Leave no amino acid unturned:
de novo design, directed evolution, and fitness landscape exploitation leads to potential Influenza therapeutics.

Tim Whitehead

Michigan State (via Baker Lab)

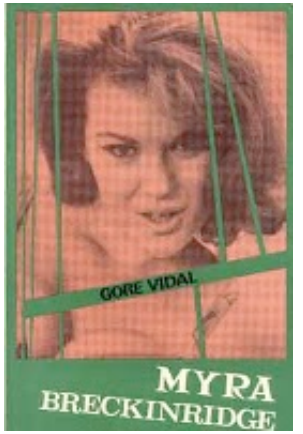
Collaborators

David

Sarel

Eva Strauch

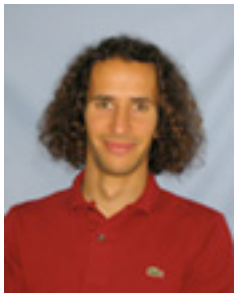
Aaron Chevalier



Jacob Corn

Ian Wilson

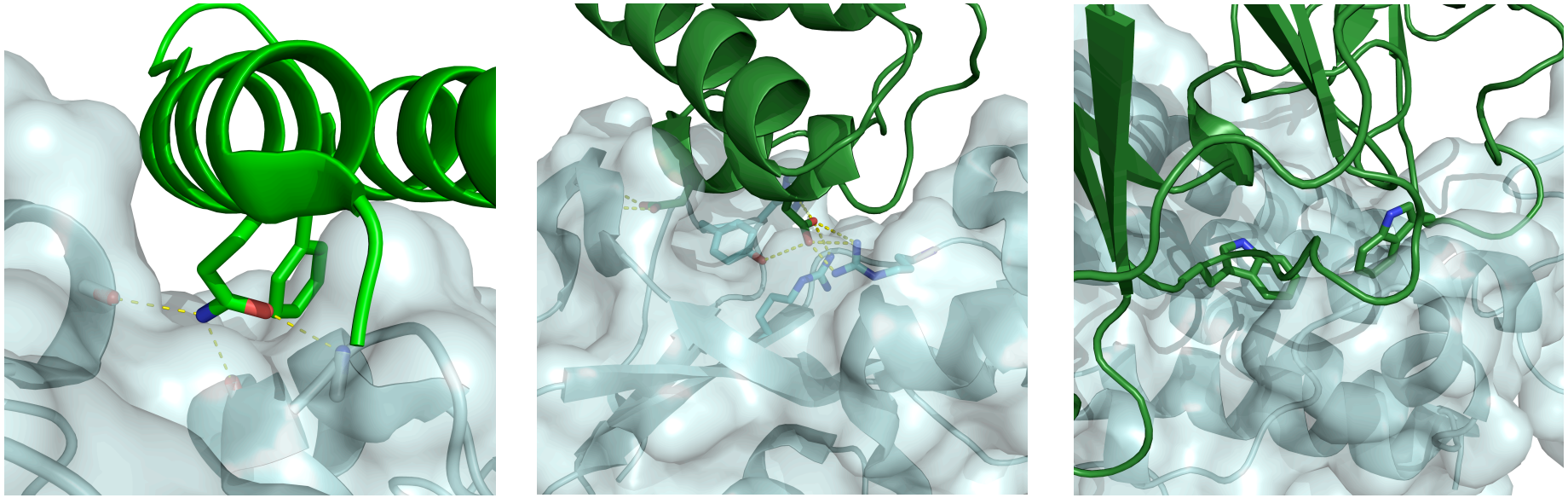
Damian Ekiert



Cyrille Dreyfus

Yifan Song

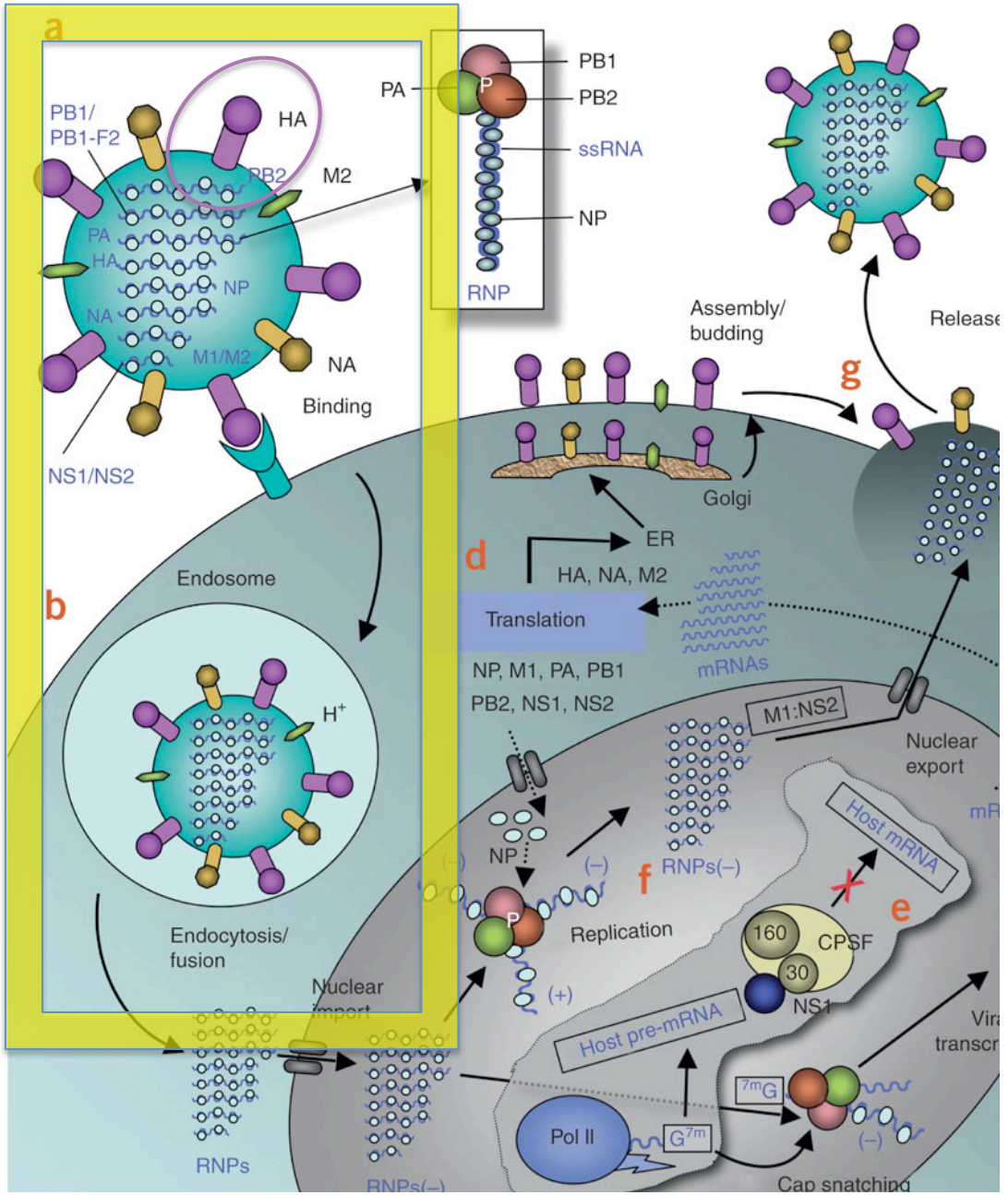
Native protein interfaces come in all shapes and sizes: how do we design new interfaces?



