

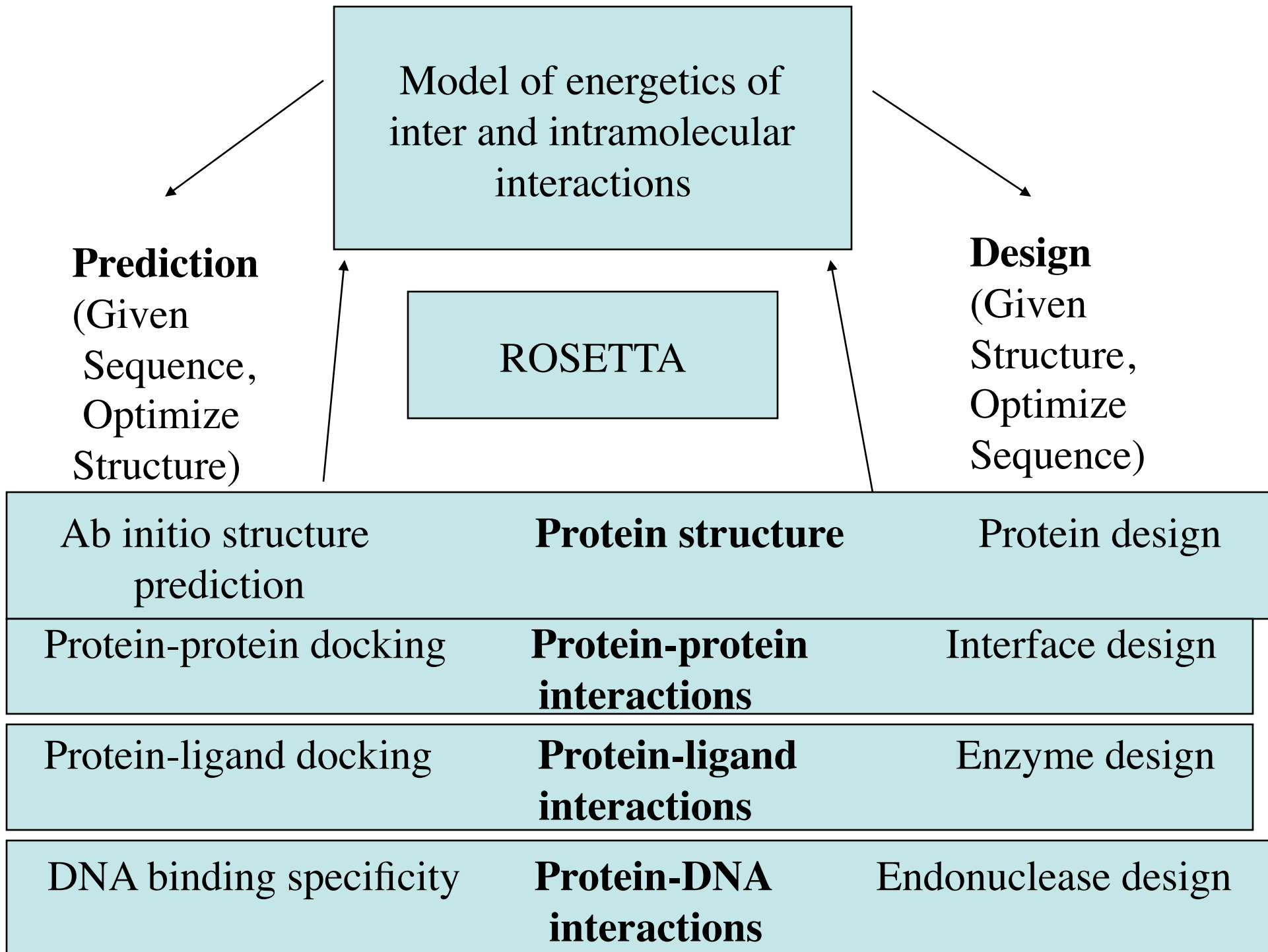
From Prediction of Structure to Design of Function

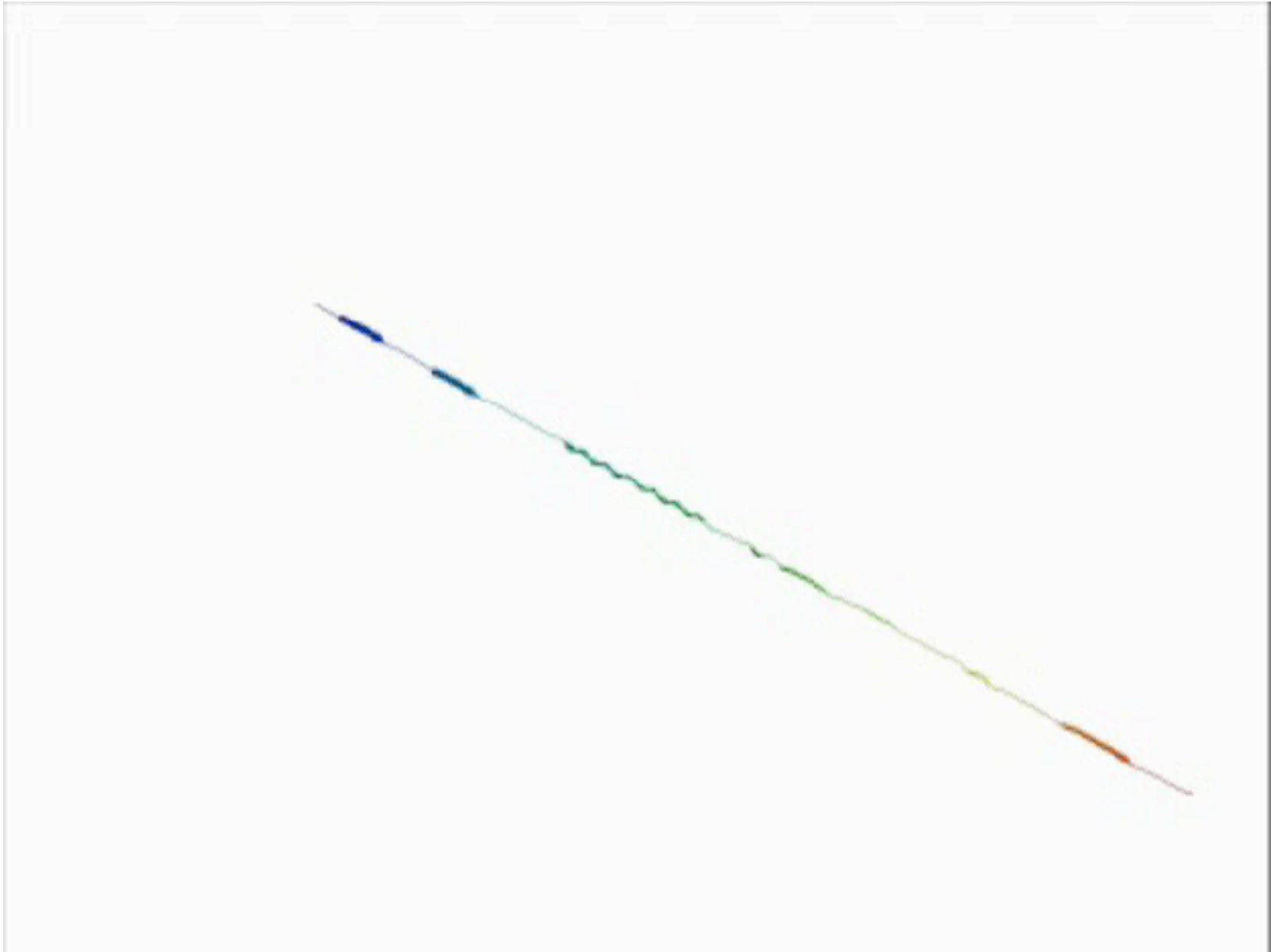
Prediction

Genome sequences => Macromolecular Structures and Interactions

Design

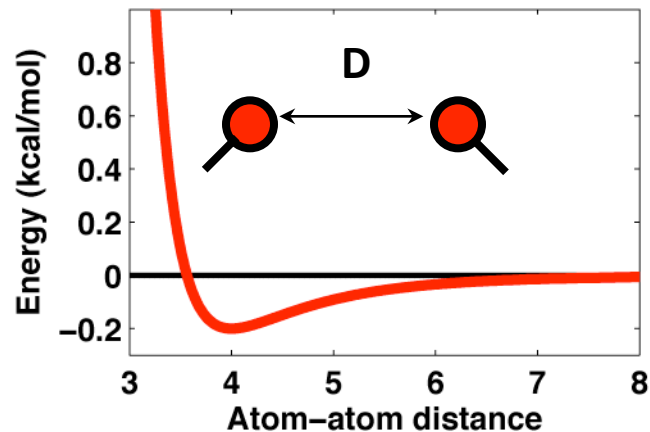
Designed sequences <= New structures, interactions, enzymes, endonucleases, vaccines



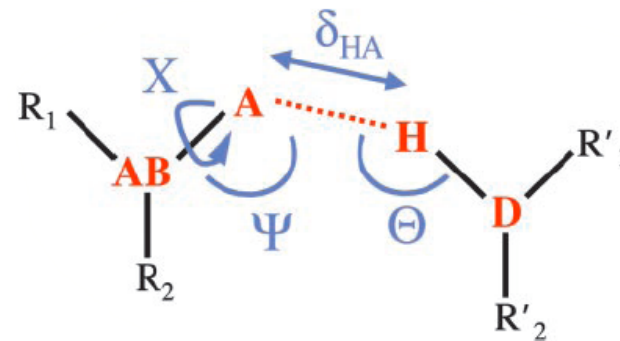


Rosetta high resolution potential

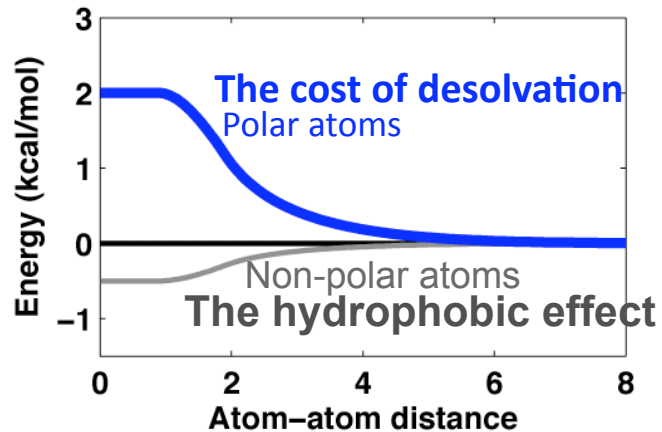
1. Van der waals packing



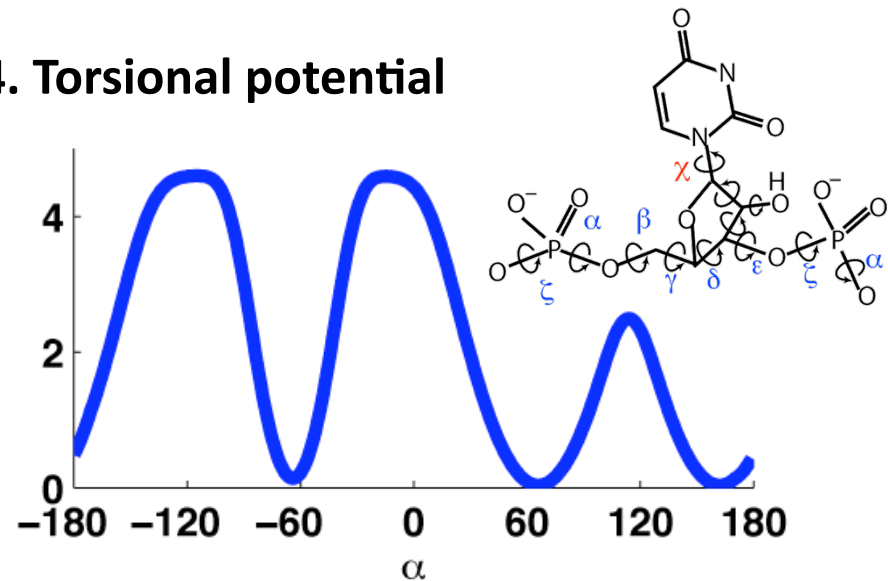
2. Hydrogen bonds



3. Solvation

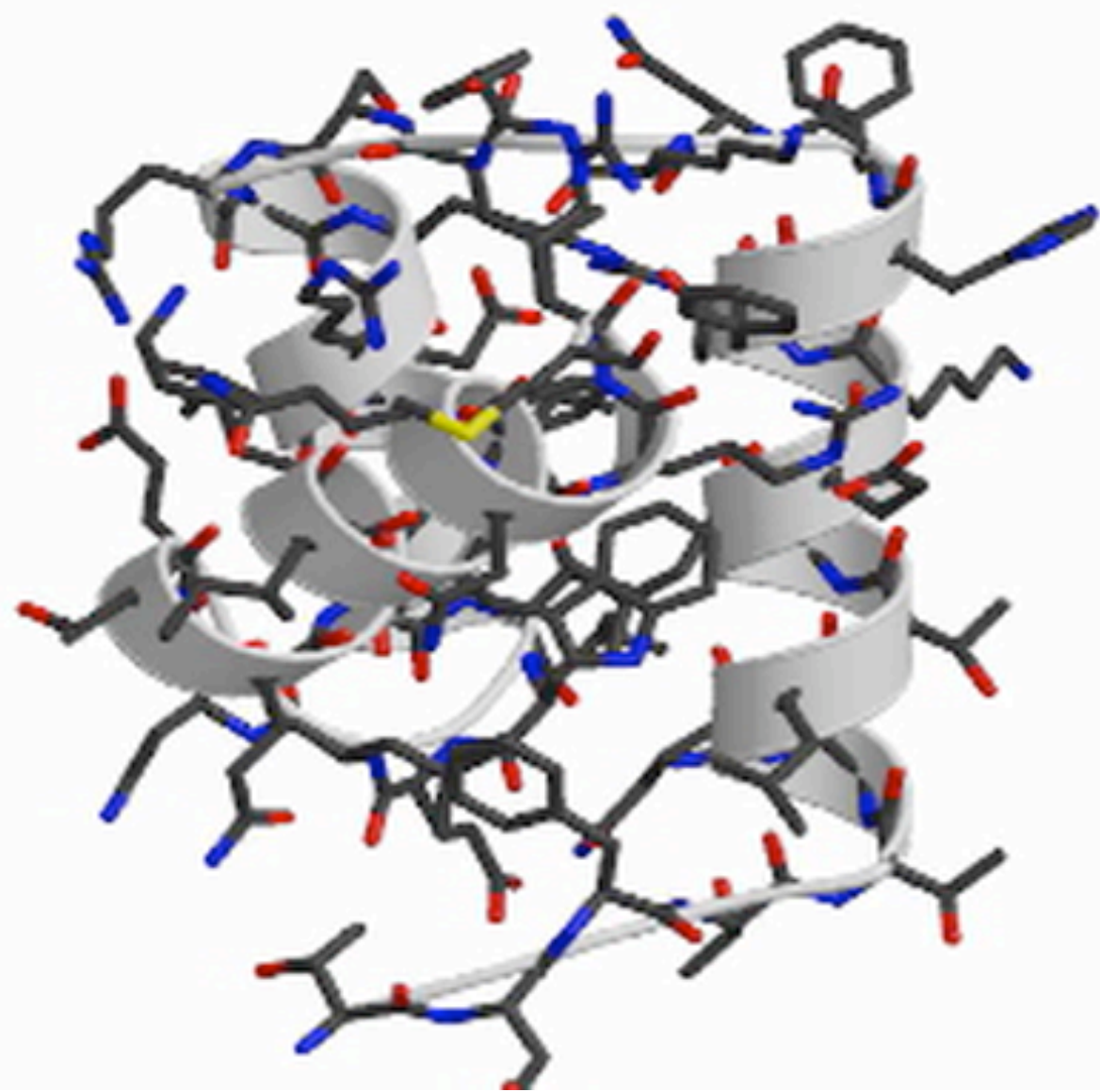


4. Torsional potential

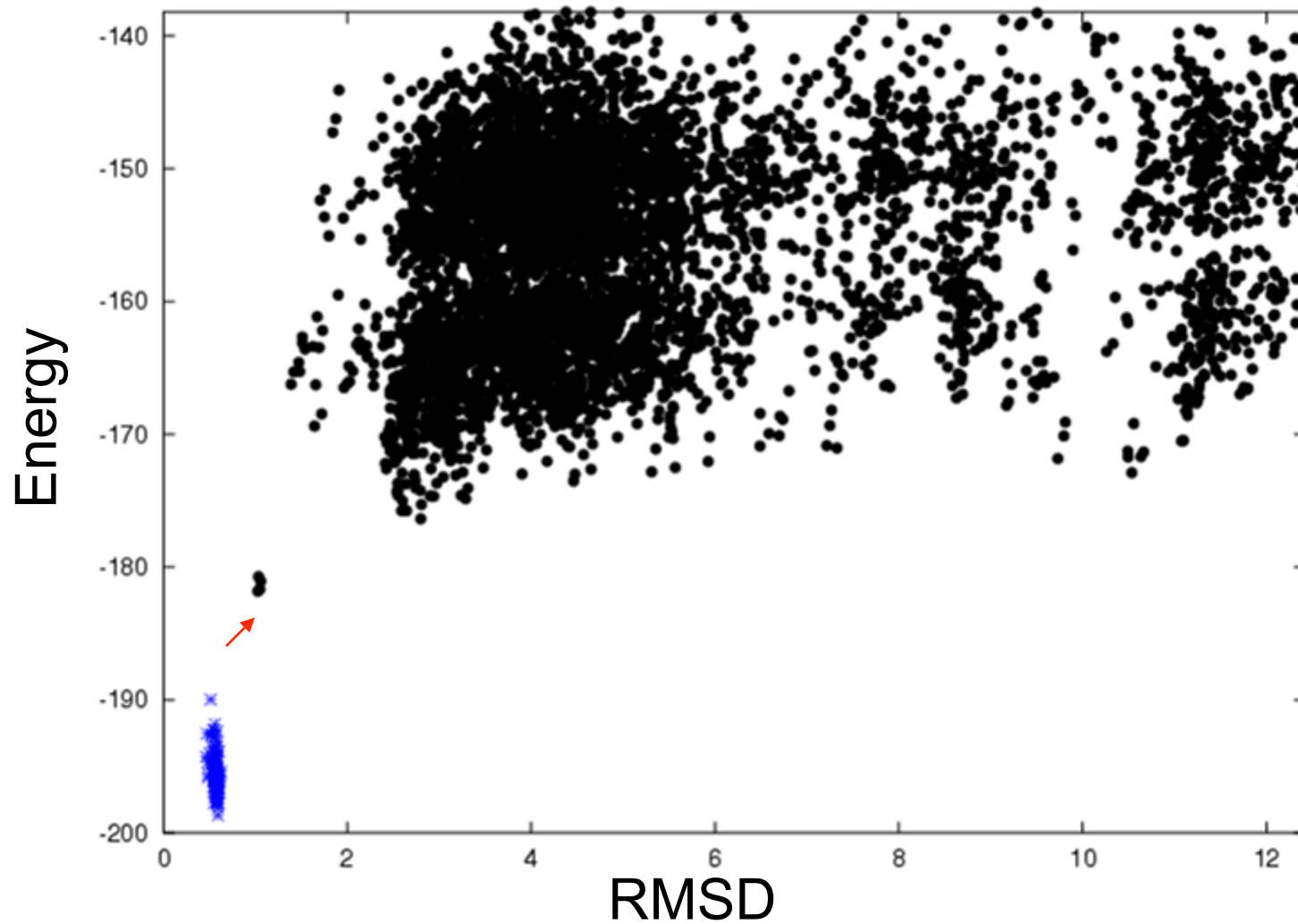


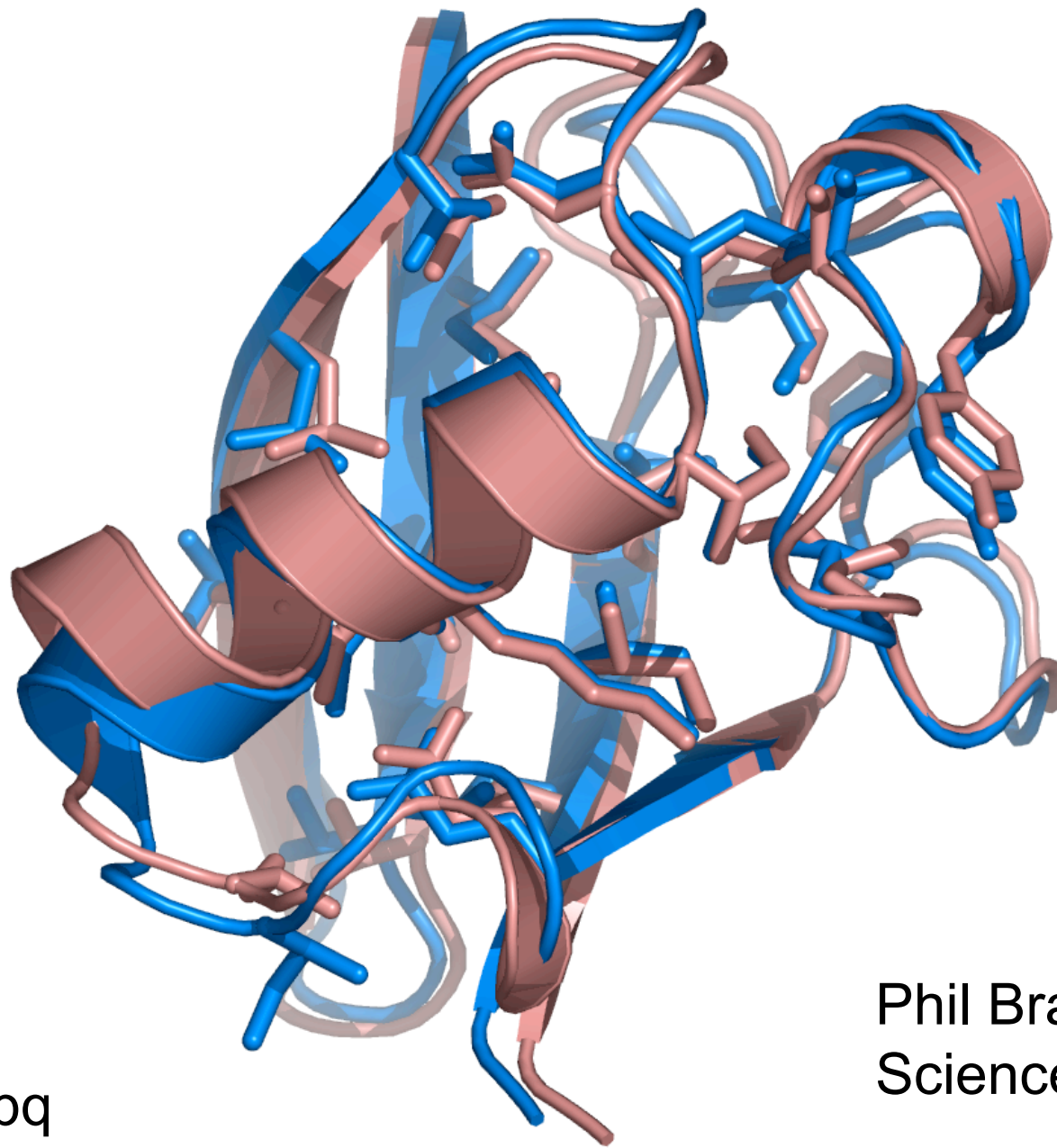
5. Electrostatic repulsion (screened)

