

Rosetta-based Computational Pipeline for Antibody Engineering

Jianqing Xu
Jeffrey J. Gray

Chemical & Biomolecular Engineering
Johns Hopkins University

Real Rosetta Stone!!!



Motivation for Antibody Modeling

- Antibodies are important and we need better antibodies.
- Antibody-based engineering can be frustrated by the difficulties of crystallizing antibody or antibody-antigen structures.
- Computational structure prediction and design provides a fast and inexpensive route to address antibody engineering problems.

Outline

- Application of Computational Pipeline (homology modeling, docking, design) to bio-defense antibody-antigen targets (Rosetta2)
- New Development and Improvement of RosettaAntibody Homology modeling (Rosetta3)

Objective of the Project:

- **Given the antibody sequence and the antigen, quickly improve the antibody (thermo stability or binding affinity)**
- **Potential usage in antibody-based biosensors or nano-devices**

**Experiment: Andy Ellington (UT Austin)
 George Georgiou (UT Austin)**

**Computation: Jeff Gray (JHU)
 Brian Kuhlman (UNC)**

- 1. Computer-aided antibody design**
- 2. Non-canonical amino acid**

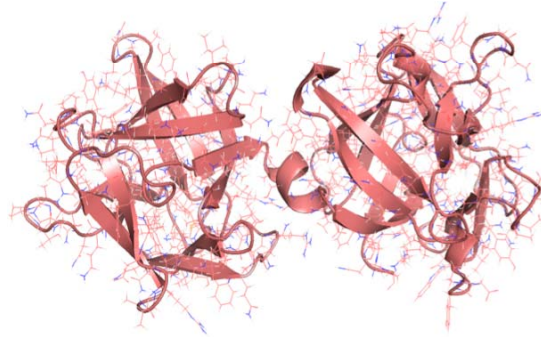
Blind Target 1: Anti-Bot56/HA33

Antibody:

Only Sequence

Antigen :

Crystal Structure
(PDB: 1YBI)



Clostridium botulinum haemagglutinin protein. The most prominent part of the protein complex that is associated with botulinum neurotoxins (35.86 kDa)

Blind Target 2: Anti-MS2/MS2

Antibody:

Only Sequence

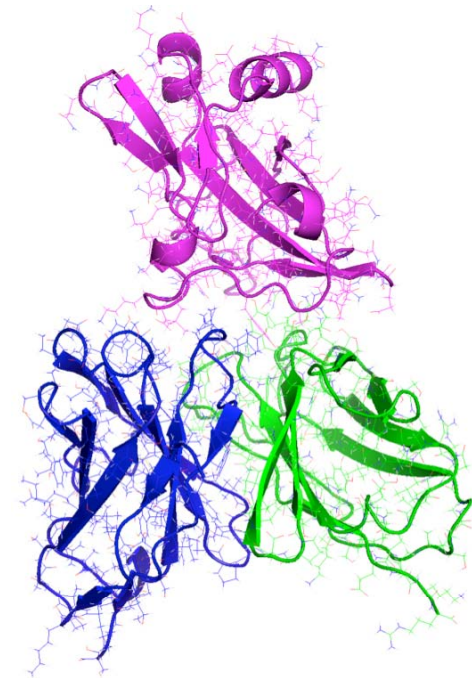
Antigen :

Crystal Structure
(PDB 1MSC)



MS2 phage coat protein, mutant version that leads to dimerization in solution (27.66 kDa)

Test Target: M18/PAD4 (PDB: 3ETB)



Antibody:

anti-anthrax (M18), variant of murine monoclonal antibody 14B7

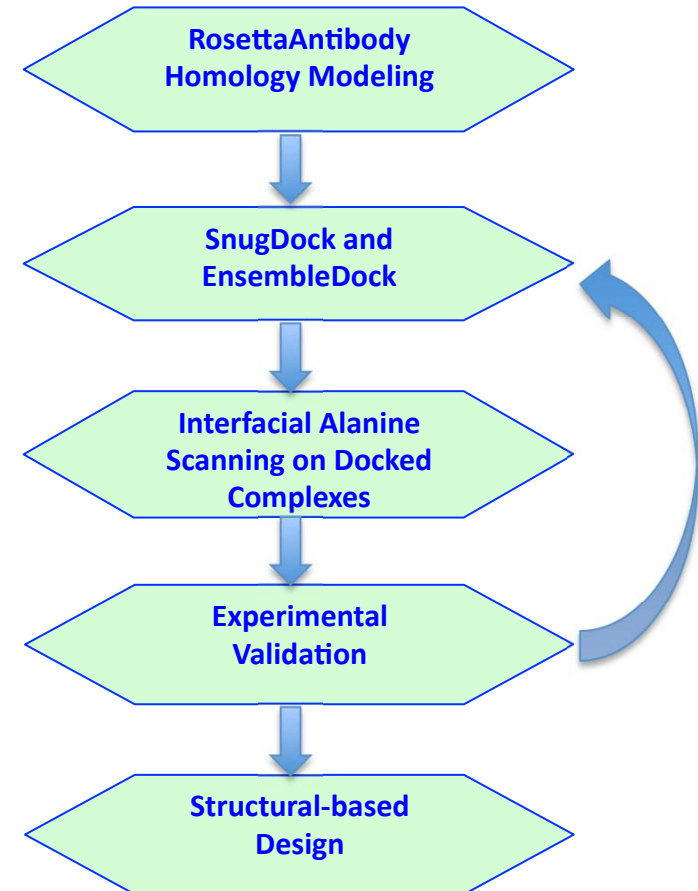
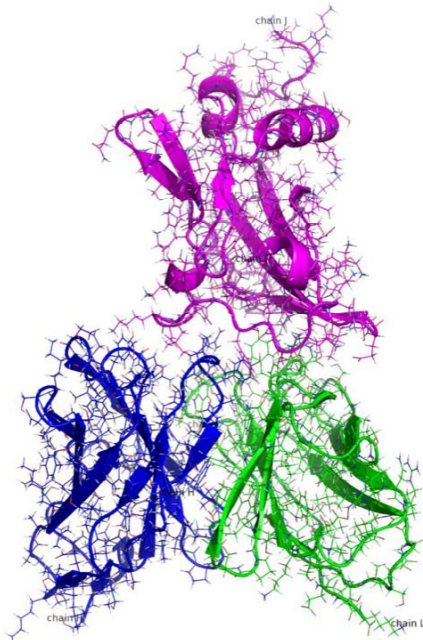
Antigen:

domain 4 of the **anthrax protective antigen**

Rosetta2 Computational Pipeline

Compare **Experimental** and **Computational Alanine Scanning** Results to Reduce Docking Sampling Space from Global to Local Search.

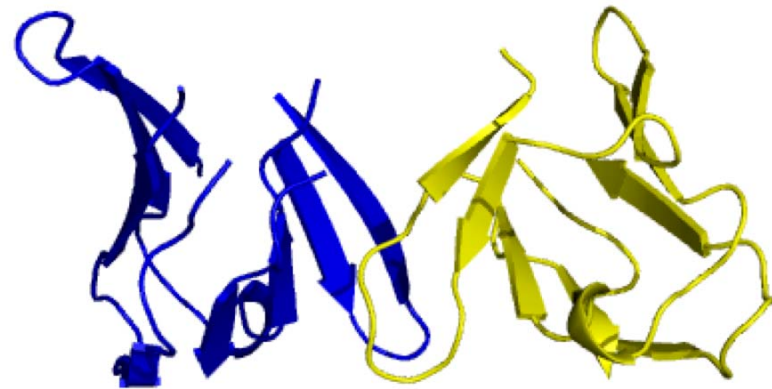
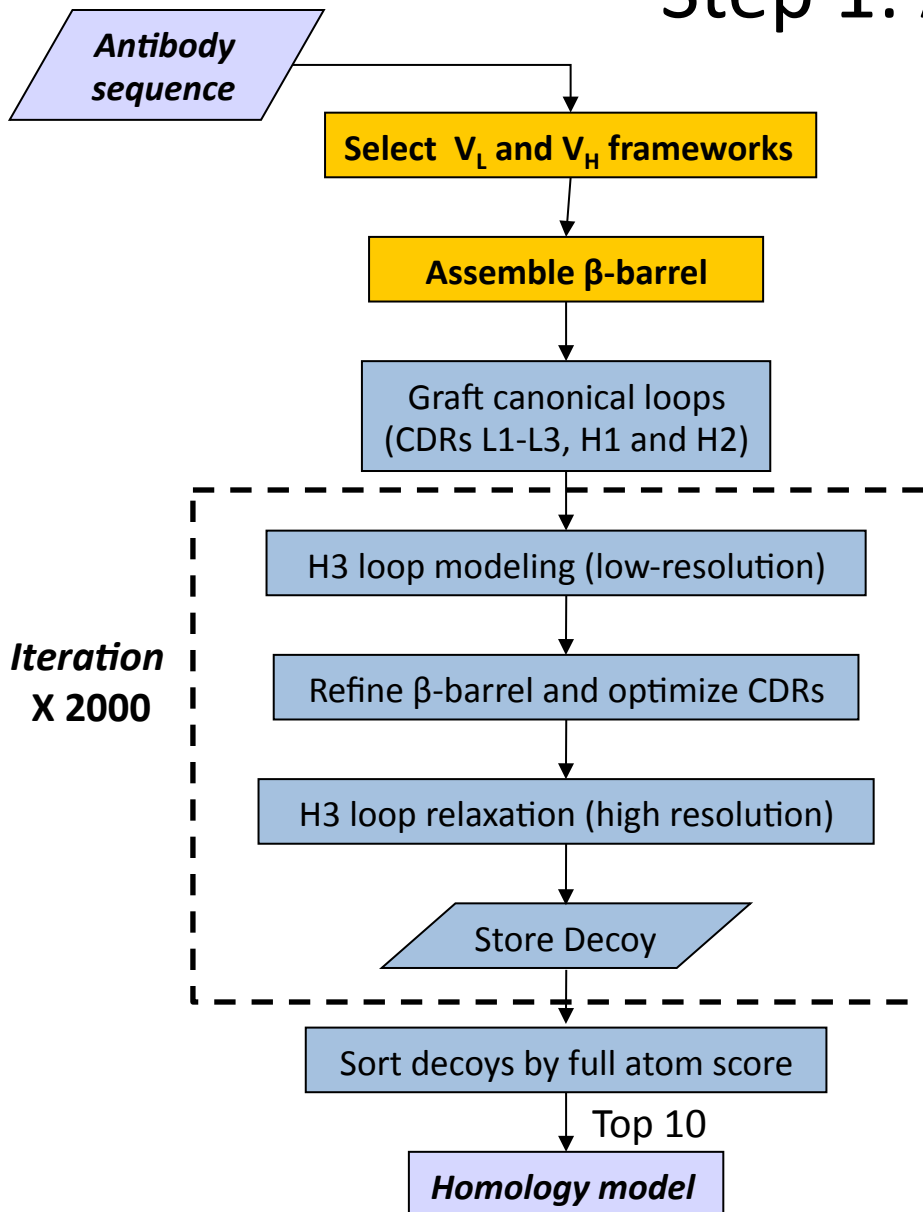
Test Target: M18/PAD4 (PDB: 3EBT)



1. A. Sivasubramanian *et. al.*, *Proteins* (2009)
2. Gray, J.J. *et. al.*, *J. Mol. Biol.* (2003).
3. S. Chaudhury, *et. al.* *J. Mol. Biol.* (2008)
4. A. Sircar *et. al.* *PLoS Comput. Biol.* (2010)
5. Kortemme, T. *et. al.* . Proc Natl Acad Sci U S A. (2002)
6. B. Kuhlman., *et. Al.*, Proc Natl Acad Sci U S A (2000)

RosettaAntibody

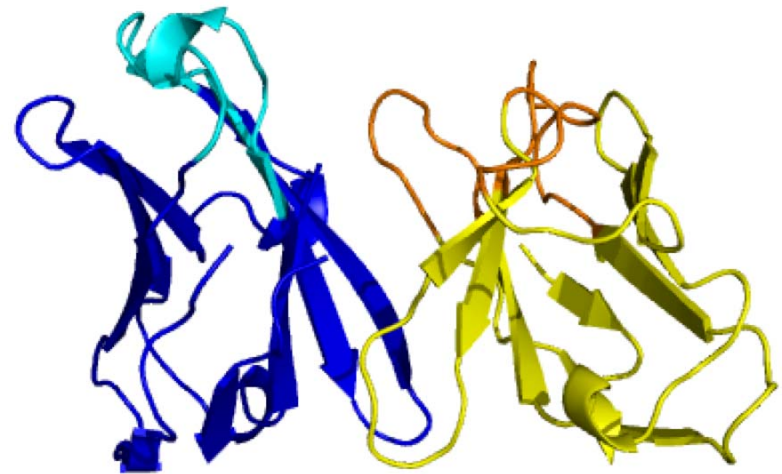
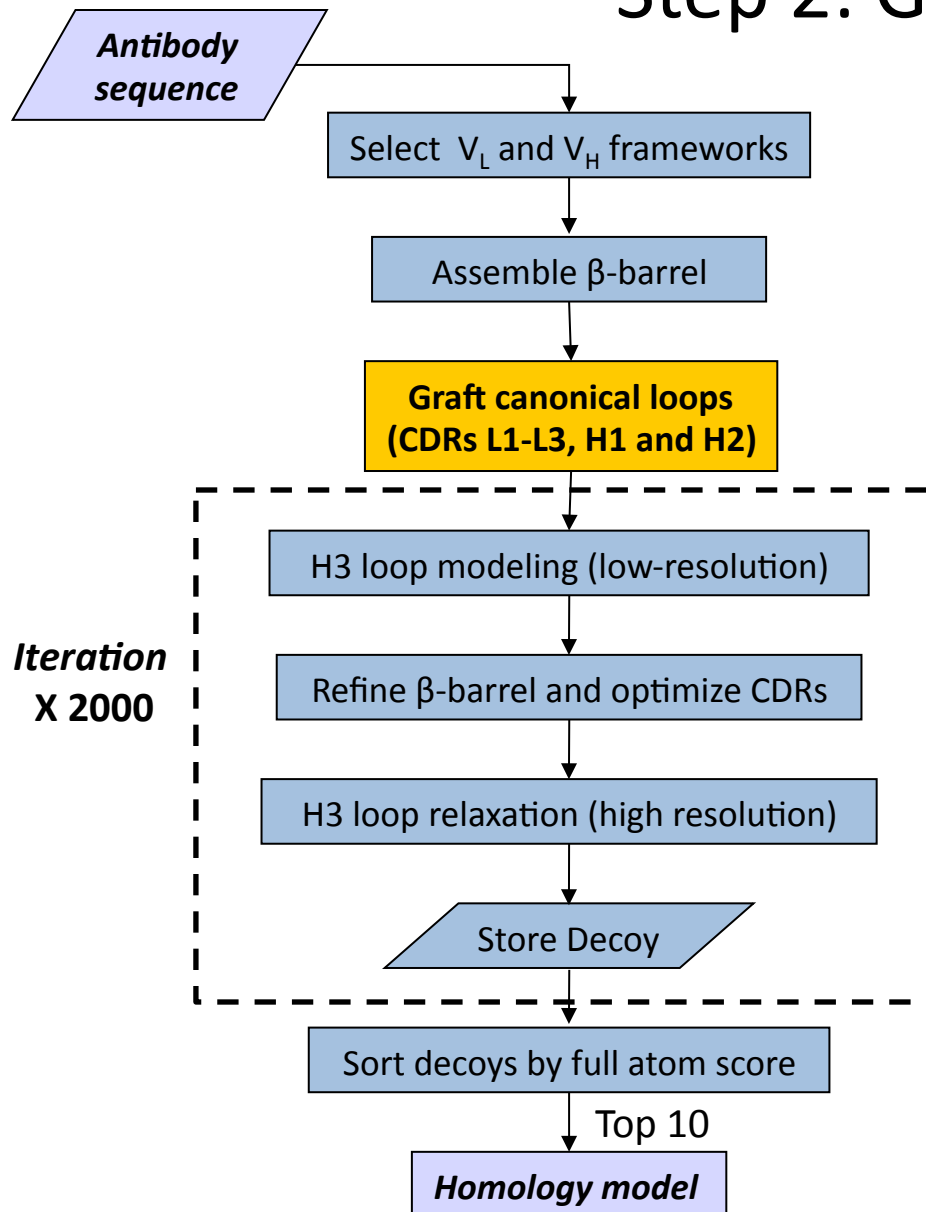
Step 1: Assemble β -Barrel