Problem statement: given a patch on a target protein surface, can one design a protein that interacts with high affinity and specificity to the targeted patch?

Influenza

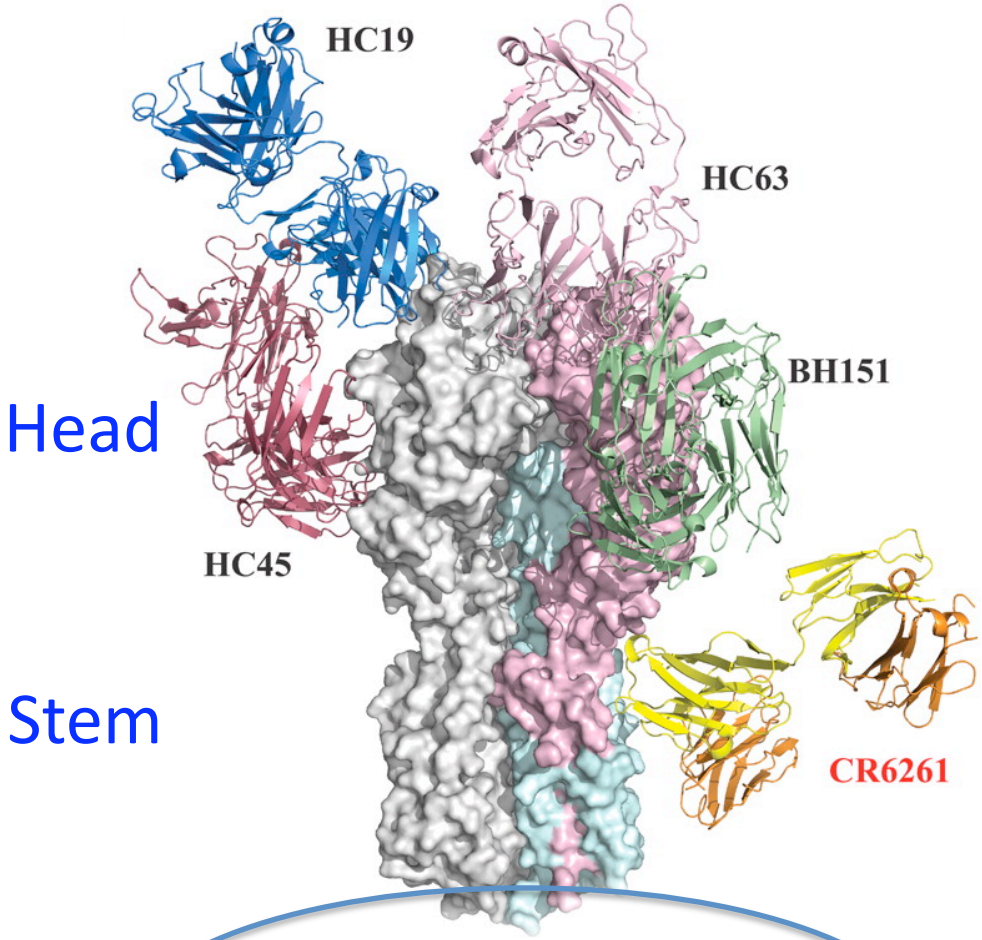
- Pandemic strains capable of millions of deaths per year
- Vaccines are strain-specific – surface proteins mutate too rapidly to develop panflu vaccine
- Most recent H1N1 (Swine flu) isolates are resistant to Tamiflu (98% in US)*
- One of two major surface proteins is ***hemagglutinin (HA)*** (e.g. H1N1, H5N1)



Nature Structural & Molecular Biology 17,530–538 (2010)

*WHO Influenza A(H1N1) virus resistance to oseltamivir – 2008/2009 influenza season, northern hemisphere 18 March 2009

Human antibodies typically bind to the *head* of influenza hemagglutinin (HA) which mutates every year. Only CR6261 binds to the conserved *stem* region, ***inhibiting*** virus infectivity.



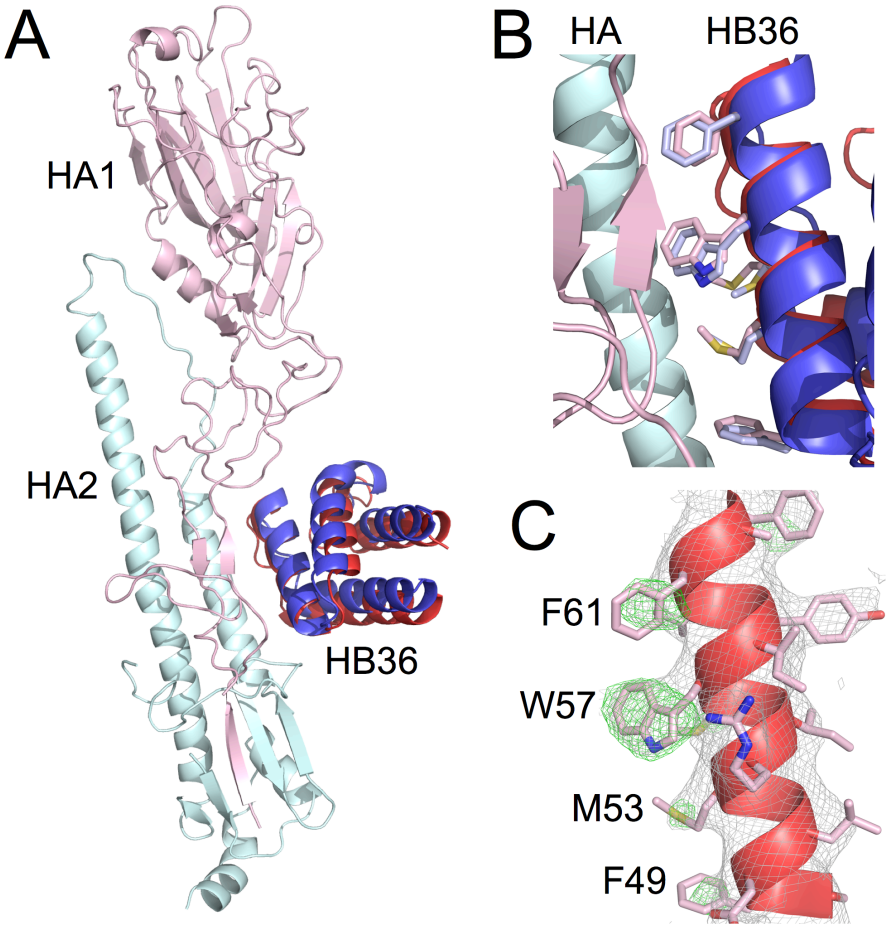
CR6261 will not be used as a frontline therapeutic

