Working with Rosetta Funnels

RosettaCon2007 Workshop

Ora Schueler-Furman
Hadassah Medical School
The Hebrew University, Jerusalem, Israel

Overview

- Introduction to Funnels
- Funnels for model selection
  - Example: entry of Anthrax into human cells
  - FunHunt: characterization of native energy funnels in energy landscape
  - Targeting of native funnels for docking and design
- Summary

Overview

- **Introduction to Funnels**
- Funnels for model selection
  - Example: entry of Anthrax into human cells
  - FunHunt: characterization of native energy funnels in energy landscape
  - Targeting of native funnels for docking and design
- Summary

Energy Landscape: Funnel around Native Structure

- Funnels describe Folding and Binding

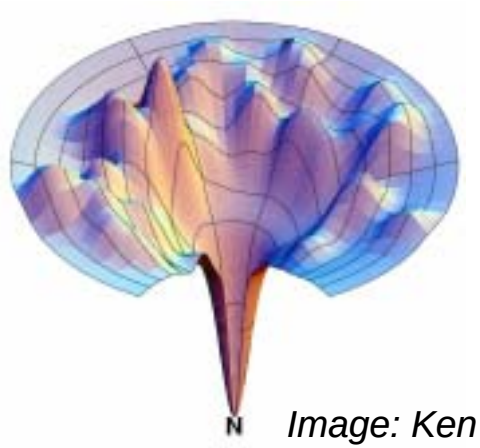
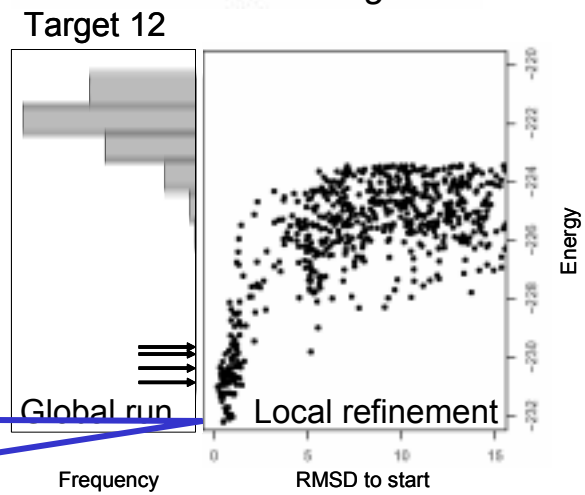


Image: Ken Dill

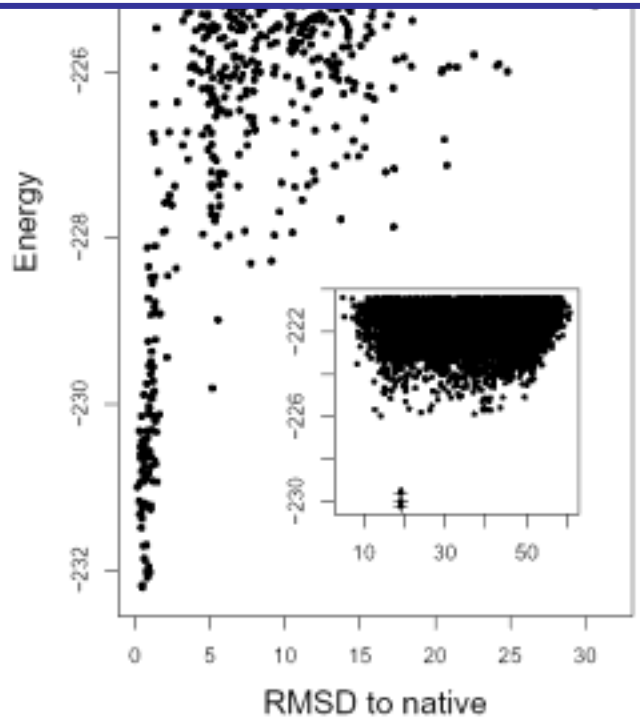
- RosettaDock high-resolution models are located at tip of funnel
- Model selection based on energy only

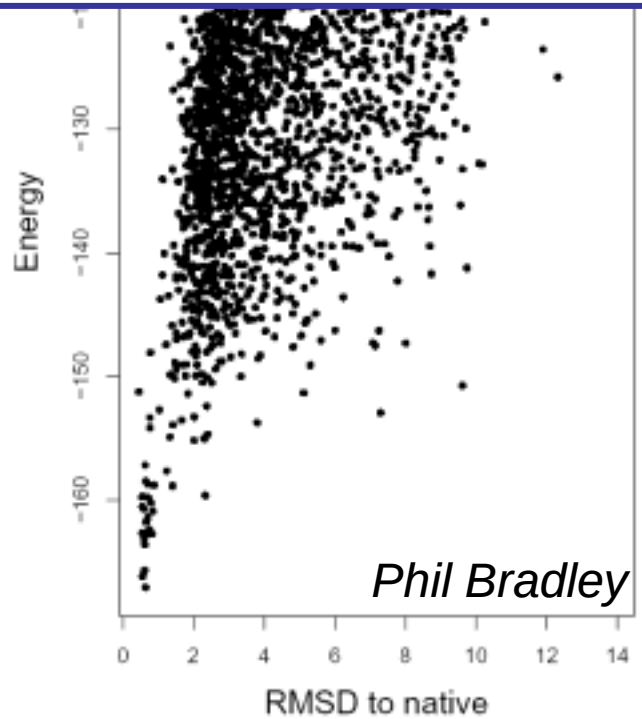


Schueler-Furman 2005

Similar Landscapes for Different Rosetta Predictions

Energy function describes well principles underlying the correct structure of monomers and complexes





Schueler-Furman et. al (2005) Science

Anthrax

with Borden Lacy & John Collier, HMS

- Introduction to Funnels
- **Funnels for model selection**
  - **Example: entry of Anthrax into human cells**
  - FunHunt: characterization of native energy funnels in energy landscape
  - Targeting of native funnels for docking and design
- Summary