Free energy - configurational entropy



Lowest energy structures sampled on independent trajectories





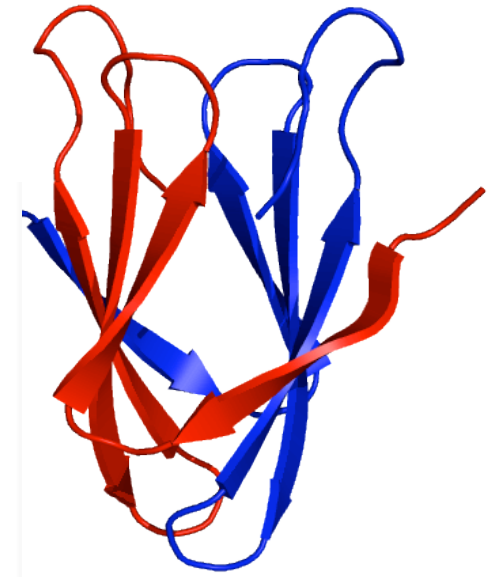
1ubq

Phil Bradley
Science 2005

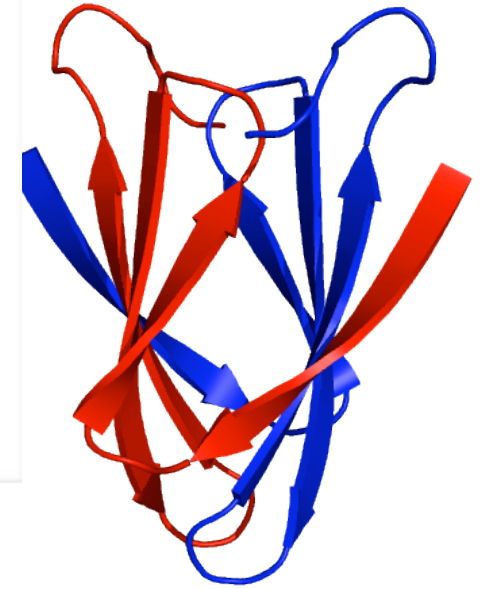
Fold-and-dock

Sequence → 

 →



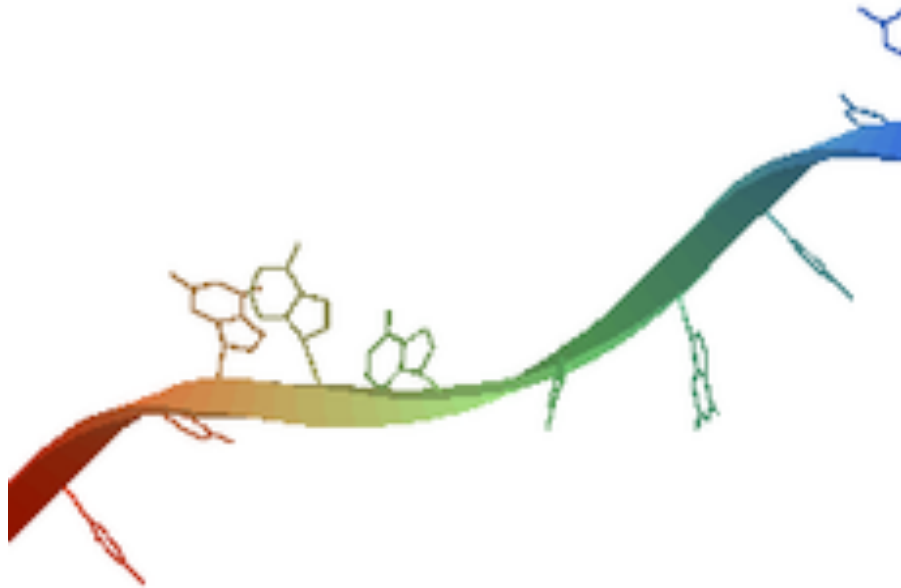
2bti: Model



2bti: Native

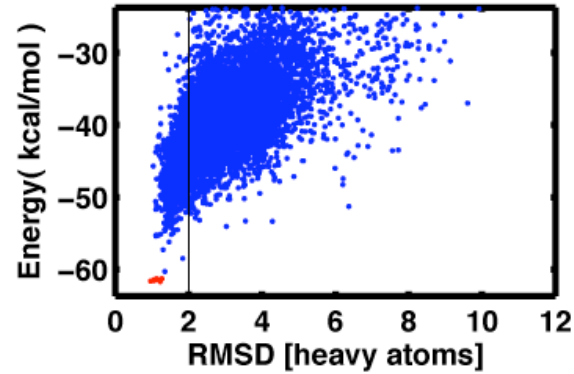
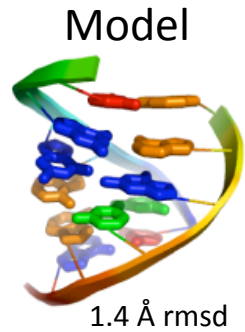
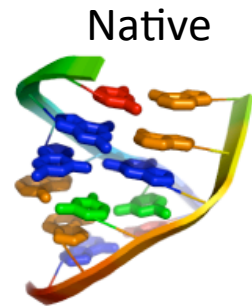
Ingemar Andre, Rhiju Das

RNA folding in Rosetta

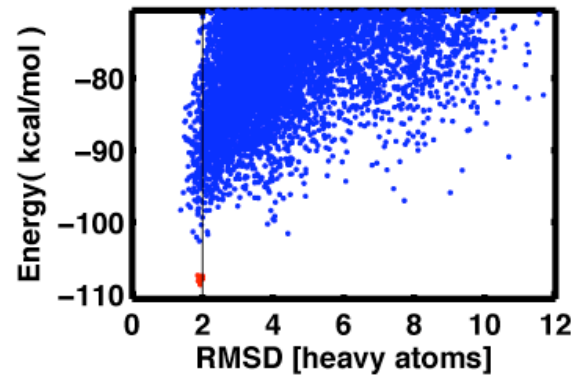
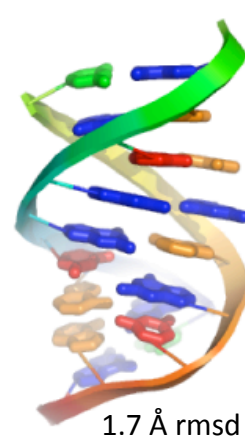
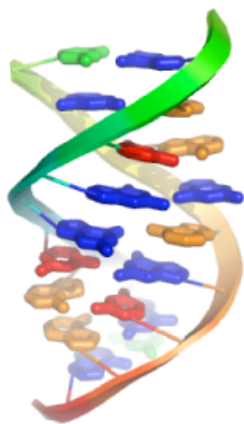
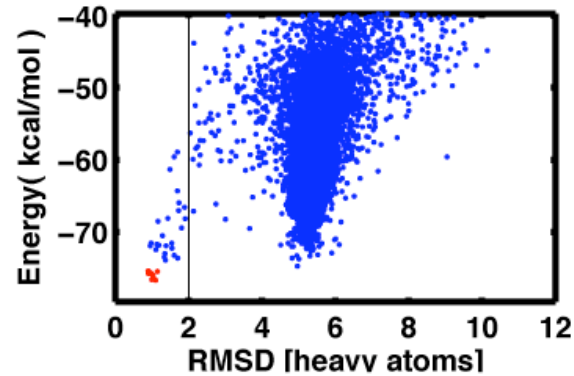
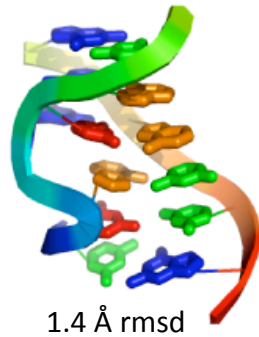
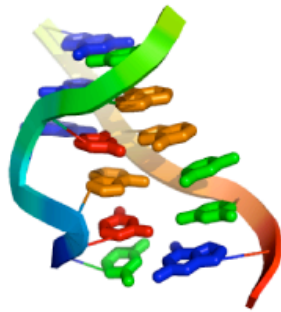


Rhiju Das

De novo modeling



In more than a third of the cases, de novo modeling achieves < 2.0 Å structures, and selects them.



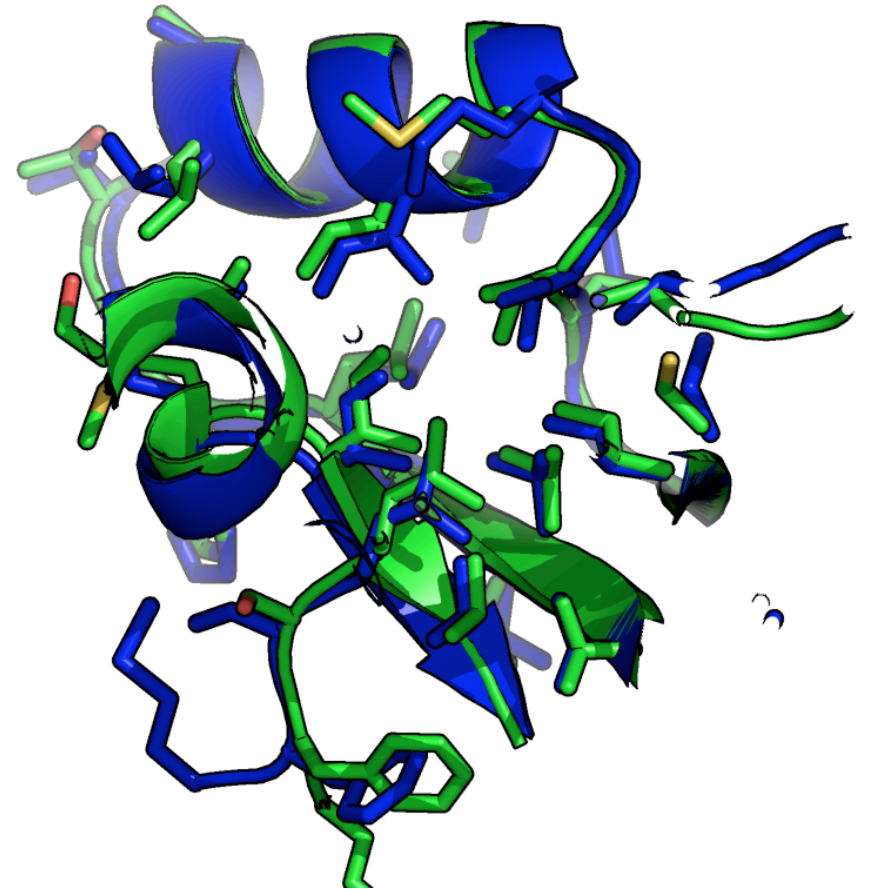
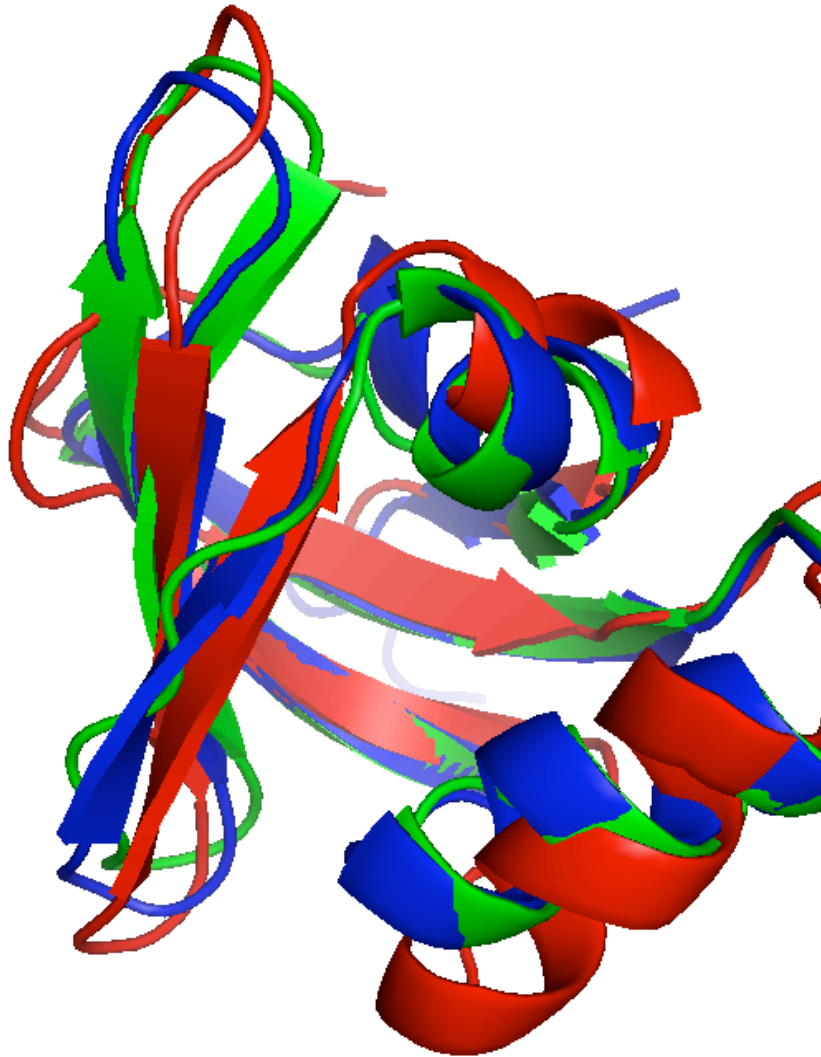
Native free energy gaps recurrent feature of structure prediction problems

- Soluble proteins, multimeric proteins, heterodimers, RNAs, membrane proteins, etc.
- Reflection of very large free energy gaps required for existence of single unique native state
- Prediction possible because (magnitude of actual free energy gap) \gg (error in free energy calculation)
- Challenge: how to sample close to native state?

How to find global minimum?

- Smarter algorithms
- Volunteer computing: [rosetta@home](#)
- Start closer: comparative modeling
- Use experimental data to limit search
- Collective brain power of game playing humans: <http://fold.it>

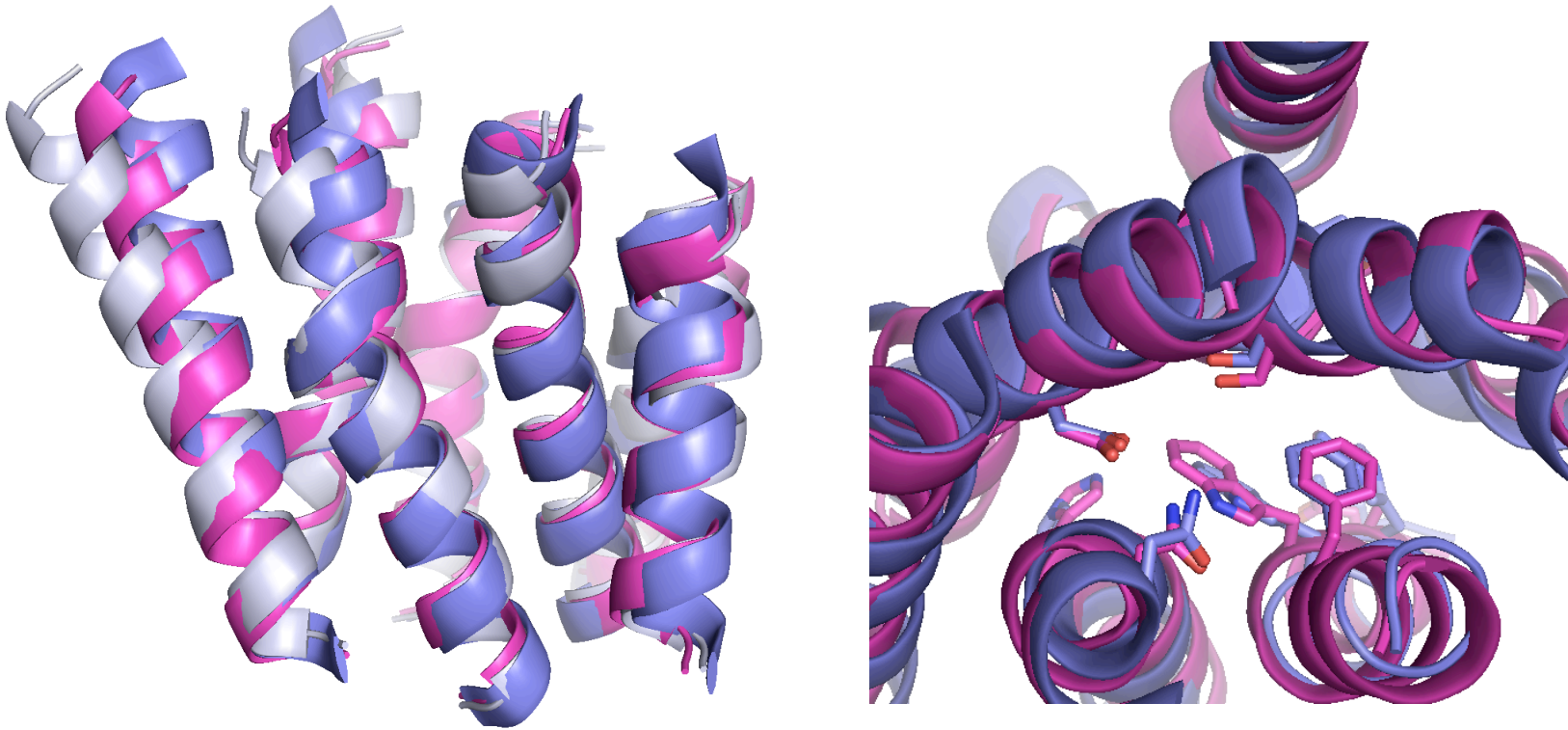
Rosetta refined comparative models often more accurate than starting template.