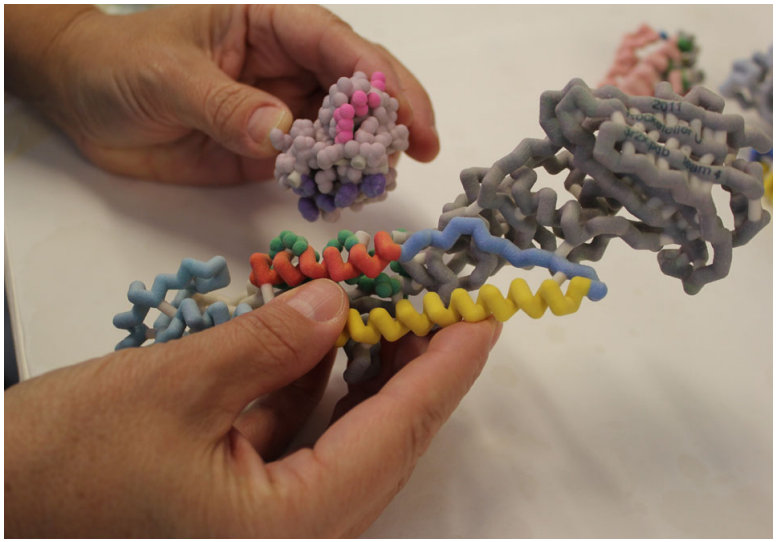
- Extremely high costs for antibody production (~$100/g)
- Extremely high dosing required (5-10 mg/kg)

Our goal is to ***design*** small proteins targeting the *stem* region, inhibiting the virus

Ekiert, Wilson *et al*. Science (2009) 324:246

Two de novo proteins were designed that target an epitope conserved among group I influenza A virus





Courtesy http://cbm.msoe.edu/

Great! But.....

With Sarel Fleishman

…both initial hit rate (2-3%) and starting affinities (>>1 uM in vitro) are discouraging

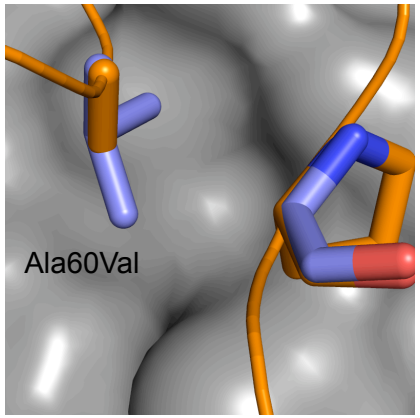
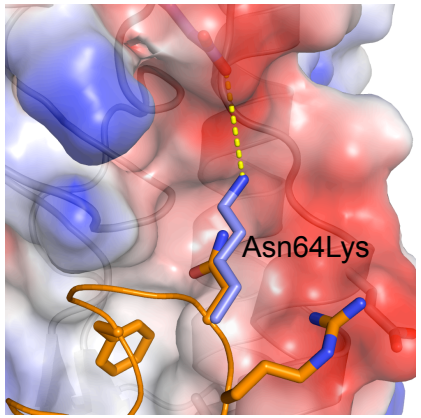
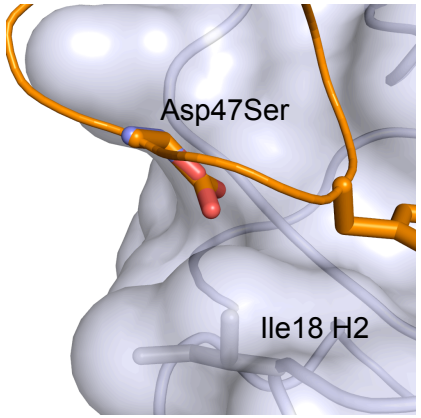
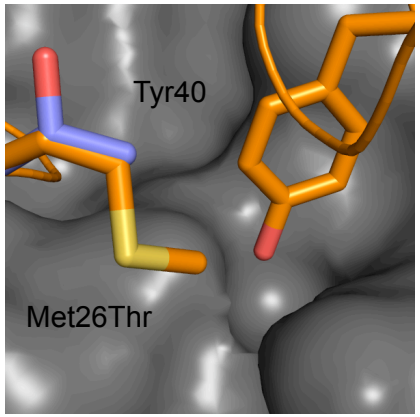
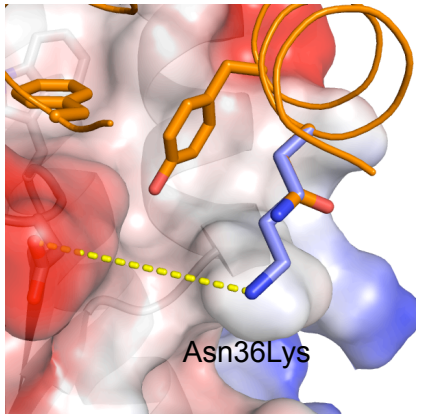
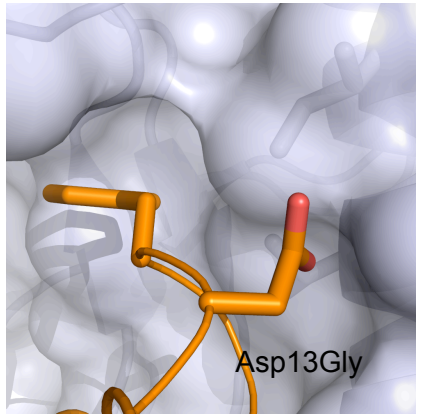
How do we fix this?

If we better understood the physics underlying molecular recognition, our success rate and starting affinities would be much higher

Other factors can undoubtedly be underlying causes for failure (i.e. protein doesn't fold correctly, binding conception is wrong, assay is crap)

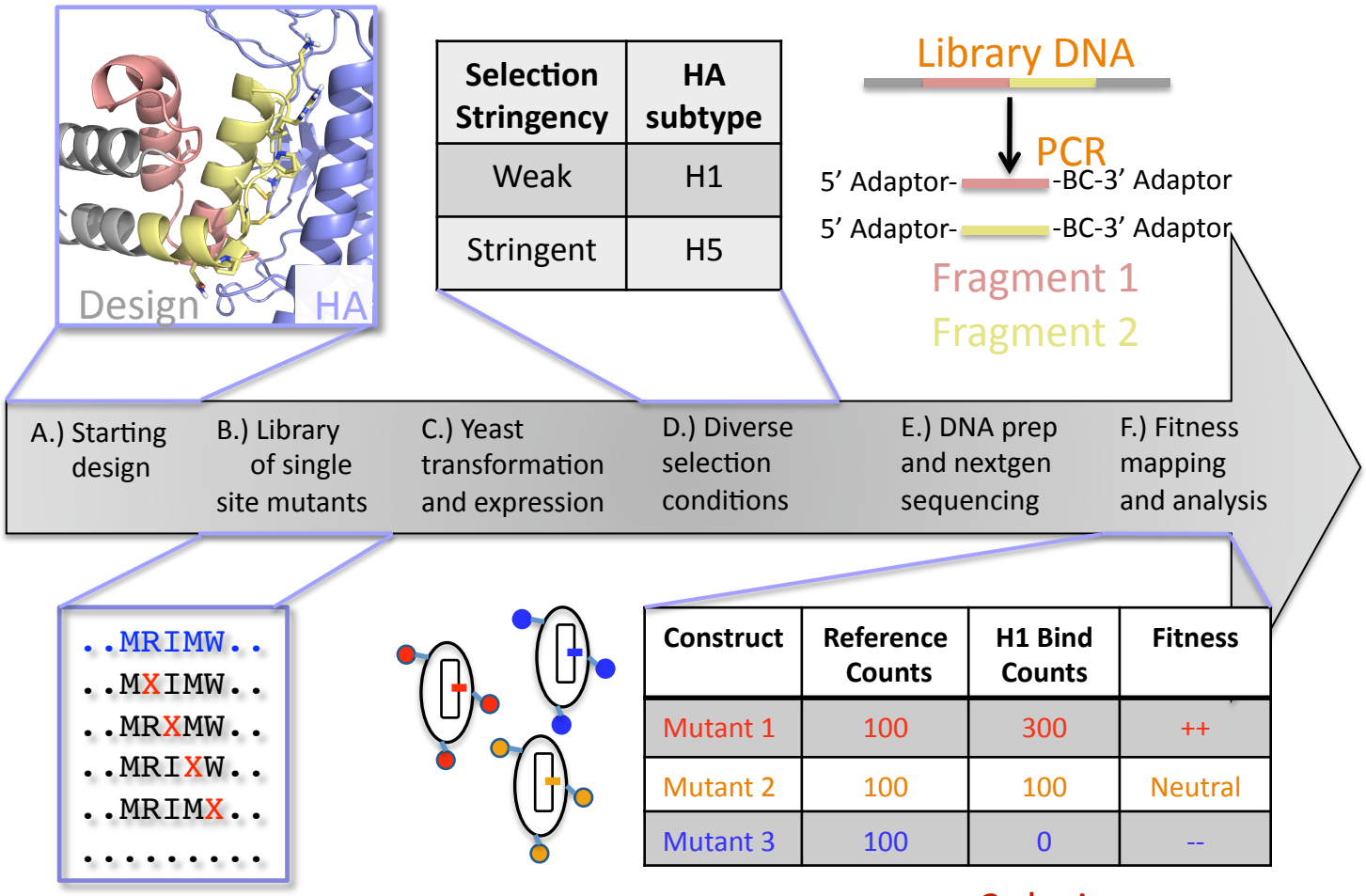
We can begin to understand the modeling deficiencies in the energy function by affinity maturation