Entry of LT Anthrax Toxin

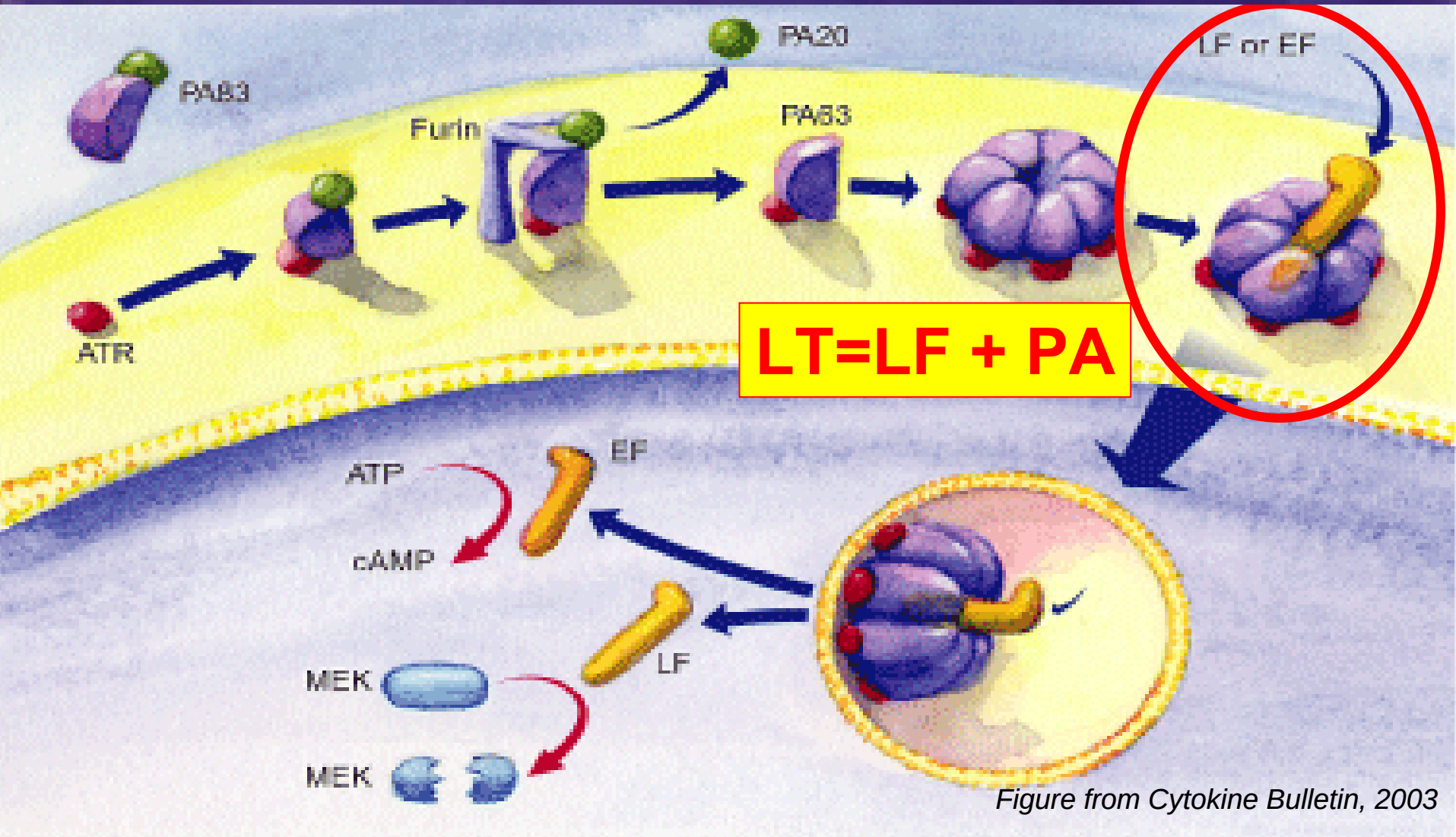
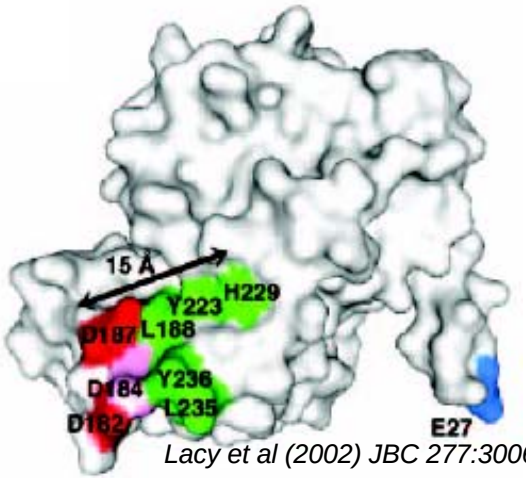


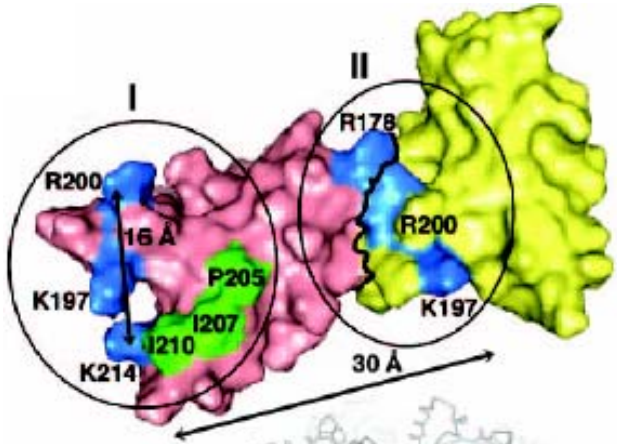
Figure from Cytokine Bulletin, 2003

Lethal Factor (LF) binds to Protective Antigen (PA)