Mike Tyka
Zscore 6.45

CASP8 T492

Blind prediction of Human A2A Adenosine Receptor TMH core region



X-ray structure

Rosetta Model 1.3 Å (over TMH region)

Beta2 adrenergic receptor 1.8 Å (over TMH region)

Patrick Barth

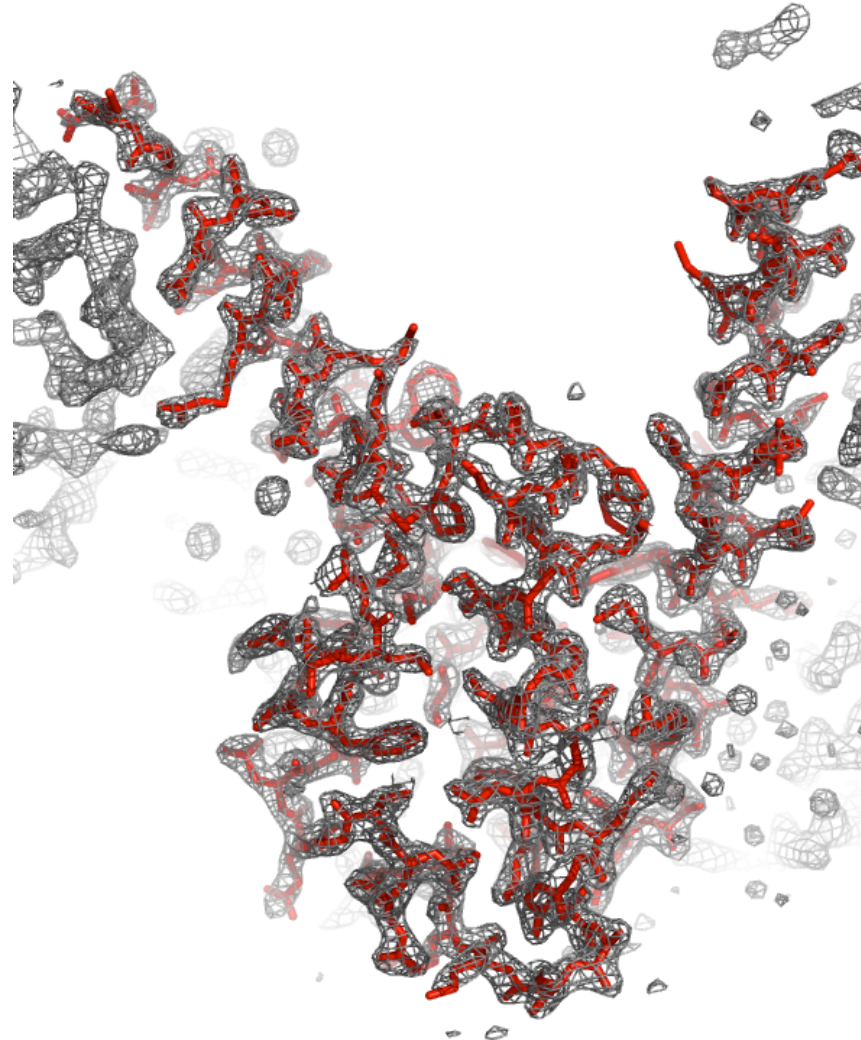
Use experimental data to help locate global minimum

- X-ray diffraction data
- NMR chemical shift assignments
- Low resolution CryoEM density
- Different from traditional approaches: data guides search, does not specify structure

"Ab initio phasing by ab initio folding"

**Red: PDB coordinates
from crystal structure
phased by selenium SAD**

**Gray: Electron density
map, phased by
molecular replacement
with ab initio Rosetta
model**

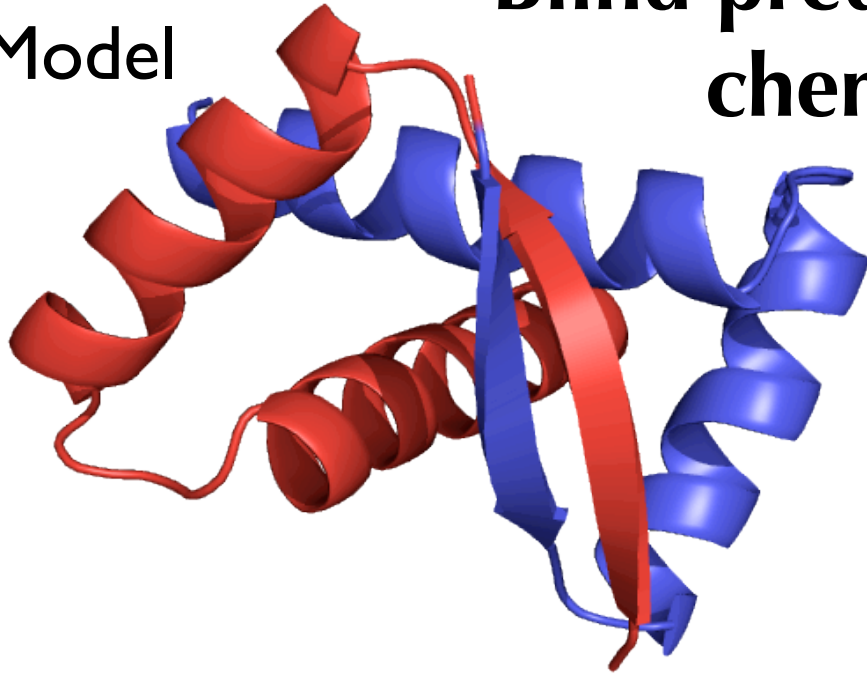


High accuracy models from limited NMR data!

- Backbone chemical shifts only
- Chemical shifts plus unassigned NOESY spectra
- Chemical shifts plus residual dipolar couplings
- Data confines search only; details from rosetta forcefield=>can be more accurate than conventional models

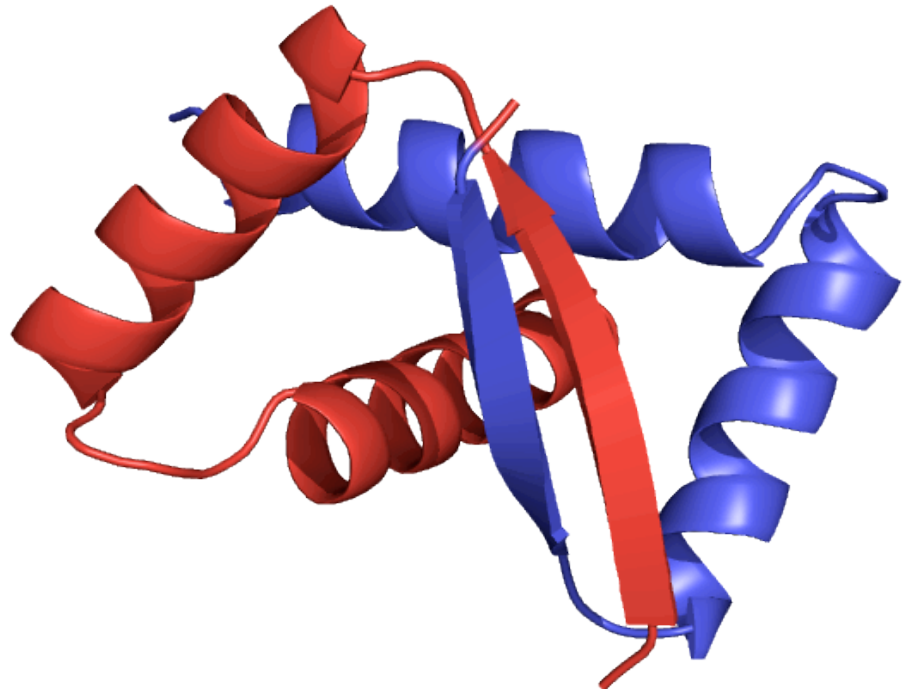
Blind prediction of SFT1 using chemical shift data

Model



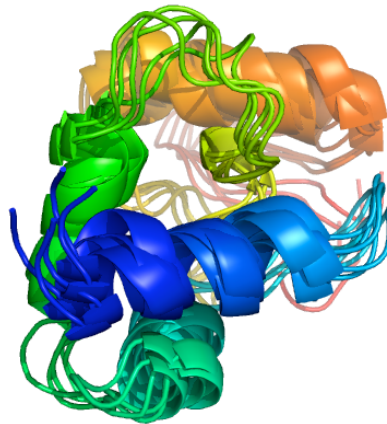
rmsd model: 1.1Å

Native

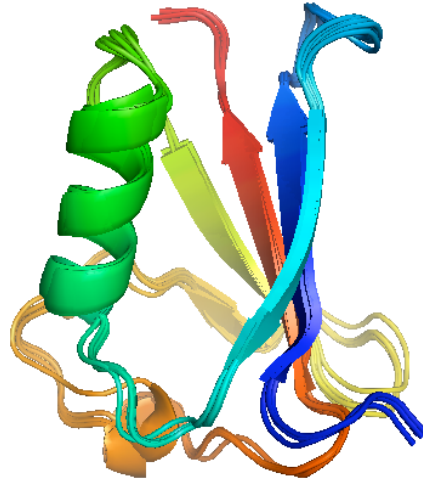
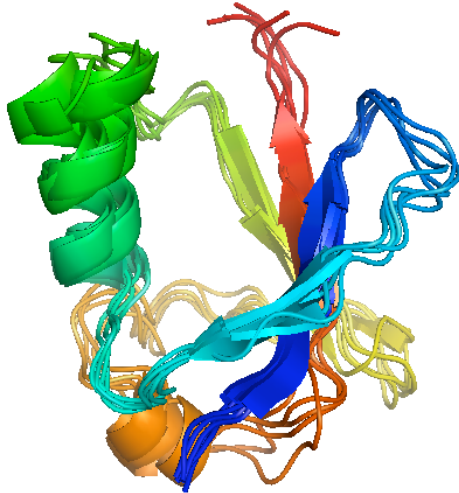


Ingemar Andre

NMR CASP Blind Targets 2009



VPR247
102aa

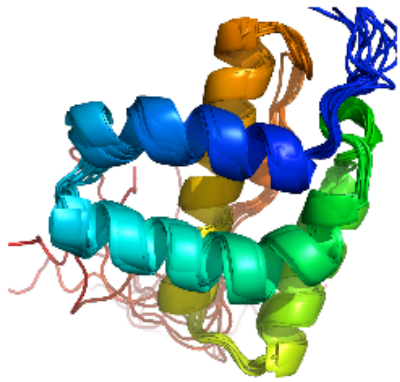


AR3436
97aa

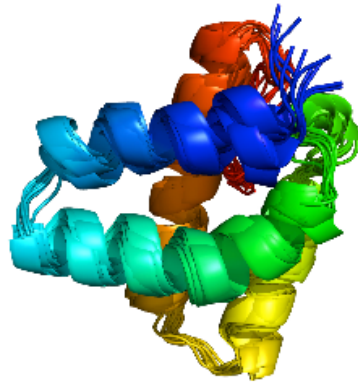
PDB from NMR

Rosetta plus chem shift plus
unassigned NOESY data

Blind Rosetta structure calculations using chemical shifts and RDCs. No sidechain assignments needed!



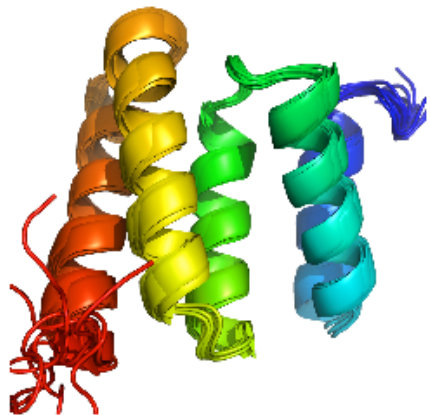
BcR268F 118 aa **0.99 Å**



DvR115G 94 aa **1.24 Å**



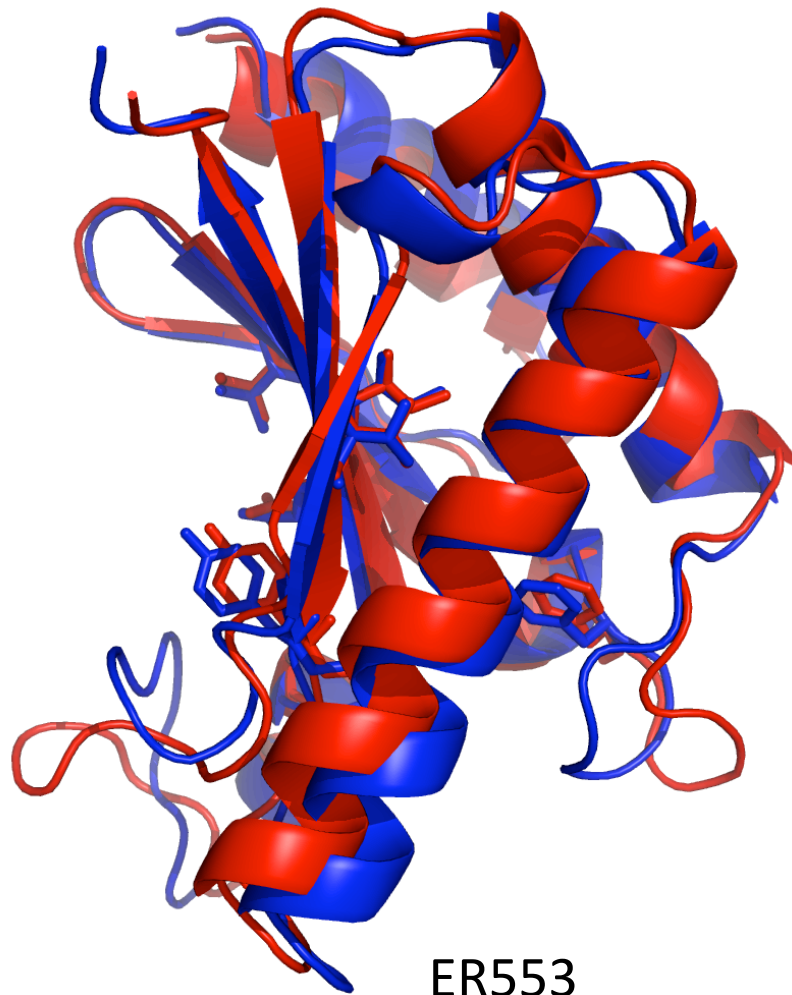
SrR115 100 aa **1.49 Å**



MaR214A 109 aa **2.54 Å**



Accurate models from chemical shifts and RDCs: new paradigm for NMR structure determination?



ER553
149 aa
1.4 Å

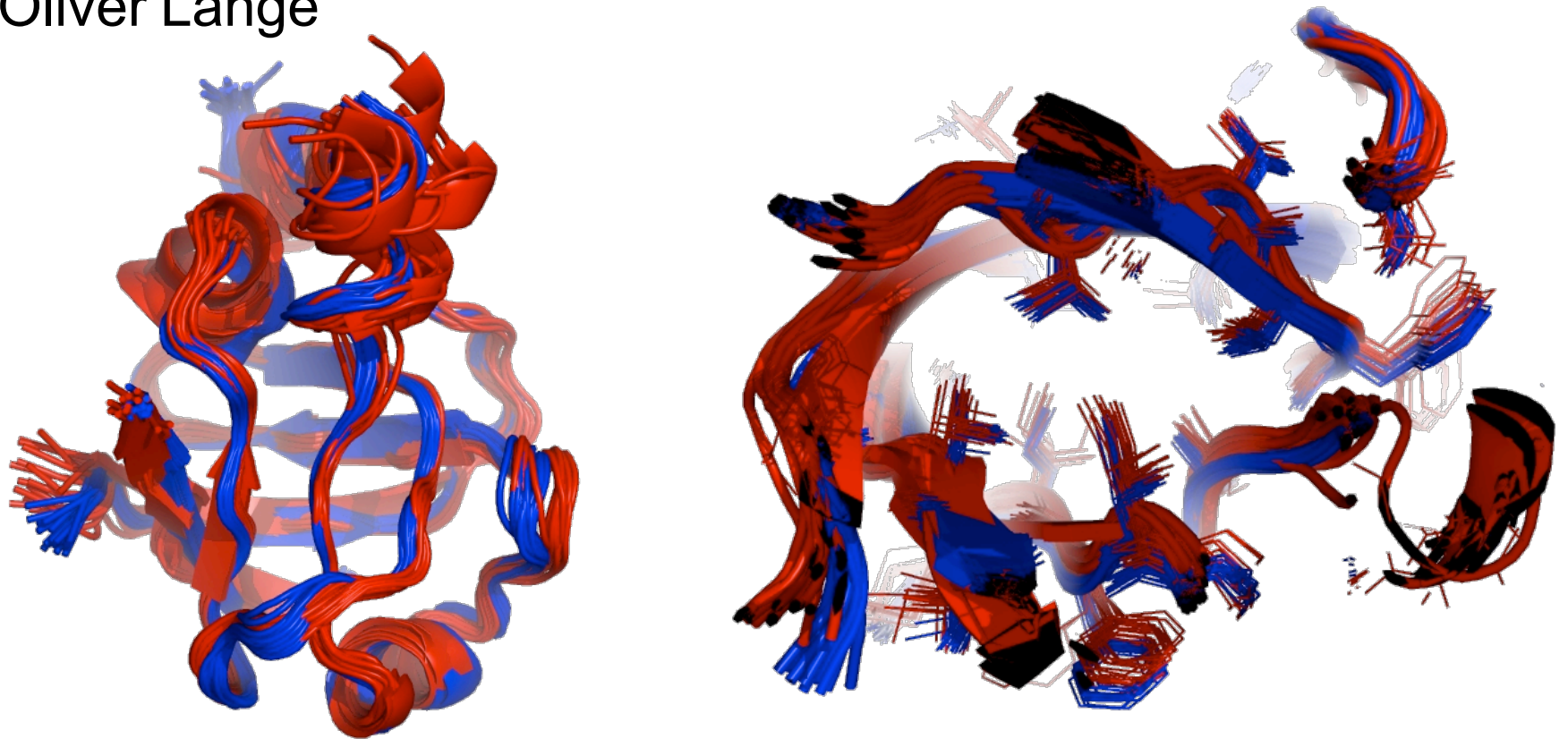
BLUE : Native structure
RED : Rosetta model



ARF1
166 aa
2.6 Å

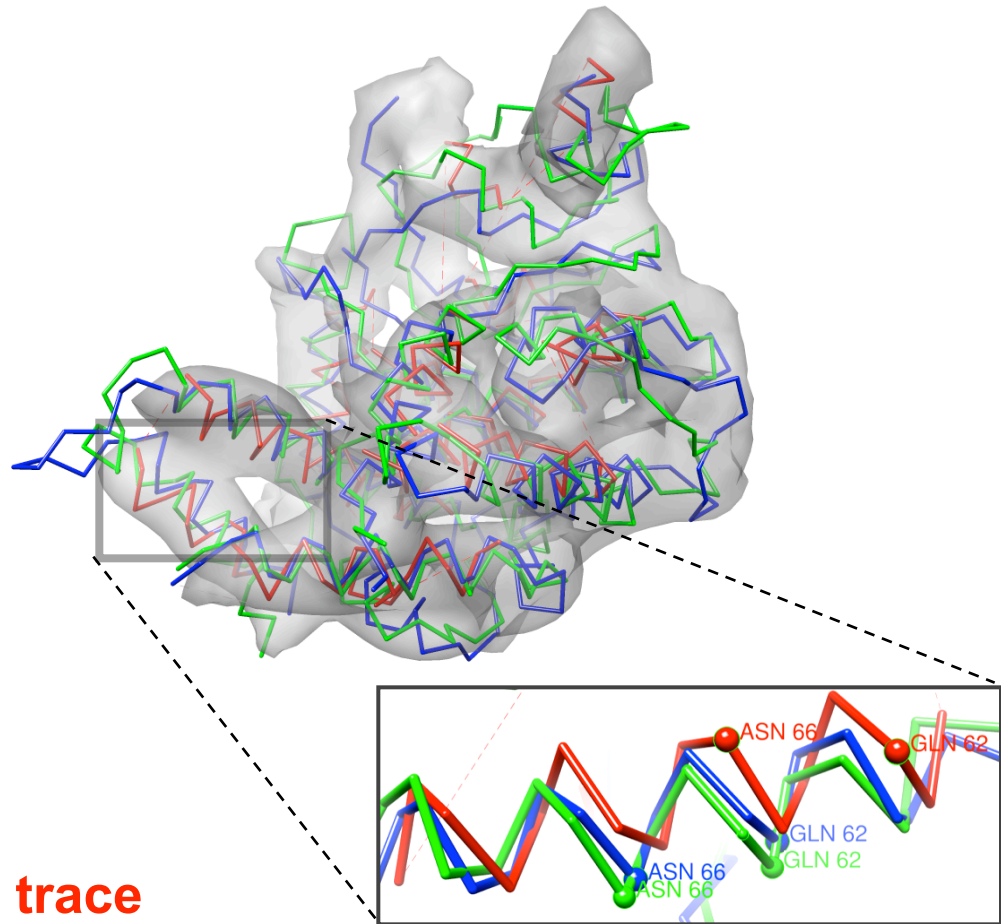
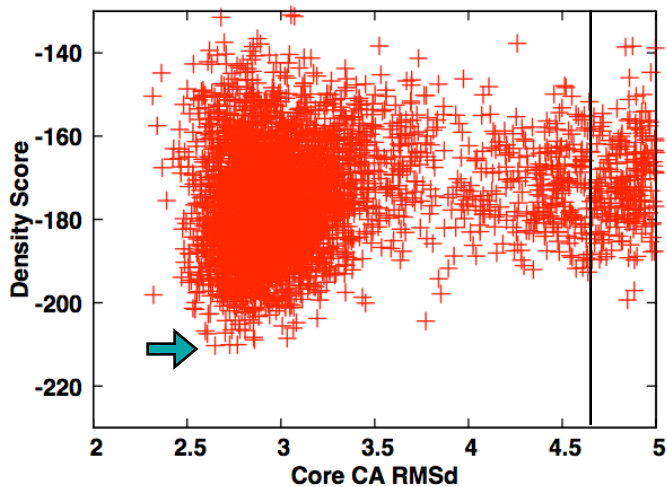
Topology-broker fold tree: allows stochastic sampling and quasi-Newton minimization of any combination of rigid body and internal degrees of freedom