Affinity maturation identifies deficiencies in modeling yet the data is sparse

Design	Repulsive Interactions	Electrostatics	Solvation
HB36			
HB80			

Can we develop an experimental method that determines the fitness of every single point mutant in a gene?

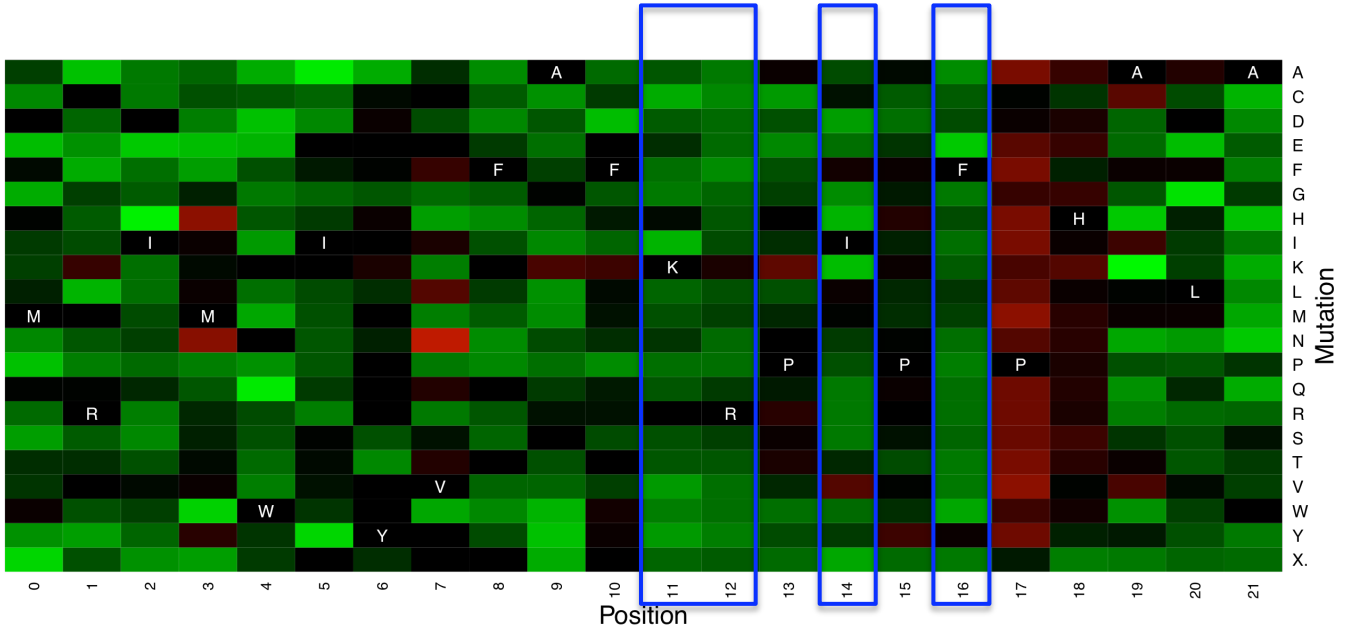
We extended deep mutational scanning procedure for use in yeast display and to cover entire genes



2 designs
30 selected populations
1.5 million reads median
Almost complete coverage (1052/1060 for HB36.4)

Fowler et al, Nat Meth 2010

Deep mutational scanning gives a fitness map for every single point mutation in a protein sequence