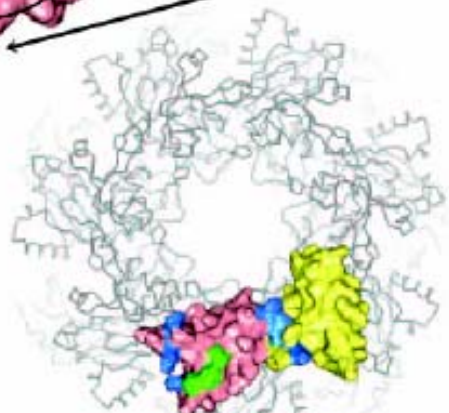
LF_n



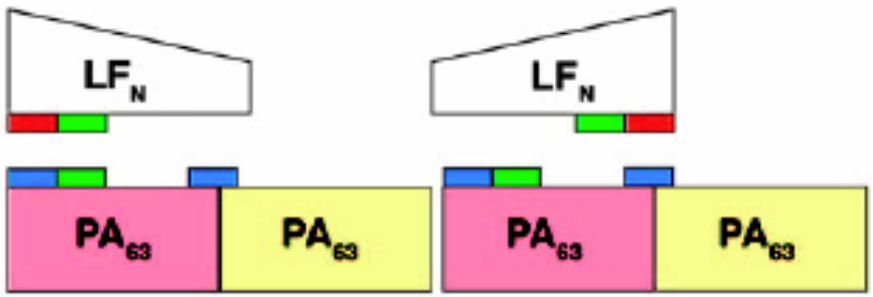
Lacy et al (2002) JBC 277:3006



PA

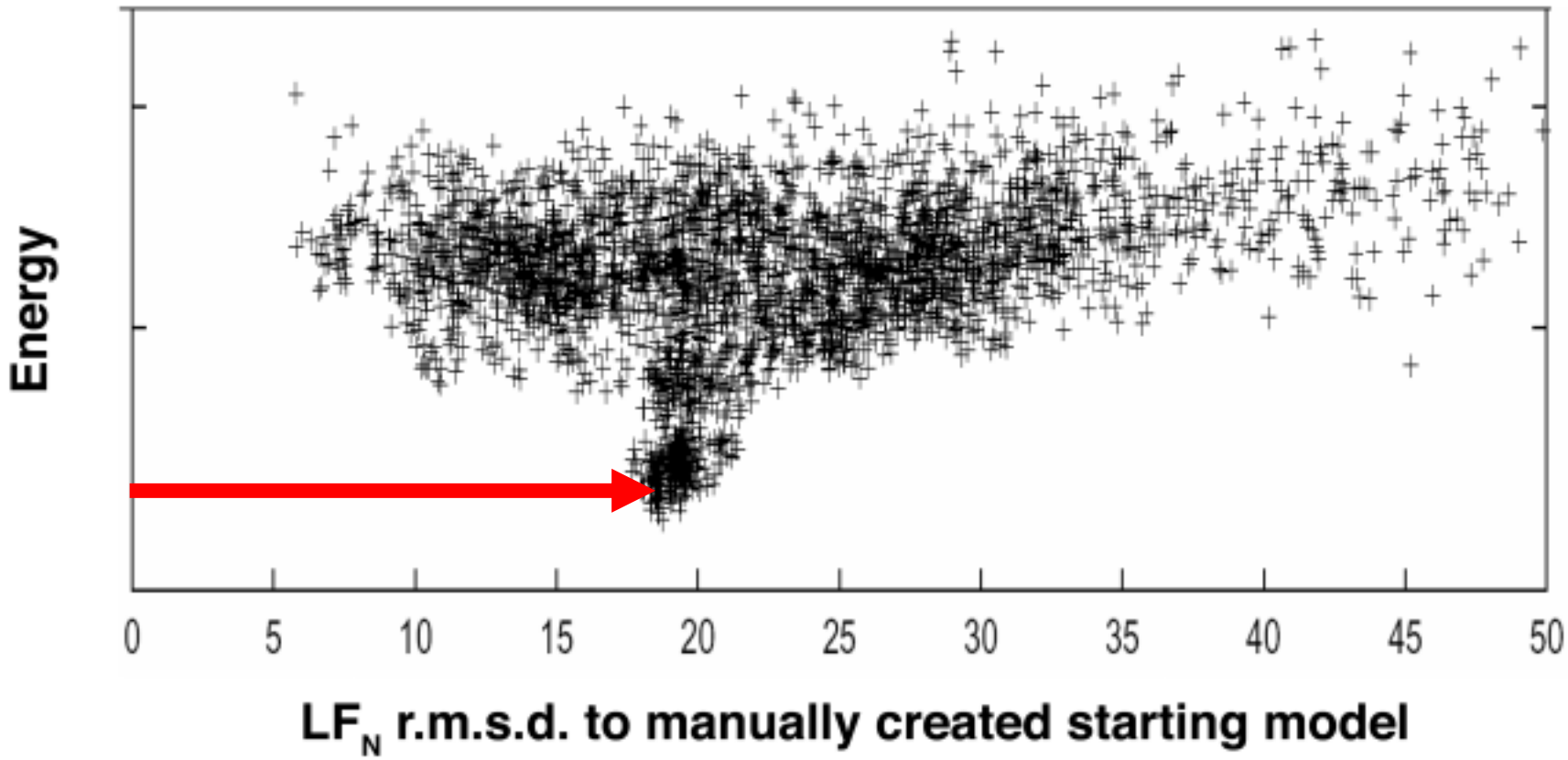


Cunningham et al (2002) PNAS 99:7049

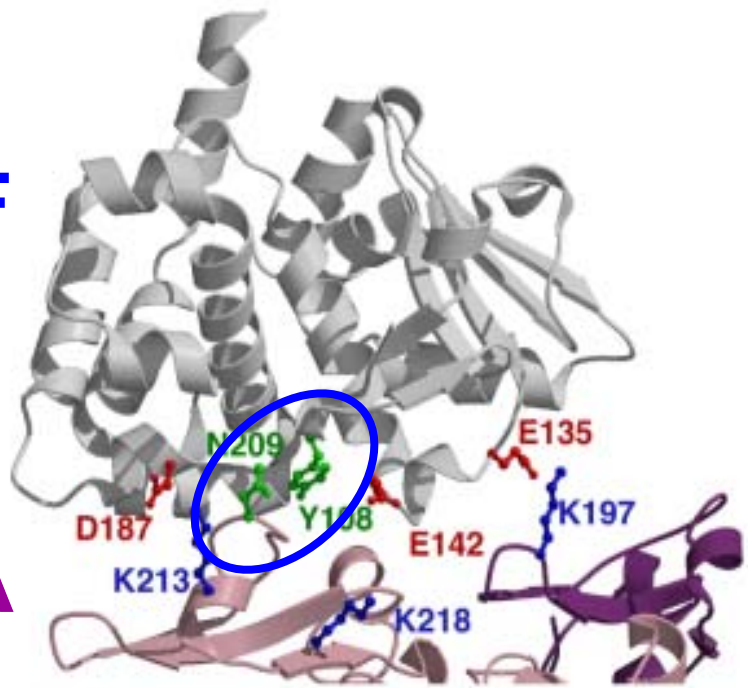


Two possibilities – which is correct?

RosettaDock Suggests Still Another Possibility…..