Oliver Lange



blue: deposited NMR structures, red: Rosetta

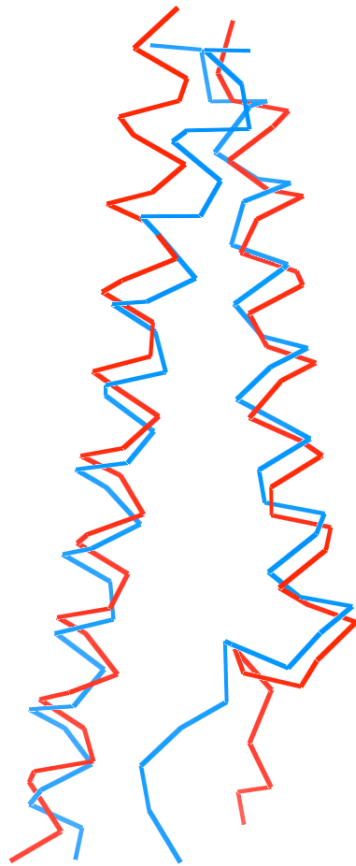
High-resolution model of RDV from 6.8Å cryoEM data



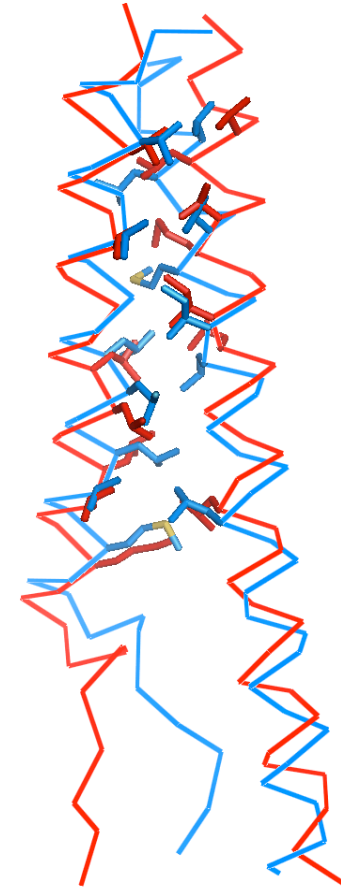
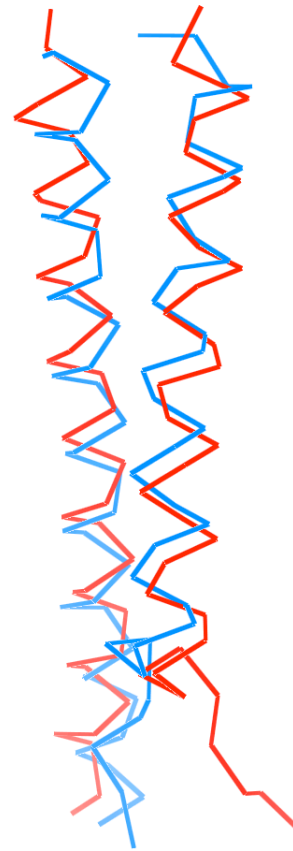
Initial α trace
Rosetta prediction
Native structure

Frank DiMaio, Wah Chiu

Integrin α IIb β 3 model based on Rosetta + disulfide constraints transmembrane section



C α rmsd: 2.1 Å

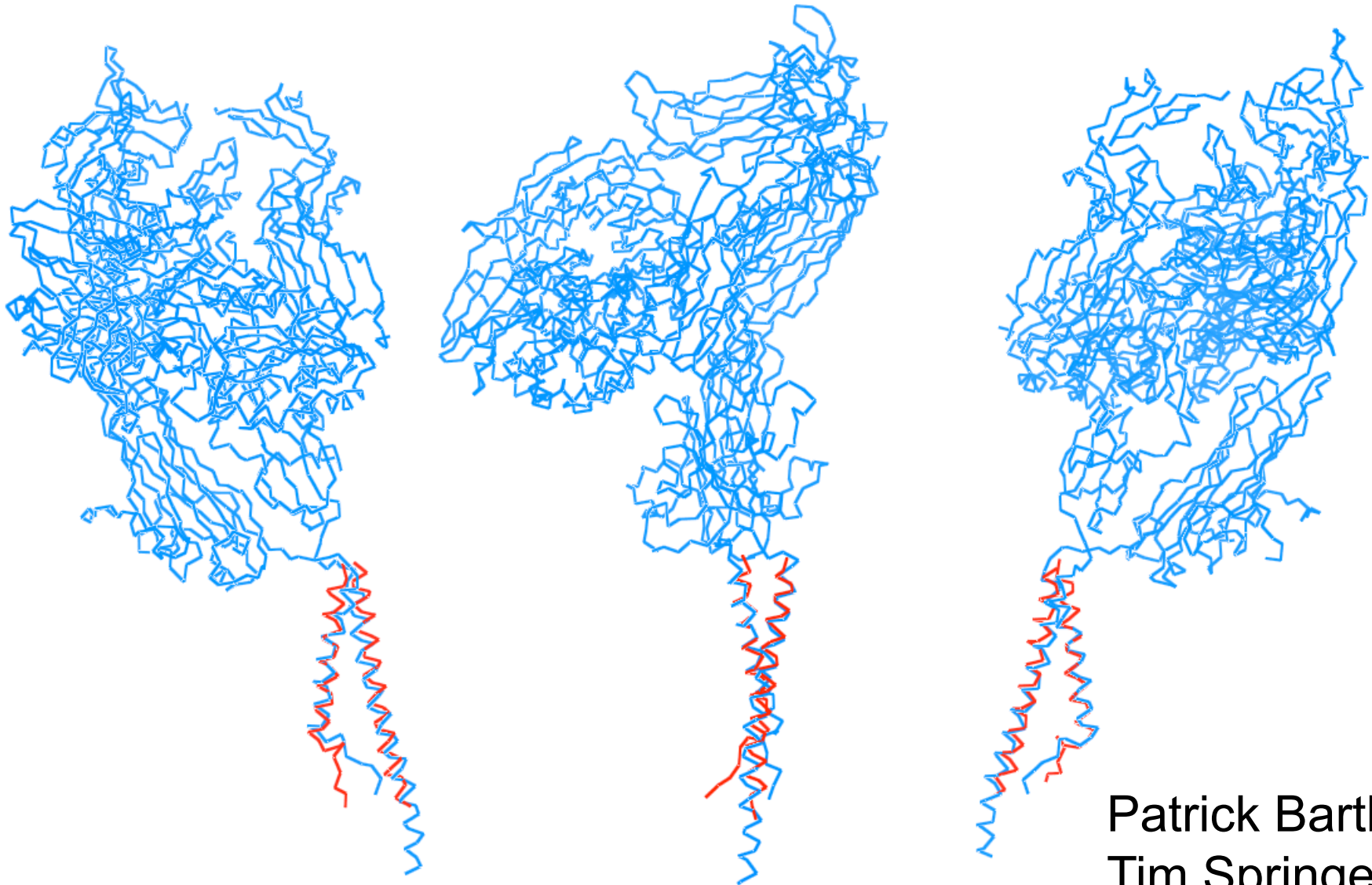


Patrick Barth
Tim Springer

Rosetta (Zhu et al., Mol.Cell in press)

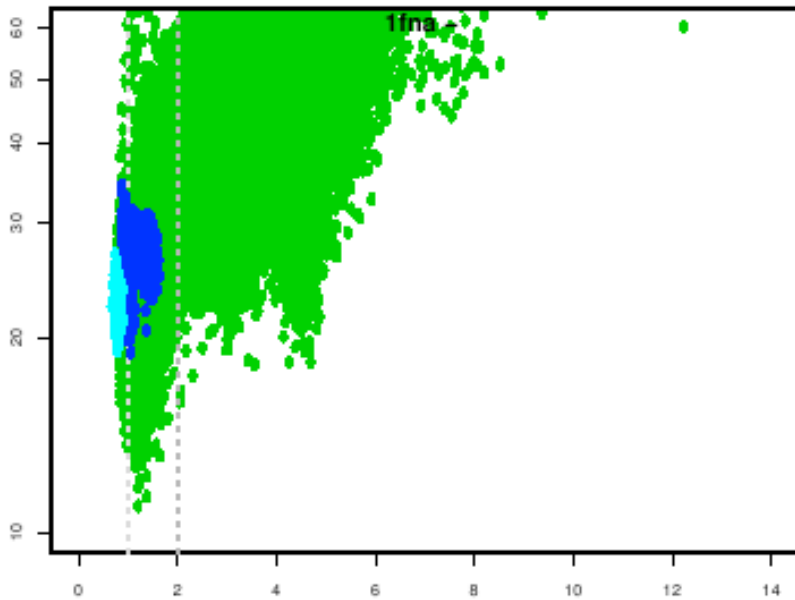
NMR (Lau et al., 2009, EMBO J., March 12)

Integrin α IIb β 3 model based on Rosetta + disulfide constraints entire heterodimer



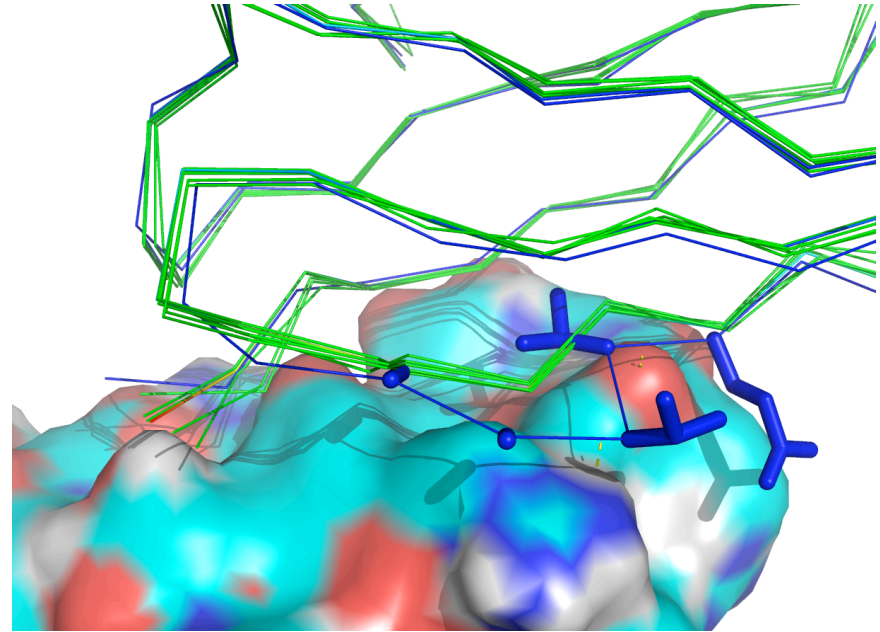
Patrick Barth
Tim Springer

Low energy Rosetta structures perhaps better models of proteins in solution than crystal structures?? Heresy!



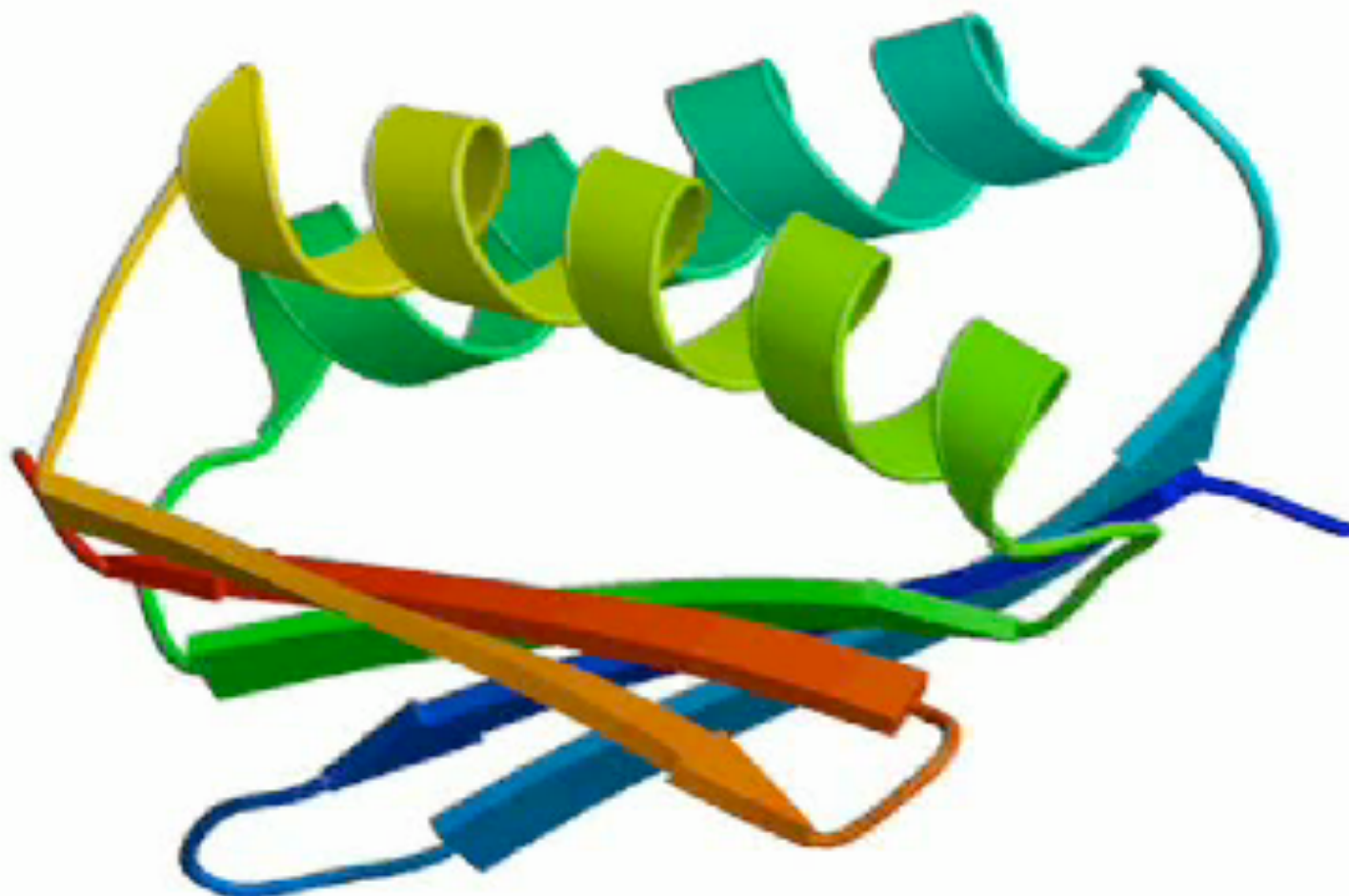
1FNA

Green: Rosetta
Blue: Native

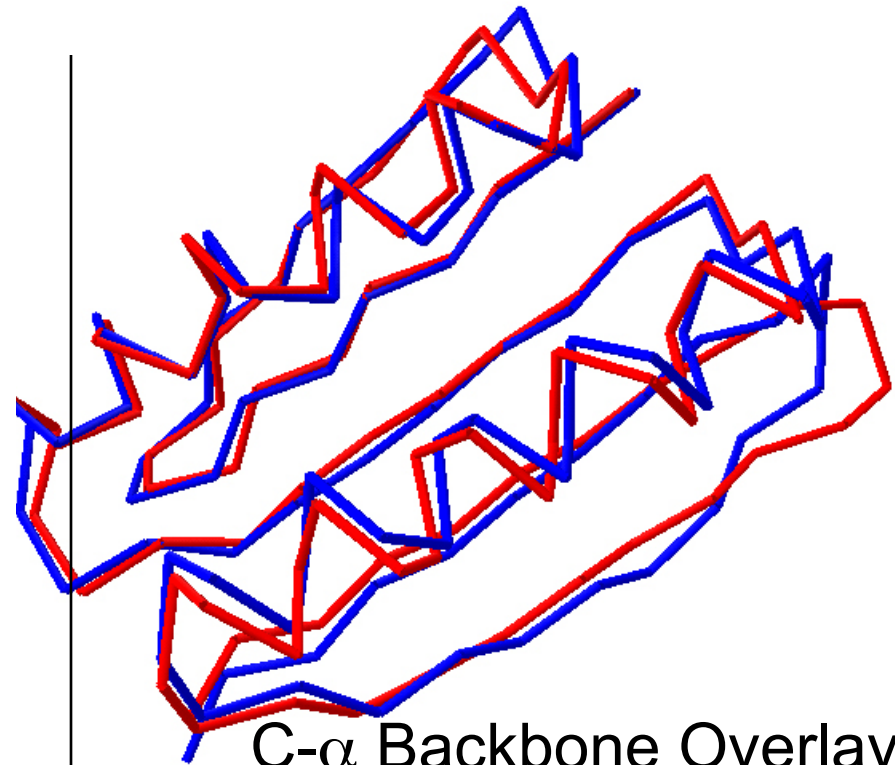
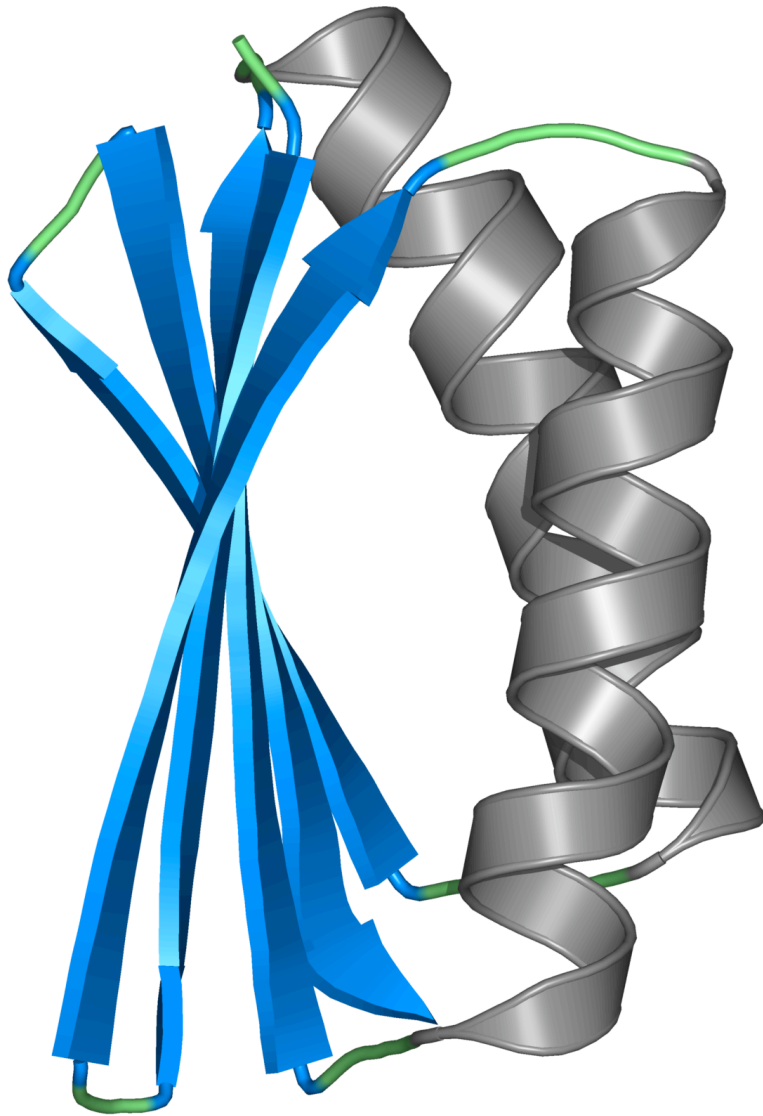


Mike Tyka
Jane Richardson

Protein Design



Top7 X-ray structure has correct topology.
Backbone RMSD to design only 1.2Å



C- α Backbone Overlay

Red : X-ray structure

Blue : Design model

Brian Kuhlman, Gautam Dantas;
Science 302 1364-8

Design of new protein functions

- Design of new protein-protein interactions
- Design of enzymes catalyzing novel chemical reactions
- Design of new DNA cutting enzymes
- Design of HIV vaccine

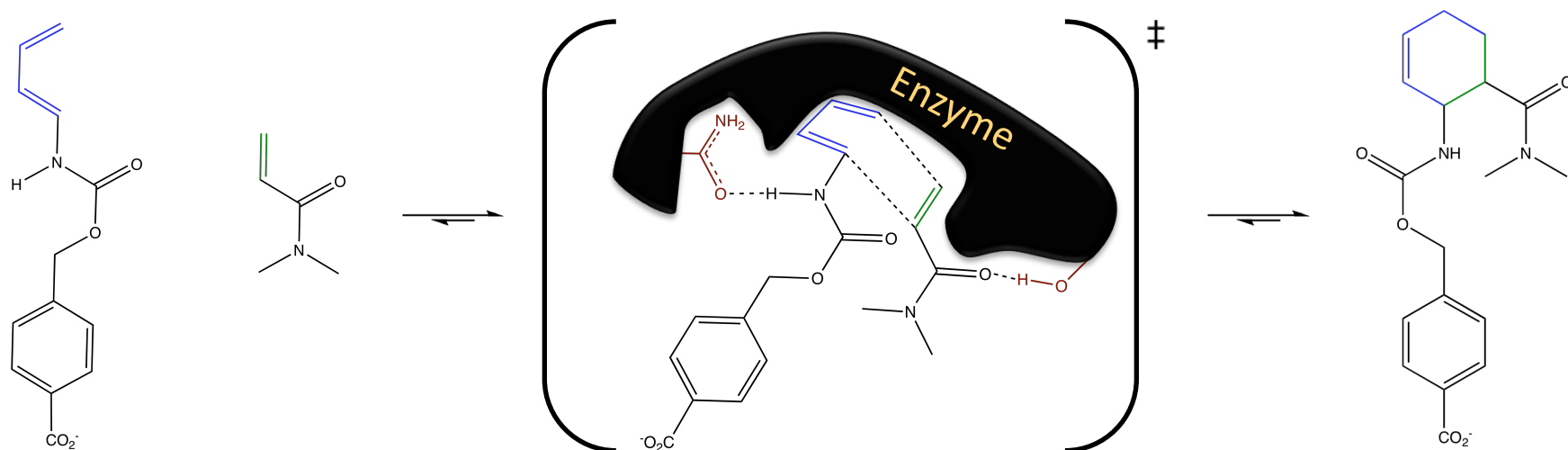
Design of Novel Enzymes

- I. Model reaction transition states and intermediates
- II. Design disembodied ideal active site around transition states and intermediates
- III. Design protein containing ideal active site