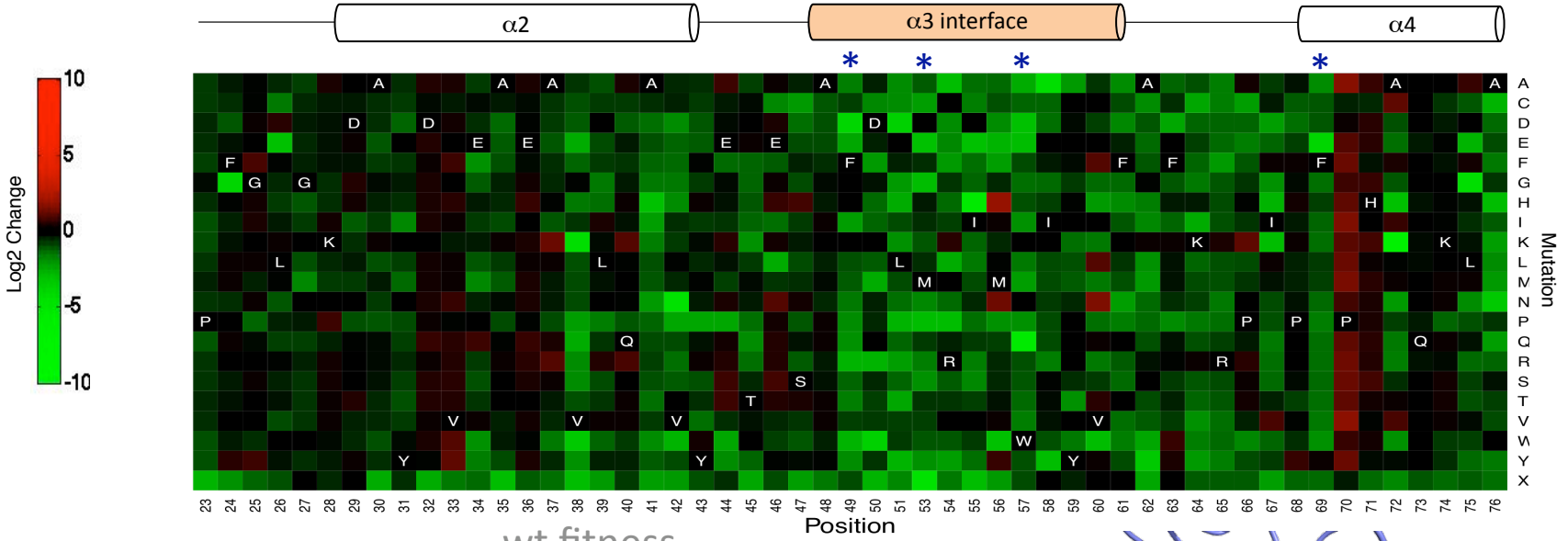
Enriched

Depleted

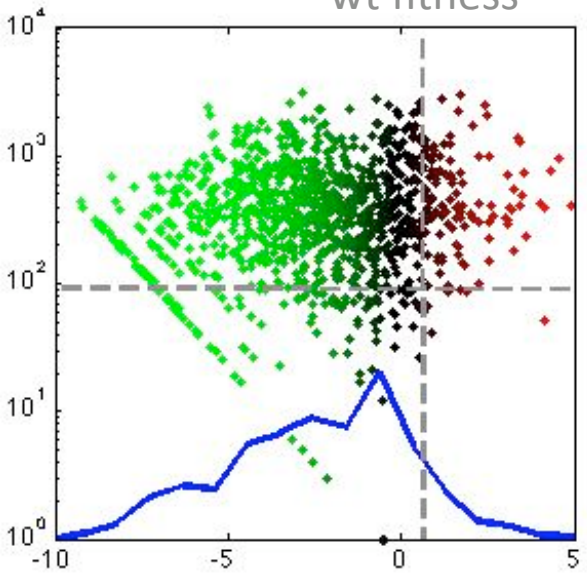
Neutral

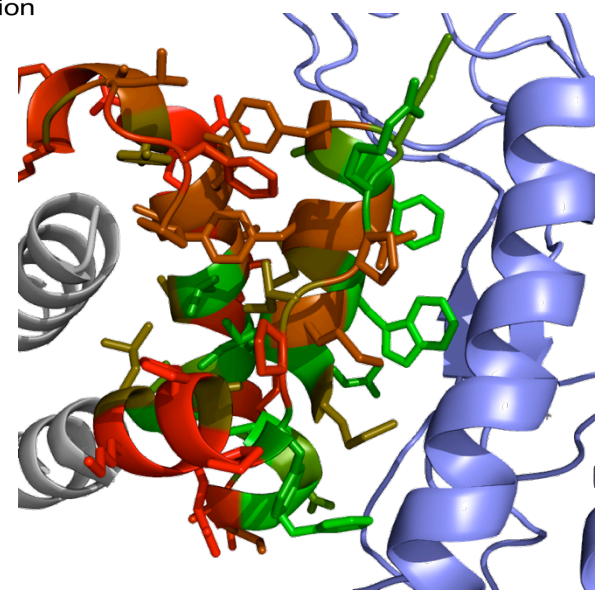
With Aaron Chevalier

Fitness landscape for HB36.4 design



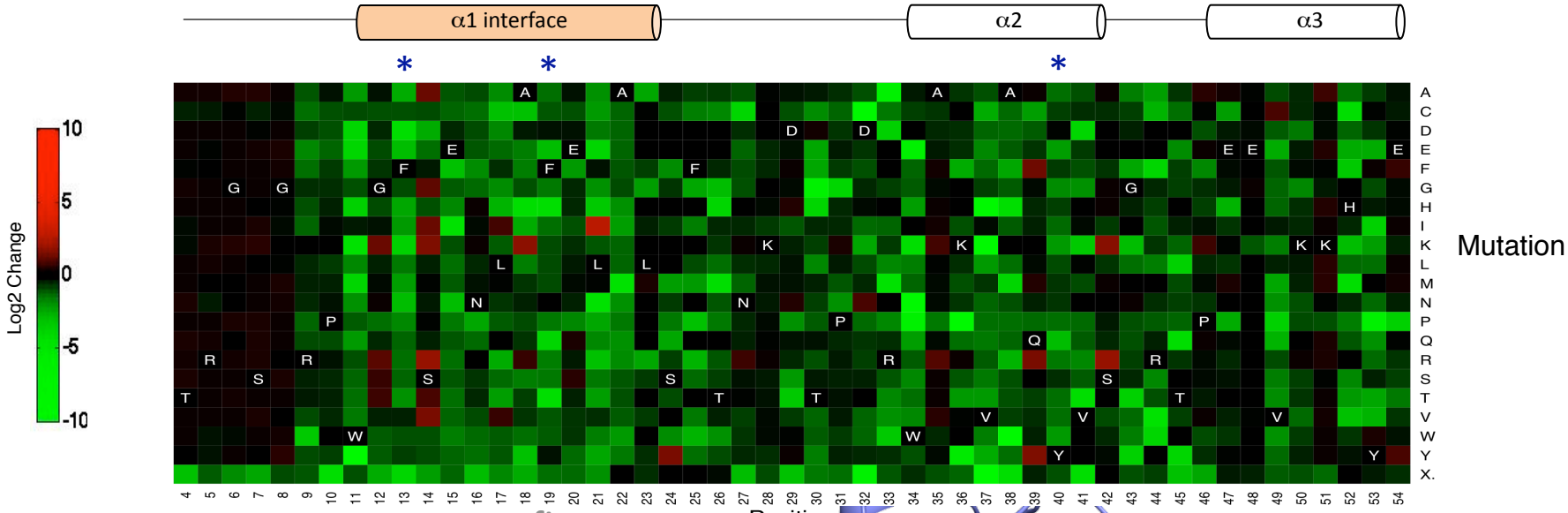
Input sequences



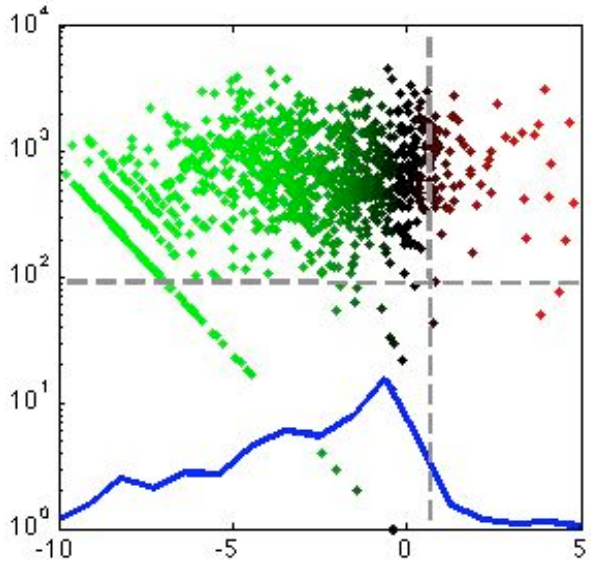


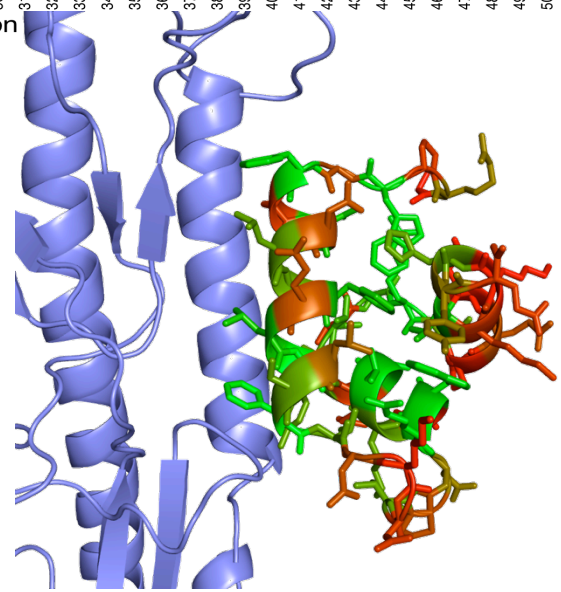
HB36.4 colored by Shannon entropy

Fitness landscape for HB80.3 design



Input sequences