- Database of Ab frameworks
- Best frame chosen by BLAST bit score
- Assembled by superposition

RosettaAntibody

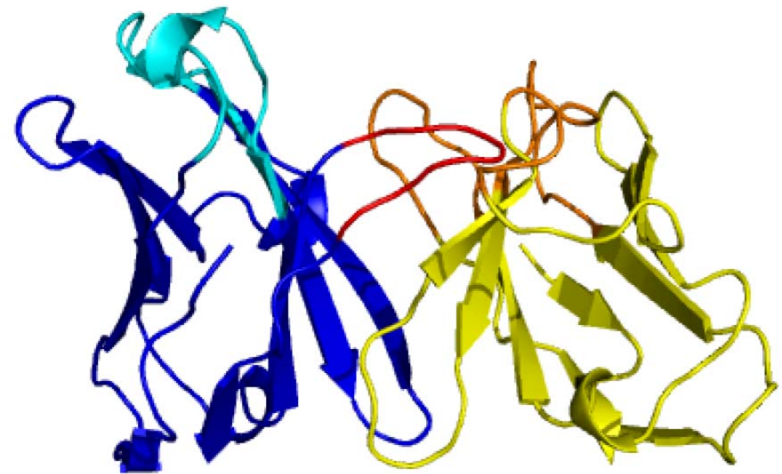
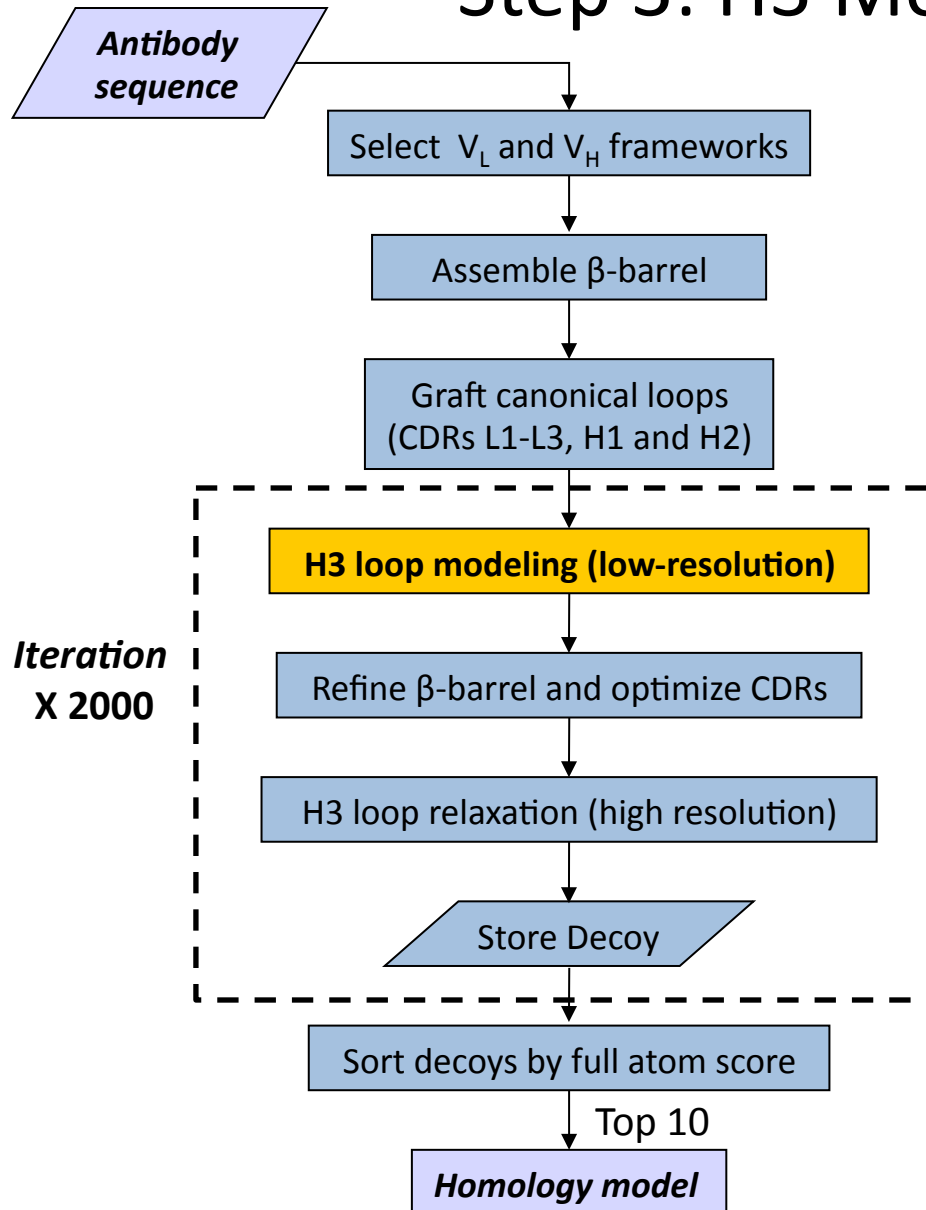
Step 2: Graft canonical loops



- Database of canonical loop structures
- Best loops chosen by BLAST bit score
- Grafted by superposing stem residues

RosettaAntibody

Step 3: H3 Modeling (Low Resolution)

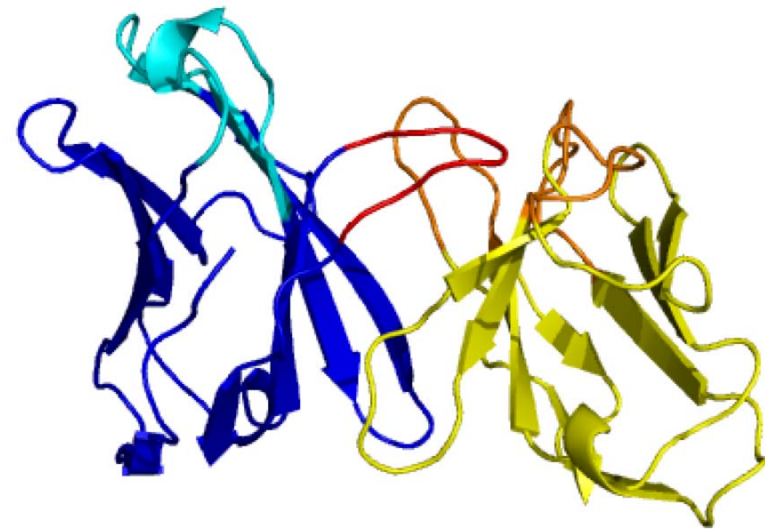
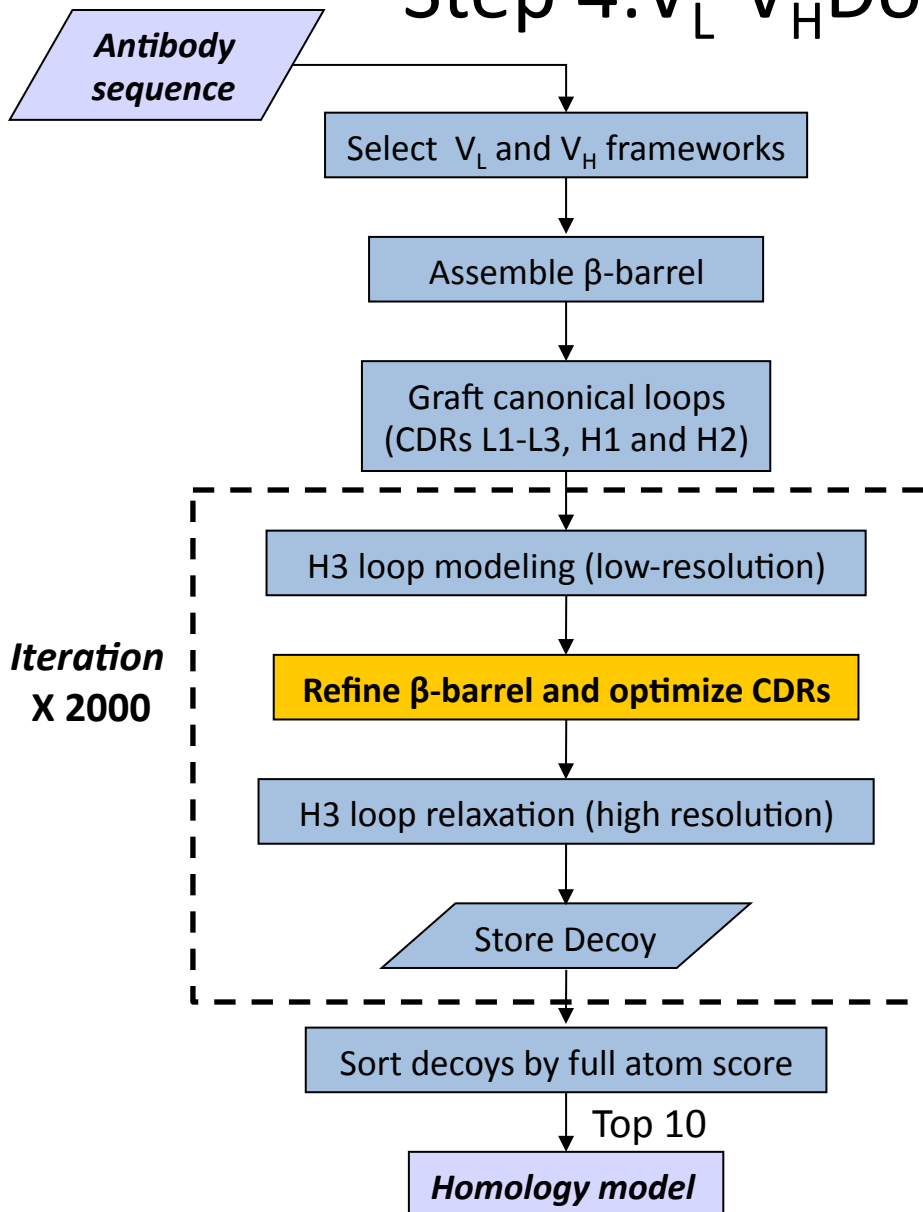


- Loop assembled with 3-residue fragments
- Kink/extended fragments used for base (Sharai *et al.*)
- Cyclic-coordinate descent (CCD) loop closure (Canutescu & Dunbrack)

Sivasubramanian, Sircar *et al.*, *Proteins* (2009)

RosettaAntibody

Step 4: V_L - V_H Docking & CDR Optimization

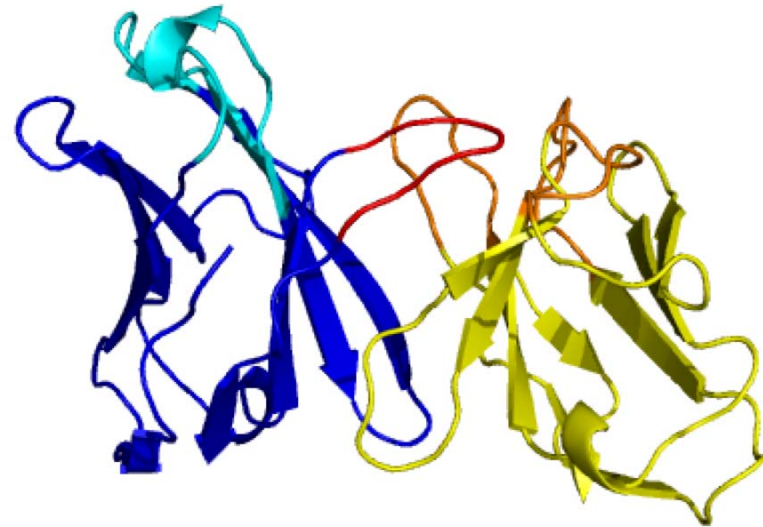
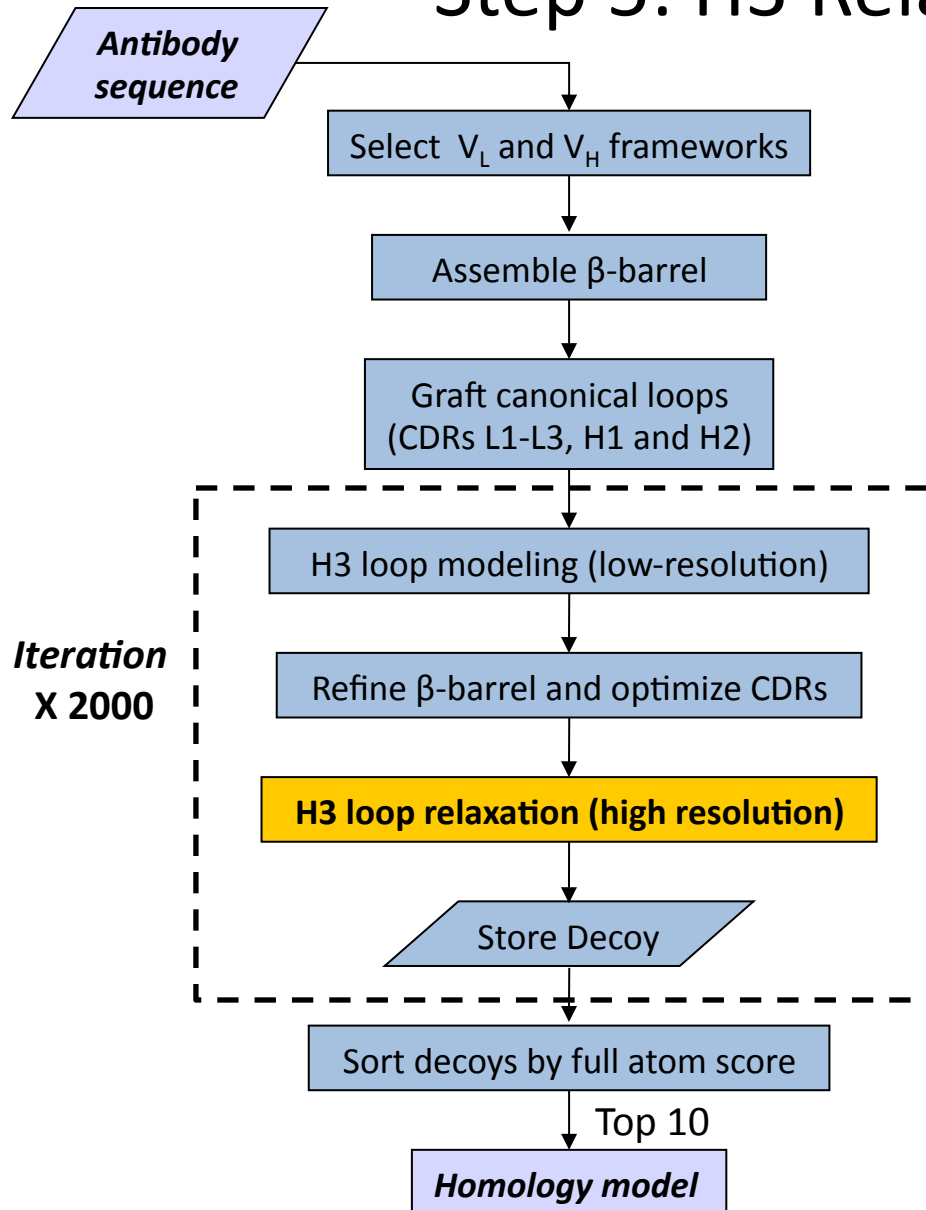