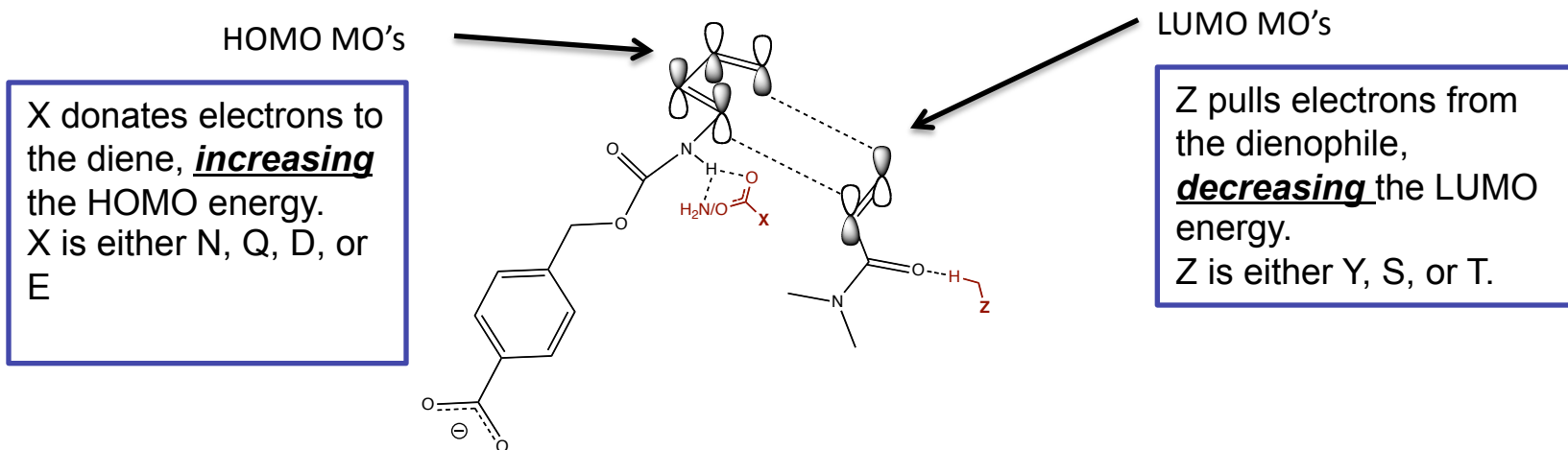
Alex Zanghellini, Daniela Roethlisberger, Lin Jiang,
Eric Althoff

de novo Computational Enzyme Design: Engineering a Stereoselective Bimolecular Catalyst

THE DIELS-ALDER REACTION



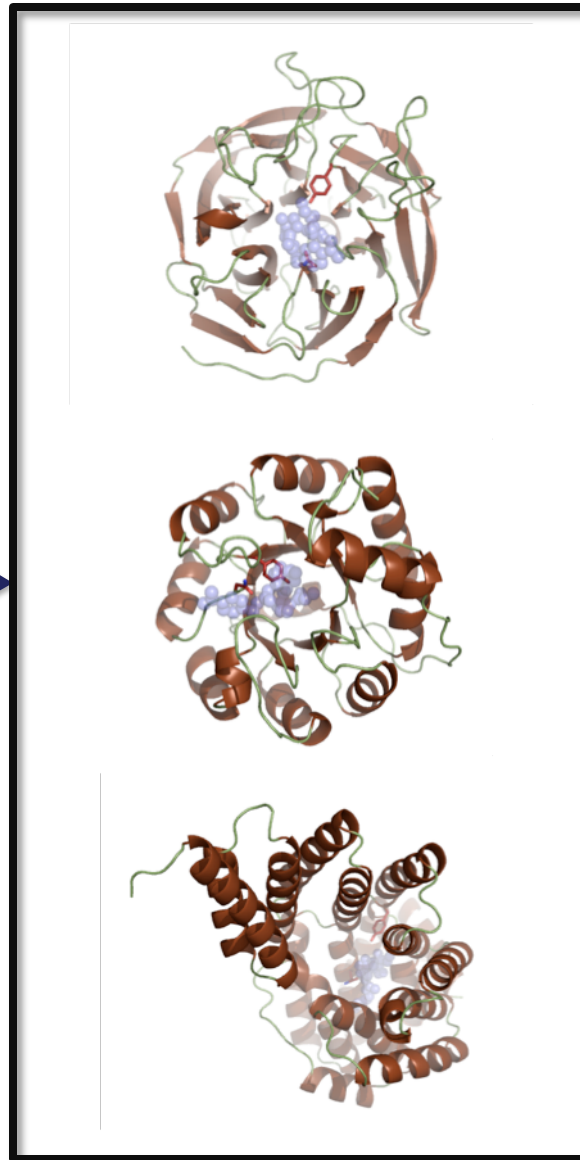
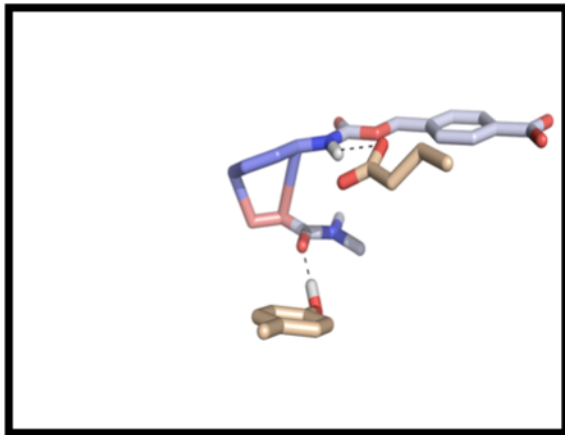
IDEAL ACTIVE SITE



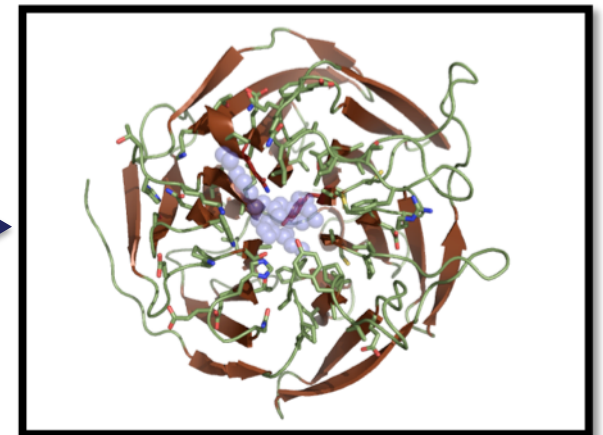
de novo Enzyme Design using Rosetta

ROSETTA MATCH

BUILD IDEAL ACTIVE
SITE



ROSETTADESIGN

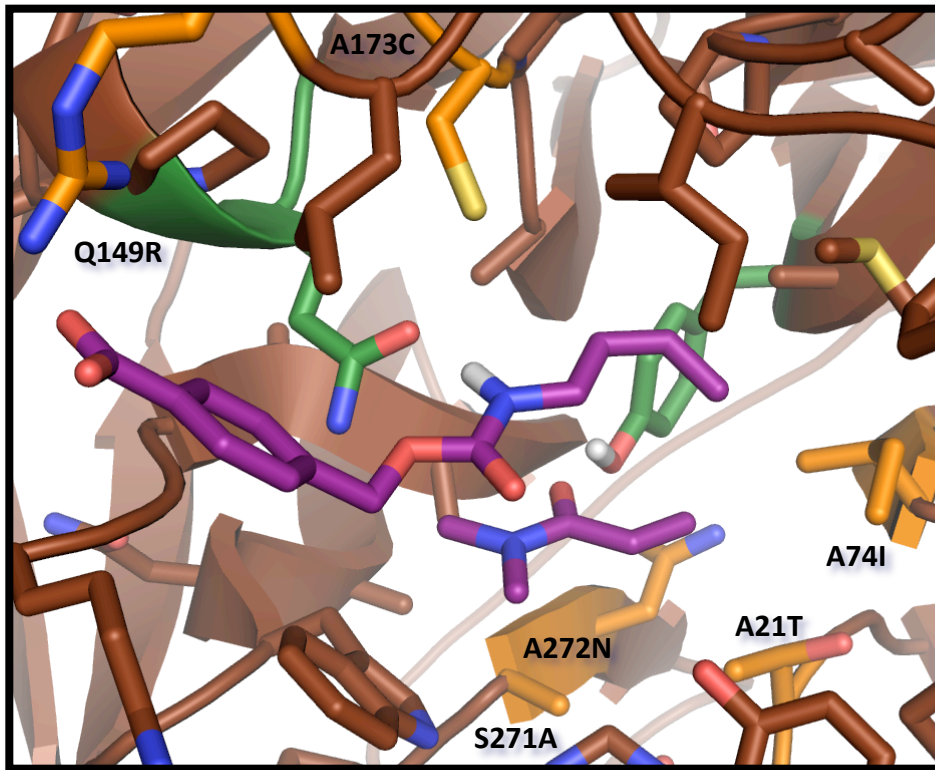


Protein Sci 2006, 15:2785-2794.
Nature 2008, 453:190-U194.
Science 2008, 319:1387-1391.

de novo designed Diels-Alderase

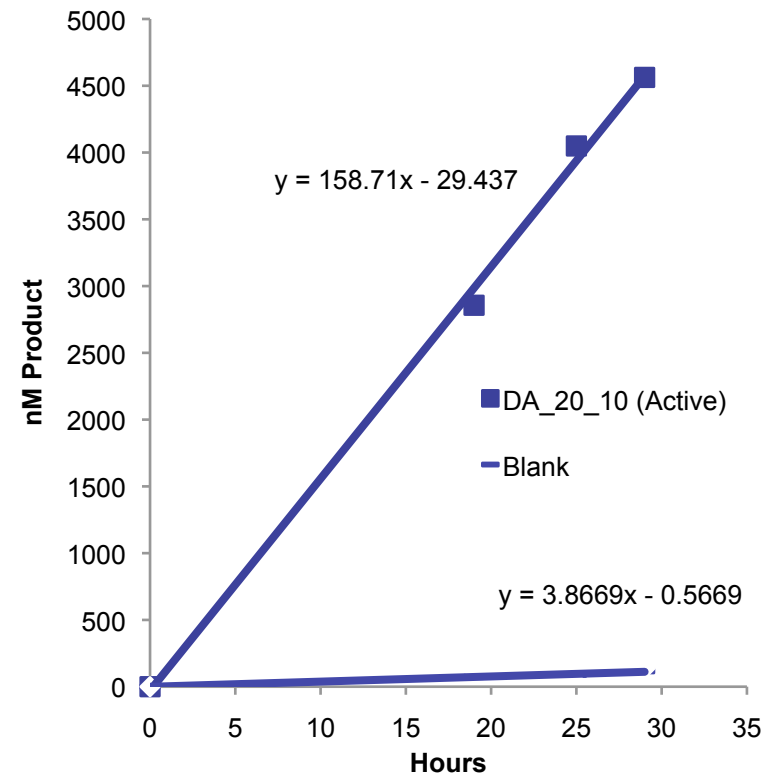
DA_20_10 ACTIVE SITE VIEW

, CATALYTIC RESIDUES



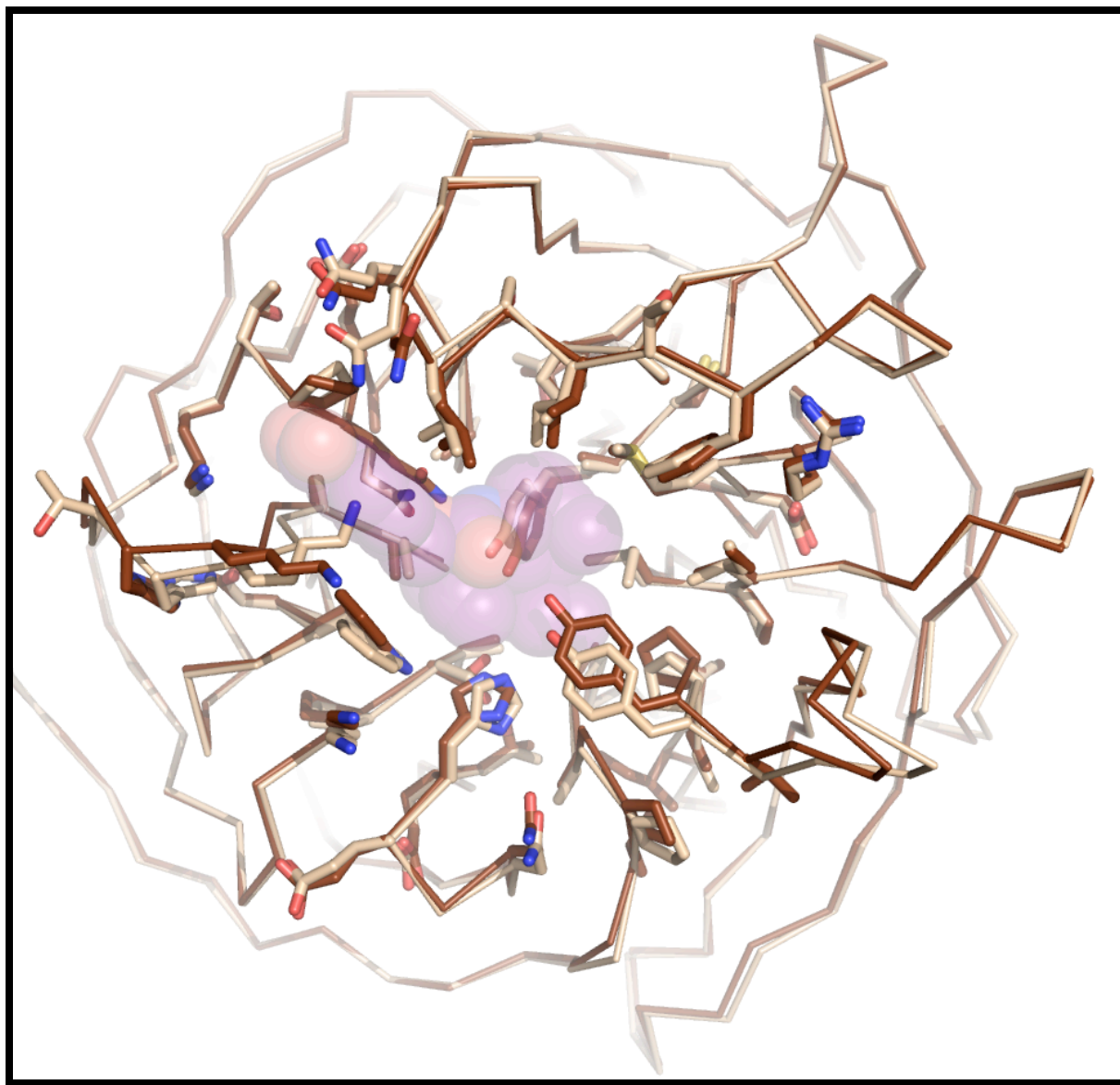
DIELS-ALDER REACTION PROGRESS CURVE

(1x PBS, 298K, 0.1MM DIENE, 3MM DIENOPHILE, 20UM PROTEIN)

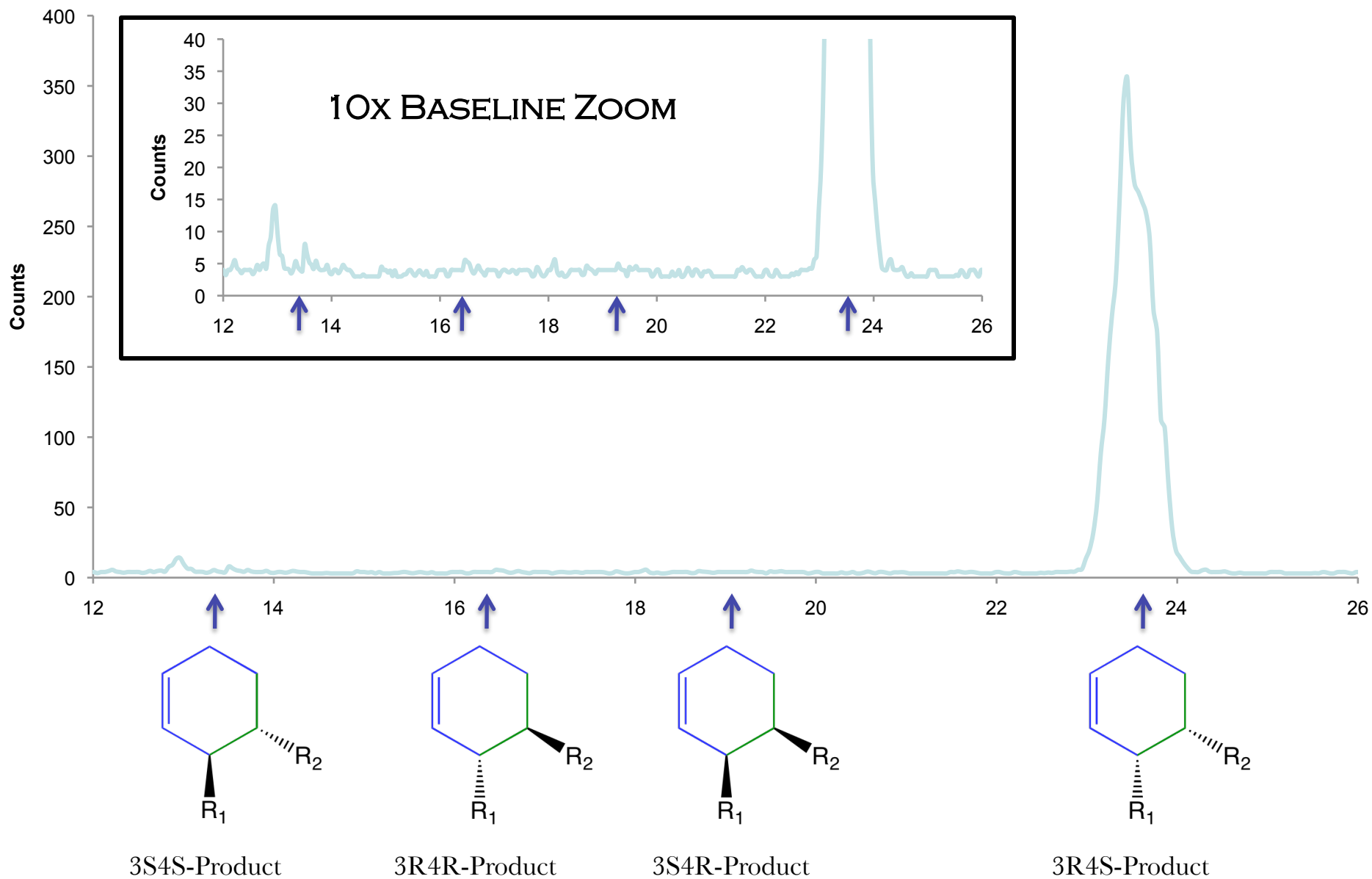


Crystal Structure of designed Diels-Alderase

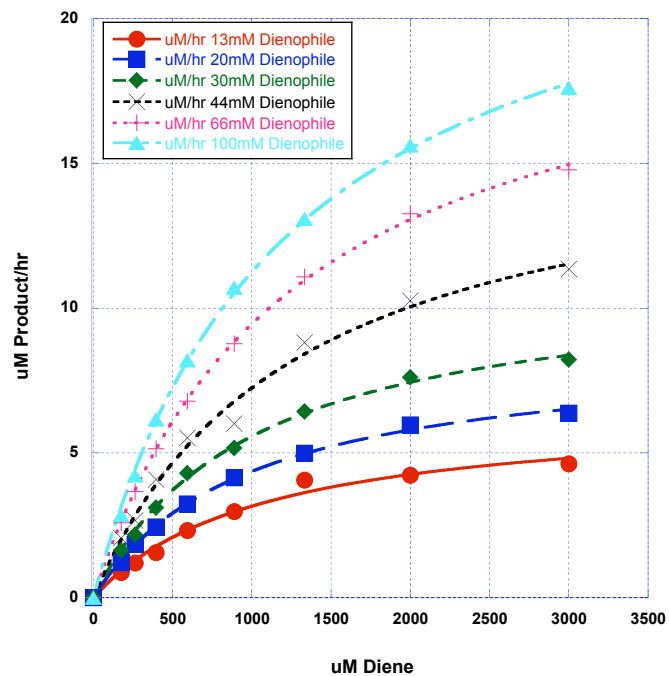
DESIGN (BROWN) vs. CRYSTAL STRUCTURE (CREAM) ALL ATOM RMSD: 0.3Å



Stereospecificity of designed Diels-Alderase



Kinetic Characterization of *designed* Diels-Alderase



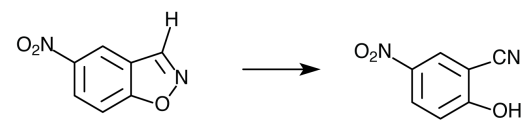
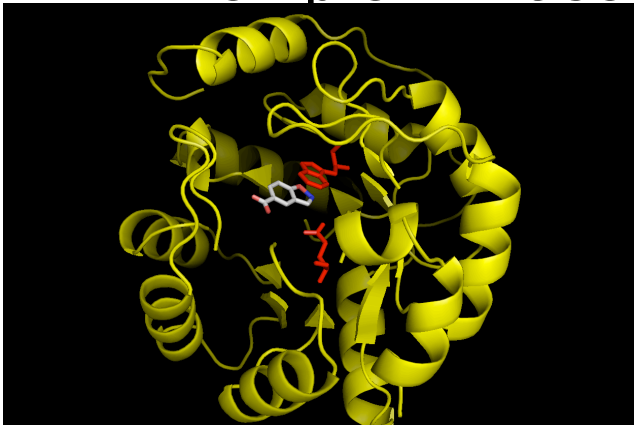
Kinetic Constants