… that Agrees with Additional Experiments …



Disulfide crosslinking

+

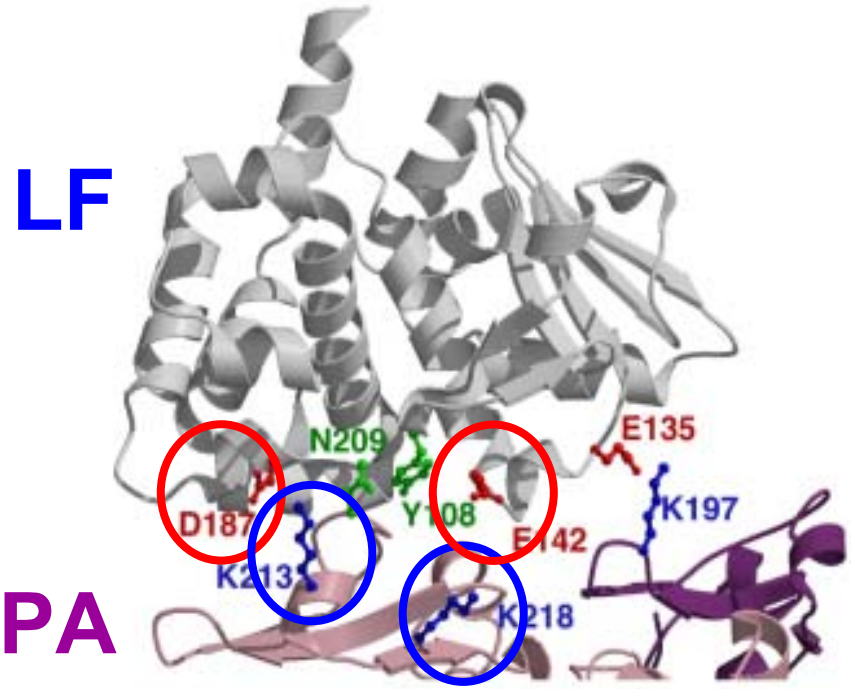
* **N209** *vs.* **Y108**

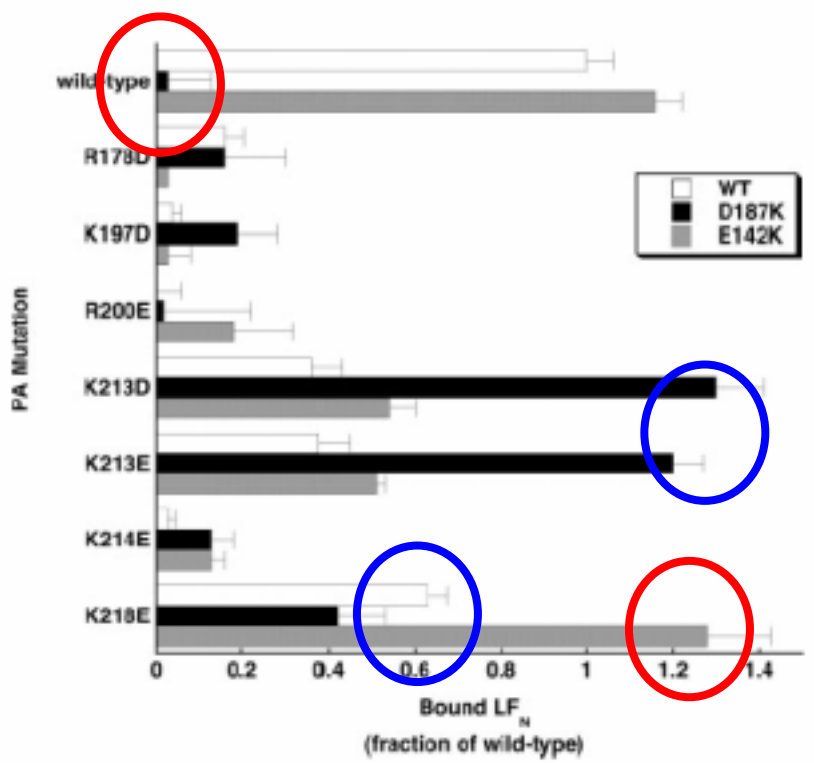
-

* **N209** *vs.* K110,Y118, Q132, S134, D136, Q228
* Y108 *vs.* S186

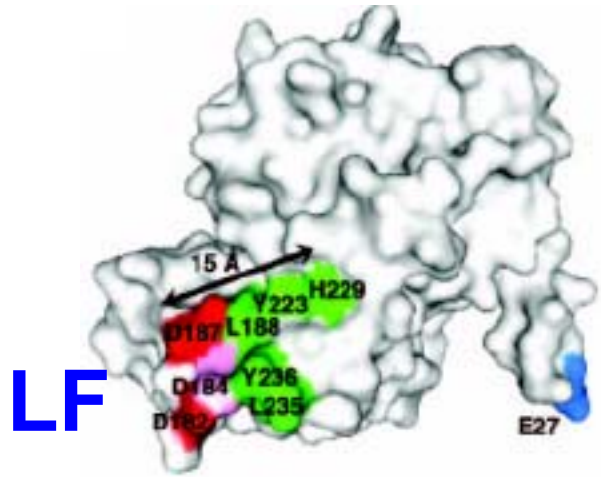
… that Agrees with Additional Experiments …

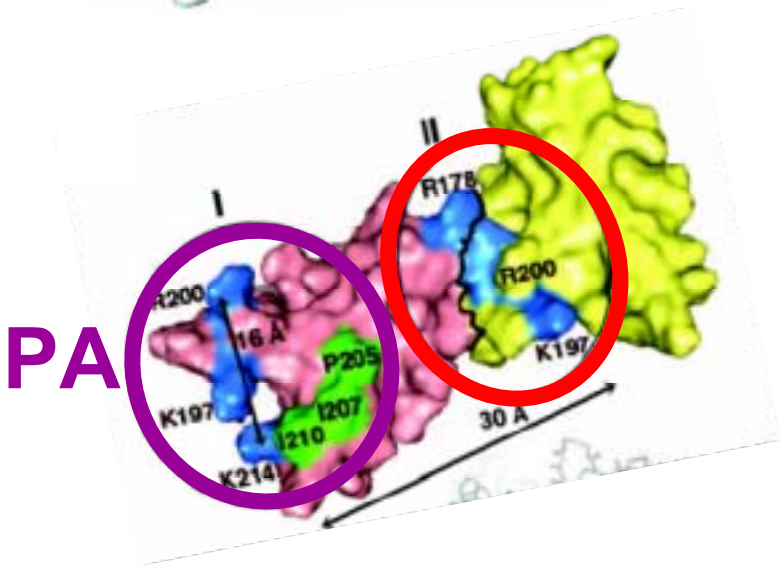
Binding assay for charge-reversal mutations

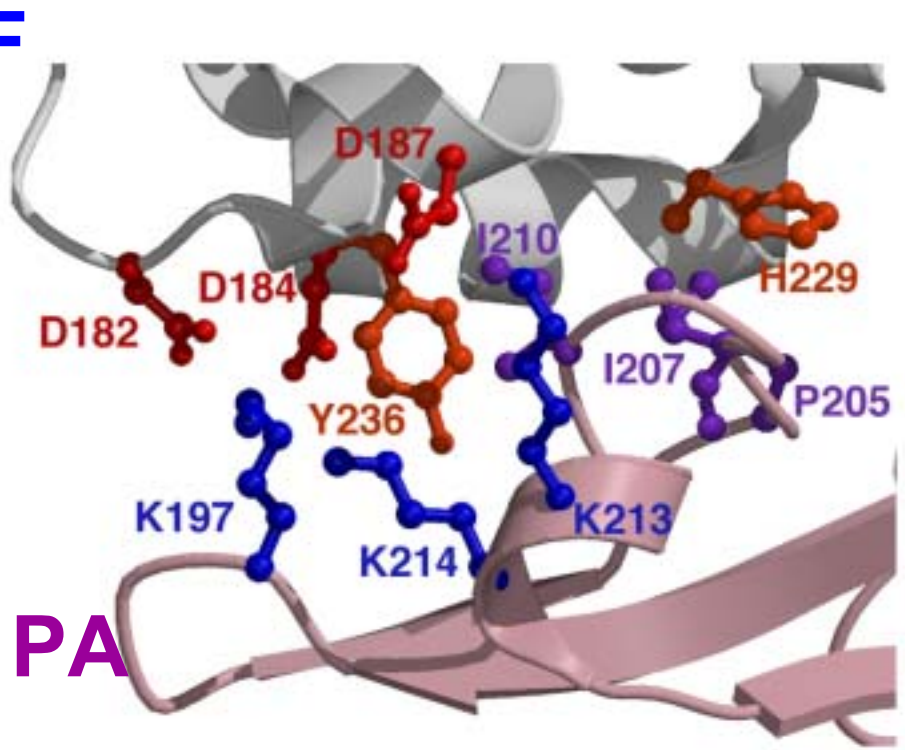




… and Changes the Concept







Site I – binds to LF

Site II – PA dimerization

LF binds to PA - Conclusions

- LF binds to PA via Site I only
- Site II is responsible for multimerization of PA

→ Purely energy-based prediction can reproduce experimental results and point at possible incorrect information

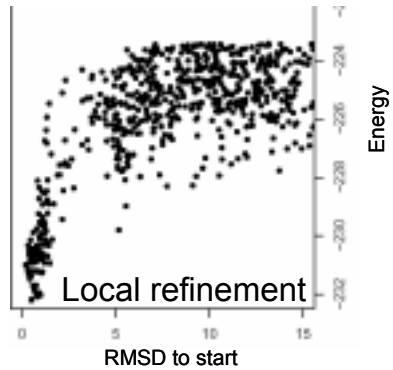
→ High-resolution prediction coupled with experimental data contributes to structural characterization of interface