- Docking of V_L to V_H
- Gradient based minimization in rigid-body space and non-H3 loop torsion angles

Sivasubramanian, Sircar *et al.*, *Proteins* (2009)

RosettaAntibody

Step 5: H3 Relaxation (High Resolution)

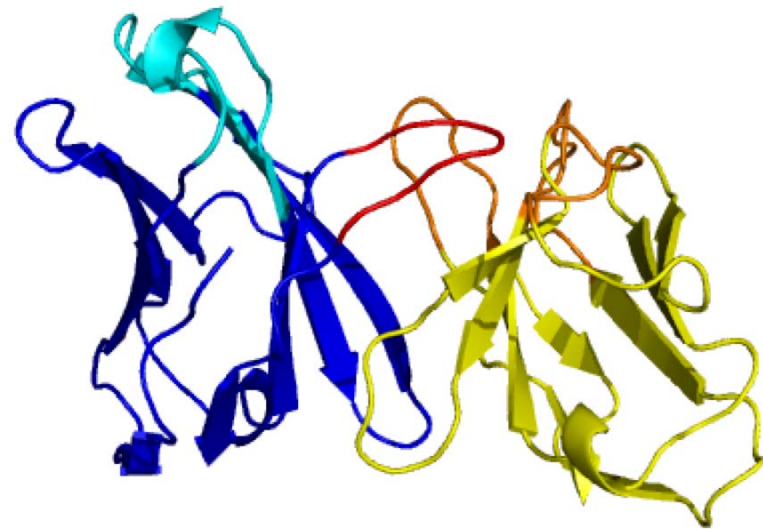
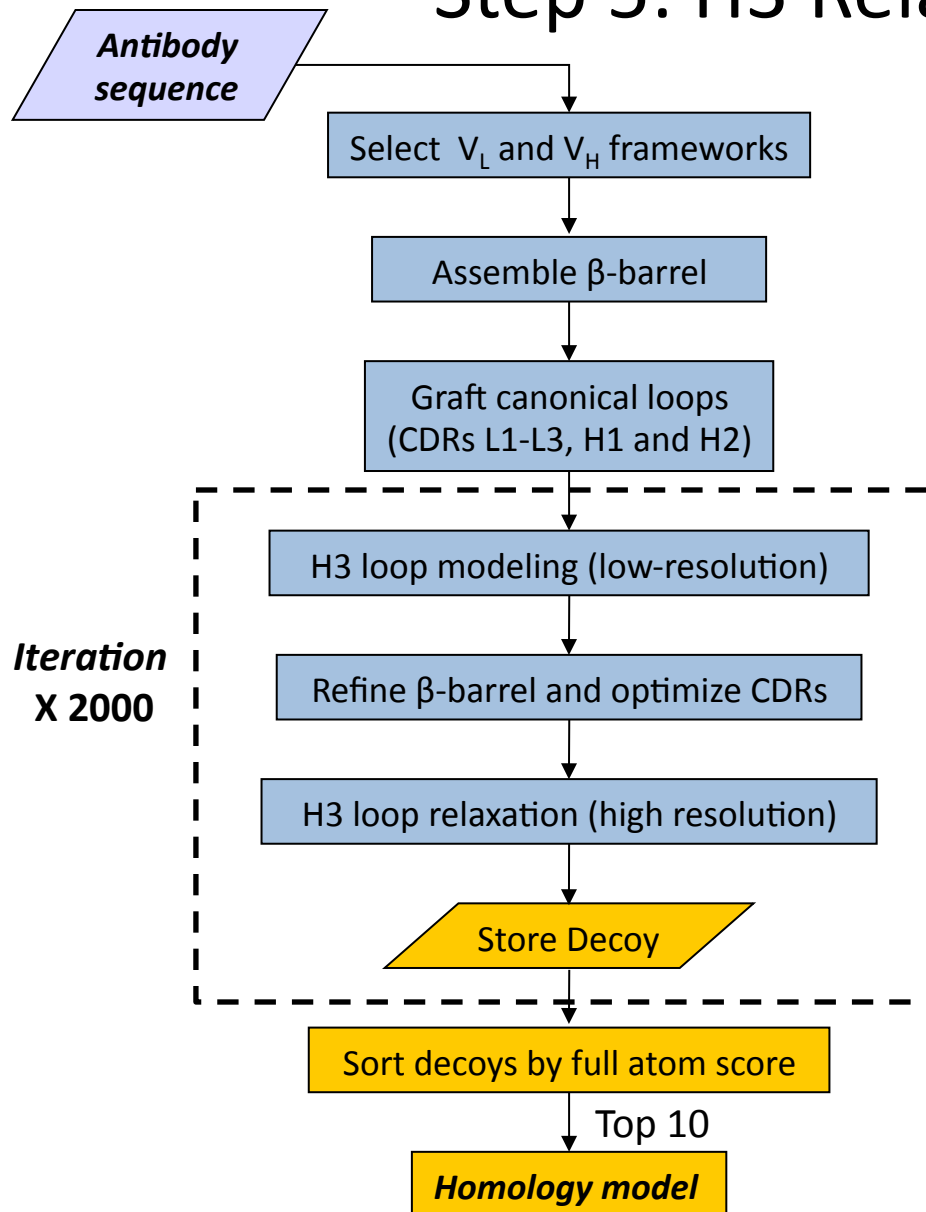


- Small perturbations of H3 (ϕ, ψ) angles
- Cyclic-coordinate descent loop closure
- Gradient-based minimization in H3 torsion angles

Sivasubramanian, Sircar *et al.*, *Proteins* (2009)

RosettaAntibody

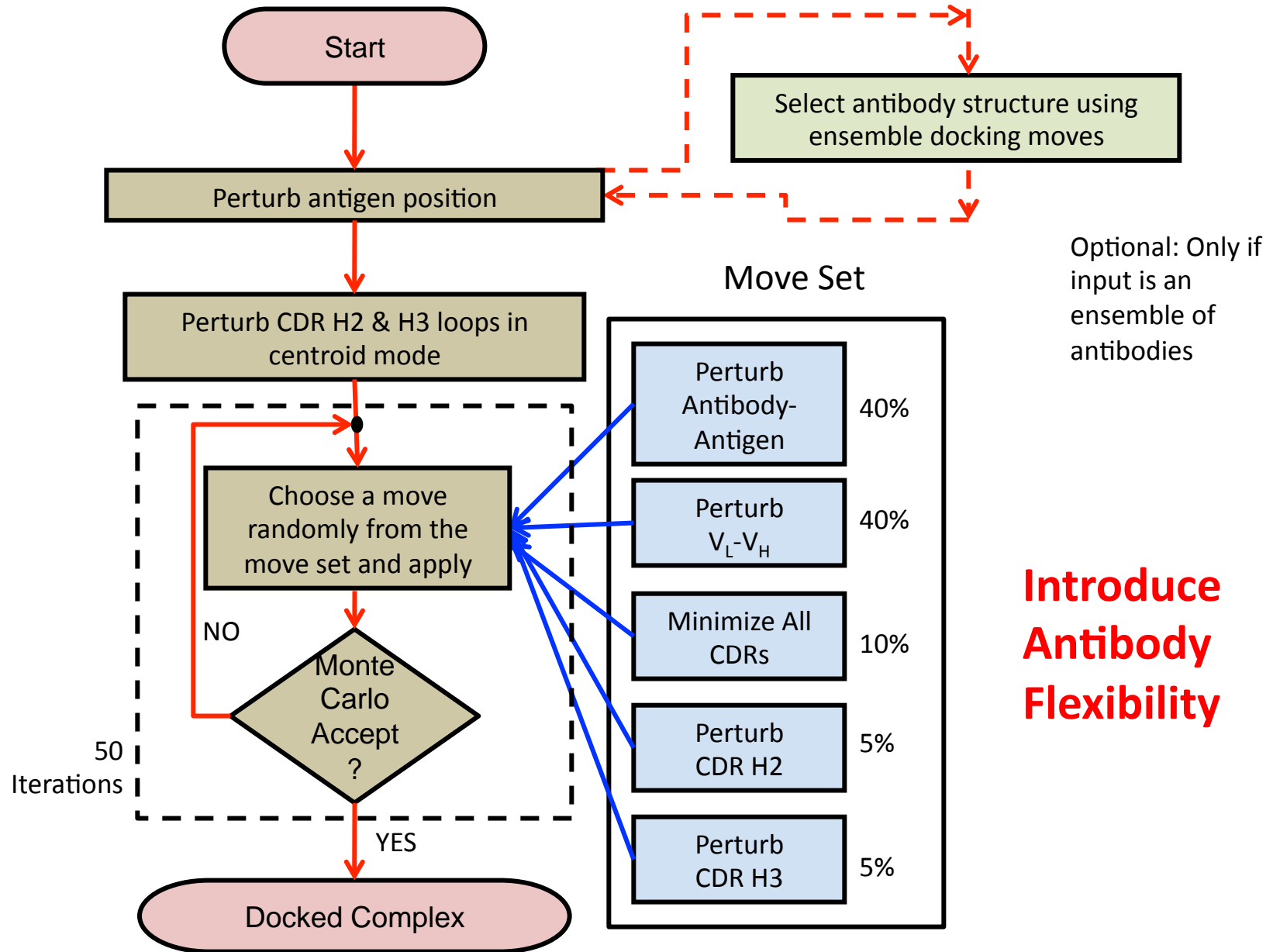
Step 5: H3 Relaxation (High Resolution)



- Small perturbations of H3 (ϕ, ψ) angles
- Cyclic-coordinate descent loop closure
- Gradient-based minimization in H3 torsion angles

Sivasubramanian, Sircar *et al.*, *Proteins* (2009)

The SnugDock + EnsembleDock Algorithm



Test Target: M18/PAD4 (PDB: 3ETB)

Top 10 (ranked by Rosetta scoring function) homology models:

11-res L1, 1A14 (91%)

7-res L2, 1B7Z (100%)

9-res L3, 1FGV (88%)

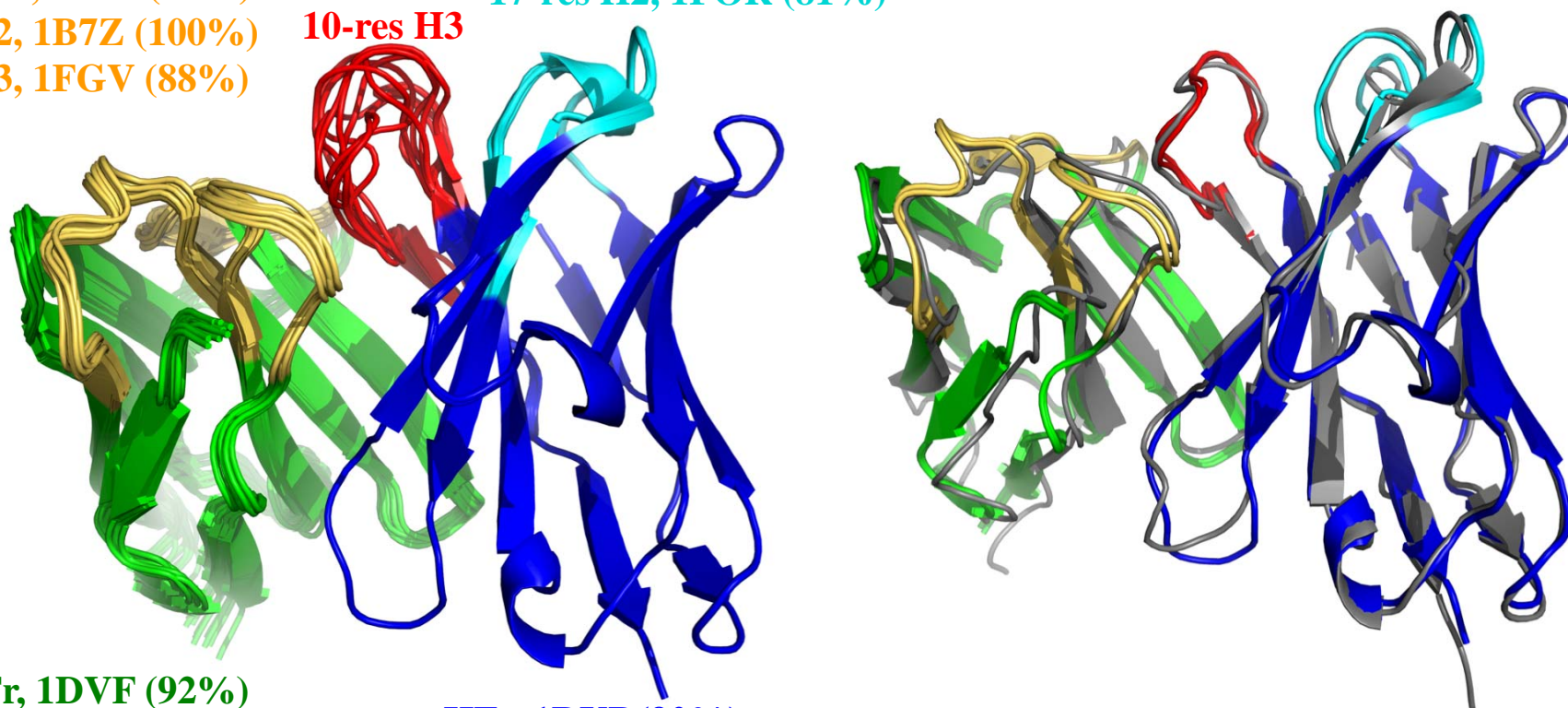
10-res H3

10-res H1, 1EMT (80%)

