement and Docking of Flexible Peptides



Barak Raveh & Nir London,
The Furman lab, the Hebrew University in Jerusalem

Peptide-Protein Interactions

- Some of the most prevalent interactions
- Mediate important processes:
  - Signaling
  - Transport
  - Transcription
  - Localization
- Systems Biology*



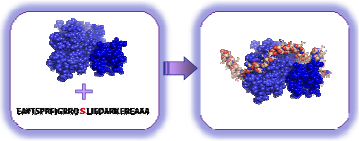
Aloy et al. PloS ONE (2008)

*Olfield et al., Proteins (2005)

Peptide Folding upon Binding

Receptor:
a globular protein

Ligand:
the flexible peptide



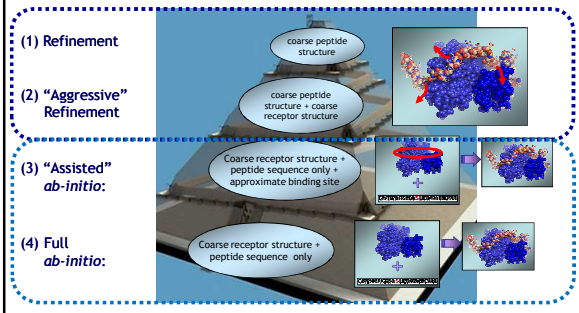
Degrees of freedom:
Peptide flexibility / Rigid body orientation / Side-chains / Receptor flexibility

Goals:

- Model complex structure → **Rosetta *FlexPepDock***
- Design peptide specificity and peptide inhibitors
- Model dynamics of peptide binding to proteins → Rosetta *PathRover**

* Raveh et al., PLOS CB 2009

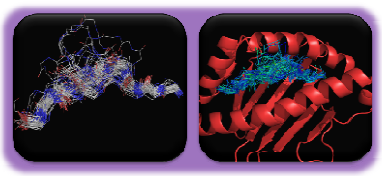
Peptide Modeling: Levels of Input Resolution



First Step: High-Resolution Refinement Protocol

- In many cases, this is the crucial step:
  - Binding site common to diverse peptides
    - *PDZ* domains, *SH3* domains, *MHC* etc.
  - Homology receptor models
  - Peptide redesign – diverse backbones in same pocket

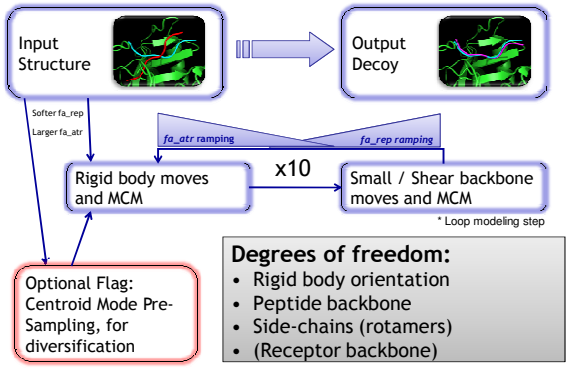
Variability in bound *MHC* peptides:



Peptide Docking in Known Binding Pockets – Previous Work

- Docking study of peptide-PDZ domains:
  Niv and Weinstein, 2005
- Docking studies of peptide-MHC proteins:
  Tzakos et al. 2004 ;
  Bordner and Abagyan, 2006 ;
  Fagerberg, Cerriotini and Michielin 2006
- Design of scoring function for flexible protein-peptide docking with known native *(n=25; Go-biased sampling)*
  Liu, Dominy and Shakhnovich, 2004

FlexPepDock High-Res Refinement Protocol



Degrees of freedom:
- Rigid body orientation
- Peptide backbone
- Side-chains (rotamers)
- (Receptor backbone)

PeptiDB Dataset

London et al., 2009 (under review, see poster)

- A curated dataset of 103 protein-peptide high-resolution (<2Å) complex structures
  - Many also in unbound conformation
- Peptides with 5-15 amino-acids
- Non-redundant (70% seq. identity)

Perturbations Benchmark: What is the Effective Range of the Refinement Protocol?

Perturbations Benchmark:

126 (89 bound ; 37 unbound) peptide-protein structures from PeptiDB

x45 perturbations of increasing magnitude:

Torsions: 0° / ±25° / ±50° / ±75° / ±100°

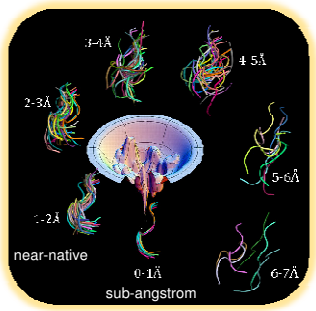
Rigid-body Translation: 0Å / 1Å / 3Å

Rigid-body rotation: 0° / 10° / 30°

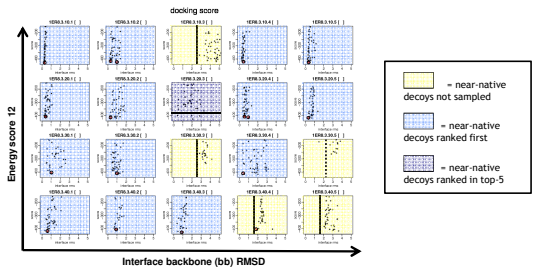
5670 starting structures **x 200** decoys for each perturbation with FlexPepDock:
→1,134,000 hi-res decoys (9,000 per complex)

Effective Range: What is the maximal deviation of start structure from native peptide structure that still allows effective refinement?