FunHunt

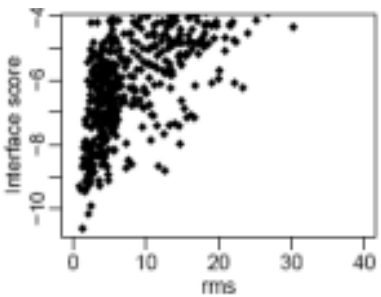
- Introduction to Funnels
- Funnels for model selection
  - Example: entry of Anthrax into human cells
  - **FunHunt: characterization of native energy funnels in energy landscape**
  - **Targeting of native funnels for docking and design**
- Summary

Energy Landscape: Many Funnels

- Global search detects more funnels
- Free monomer structure → backbone inaccuracies → TRUE funnel cannot be selected by energy or "funnel quality"



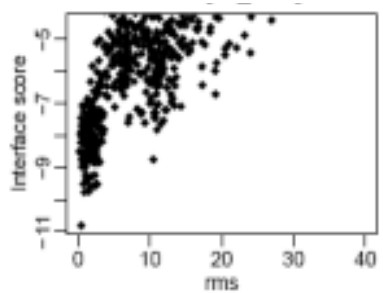
TRUE Funnel



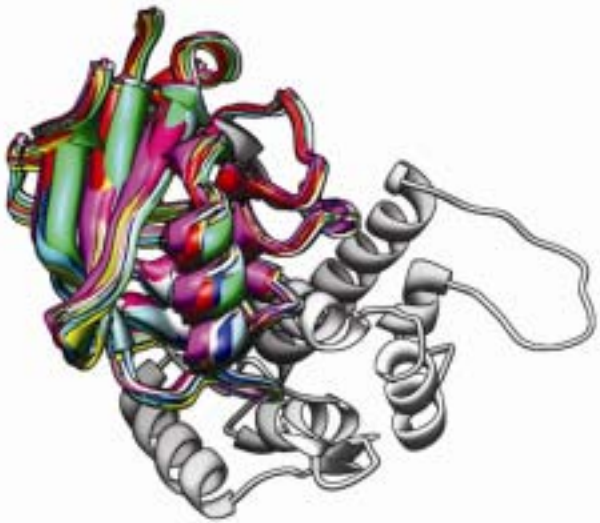
How can TRUE Funnel be selected

?

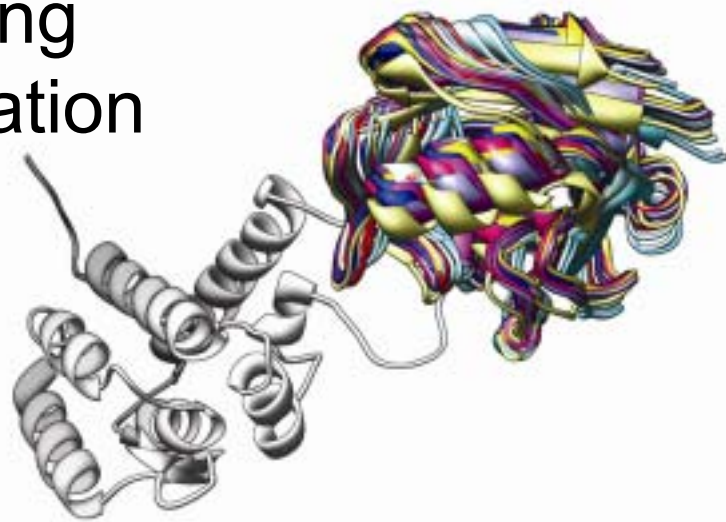
FALSE Funnel



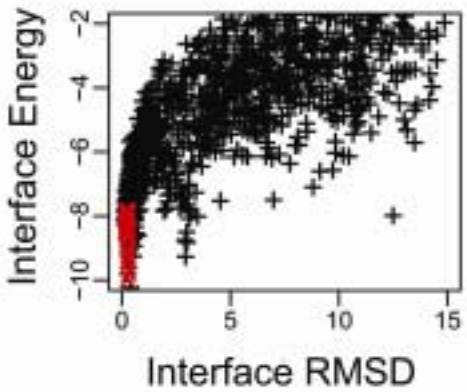
FunHunt: Feature-based Classification of Native Funnels



- Dense sampling around orientation
  - ✓ accounts for fluctuations
  - ✓ improves signal detection