17-res H2, 1FOR (81%)

LFr, 1DVF (92%)

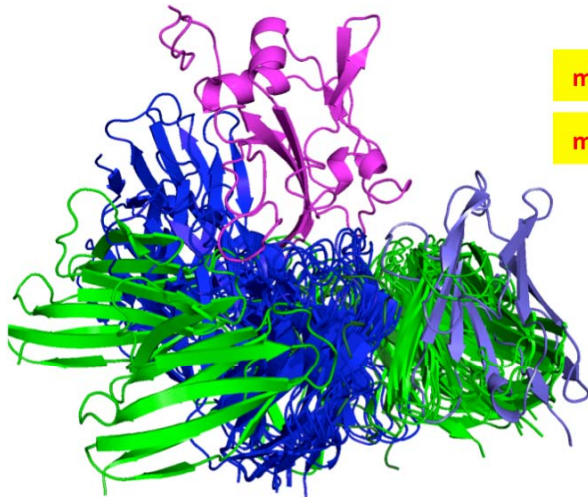
HFr, 1RUR(90%)



Test Target: M18/PAD4 Complex (PDB: 3ETB) (EnsembleDock+SnugDock)

Docking (**Global** Search)

Antigen Crystal Structure
+
Ensemble of Antibody
Homology Models



Top 10 complexes, interface score (out of **100 K** decoys)

Alanine Scanning
(computational ddG & experimental ELISA)

Docking Model	19326 ddG (kcal/mol)	ELISA	
		Absorbance	Fold Increase
		<i>WT</i>	1
		<i>background</i>	0.38
ILE 646	0.35		
ASP 648	-0.19	0.1	0.77
GLU 650	0.06	0.0985	0.76
LEU 652	3.2	0.0905	0.69
LYS 653	0.79		
match GLU 654	4.27	0.0605	0.46
VAL 655	0.05		
match ASP 658	2.29	0.081	0.62
ARG 659	1.9		
ASN 682	0.18	0.058	0.44
ASP 683	-0.04	0.1035	0.79
LYS 684		0.067	0.51
LEU 685	0.25	0.0645	0.49
TYR 688	1.19	0.083	0.64
ILE 689	0.03		
SER 690	0.07	0.171	1.31
ASN 691	0.05		
TYR 694	0.62		
ASN 713	0.02		
ASP 715	-0.17		
THR 716	0.37		
SER 717	1.6	0.0985	0.76
THR 718	0.51		
ASN 719	0.01	0.093	0.71

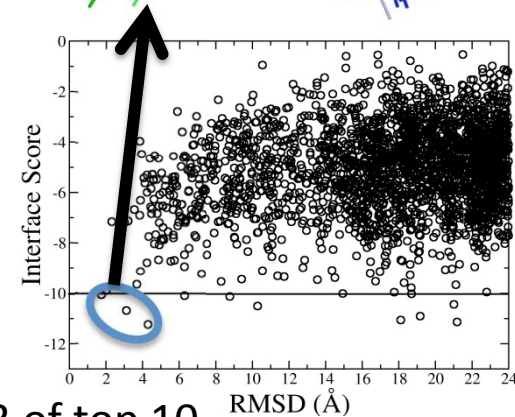
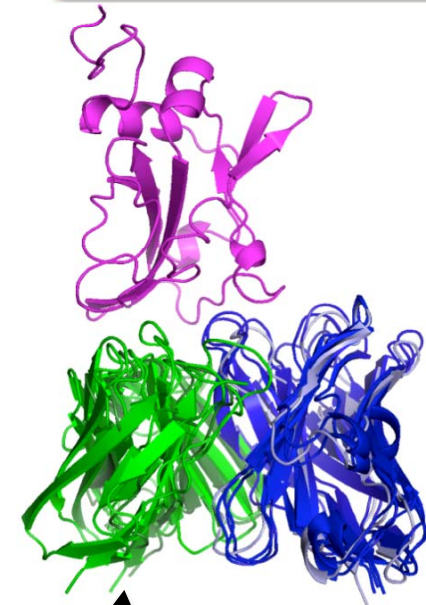
Define Hot Spots:

ddG > 1kcal/mol

Absorbance < 0.083

(Fold Increase < 0.64)

Docking (**Local** Search)

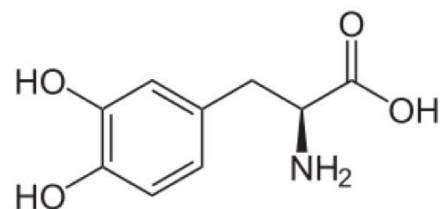


3 of top 10

Unnatural Amino Acid Design

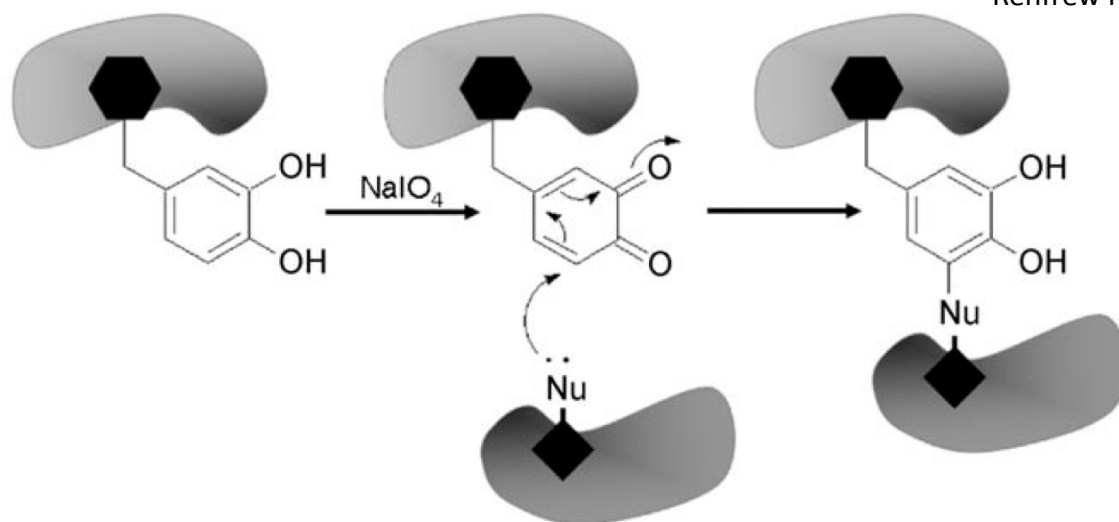
(with people from Ellington, Georgiou, Kuhlman's labs)

L-3,4-dihydroxyphenylalanine (L-DOPA)



Use the predicted M18/PAD4 complex structure to predict the best location on antibody to mutate to L-DOPA to crosslink the antigen.

Renfrew PD, Choi EJ, Bonneau R, Kuhlman B, PLoS One (2012)

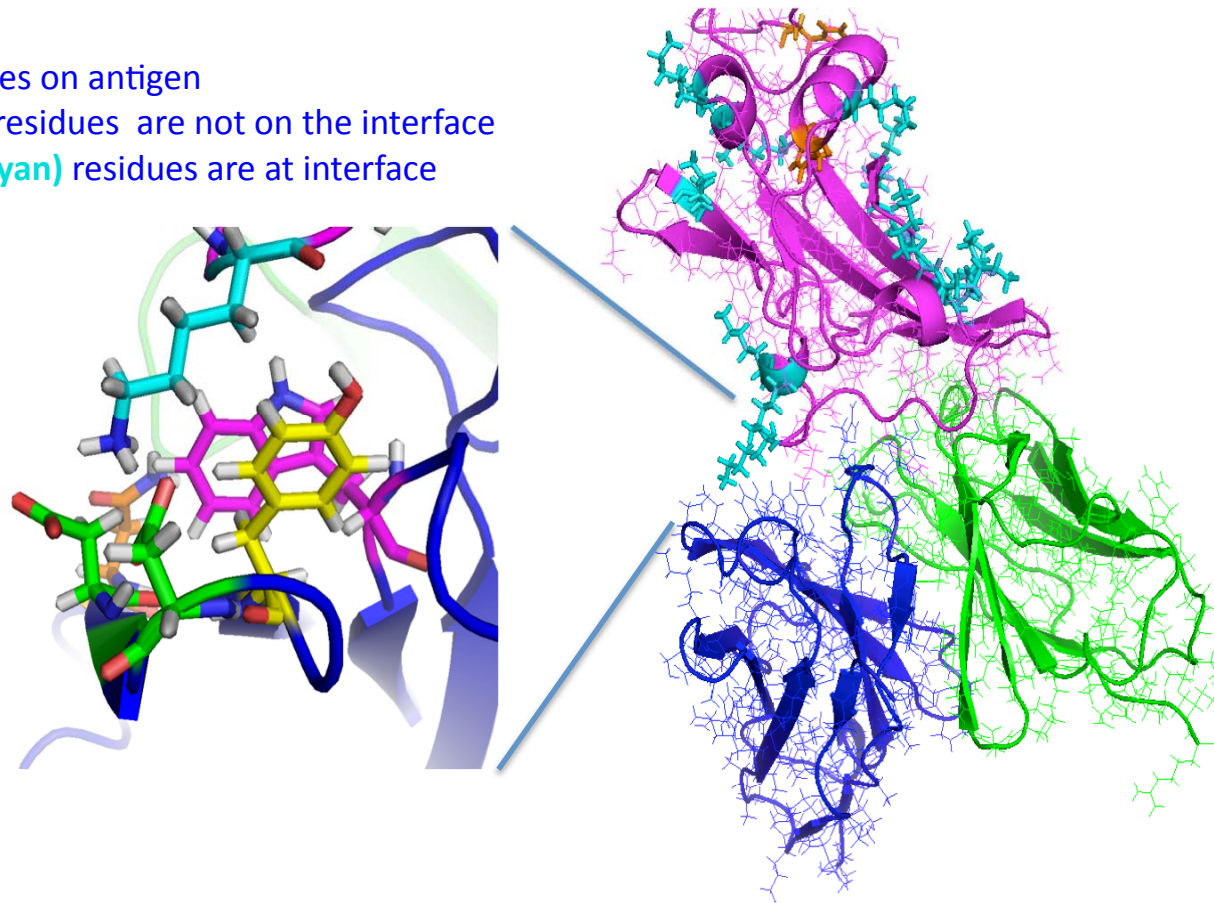


Test Crosslinking Prediction in Rosetta

L-DOPA reacts with CYS, HIS, or LYS residues
on antigen

1. No CYS residues on antigen
2. **HIS (Orange)** residues are not on the interface
3. Several **LYS (Cyan)** residues are at interface

Test Target: M18/PAD4 (PDB: 3EBT)



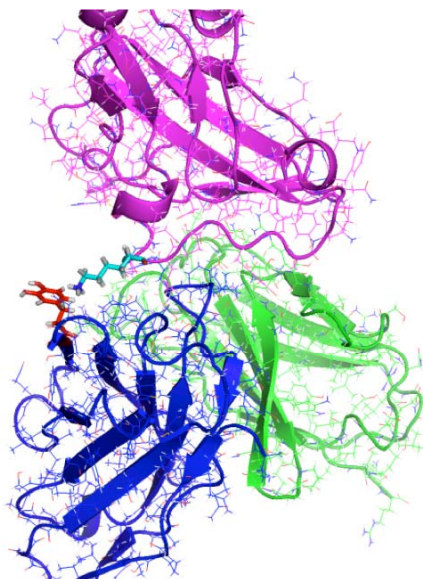
Choose a residue on Antibody to mutate to L-DOPA:

- a. **Accessible Distance** between Lys and L-DOPA
- b. **Exposed to the Solvent** (IO_4^- accessible)
- c. **Interface Compatible**

Rosetta L-DOPA Crosslinking Prediction Test on M18/PAD14 Complex

Computational Prediction:

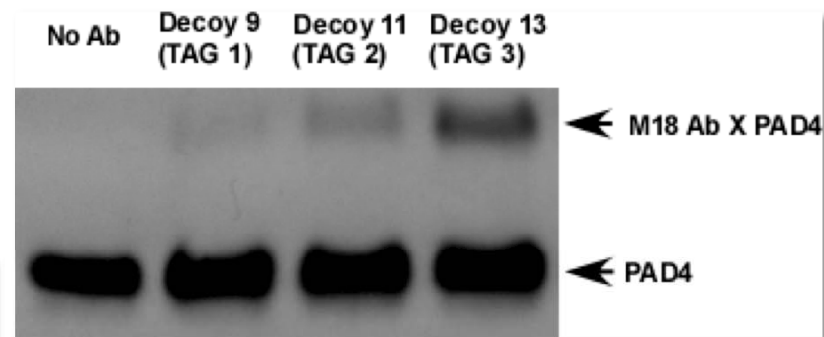
Decoy 13



Ranking	Decoy ID	Antigen	Antibody	a). Distance	b). nbr_atoms (<3 Å)	c). Score (with constraint)
1	13	J_684_LYS	H_56_ASP	3.58	13	6.90
2	11	J_684_LYS	H_54_ASP	2.85	20	53.91
3	9	J_684_LYS	H_52_TYR	3.01	35	17.18
4	7	J_684_LYS	H_33_TRP	3.60	40	53.57
5	10	J_684_LYS	H_53_GLY	4.08	22	13.59
6	15	J_684_LYS	H_58_ASN	5.12	21	6.40
7	3	J_679_LYS	H_31_SER	5.74	15	2.29
8	4	J_679_LYS	H_53_GLY	6.09	23	2.28

Experimental Validation (Western Blot):