Peptide Backbone RMSDs Illustrated:

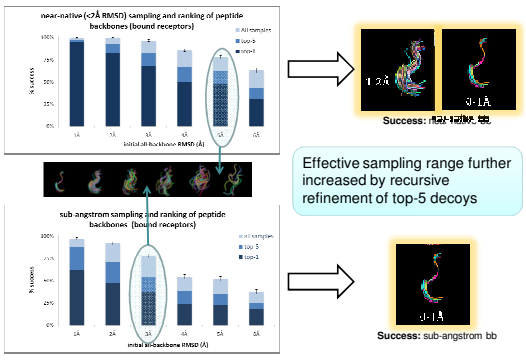


Assessing the Effective Range: Sampling + Ranking

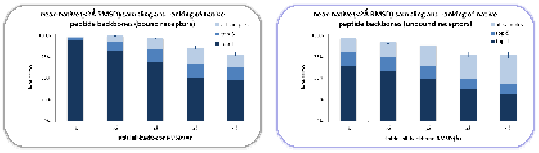


Effective Range of Refinement: Sampling & Ranking

(bound docking, w/o native side-chains)



Bound vs. Unbound Docking



Bound Unbound

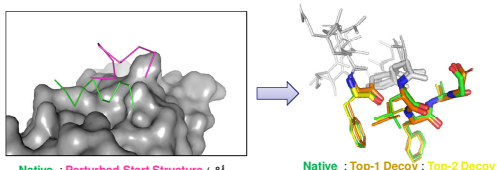
- Sampling performance remains high, but ranking quality is worse (still reasonable)
- Complies with observation that backbones of peptide binding sites are rigid (London et al., 2009, in review)

The Furman Lab Presents:

"In-Lab Blind Motif Predictions"

1T7R (Androgen growth hormone receptor)

- Residues of the binding motif accurately recovered (**FxxLF** motif)
- Decoys diverge on non-motif residues

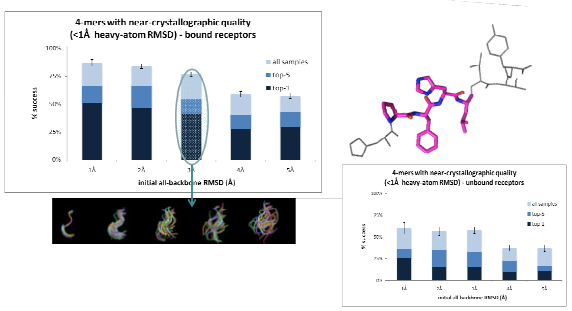


Native ; Perturbed Start Structure (~8Å bb-RMSD ; significant backbone deviation)

Native ; Top-1 Decoy ; Top-2 Decoy
➔ Our guess: FxxLFAG

Baseline for specificity prediction

➔ 4-mers test: <1Å All-Atom Modeling of *Four* Consecutive Residues

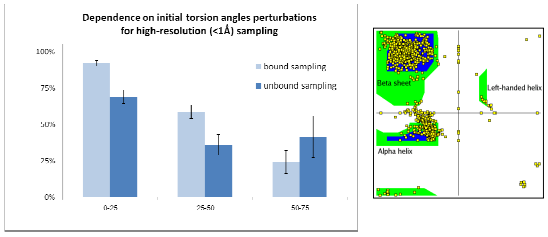


Unbound Case

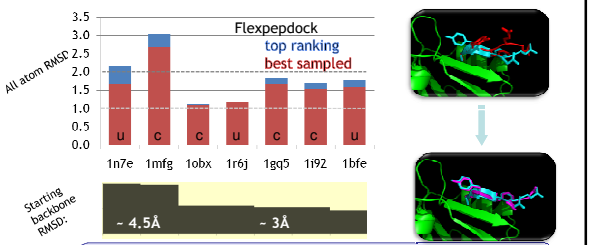
Effect of Initial Deviation in Internal Coordinated - Initial RMSD Doesn't Tell the Whole Story

•When the initial deviation in internal coordinates (φ,ψ) is large, it is harder to find a sub-angstrom decoy for the same initial RMSD

•The protocol effective range is ~±45° in torsion space (a 90° by 90° window in the Ramachandran plot)