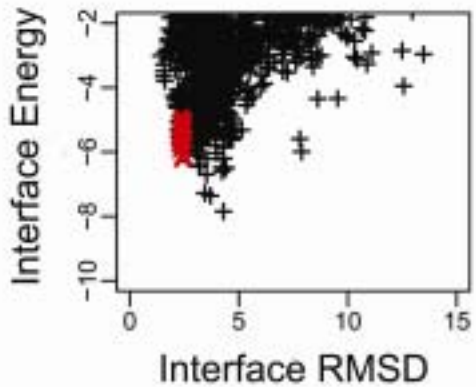


- ➢ Characterize each by set of features
- ➢ Define distinctive features

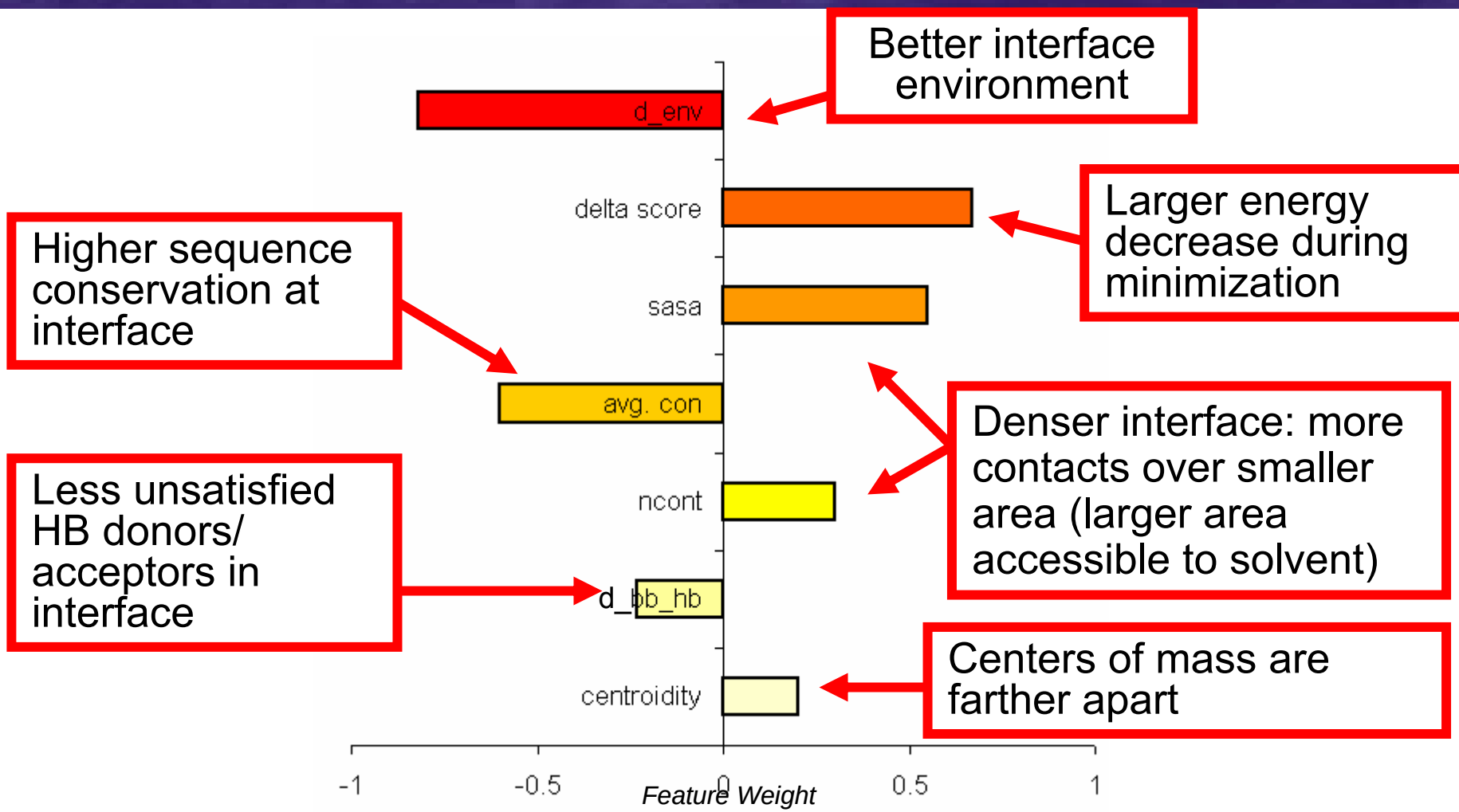


FALSE

TRUE



FunHunt Features (Selected by SVM Classifier)



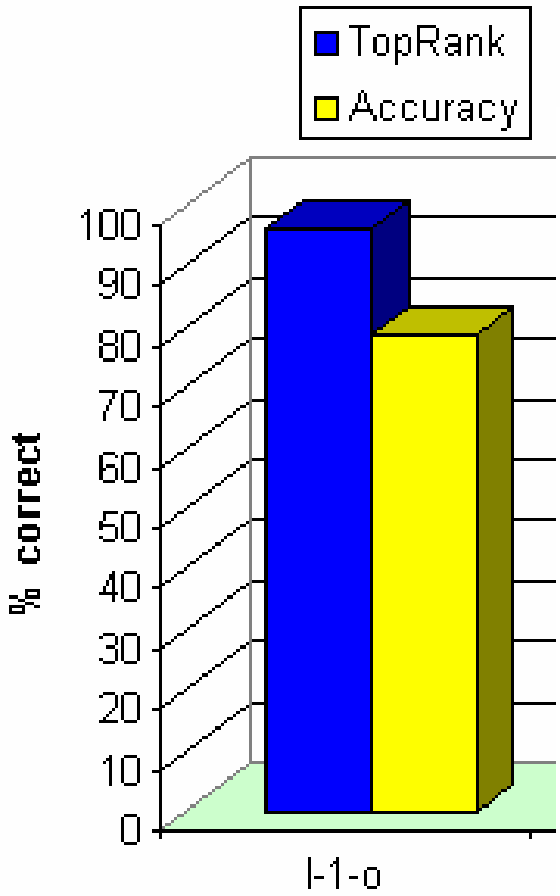
FunHunt Performance

Successful classification

- ✓ 50/52 correctly predicted: top-ranking model is from TRUE Funnel
- ✓ 80% models correctly classified (TP+TN)

FunHunt is robust

- ✓ *Leave1out (L1O) ≈ Leave8out (L8O)*



CAPRI Targets

- **12/12 - works**

Target	Complex Type	Correct ?	Accuracy	TRUE score	FALSE score	FALSE rmsd
T11	U-H(NMR)	√	88	-5.53	-7.11	10.79
T12	U-B	√	90	-9.28	-7.14	15.32
T13	*U-B*	√	*88*	*-7.09*	*-6.93*	*18.73*
T14	H-B	√	40	-14.37	-11.90	34.77
T15	B-B	√	10	-14.07	-8.87	9.09
T18	U-U	√	94	-6.63	-6.99	22.65
T19	*H-B*	√	*74*	*-9.33*	*-6.74*	*15.45*
T21	U-U	√	100	-6.32	-8.39	8.65
T24	U-H	√	100	-5.48	-8.09	16.53
T25	U-B	√	92	-7.04	-9.59	16.19
T26	U-U	√	100	-9.11	-6.65	12.56
T27	U-U	√	100	-6.66	-9.03	11.32

Location of Starting Orientation for Flexible Docking *(Capri Target T18)*

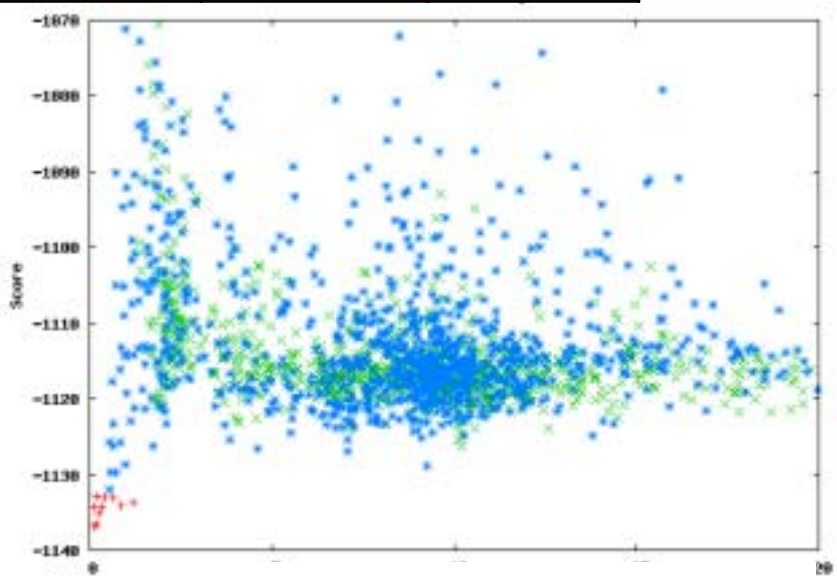
22.65	-6.99	-6.63	94	√	U-U	T18

- T18 failed with regular RosettaDock

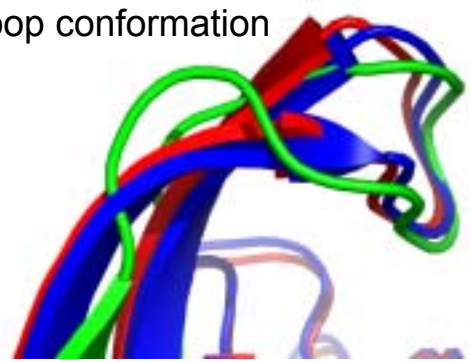
Backbone flexibility necessary

- But Where?