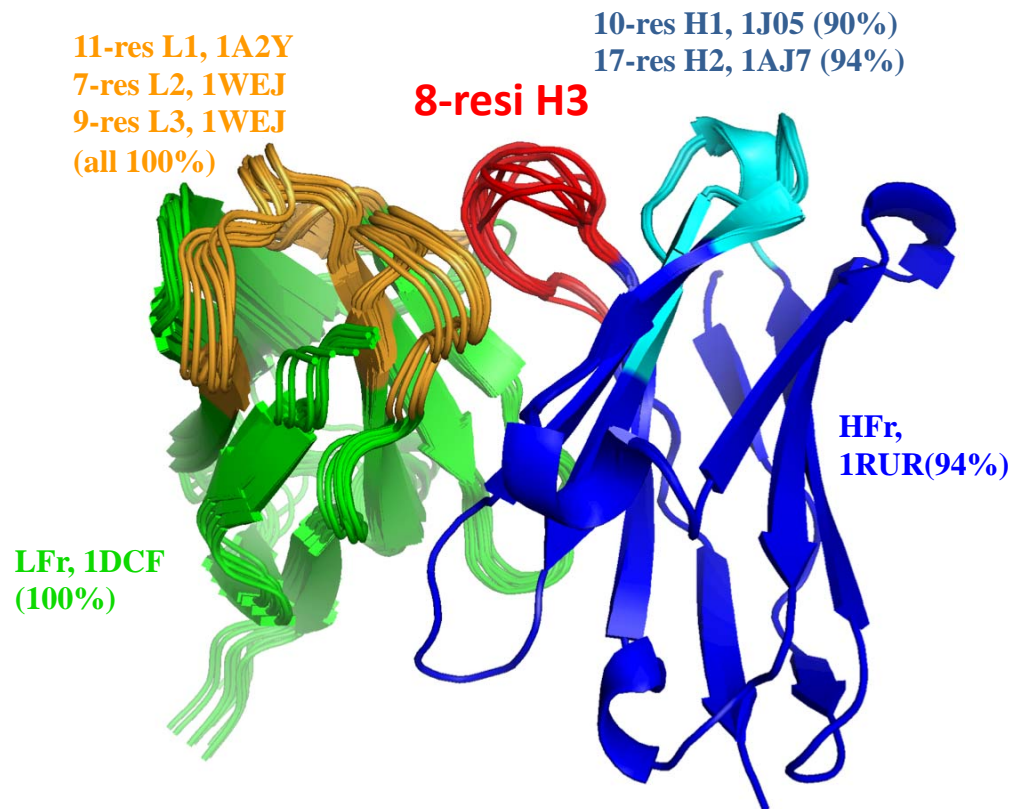
NaIO_4 was added to a mixture of antibody-antigen to oxidize the L-DOPA.



The validation of top 3 predictions:

Crosslinking efficiency increases in cases having lower number of neighboring side chains (criterion b) near L-DOPA

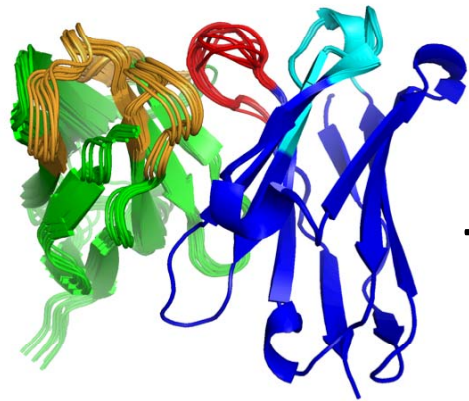
(Blind Target 1) Anti-Bot56/HA33 (in progress)



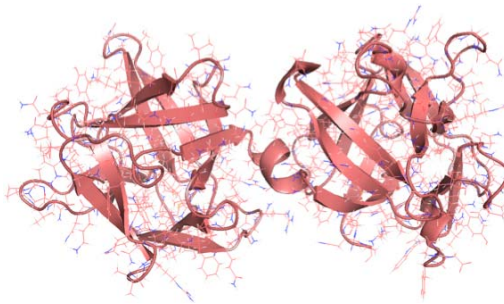
- Find best **templates** for non-H3 CDRs and light/heavy chain frameworks.
- **Graft** non-H3 CDRs templates on framework templates.
- **De novo loop modeling** of **8-residue** CDR H3 loop. (7-9 residue H3 was benchmarked to get median global RMSD 1.9 Å)
- **Docking** of L-H chains in high-resolution.

Docking (EnsembleDock + SnugDock) With *Global Search*

Ensemble of Anti-Bot56 Homology Models

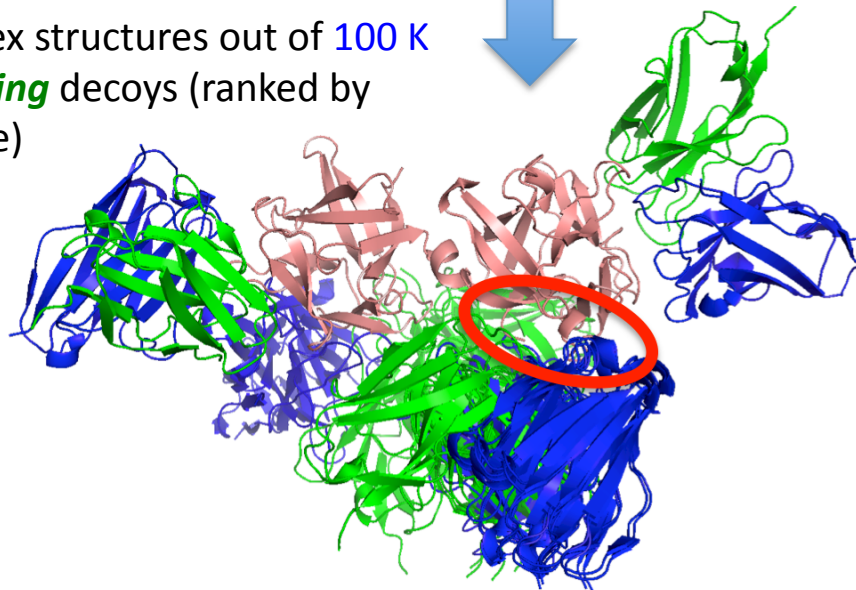


+



Clostridium Botulinum
haemagglutinin (HA33)
(1YBI)

Top 10 complex structures out of 100 K
global-searching decoys (ranked by
interface score)

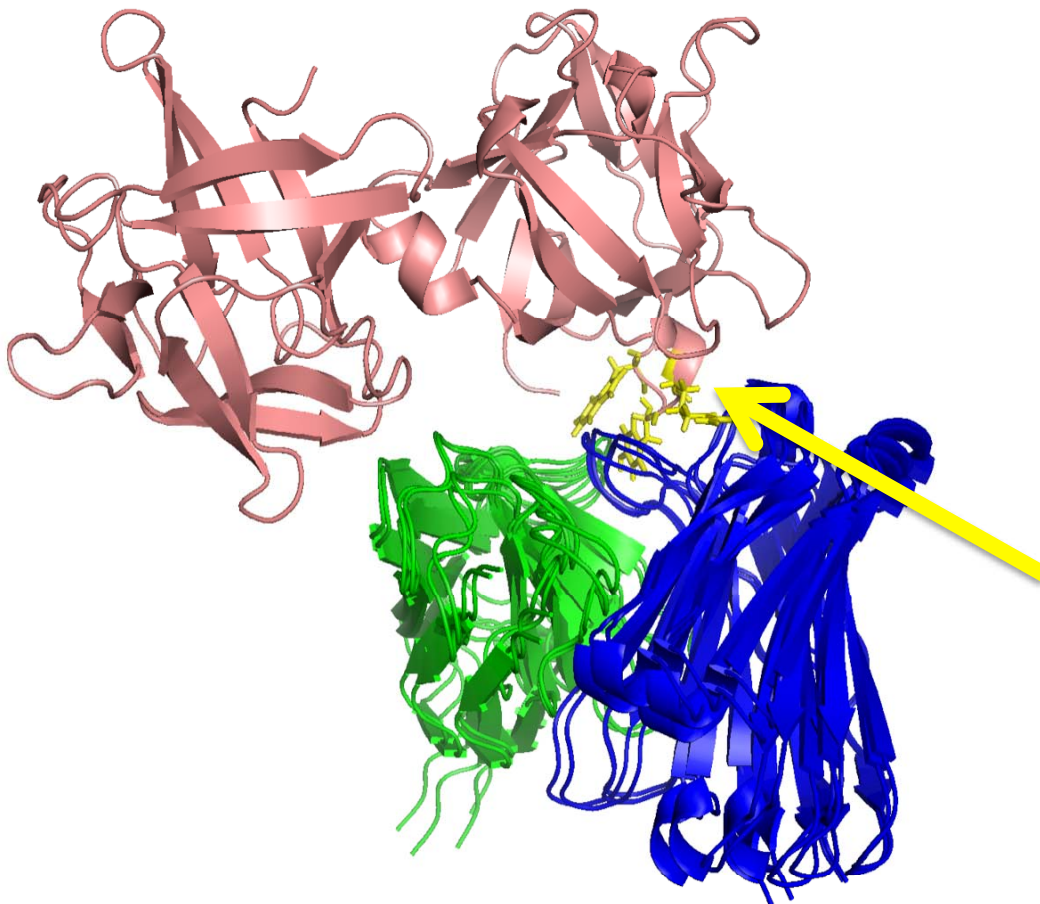


Potential Epitopes:

4/10 structures have their
binding regions converged
to the C-terminal of HA33

Computational Interface Alanine Scanning (Docked Anti-Bot56/HA33 Complex from *Global Search*)

- Experimental validation (*even just one residue on the antigen*) can localize the epitope position.
- Docking with *local search* can significantly decrease the sampling space and increase the prediction success rate.

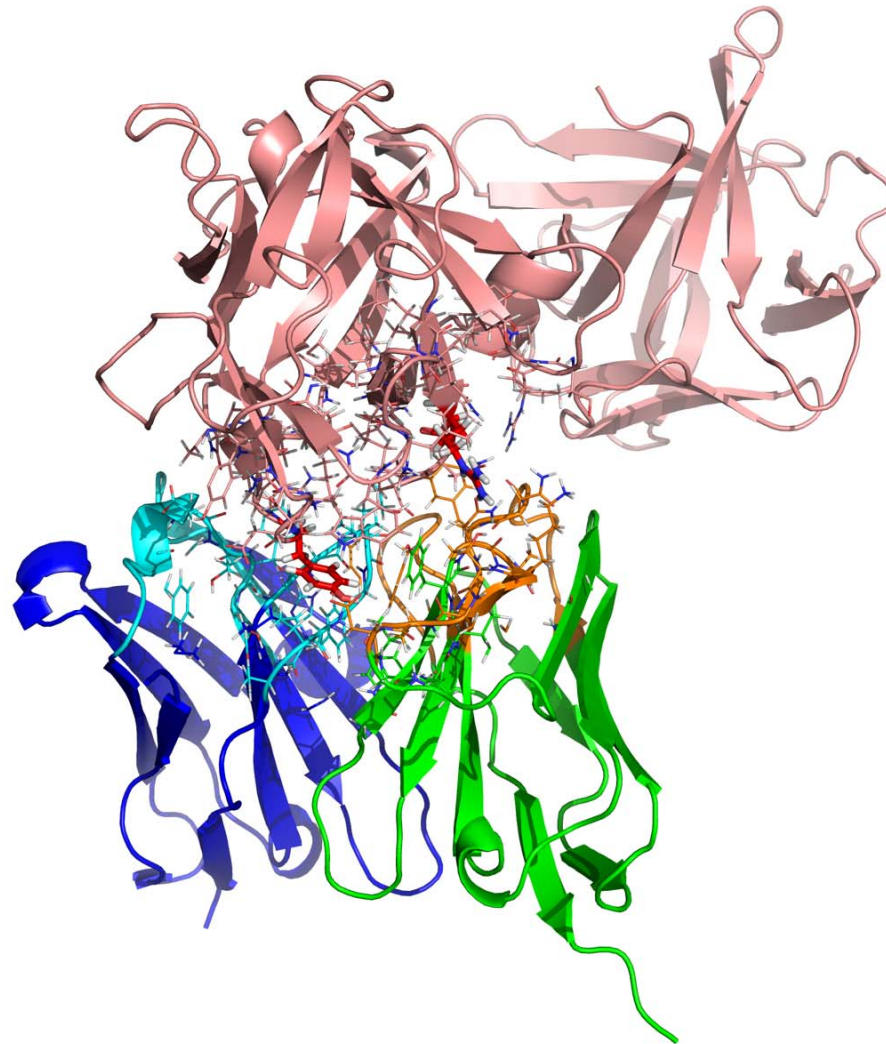


Potential Antigen Epitope Residues: Computational ddG > 1 kcal/mol

PDB #	Residue	Model 4553	Model 15013	Model 56965	Model 27190	Model 29830	Model 62105	Model 23819	Model 16700	Model 55862	Model 44238
30	GLN								1.13		
31	VAL									1.35	
34	ASN									1.14	
39	GLN	1.43				1.02					
41	THR	1.32					1.29				
42	ARG	1.54					1				
44	TYR				1.01				1.81		
66	ILE								2.11	1.39	
67	HIS								3.19	1.66	
68	ASN								2.44	1.21	
74	THR									1.24	
76	ASN									3.04	
80	HIS										2.08
83	SER									1.36	
85	GLN								1.05	1.44	
121	ASP										2.35
128	LYS										1.51
131	THR										1.94
133	ASN										1.2
152	ASN					1.45					
154	THR					1.94					
163	LEU				1.38						
182	TRP	1.36	1.87	1.47	4.05	1.25					
174	ASN									1.87	
184	TYR	3.28	3.52	3.55	1.44	2.98	3.61				
186	ARG	1.3	8.13	3.53			2.79				
189	LYS			1.03							
191	THR					2.13					
221	ASN								1.45		
224	ARG								1.18		
265	TYR								3.13		
276	GLN								1.74		
278	PHE								2.64		
289	ASN				1.21						
291	ARG	1.26				1.01					
292	ASN	1.59									

Experimental Validation on Top 1 structure

		Docking Model	Experiment ELISA	sd
		Kcal/mol	0.297 (wild type)	0.037
GLN	39	1.43	0.260	0.041
THR	41	1.32	0.268	0.002
ARG	42	1.54	0.256	0.001
PHE	153	0.57	0.306	0.018
VAL	172	0.59	0.325	0.012
TRP	182	1.36	0.251	0.038
TYR	184	3.28	0.186	0.006
ARG	186	1.3	0.326	0.004
ASN	187	0.08	0.205	0.024
GLN	188	0.02	0.133	0.016
LYS	189	0.29	0.210	0.007
ILE	207	0.67	0.253	0.043
ARG	291	1.26	0.240	0.008
ASN	292	1.59	0.274	0.013