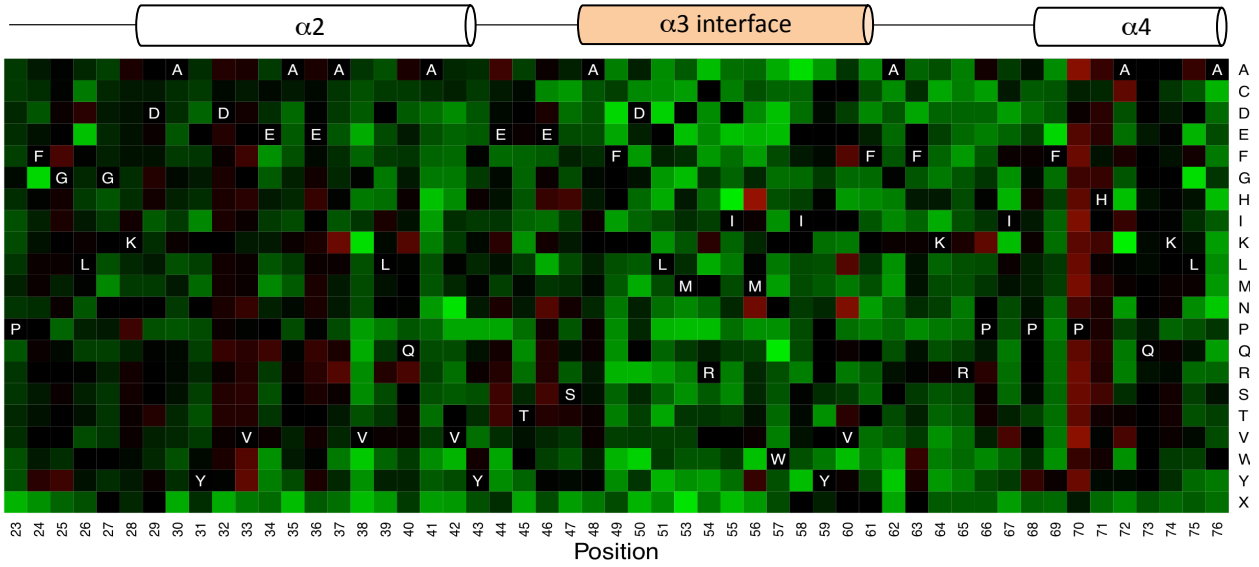




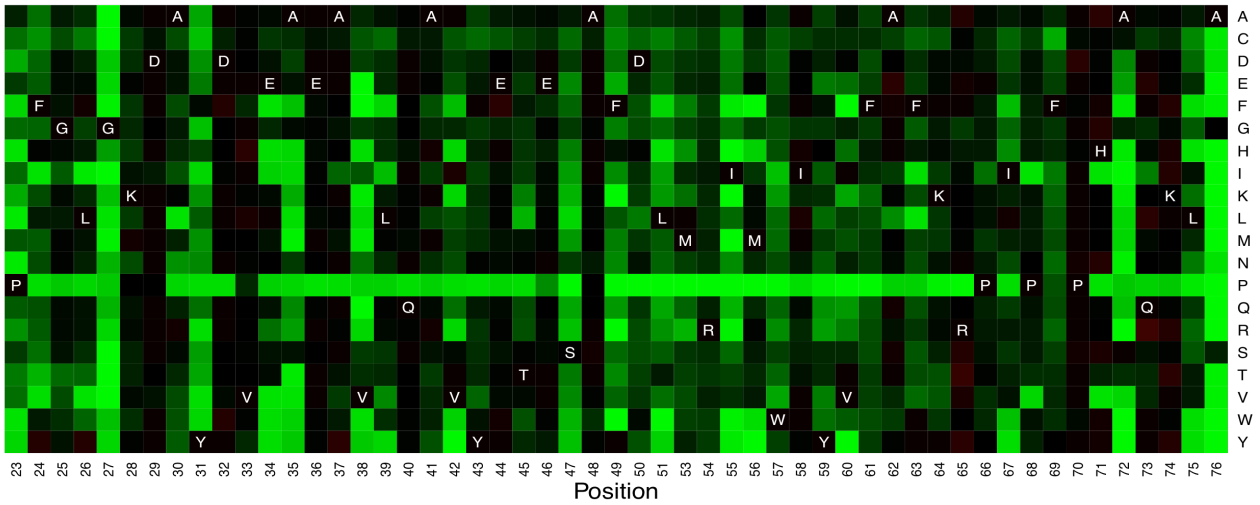
HB80.3 colored by Shannon entropy

Rosetta cannot identify beneficial substitutions in HB36

Expl Fitness

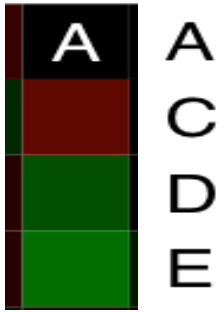


Rosetta sc12

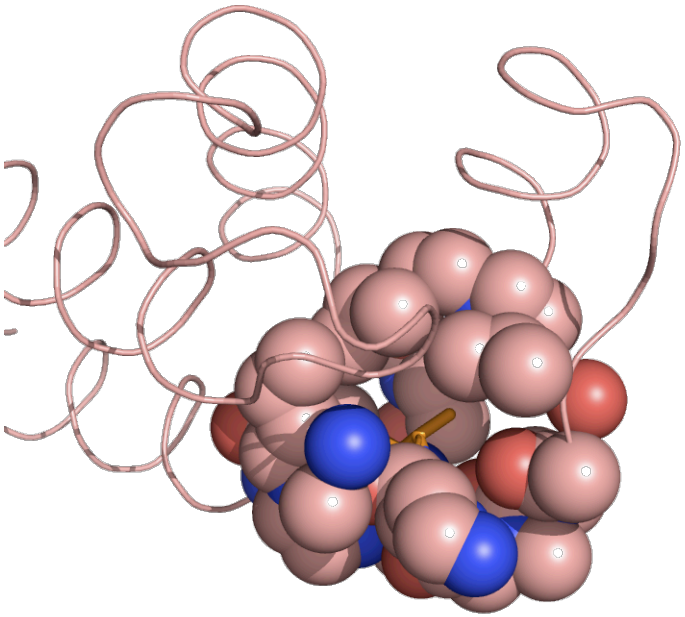


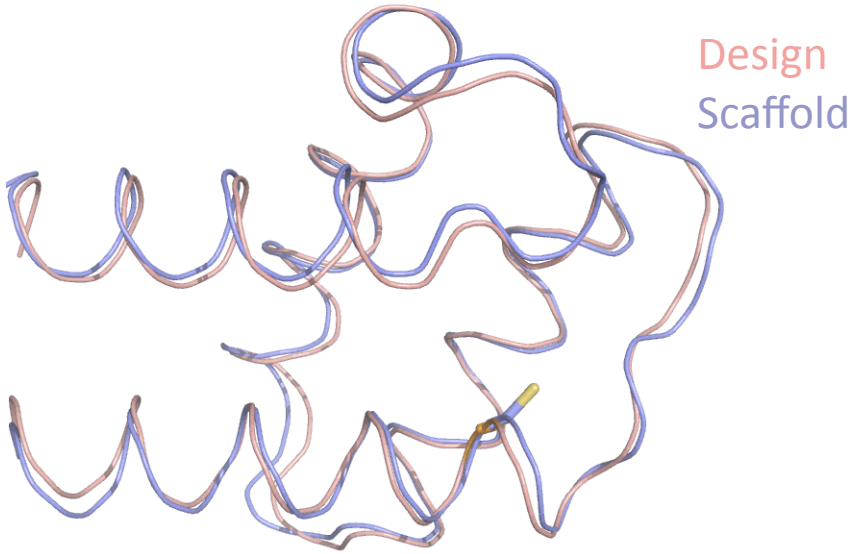
Lesson #1: don't mutate the core indiscriminately!

Deep sequencing found a A72C core mutation as beneficial which Rosetta hates.