Enzyme	k_{cat} (hr^{-1})	$K_{M-diene}$ (mM)	$K_{M-dienophile}$ (mM)
DA_20_00 (298K)	0.10	3.53	146.3
DA_20_10 (298K)	2.39	0.95	56.1
mAb 7D4 (310K)	0.21	0.96	1.7
mAB 4D5 (310K)	0.21	1.6	5.9

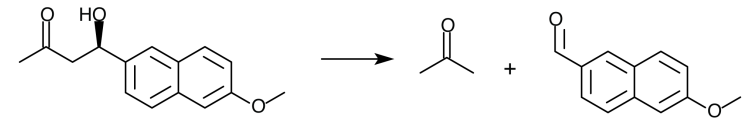
De novo enzyme design-- Successes thus far

- General acid-base catalysis: Kemp elimination
- Covalent catalysis: novel aldol and Michael condensation catalysts (dozens of active retroaldol designs on several different scaffolds)
- Bimolecular reactions: Diels Alder
- Polar transition state stabilization: ester hydrolysis

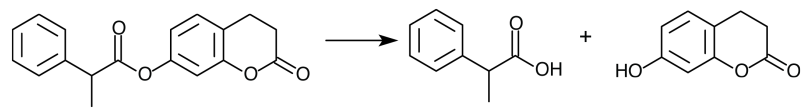
Kemp eliminase



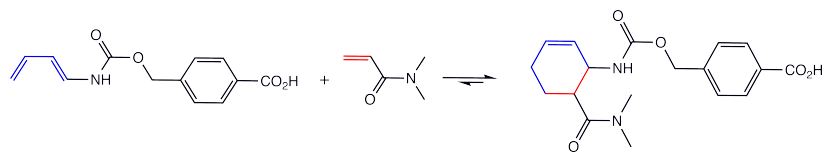
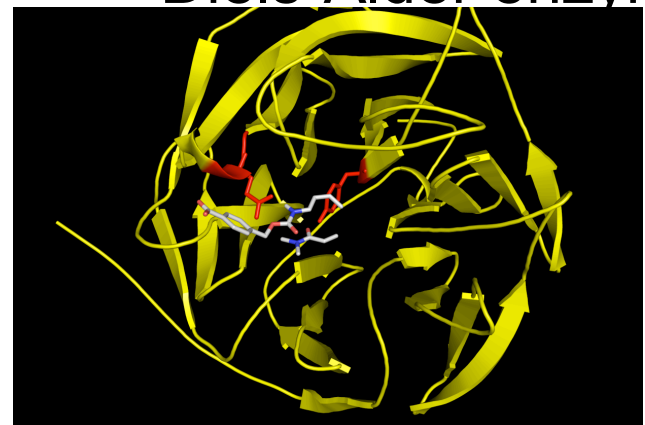
Retro-aldolase



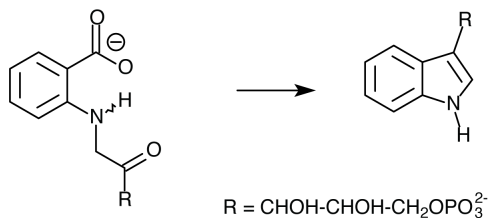
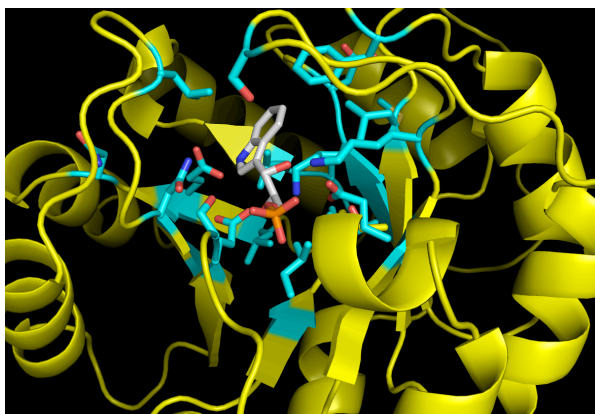
Esterase



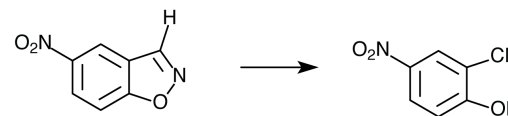
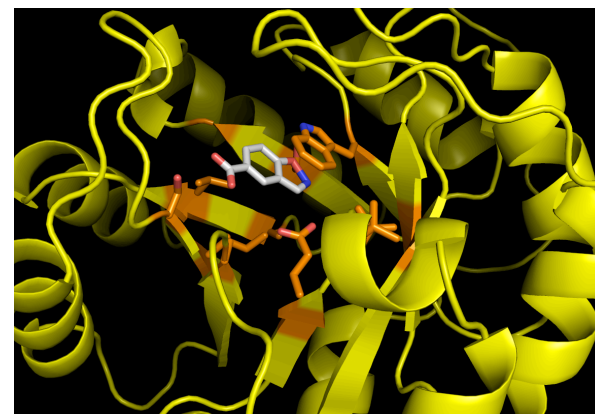
Diels-Alder enzyme



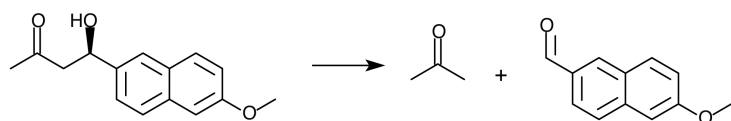
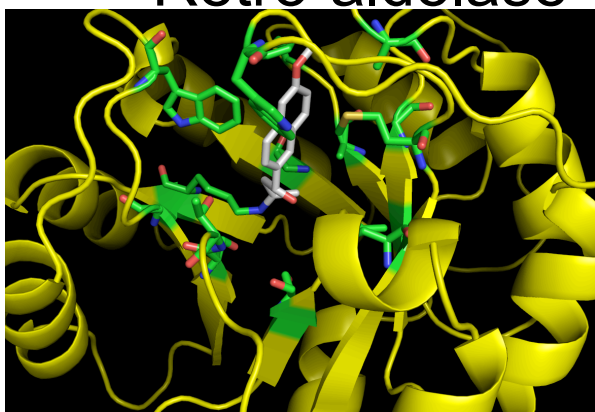
Indole-3-glycerol phosphate synth.



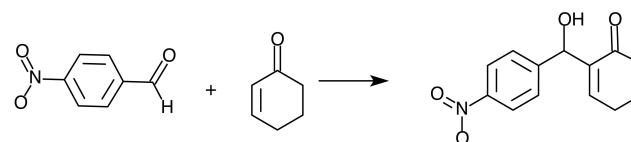
Kemp eliminase



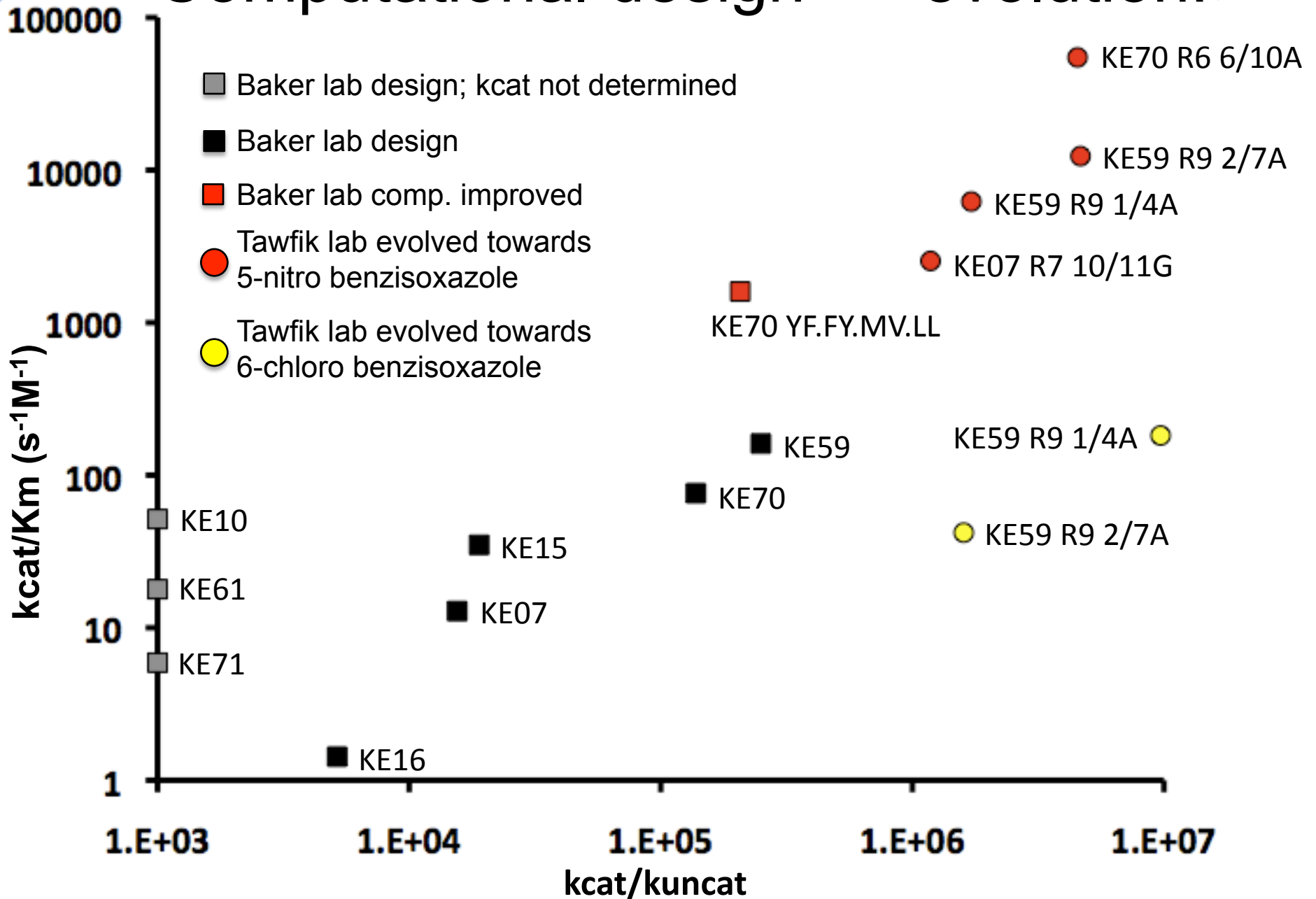
Retro-aldolase



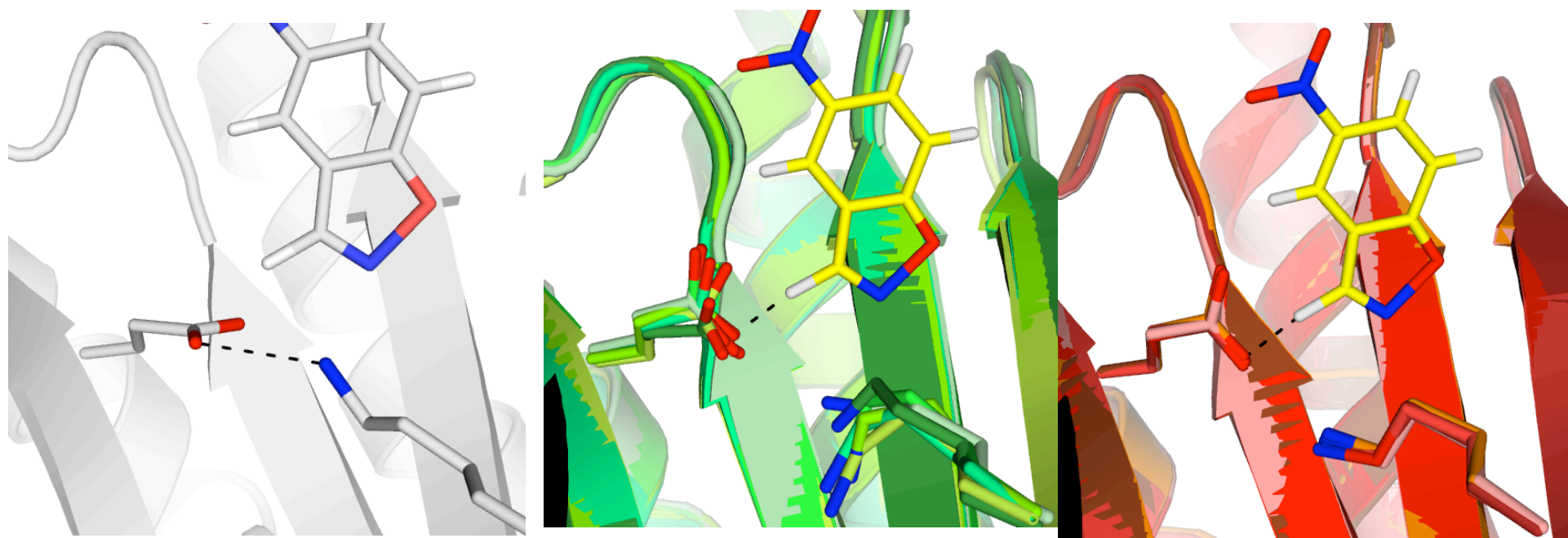
Baylis-Hillman enzyme



Computational design => evolution!



Structures of evolved variants illustrate shortcomings of design round 0 - round 4 - round 6



Precise positioning of catalytic groups critical!

Olga Kheronsky, Orly Dym, Danny Tawfik

De novo enzyme design-- lessons and questions

- Can design active enzymes from scratch!
- Starting activities low, but can be increased readily by directed evolution
- Need more precise positioning of catalytic groups, elimination of competing reactions (aldolase trapped intermediates), etc.
- Enzymes are masters of art of compromise--have to do everything well!
- Critical question is about evolution--what fraction of nascent enzymes have the potential to become highly active catalysts??

	Search problem?	Low accuracy?	Solution
Structure calculation	Yes	No	Experiment then Computation
Function design	No	Yes	Computation then Experiment

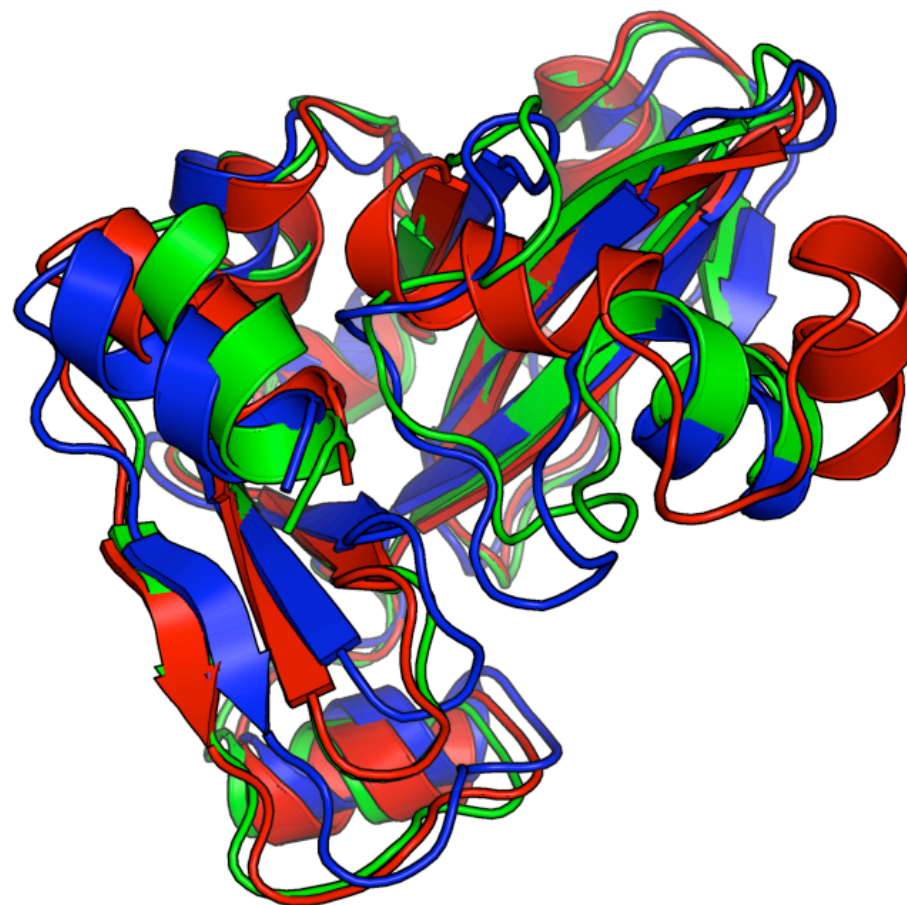
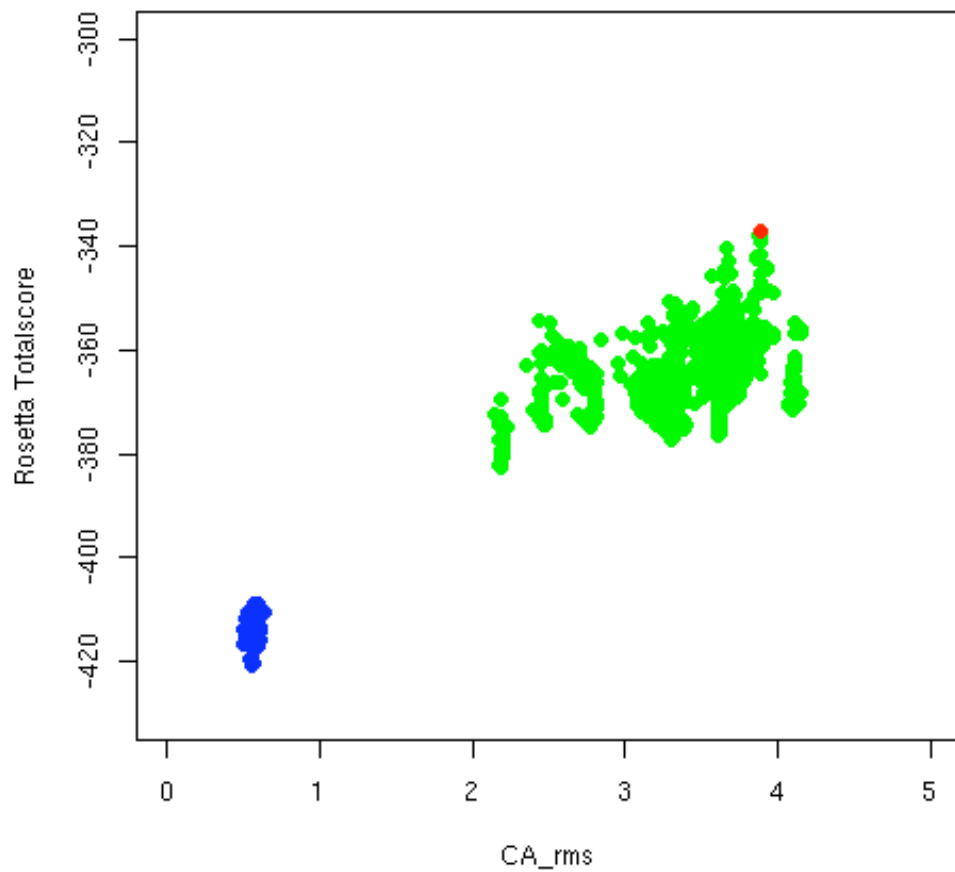
Accuracy high for structure calculation:
 Evolved energy gap for folded macromolecules

Accuracy low for enzyme design:
 No evolved energy gap for designed macromolecules
 Don't have complete understanding of requirements for catalysis. Will learn in the process!

Rosetta@home puts people's
computers to work to solve problems;
how to enlist their brains as well?