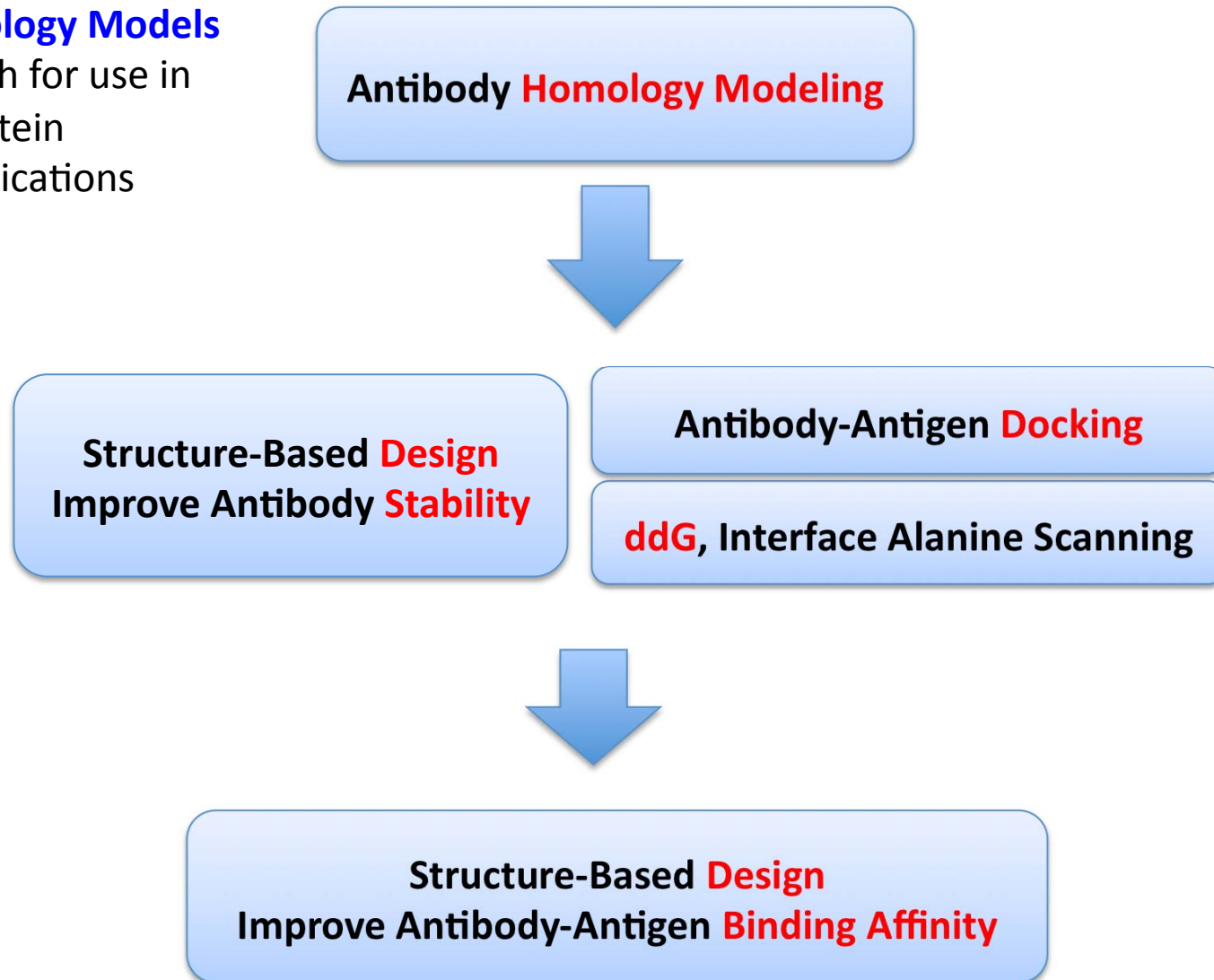


Doing Local Docking for better match for alanine scanning

Pipeline Challenges

Quality of Homology Models

must be very high for use in downstream protein engineering applications

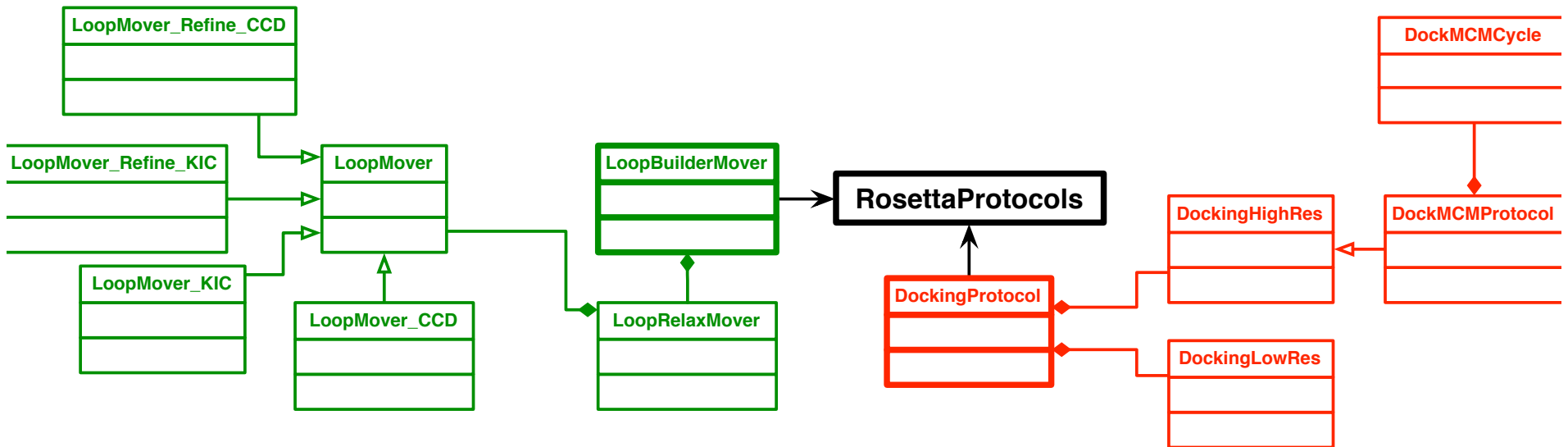


Improving RosettaAntibody

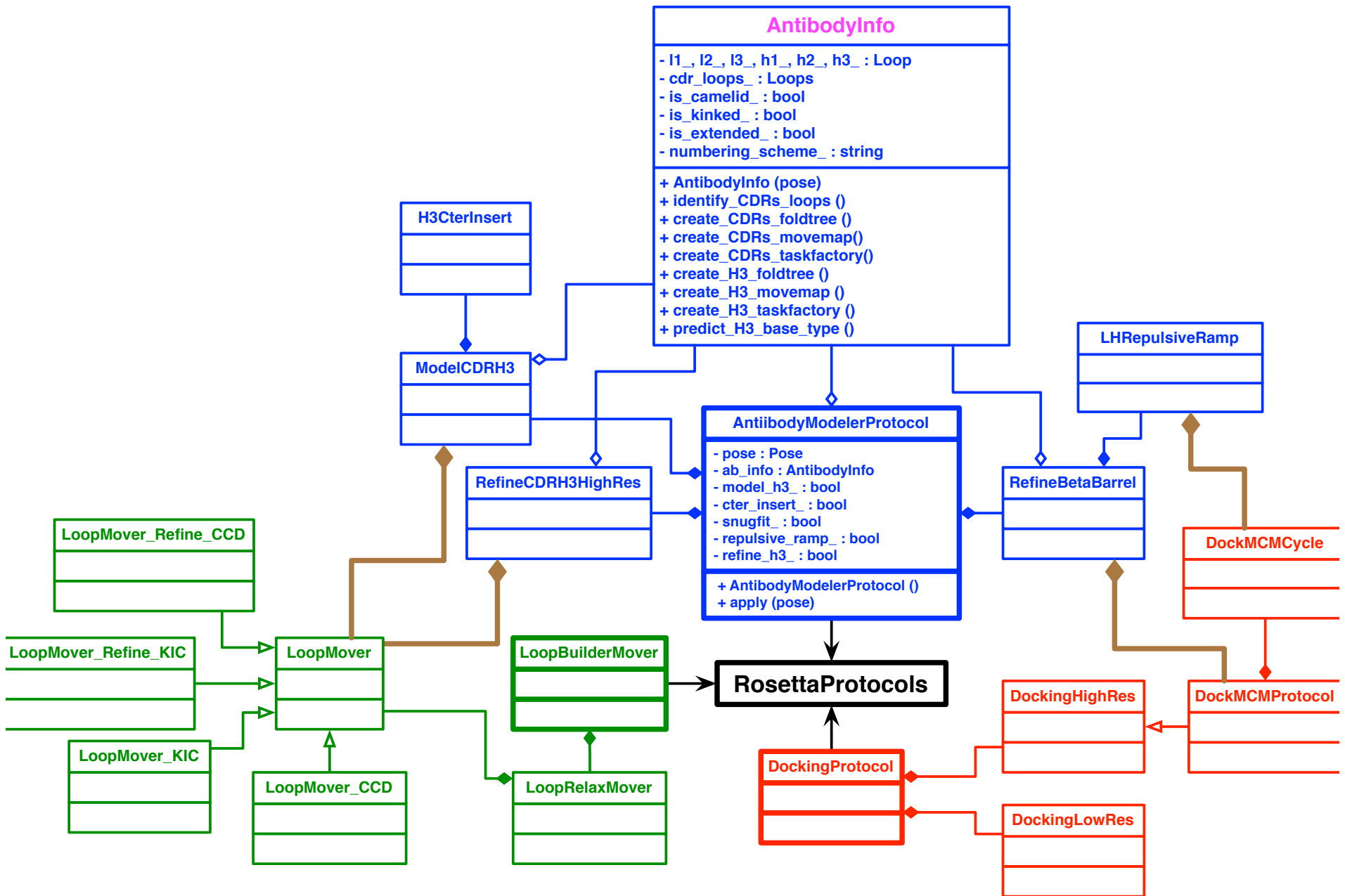
- **Rosetta3 Scoring Function and Sampling Strategy**
- Antibody Database
- Template Selection (CDR, framework and V_L - V_H) and Loop Grafting
- **V_L - V_H Docking (Docking Protocol in Rosetta3)**
- **H3 Loop Modeling (loop modeling protocols in Rosetta3)**
 - Rosetta fragments with cyclic coordinate descent (CCD) loop closure
Canutescu, A. A.; Dunbrack, R. L.. Protein Science (2003)
 - **improved loop modeling method: kinematic closure (KIC)**
Already in Rosetta3 Mandell DJ, Coutsias EA, Kortemme T. (2009). Nat. Meth.
 - **improved rule of H3 base prediction**
Kuroda D, Shirai H, Kobori M. Nakamura (2009) Proteins

Object-Oriented Code !

New Object-Oriented Code



New Object-Oriented Code

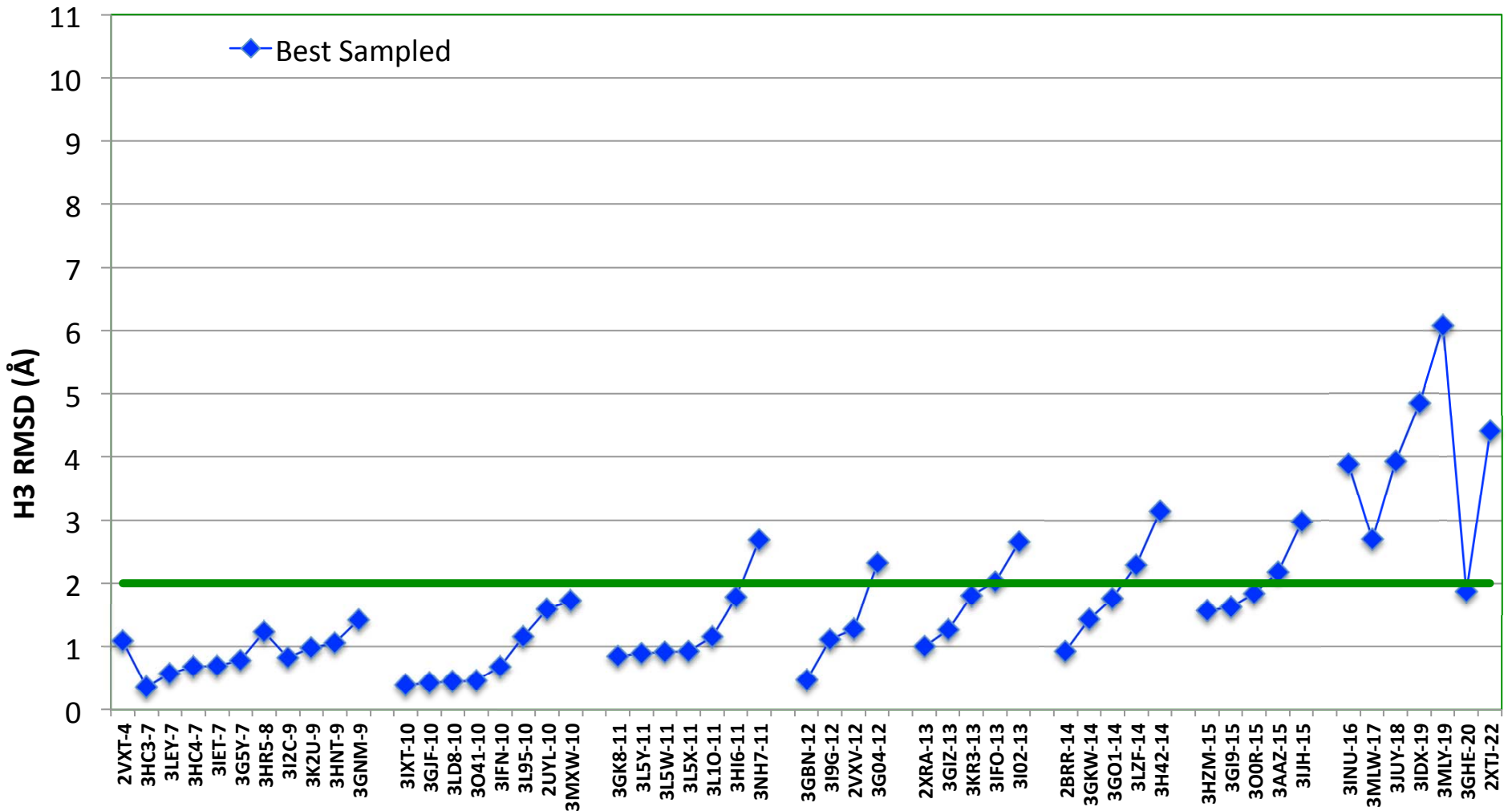


H3 Recovery on Crystal Structure Framework

Performance of KIC in Rosetta Loop Mover Protocol

Sampling: What's the best RMSD KIC can sample? (benchmark 52 antibodies)

Scoring: Can Rosetta scoring function recognize the best sampled structure?