What was going on here?

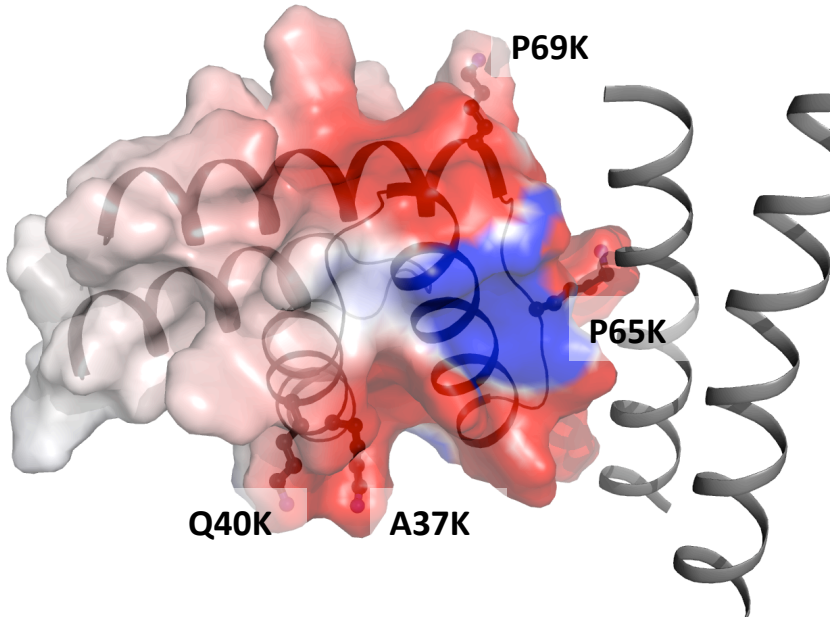


Design moves with backrub and relax, filling void left by cys from scaffold



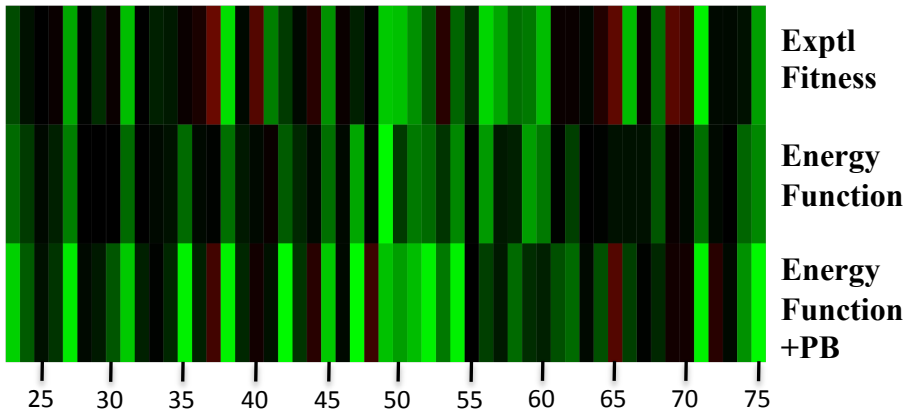


Lesson #2: electrostatics plays a much larger role in binding apolar complexes than we appreciated



HB36 HA

All lysines substitutions in HB36



Yifan Song

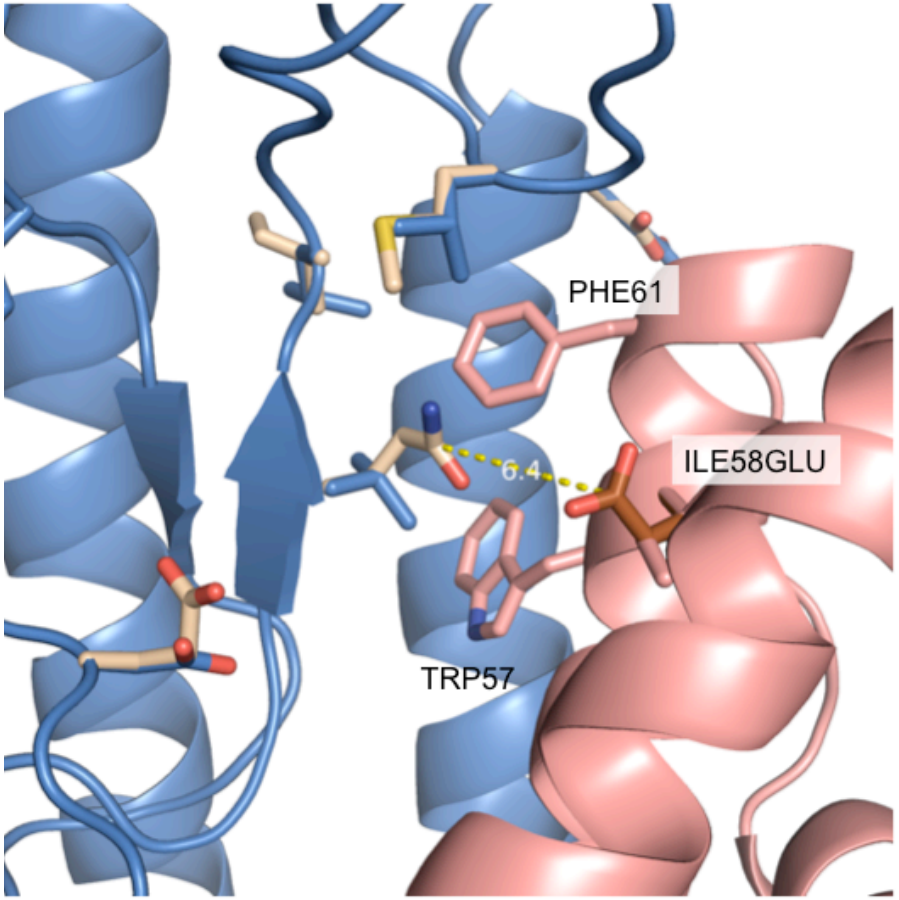
What can we do with these fitness landscapes besides learn how to better design?