FunHunt can locate initial orientation and suggest region



Correctly predicted loop conformation



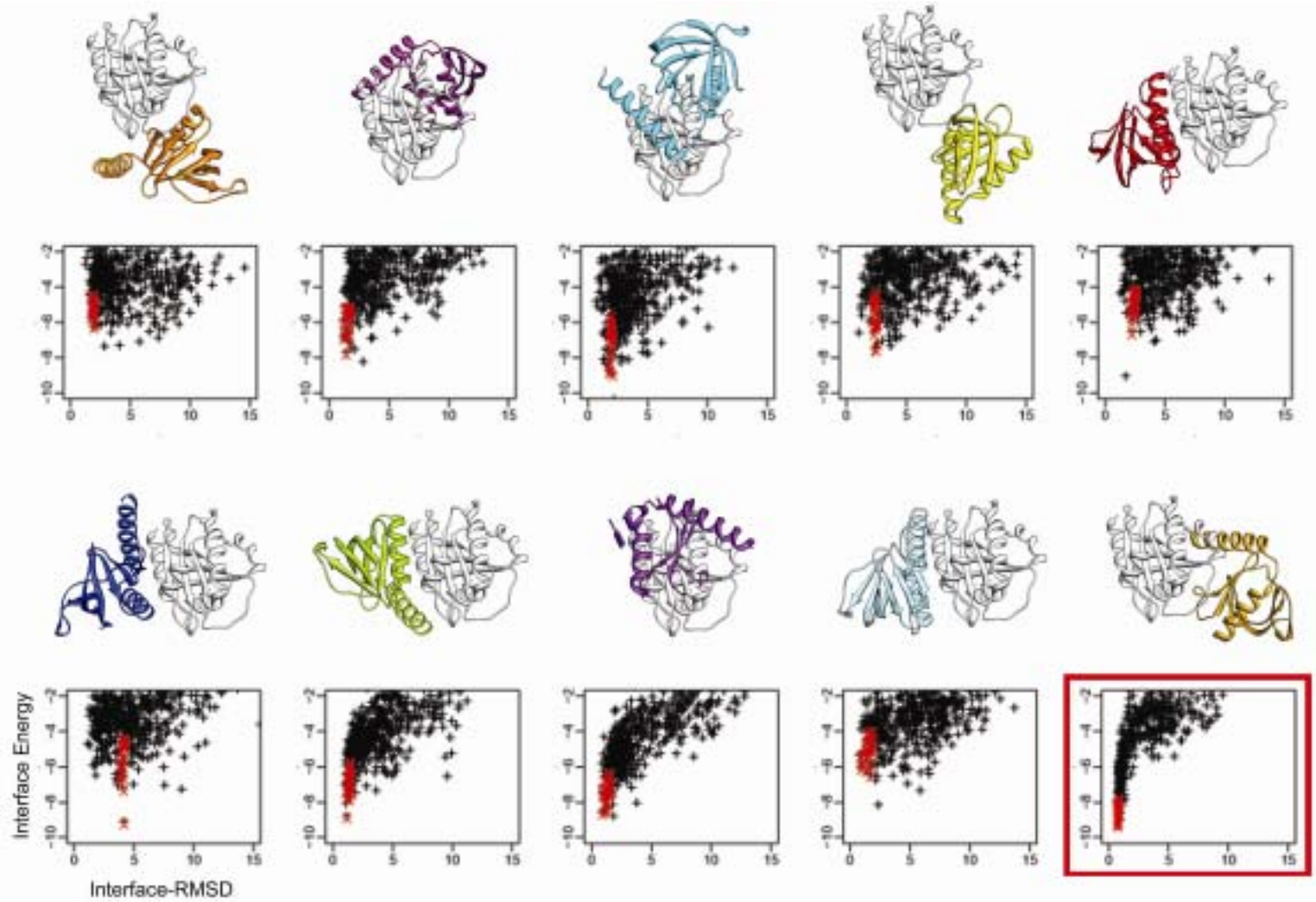
Red, orange – bour

Green – unbound;

Blue – model

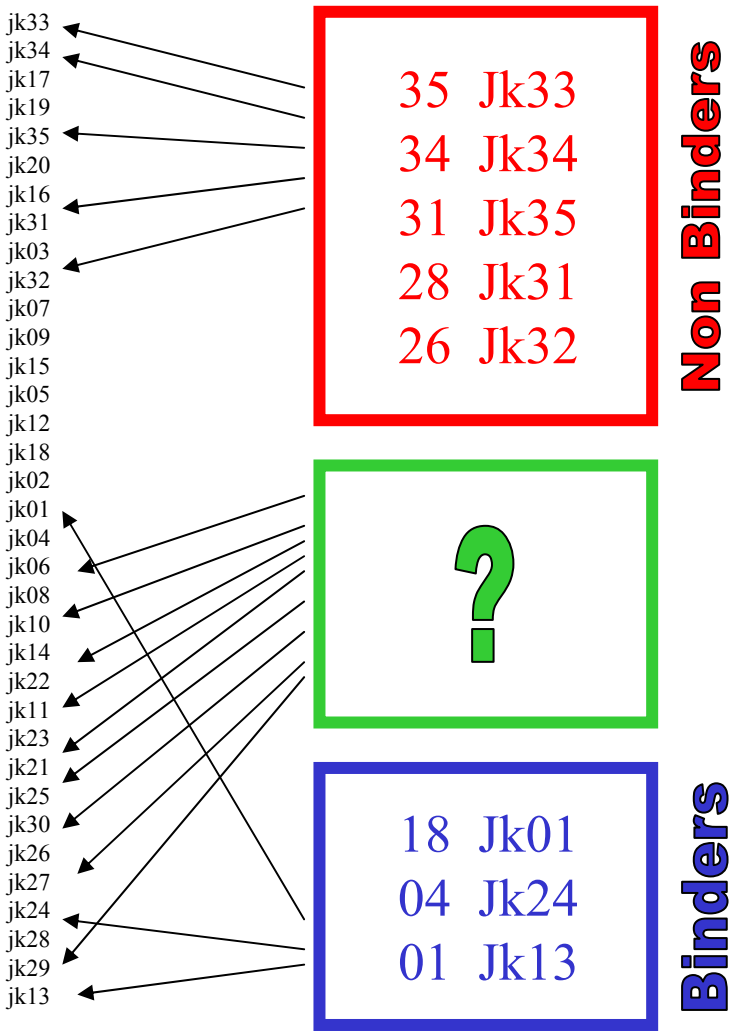
Flexible Docking by Chu Wang et al.