FoldIt--Multiplayer online
computer game for research and
education

Adrien Treuille, Seth Cooper,
Zoran Popovic, Firas Khatib



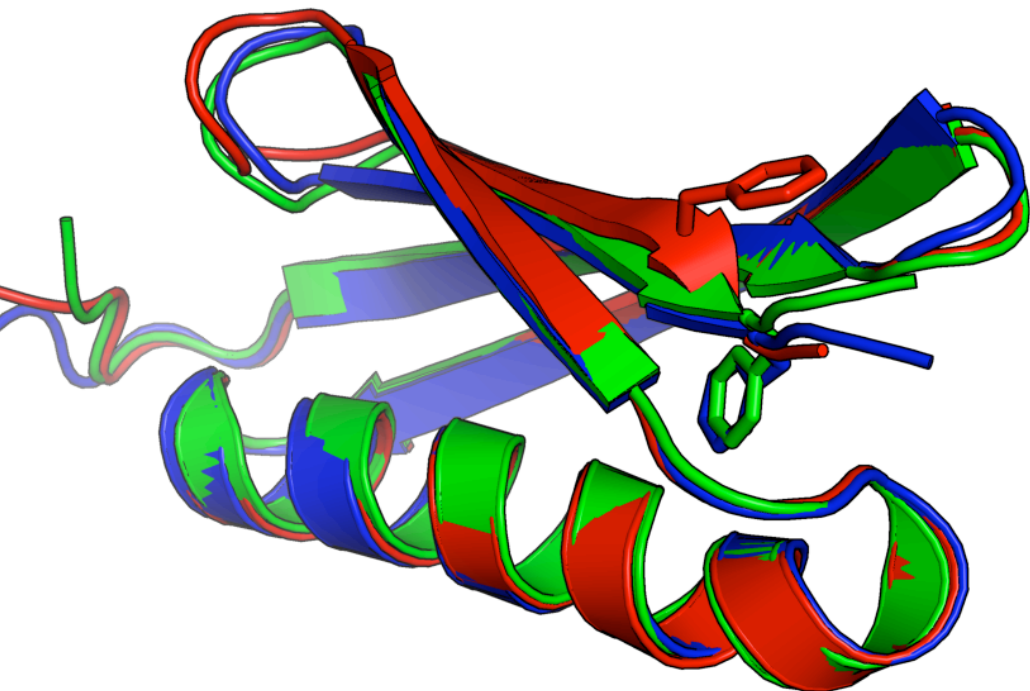
Blue = Native

Red = Foldit Puzzle

Green = Highest Scoring Foldit Solution

Player name: bzipitidoo

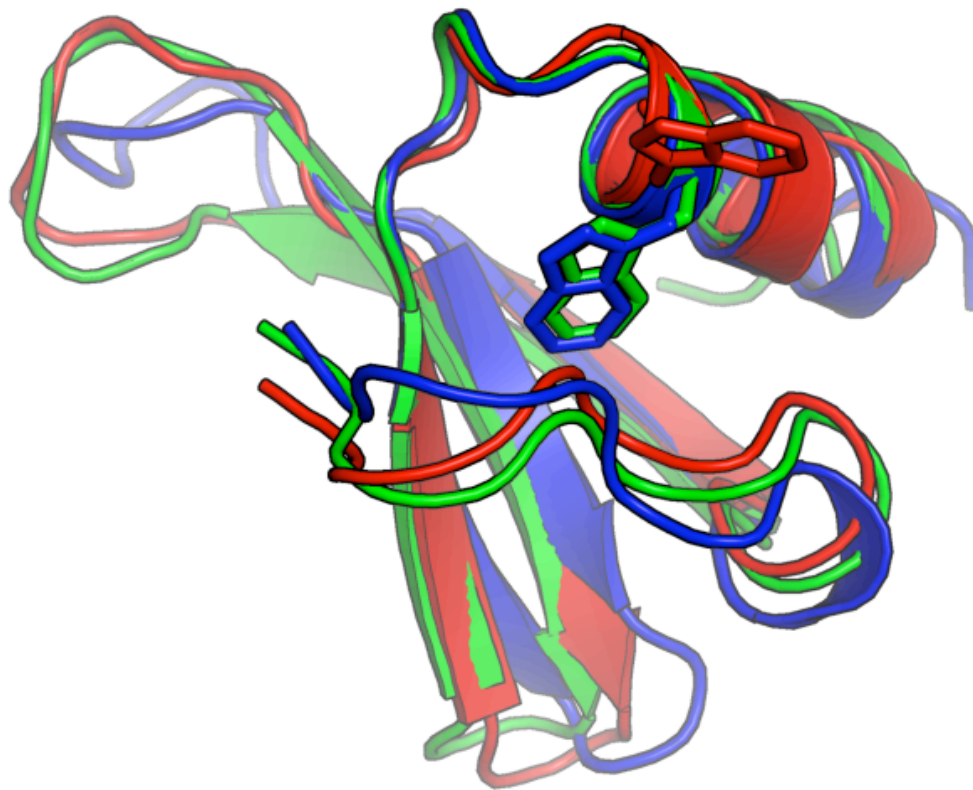
Foldit team name: Void Crushers



Blue = Native

Red = Foldit Puzzle

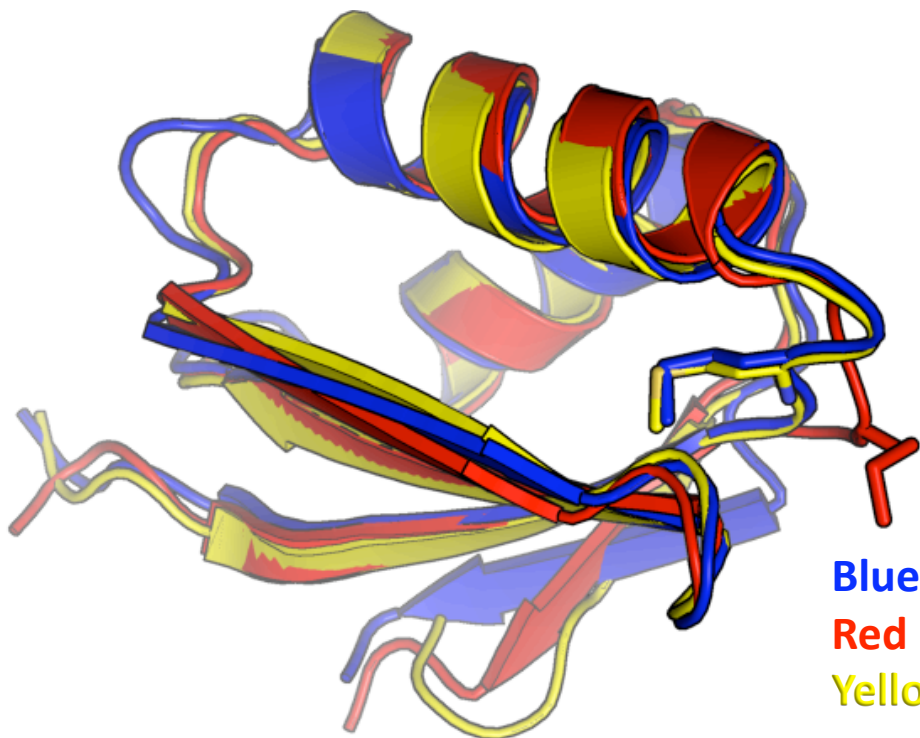
Green = Highest Scoring Foldit Solution



Blue = Native

Red = Foldit Puzzle

Yellow = 2nd Highest Scoring Foldit Solution



Acknowledgements

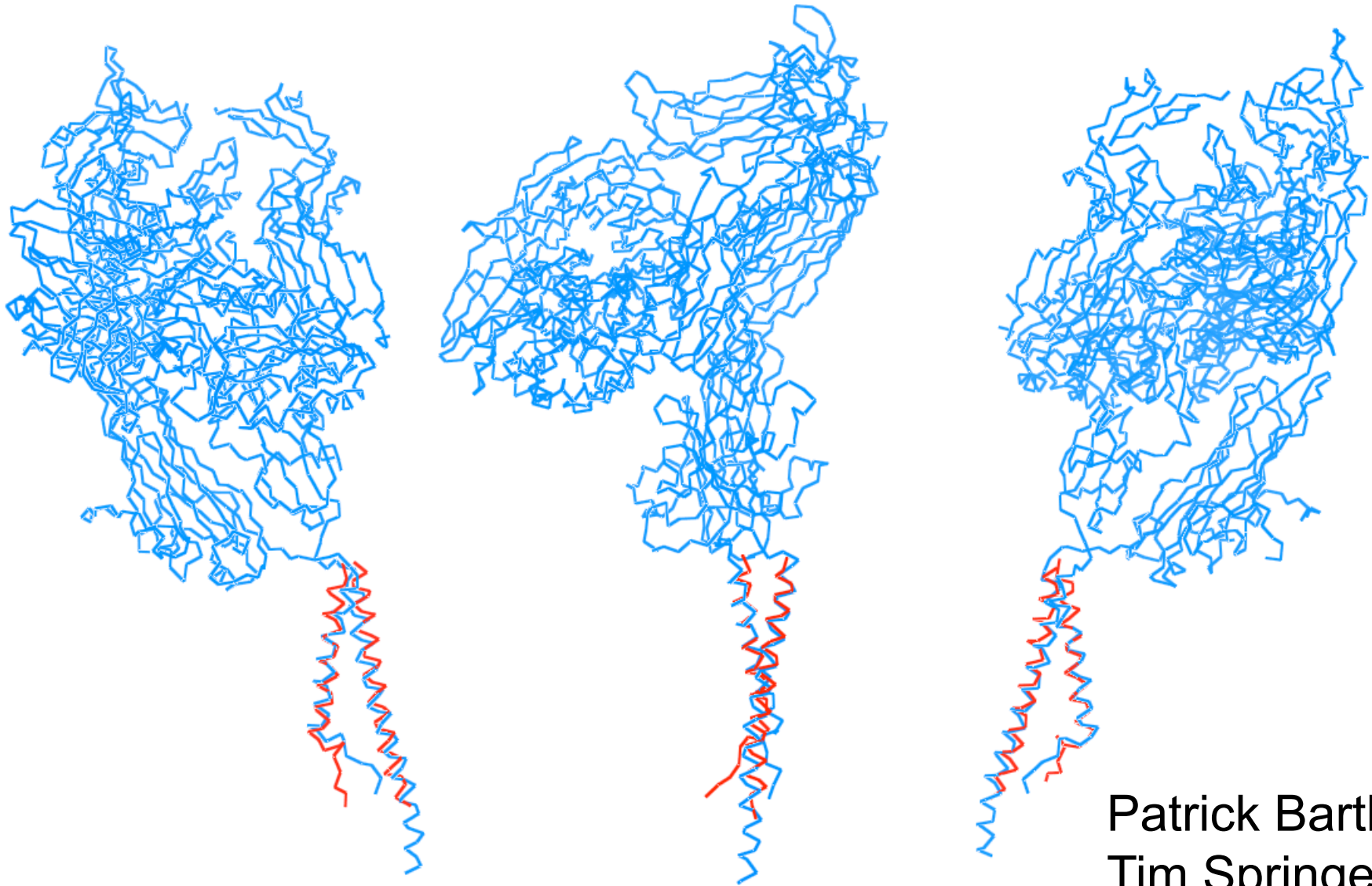
- Structure prediction:
Mike Tyka
Ingemar Andre
Patrick Barth
Oliver Lange
- Incorporation of experimental data
Watson Raman Ad Bax
Rhiju Das Wah Chiu
- Enzyme design
Justin Siegal Danny Tawfik and Olga Kheronsky
Alex Zanghellini Don Hilvert
Daniela Roethlisberger
Eric Althoff

Rosetta@home puts people's
computers to work to solve problems;
how to enlist their brains as well?

FoldIt--Multiplayer online
computer game for research and
education
<http://fold.it>

Adrien Treuille, Seth Cooper,
Zoran Popovic, Firas Khatib

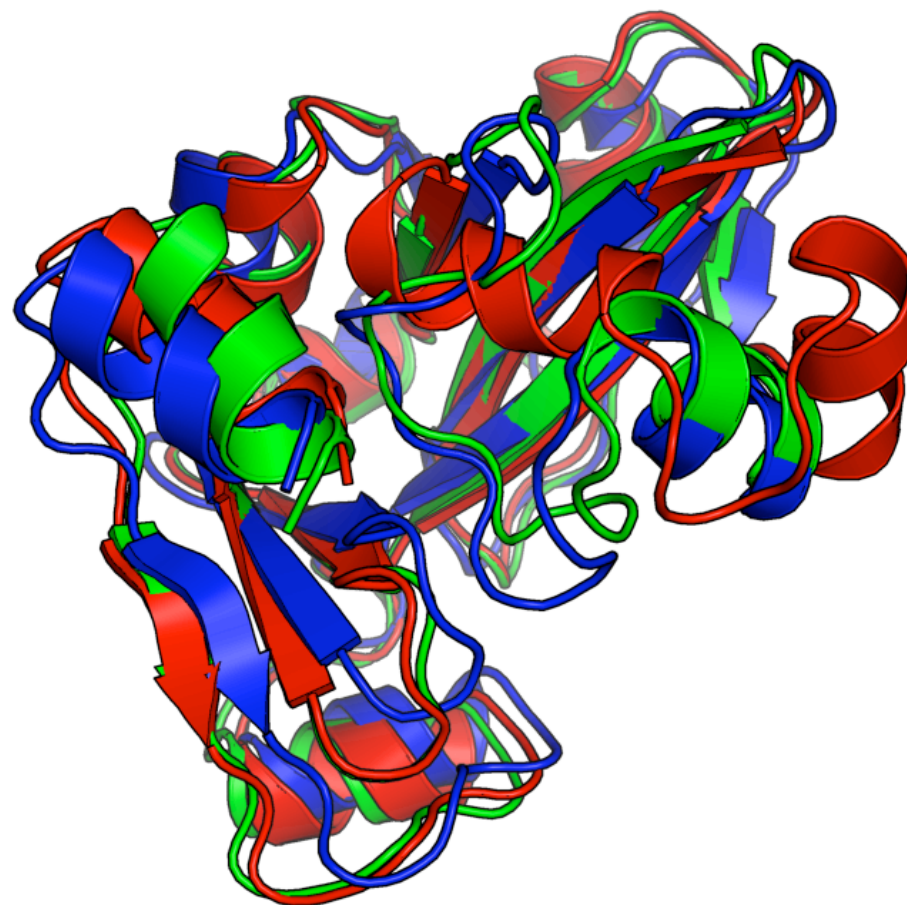
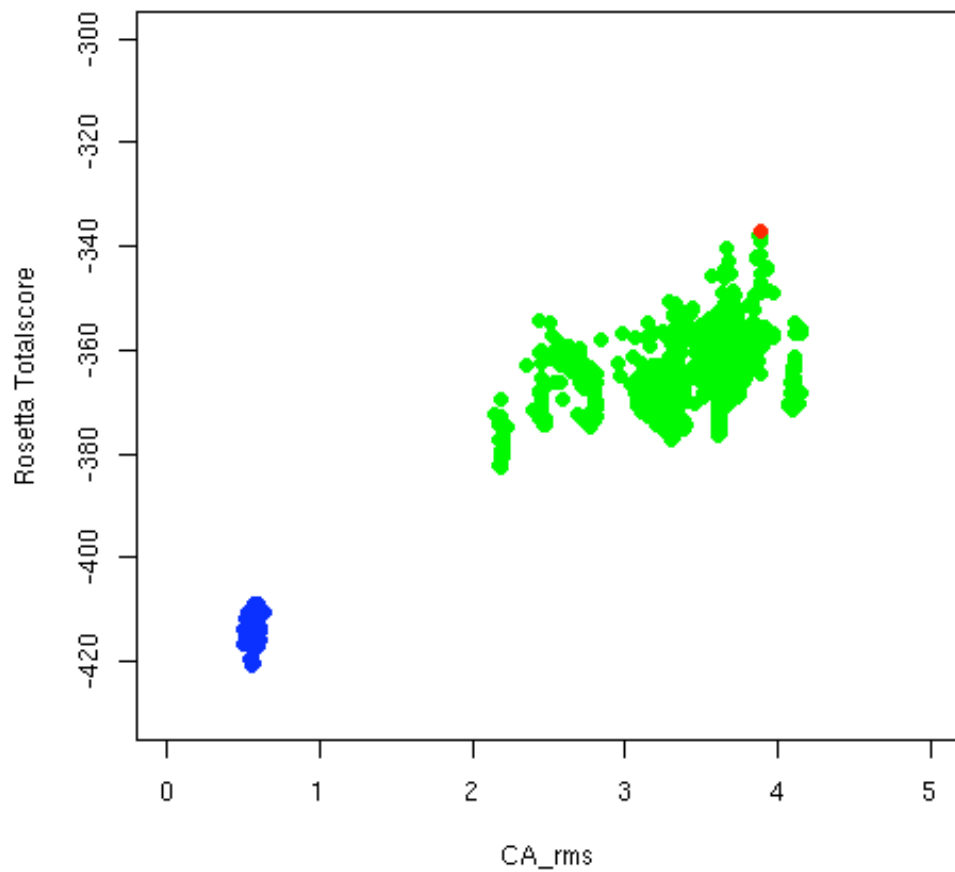
Integrin α IIb β 3 model based on Rosetta + disulfide constraints entire heterodimer



Patrick Barth
Tim Springer

- Structure determination:
experiment=>computation=>global
minimum
- Function design:
computation=>experiment=>high
activity

Problems are opposite, in structure determination have high accuracy but search problem; in enzyme design, no search problem but low accuracy