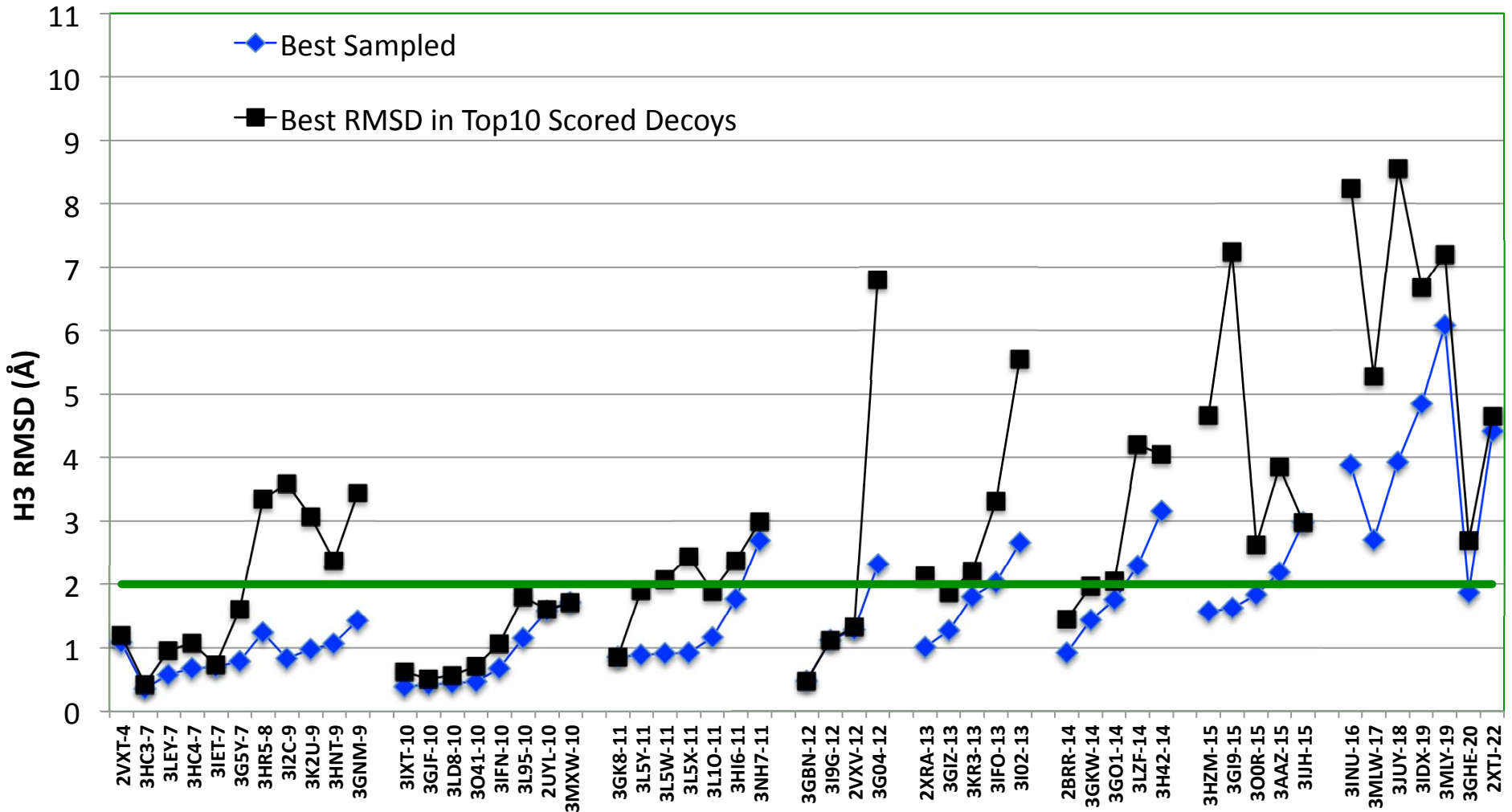


H3 Recovery on Crystal Structure Framework

Performance of KIC in Rosetta Loop Mover Protocol

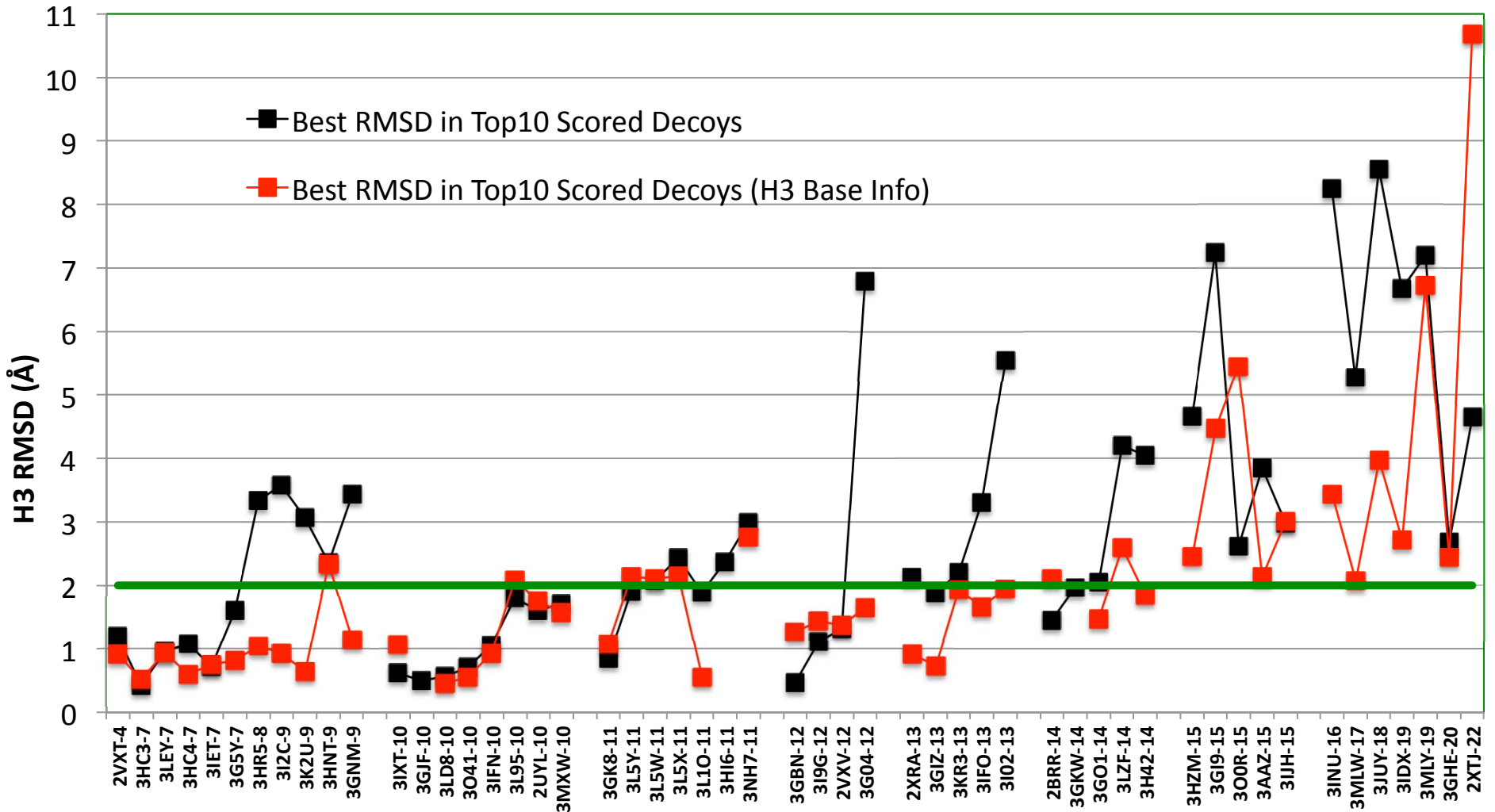
Sampling: What's the best RMSD KIC can sample? (benchmark 52 antibodies)

Scoring: Can Rosetta scoring function recognize the best sampled structure?



H3 Recovery on Crystal Structure Framework

Performance of **New RosettaAntibody** (with **H3 Kink Information**)

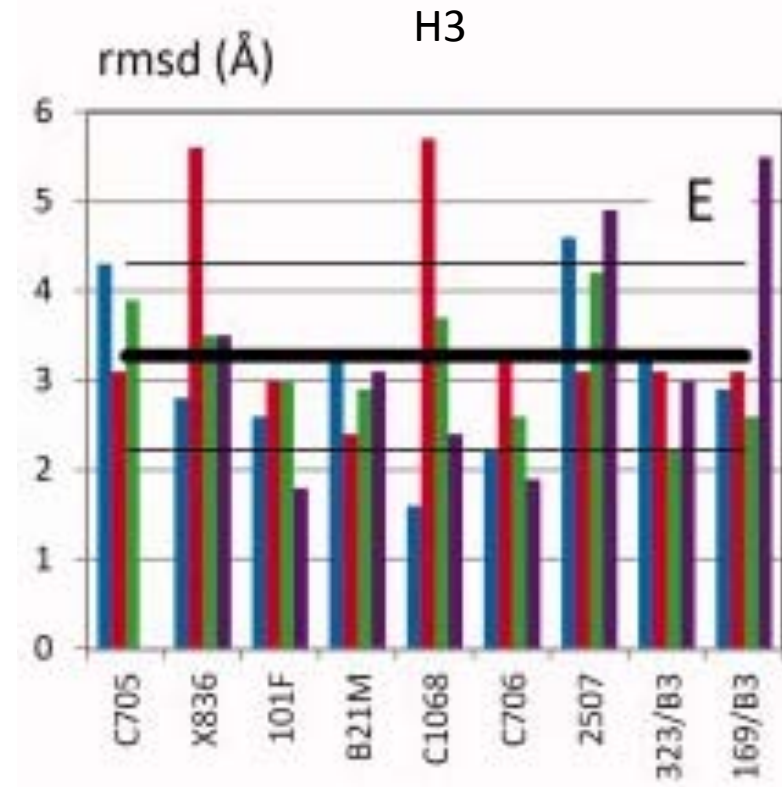


Build H3 on Homology Models

1. Run the **whole protocol of new RosettaAntibody code**, build a complete homology model
2. Compare the performance to other antibody homology modeling methods

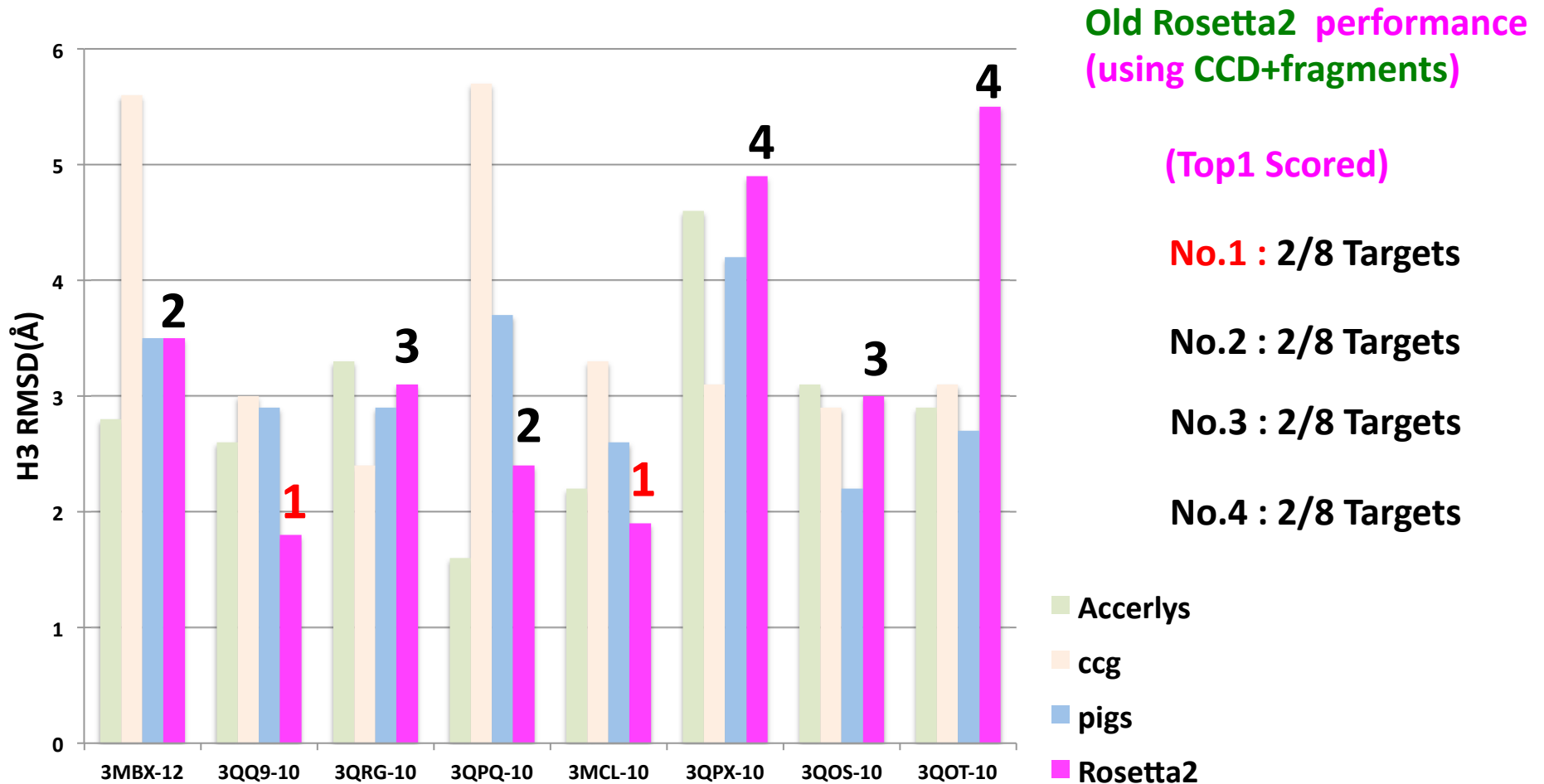
Almagro *et al.* Antibody Modeling Assessment (*Proteins* 2011)

- 9 Blinded Ab structure predictions (H3 10-13 residues)
- RosettaAntibody (Old Rosetta2), PIGS, CCG, Accelrys
- RMSD ~ 1.2 Å overall, 1 Å CDR Loops, 3 Å CDR H3 (Consistent across approaches)
- H3 is the biggest Challenge