Selection of Native Orientation from PATCHDOCK Models *(CAPRI Target T25)*



Original orientations from PatchDock, Schneidman-Druhovny et al

FunHunt as a Tool for Interface Design Selection



- Can FunHunt discriminate binders from non-binders ?
- 35 Unlabeled designs

FunHunt - Conclusions

- ✓ **FunHunt** selects ***native funnel*** among funnels in RosettaDock energy landscape
  - ✓ Based on small set of features selected by SVM
- ✓ **FunHunt** locates regions that need modeling of ***backbone flexibility*** in RosettaDock
- ✓ **FunHunt** can be used for ***interface design*** selection

Acknowledgments

Furman Lab *Hebrew U Jerusalem, Israel*

- **Nir London**
- Dan Reshef, Barak Raveh, Vered Fishbain & Dana Attias

Rosetta Community

- **Chu Wang**, David Baker *UW*
- Baker lab computer clusters
- Jeff Gray *JHU*

Anthrax

Harvard Medical School

- Borden Lacy
- John Collier

SVM

- Asa Ben Hur *Col. State*
- Yael Mandel *Technion*

Miro: Catalan Landscape: The Hunter

Dataset of Protein Complexes

- Set52: <70% seqid between both partners **n=52**
- Set32: <70% seqid between any partner **n=32**

Enzyme – Inhibitor (EI) n=19/11		
1ACB	UU	α-chymotrypsin/ Eglin C
1CGI	UU	α-chymotrypsinogen/ Pancreatic secretory trypsin inhibitor
1CHO	UU	α-chymotrypsin/ Ovomucoid 3rd Domain
1PPE	BU	Trypsin/ CMT-1
2PTC	UU	β-trypsin/ Pancreatic trypsin inhibitor
1TAB	BU	Trypsin/ BBI
1AVW	UU	Trypsin/ Soybean trypsin inhibitor
1V5I	UU	Subtilisin BPN / Serine protease inhibitor POIA1
2SNI	UU	Subtilisin Novo/ Chymotrypsin inhibitor 2
2SIC	UU	Subtilisin BPN/ Subtilisin inhibitor
1BRS	UU	Barnase/ Barstar
1MAH	UU	Mouse Acetylcholinesterase/ Fasciculin 2
1UGH	UU	Human Uracil-DNA glycosylase/ Inhibitor
1DFJ	UU	Ribonuclease A/ Ribonuclease inhibitor
1STF	BU	Papain/ Stefin B
1BTH	UU	Thrombin mutant/ Pancreatic trypsin inhibitor
4HTC	BU	α-Thrombin/Hirudin
1TMQ	UU	α-amylase/ RagI inhibitor
1TE1	UU	Xylanase/ XIP-I inhibitor
Antibody – Antigen (Ab-Ag) n=9/1		
1BVK	UU	Fv/ Lysozyme
1MLC	UU	IgG1 D44.1 Fab fragment/ Lysozyme
1WEJ	UU	Fab fragment/ Cytochrome C
1AHW	UU	5G9/ Tissue factor
2JEL	BU	Jel42 Fab Fragment/ A06 Phosphotransferase
1NCA	BU	Fab NC41/ Neuraminidase
1EO8	BU	Bh151 Fab/ Hemagglutinin
1IAI	BU	IgG1 Idiotypic Fab/ Igg2A Anti-Idiotypic Fab
1IGC	BU	IgG1 Fab Fragment/ Protein G

OTHER n=24/20		
1AK4	UU	Cyclophilin/ N' domain of HIV1 capsid
1EER	UU	Erythropoietin/ erythropoietin receptor
1Z8U	UU	α-hemoglobin/ α-hemoglobin stabilizing protein (AHSP)
1FQJ	UU	Rgs9 (rgs domain)/ gt-i1 chimera alpha unit
1GOT	UU	Transducin Gt-α, Gi-α chimera/ Gt- β–γ
1GUA	UU	Rap1/ Raf1 (ras binding domain)
1I4D	UU	Arfaptin/ Rac1
1HE1	UU	ExoS gap domain/ Rac1
1ATN	BU	Actin/ Deoxyribonuclease I
2BTF	BU	β-actin/ Profilin
1NMU	UU	MBP/ L30
1S1Q	UU	TSG101(UEV) domain/ Ubiquitin
1SYX	UU	U5/ snRNP
1F80	UU	holo-acyl-carrier-protein synthase/ holo-acyl-carrier-protein
1WQ1	UU	Ras GAP/ Ras
1AVZ	UU	HIV-1 NEF/ FYN tyrosine kinase SH3 domain
1MDA	UU	Methylamine dehydrogenase/ Amicyanin
1GLA	BU	Glycerol kinase/ GSF III
1A0O	BU	Che A/ Che Y
1FIN	UU	CDK2 cyclin-dependant kinase 2/ Cyclin
1FQ1	UU	CDK2/ KAP
3HHR	BU	Human growth hormone/ Receptor
2PCC	UU	Cytochrome C Peroxidase/ Iso-1-Cytochrome C
1EFU	BU	E. coli EFtu/ Efts

Features Considered Initially for Classification: Details

Feature	Details	*Name*
Environment	Interface propensity of residues (*vs* surface)	*D_env*
Residue pairs	Interface contact propensity of residue pairs	*D_pair*
Full atom energy	Energy function of RosettaDock protocol	*Score*
Lennard Jones	Attractive and Repulsive VdW-forces	*Fa_atr, Fa_rep*
Solvation	Solvation free energy (Lazaridis-Karplus)	*Fa_sol*
Hydrogen bonds	Orientation-dependent hb	*Hb*
Side chain conformations	Propensity based on backbone dependent library (Dunbrack). Free monomer conformation is favored.	*Fa_Dun*
electrostatic energy	Coulomb energy of interface atoms	*D_elec*
Trajectory	Drop in energy during full atom Monte-Carlo minimization	***Δenergy***
Full atom energy	Energy function of RosettaDesign protocol for interface scoring	*Eres*
Softened Lennard Jones	Attractive and repulsive terms across interface: Softened repulsive	*Eatr, Erep*
Solvation	Interface solvation free energy	*Esol*
Hydrogen bonds	Interface hydrogen bonds	*Ehbnd*
"pair" energy	Propensity of charged side chain atoms to contact each other - Rosetta approximation of the electrostatic effect	*Epair*
Interface contacts	Atom-atom contacts across interface within 5Å.	*Ncont*
Protein surface	Solvent Accessible Surface Area (SASA) of protein	*SASA*
Polar interface	Polar SASA buried upon binding	*ΔSASA_POL*
Apolar surface	Apolar SASA buried upon binding	*ΔSASA_APOL*
Packing density	Quality of packing (relative to values from statistic analysis of packing density in similar environment)	*SASApack*

Feature	Details	*Name*
Sequence conservation	Degree of sequence conservation of interface residues (based on CONSURF); total score of interface residues	***conScore***
	Average conservation score of interface residues	***avgCon***
	Maximal conservation score among interface residues	***maxCon***
Solvation	SASA-based calculation of solvation	***Gsolt***
Uniformity	Variance of the energy contributions of interactions across the Interface	***varIfCont***
Centroidity	Distance from center of mass of interface atoms	***Centro***
	Distance from center of mass. Averaged over interface	***CentroAvg***
Secondary structure	Content of Helix, Strand, Turn, Coil at interface (based on STRIDE)	***H,S,T,C Cont.***
	Average secondary structure content	***H,S,T,C Avg.***
Structural definition	Total temperature factor of interface residues	***totB***
	Average temperature factor of interface residues	***avgB***

Hydrogen bonds at interface	Number of *side-chain* hydrogen bond donors/ acceptors that are unsatisfied upon binding	*ΔSC_HB*
	Number of *main-chain* hydrogen bond donors/acceptors that are unsatisfied upon binding	*ΔBB_HB*
Satisfaction of donors/ acceptors	hydrogen bonding - unsatisfied buried hydrogen bond donors and acceptors upon binding	*ΔGU_TOT*
	"", weighted according to different donor/acceptor types	*ΔWGU_TOT*