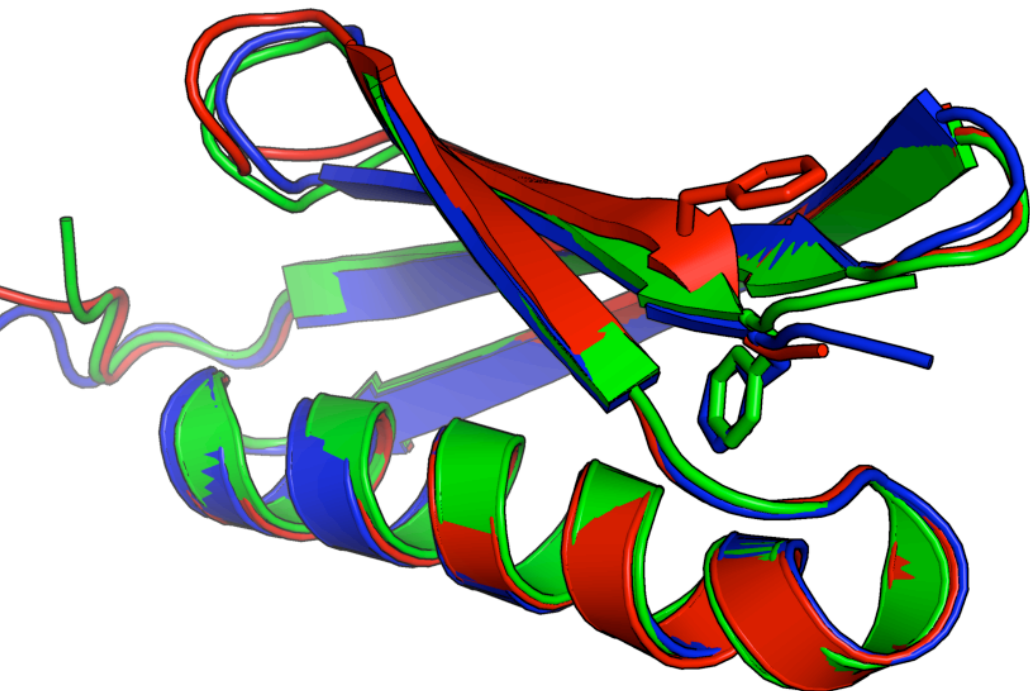
Blue = Native

Red = Foldit Puzzle

Green = Highest Scoring Foldit Solution

Player name: bzipitidoo

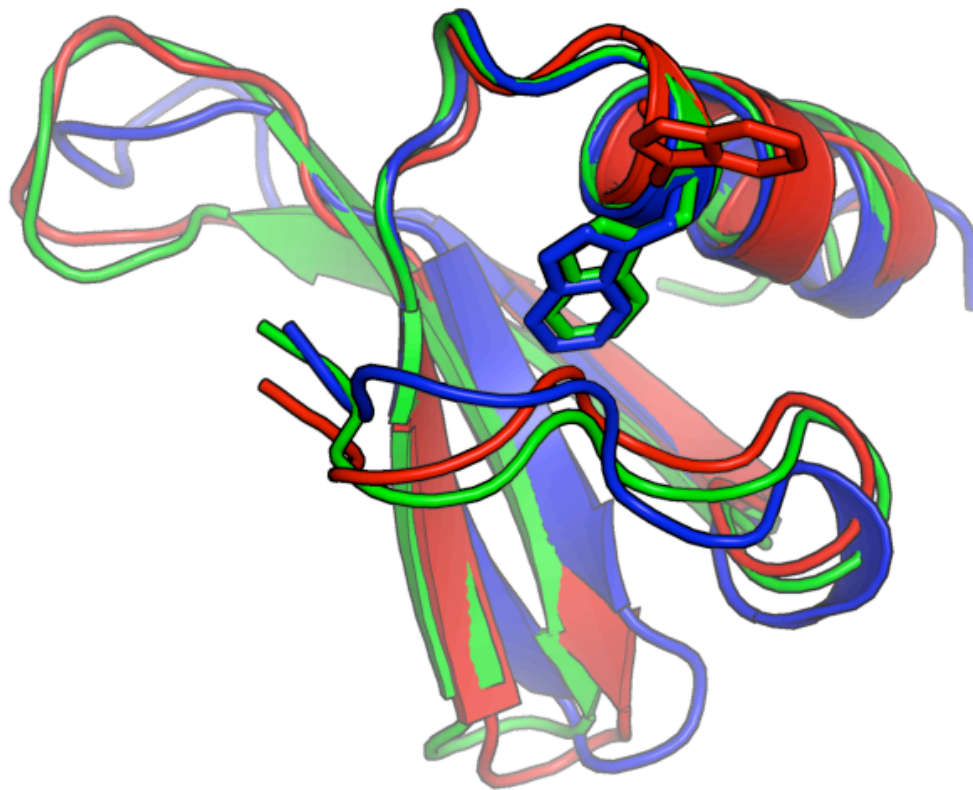
Foldit team name: Void Crushers



Blue = Native

Red = Foldit Puzzle

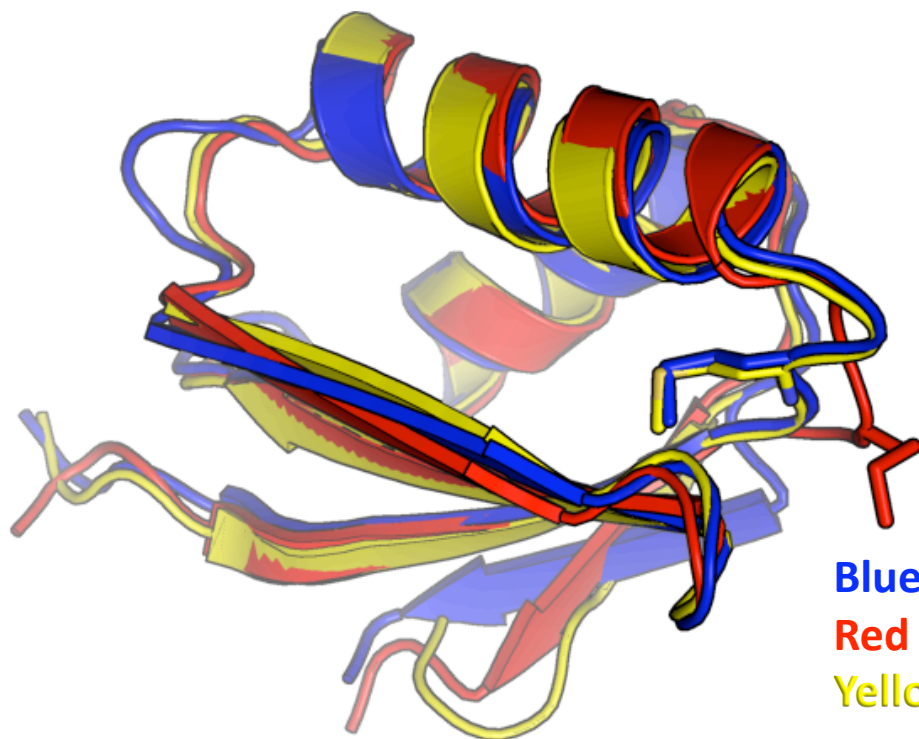
Green = Highest Scoring Foldit Solution



Blue = Native

Red = Foldit Puzzle

Yellow = 2nd Highest Scoring Foldit Solution



Improving autobuilt model in 4Å crystallographic data

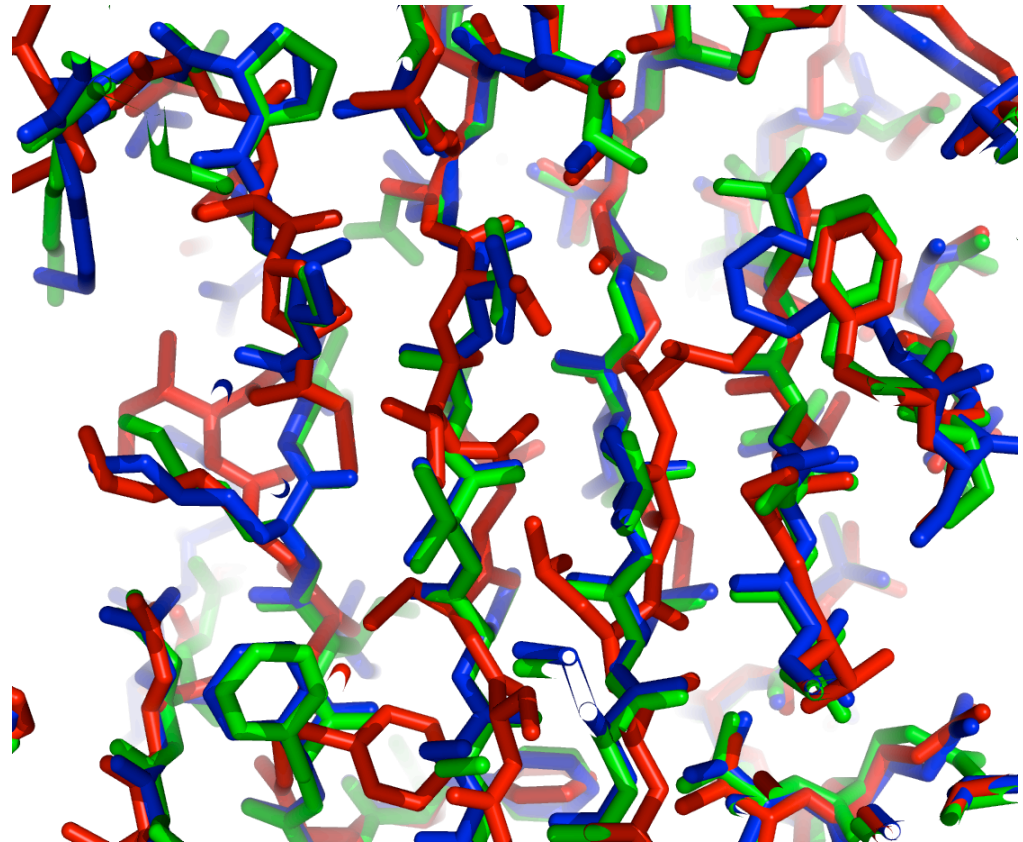
Autobuilt model

- 1.12Å RMS
- 85% C α within 1Å of native

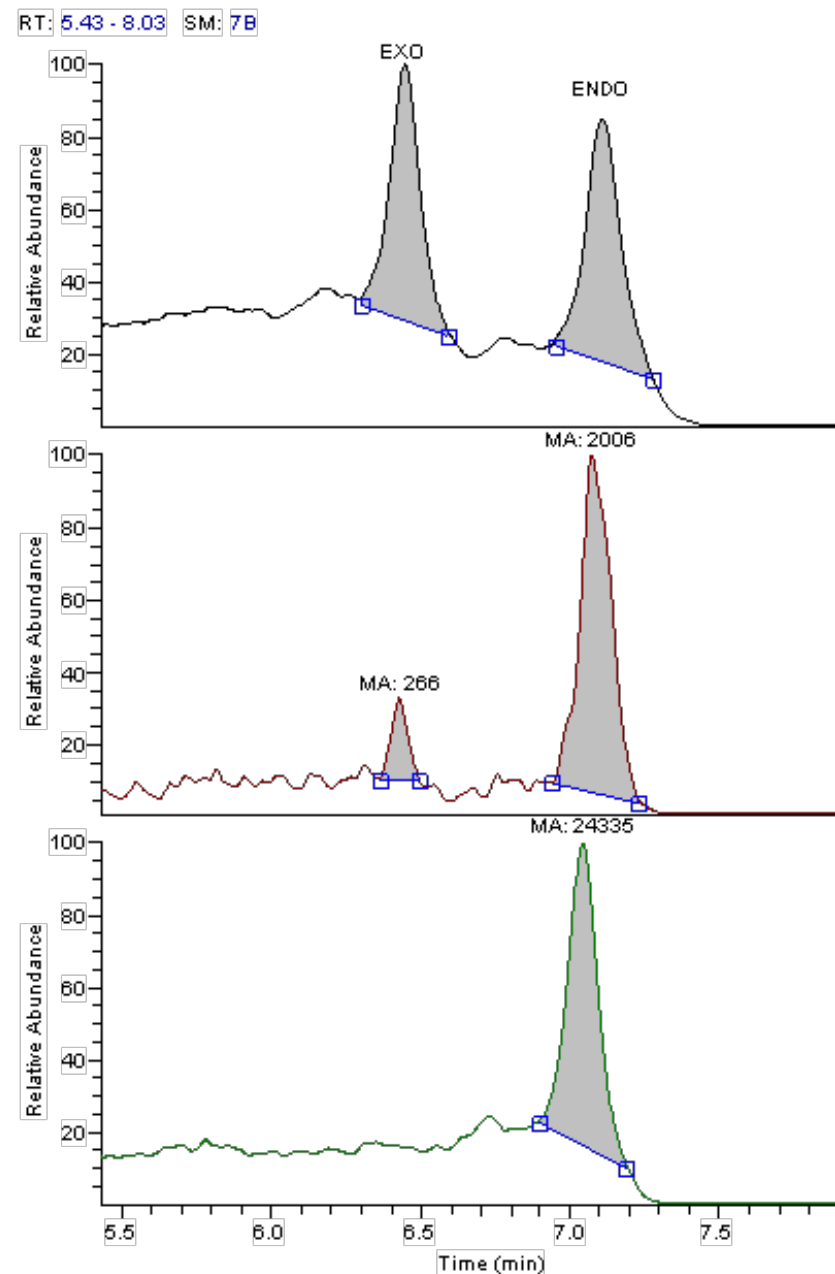
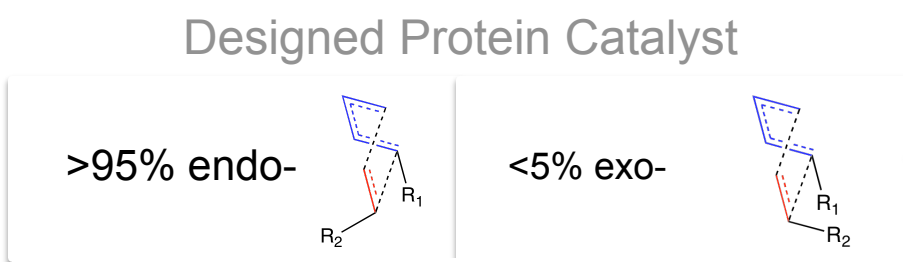
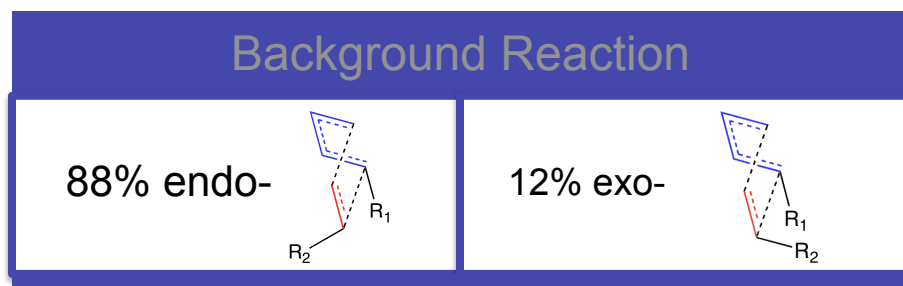
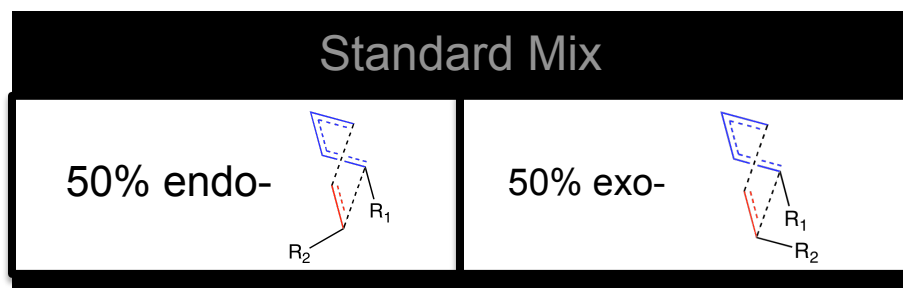
Rosetta prediction

- 0.88Å RMS
- 92% C α within 1Å of native

Native structure



Designed enzyme is >95% Stereoselective for the Endo Diastereomer!



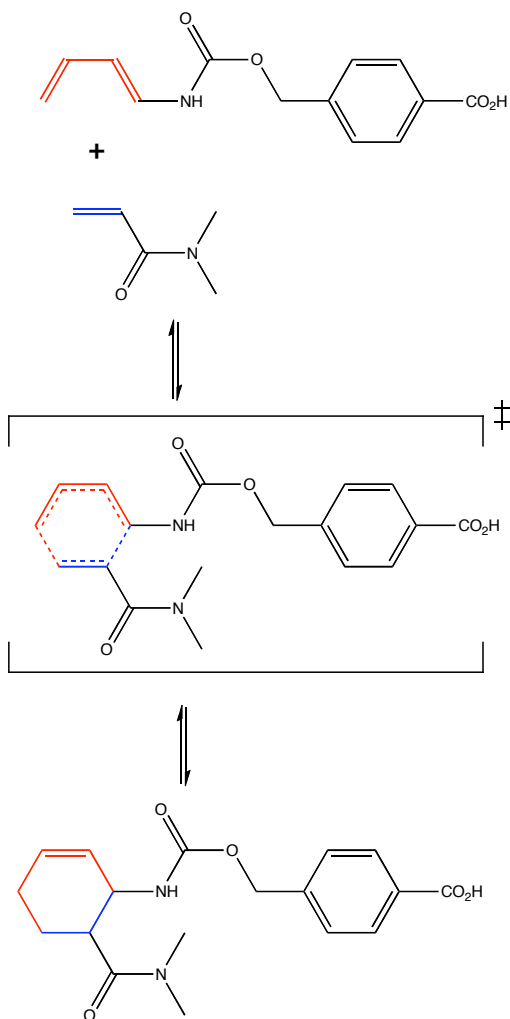
Rate enhancement greater than 10^4 (depending on definition)

	Description	Units	DA ₋₂₀ - ₁₀	7D4
$(k_{\text{cat}}/K_{\text{M-Diene}}K_{\text{M-Dienophile}})/k_{\text{uncat}}$	rate enhancement per mole of enzyme	M ⁻¹	1.11 x 10 ⁶	2.95 x 10⁶
$(k_{\text{cat}}/K_{\text{M-Diene}})/k_{\text{uncat}}$	rate enhancement saturating dienophile	-	4.03 x 10⁴	5.01 x 10 ³
$(k_{\text{cat}}/K_{\text{M-Dienophile}})/k_{\text{uncat}}$	rate enhancement saturating Diene	-	1.30 x 10 ³	2.83 x 10³

Justin Siegal and Alex Zanghellini

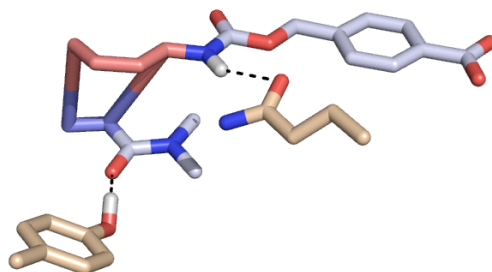
Computational Enzyme Design of A Novel Intermolecular Diels Alderase

Select Reaction

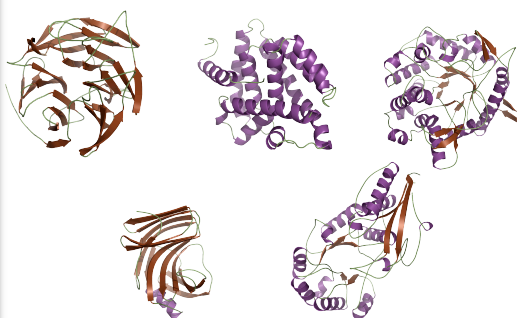


Build Enzyme *in silico*

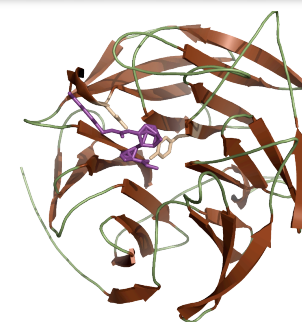
3D Model of Ligand and Catalytic Amino Acids



Protein Scaffold Library

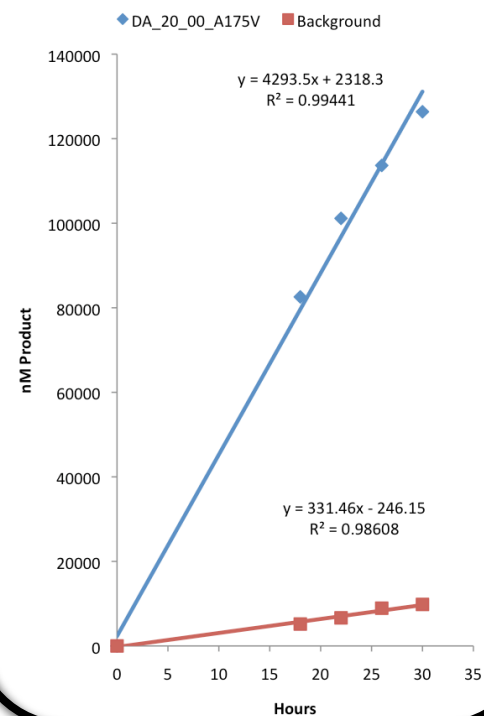


Validate Novel Enzyme



Progress Curve
for Diels Alder Reaction

(1x PBS, Room Temperature, 200microM Protein)



Justin Siegal, Alex Zanghellini

De novo enzyme design-- lessons

- Can design active enzymes from scratch!
- Starting activities low, but can be increased readily by directed evolution
- Need more precise positioning of catalytic groups, elimination of competing reactions, etc.
- Enzymes are masters of art of compromise--
have to do everything well!

Acknowledgements

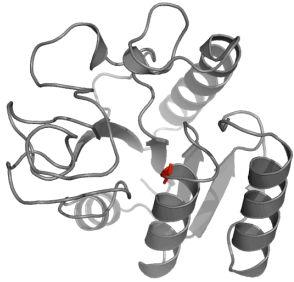
- Structure prediction:
 - Mike Tyka
 - Ingemar Andre
 - Patrick Barth
 - Nick Grishin
- Incorporation of experimental data
 - Vatson Raman
 - Rhiju Das
 - Ad Bax
 - Yang Shen
- Enzyme design
 - Justin Siegal
 - Alex Zanghellini
 - Daniela Roethlisberger
 - Eric Althoff
- Foldit
 - Adrien Treuille
 - Seth Cooper
 - Zoran Popovic

De novo enzyme design-- Successes thus far

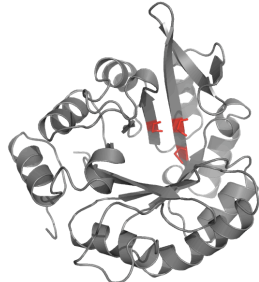
- General acid-base catalysis: Kemp elimination
- Covalent catalysis: novel aldol and Michael condensation catalysts
- Bimolecular reactions: Diels Alder
- Polar transition state stabilization: ester hydrolysis

Aldolase Design Diversity

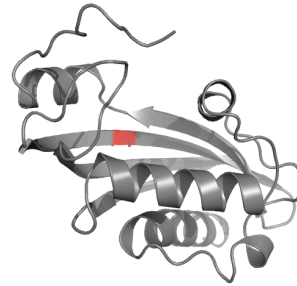
Red shows **Imine-Lysine positions** of active designs. Wide range of positions and scaffolds!



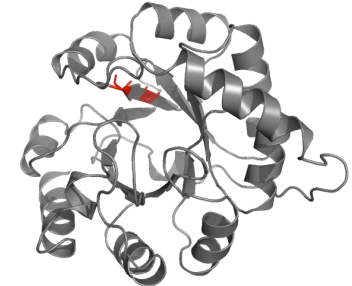
Rossman-1ilw (1)



TIM-1thf (3)



KSI-1oho (1)



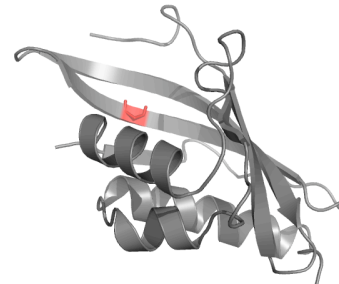
TIM-1i4n (3)



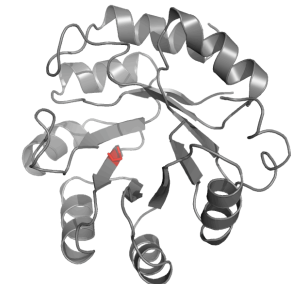
TIM-1a53, 1lbl,
1lbf, 2c3z (32)



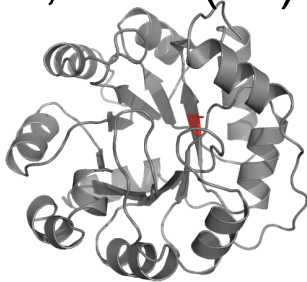
BetaBarrel-1v04 (1)



NTF2-1sjw (1)



TIM-1dl3 (3)



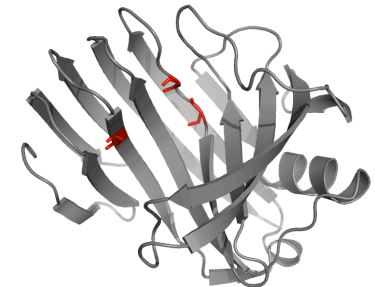
TIM-1igs (1)



Jelly Roll-1pvx (2)



Jelly-1m4w, 3b5l (10)



Jelly- 1f5j (4)