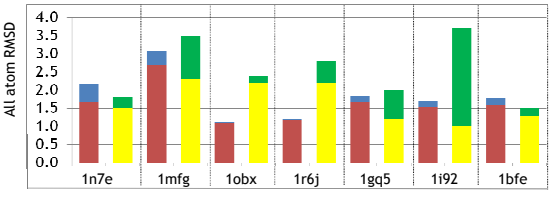
Application: Cross/Unbound Docking of PDZ Domains



- Models within ~2.0Å **all atom** RMSD (~1.2Å **backbone** RMSD), or even 1Å
- *Score12 does well:* Selected model similar to best sampled model
- *Effective Range:* starting RMSD → final RMSD

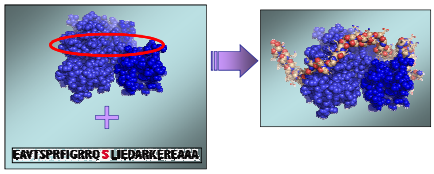
Application: Cross/Unbound Docking of PDZ Domains

Comparison to PDZ Doc Scheme* **Niv & Weinstein, 2005*

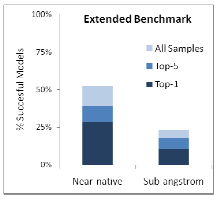


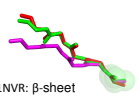
Towards *Ab-Initio*?

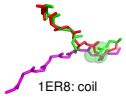
- Assume the peptide structure is not known, but the approximate binding site is known

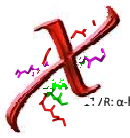


Towards *ab-initio*: Docking from an Anchored Extended Conformation



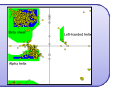






Color code:
native peptide ;
extended peptide ;
top-1 decoy

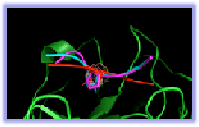
Possible solution: use ideal helices / fragments as start structures (recall that effective range of protocol is ~90°x90° window in Ramachandran plot)



"Assisted" *ab-initio:*

FTMap: prediction of ligand binding sites by saturation of small molecules (Kozakov *et al.*, 2009)

(1) Use FTMap, and other measures, to predict peptide binding sites
(Dana Attias & Nir London)
(2) Docking extended peptide with FlexPepDock

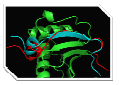


Rosetta FlexPepDock – What's next?

- Loop modeling (multiple anchors, e.g., from FTMap)
- Peptide design
- Prediction of specificity motifs
- Longer peptides
- Aggressive refinement (receptor relax, fragments, etc.)

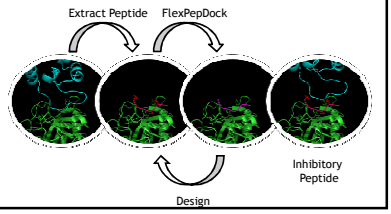
Design Protocol

Redesign peptide interactions:



Entreropathogenic type III secretion system (joint work with Ilan Rosenshine)

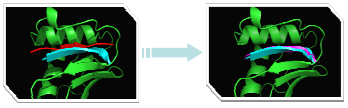
Design novel Inhibitory Peptides:



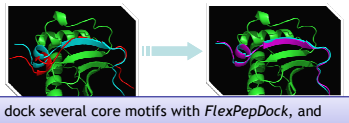
Moving to Longer Peptides – 2 Steps Protocol

- *InvB* (Effector Proteins from *Salmonella* type III secretion system)
- Long intrinsically disordered domain, with a core motif (4-5 amino-acids) revealed by sequence alignment *(Lilic et al., 2006)*

1. Dock core motif



2. Adding perturbed flanks (22 amino-acids):



Or with loop closure: dock several core motifs with *FlexPepDock*, and connect them using loop modeling modules (particularly relevant for this system)

Acknowledgments

Hebrew University, Jerusalem (Hadassah Medical Research Institute)
Nir London (joint work)
Ora Schueler-Furman
Dana Attias

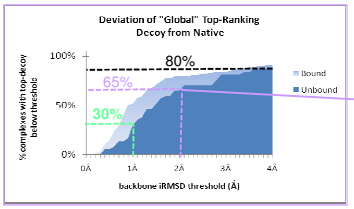
Tel-Aviv University (School of Computer Science)
Dan Halperin

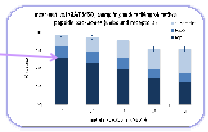
Boston University:
Dima Kozakov
Sandor Vajda

Rosetta Community:
Phil Bradley (FHCRC, Seattle)
And everybody…

Bound vs. Unbound Energy Landscape:
RMSD of "Global" Top-Ranking Decoy

Recall we created 9,000 decoys for each complex, for a very wide range of starting perturbations (0Å-15Å)
→ In how many of the complexes, the "global" top-decoy has a good RMSD?





Unbound refined to <2Å RMSD