Build H3 on Homology Models

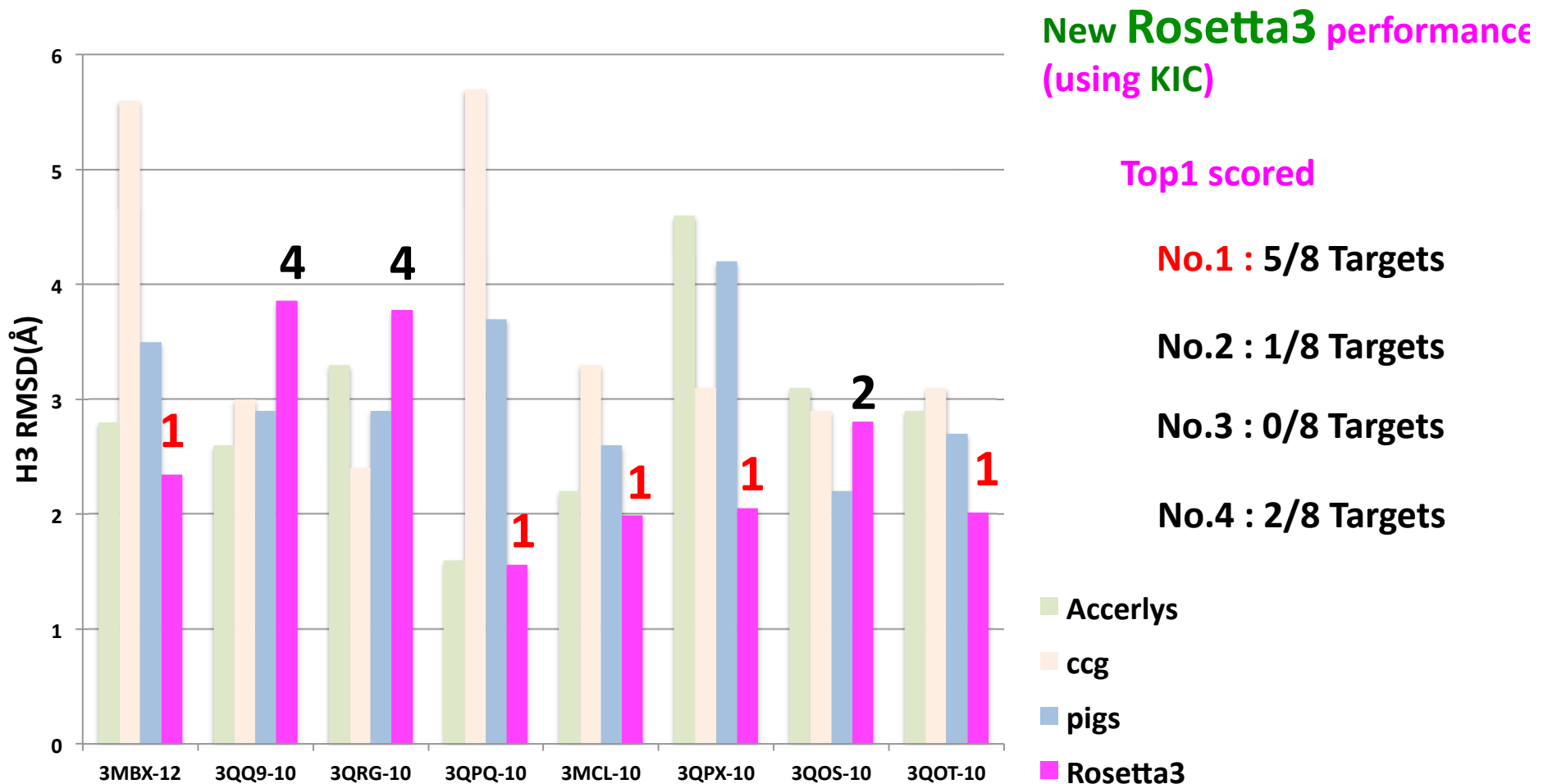
Performance of the **Old RosettaAntibody** Protocol (**whole protocol**)



(The Original Data in Almagro *et al's* paper)

Build H3 on Homology Models

Performance of the **New RosettaAntibody** Protocol (**whole protocol**)



**Only H3 loop modeling has been specifically improved in current study.
More improvements are in progress**

Summary

- **Rosetta Computational Pipeline**
 - With the input from experimental data, SnugDock + EnsembleDock using antibody homology models can successfully identify antigen epitopes.
 - Non-canonical amino acid design can be successfully conducted in Rosetta.
 - RosettaAntibody homology models were used to successfully design highly thermo resistant antibodies.
- **Refactored New RosettaAntibody Homology Modeling**
 - Object-oriented code facilitated the development and improvement of the new RosettaAntibody.
 - KIC with antibody information provides better H3 loop modeling.
 - Compared to other antibody homology modeling methods, new RosettaAntibody protocol has better performance.

Acknowledgments

Gray Lab (JHU)

Dr. Jianqing Xu

Dr. Daisuke Kuroda

Dr. Jason Labonte

Krishna Praneeth Kilambi

Brian Weitzner

William Proffitt

Nick Marze

Mike Pacella

Sergey Lyskov

Other Group Members

Collaborators

U-Texas Austin

Prof. Andy Ellington

Prof. George Georgiou

Dr. Aleksandr Miklos

Dr. Randall Hughes

Kam Hon Hoi

UNC

Prof. Brian Kuhlman

Bryan Der

Dr. P. Douglas Renfrew (NYU)

Servers

antibody.graylab.jhu.edu

rosettadock.graylab.jhu.edu



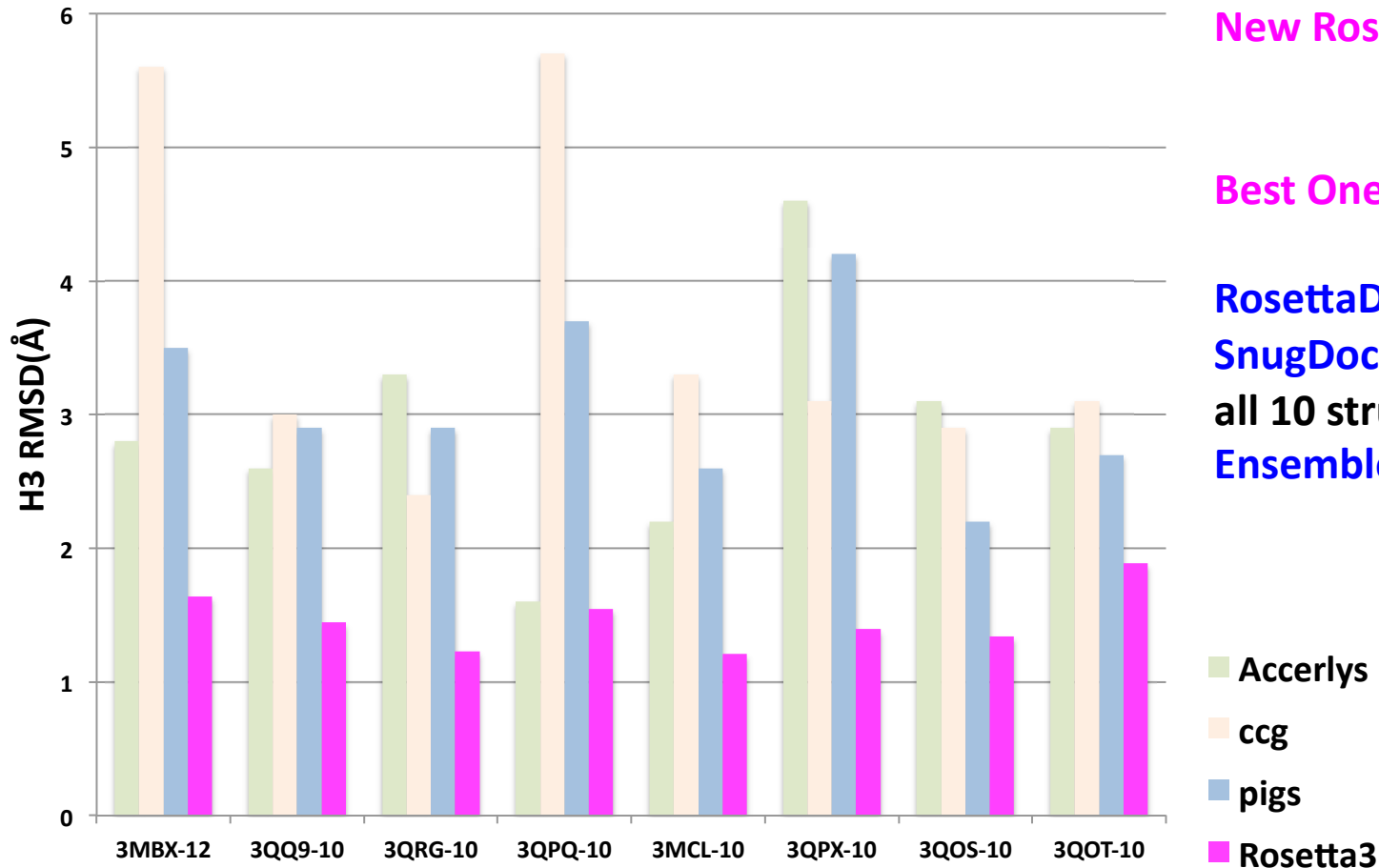
graylab.jhu.edu / pyrosetta.org / rosettacommons.org



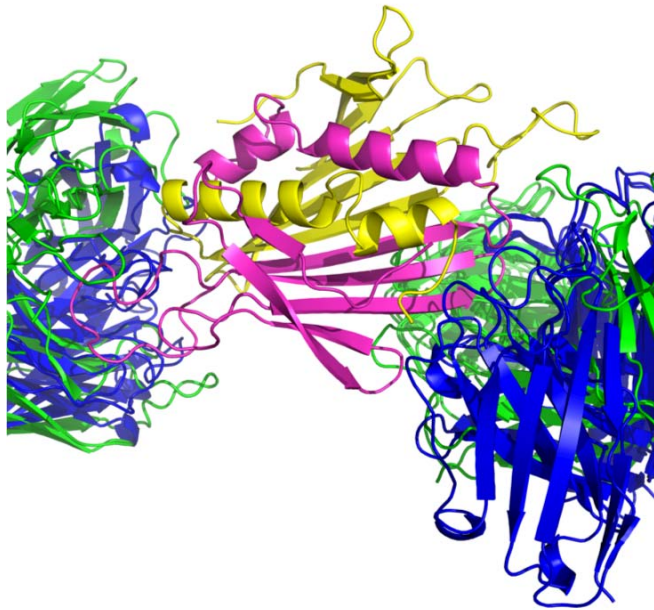
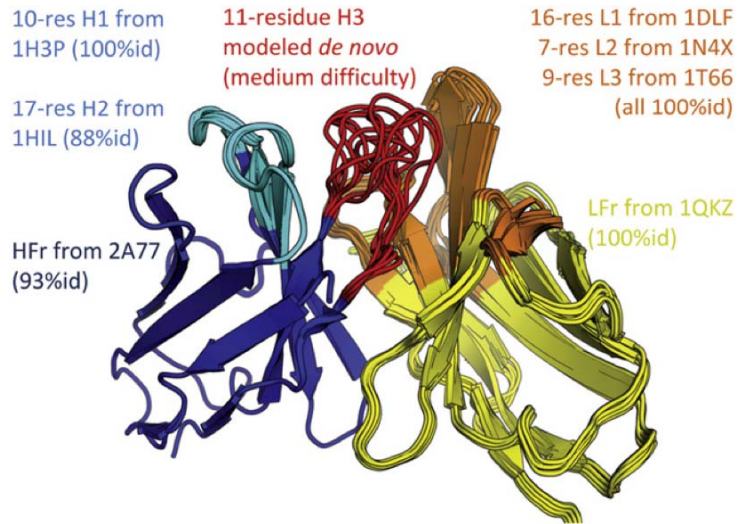
Thanks!

Build H3 on Homology Models

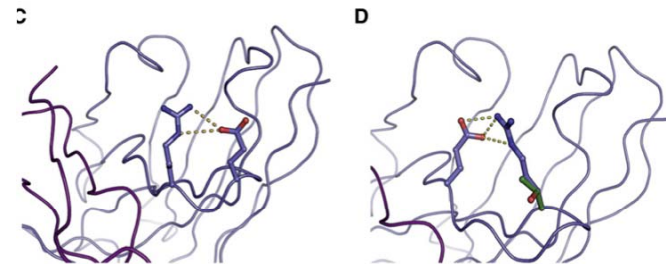
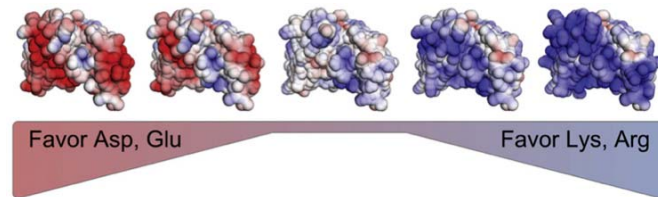
Performance of the **New RosettaAntibody** Protocol (**whole protocol**)



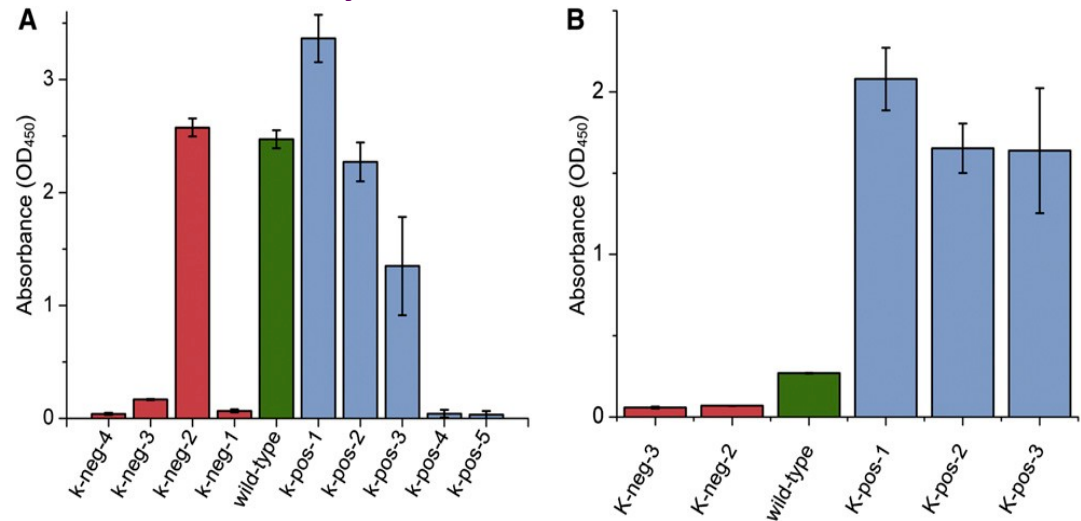
(Blind Target 2) Anti-MS2/MS2 (in progress)



Design of super-charged highly thermoresistant antibodies



Bryan Der, Brian Kuhlman's lab



Alex Miklos, Andy Ellington's lab