- Diagnostics: engineering specificity

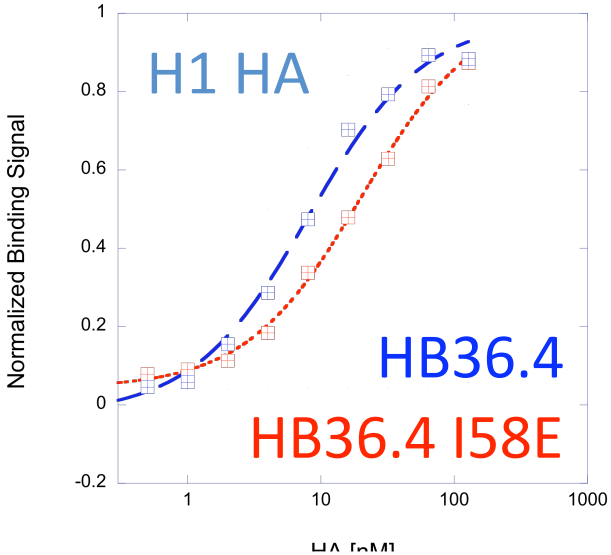
- Therapeutics: engineering affinity

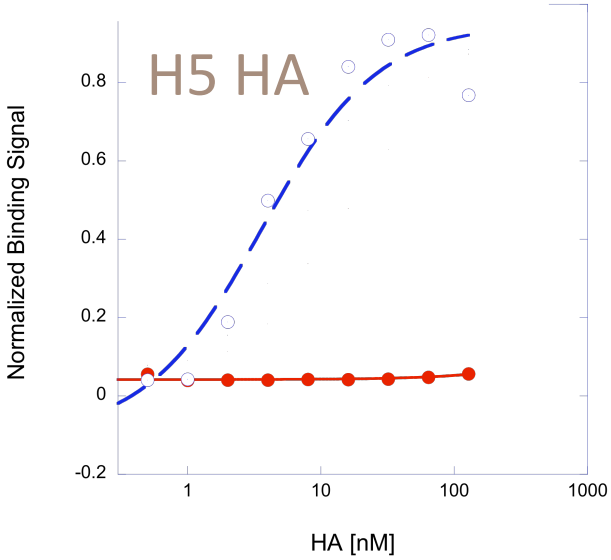
Engineering Specificity: Deep sequencing data shows us how to make H1-specific binders

Difference in HA epitope region between H1 and H5-specific subtypes



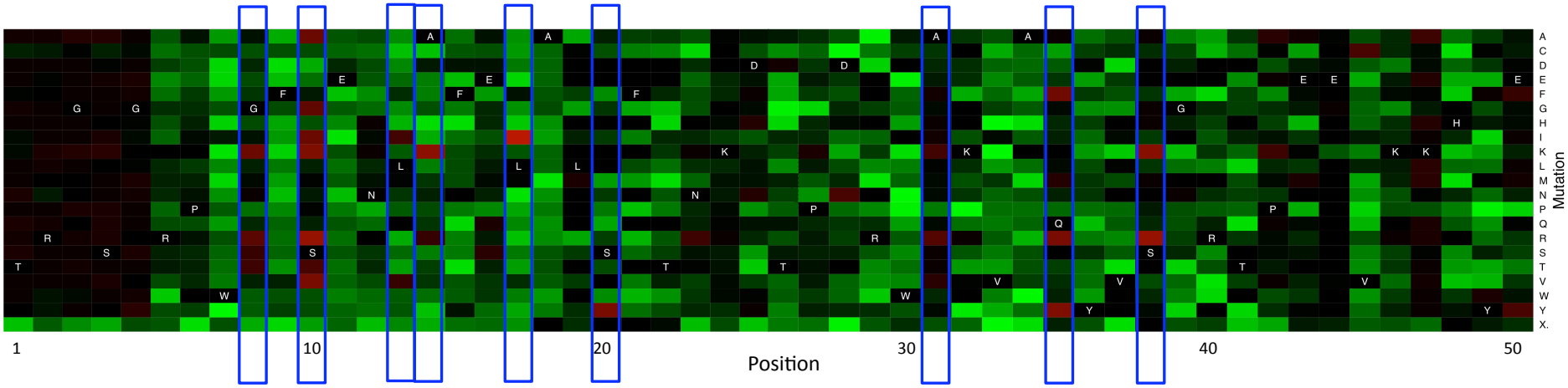
HA HB36.4





With Aaron Chevalier

Engineering affinity: beneficial substitutions are to a first approximation pairwise additive



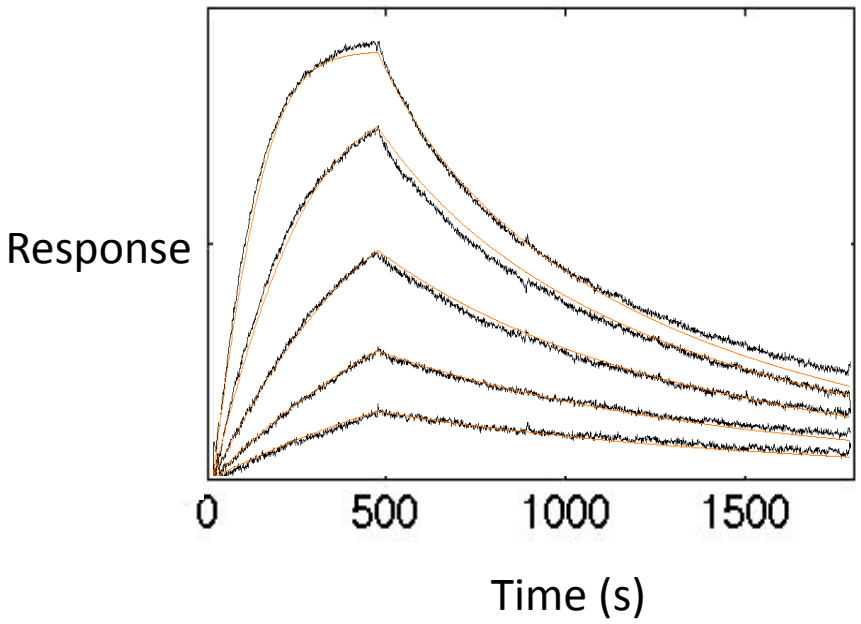
1. Create library where boxed positions can be either wt or enriched substitutions
2. Average 6 mutations per gene (range 0-9)
3. Theoretical library size < $1x10^6$
4. Stringent selection for 5 sorts using yeast display

HB80.3 Position	12	14	17	18	21	24	35	39	42
		K							
	K	R	I				K	Y	K
	R	N	V	K	I	Y	R	R	R
init. Library	G	S	L	A	L	S	A	Q	S
	E			T			E	Stop	N
	G			E			G	C	
							T	W	
							K	H	
best variant	K	S	I	A	I	S	K	Q	K

With Aaron Chevalier

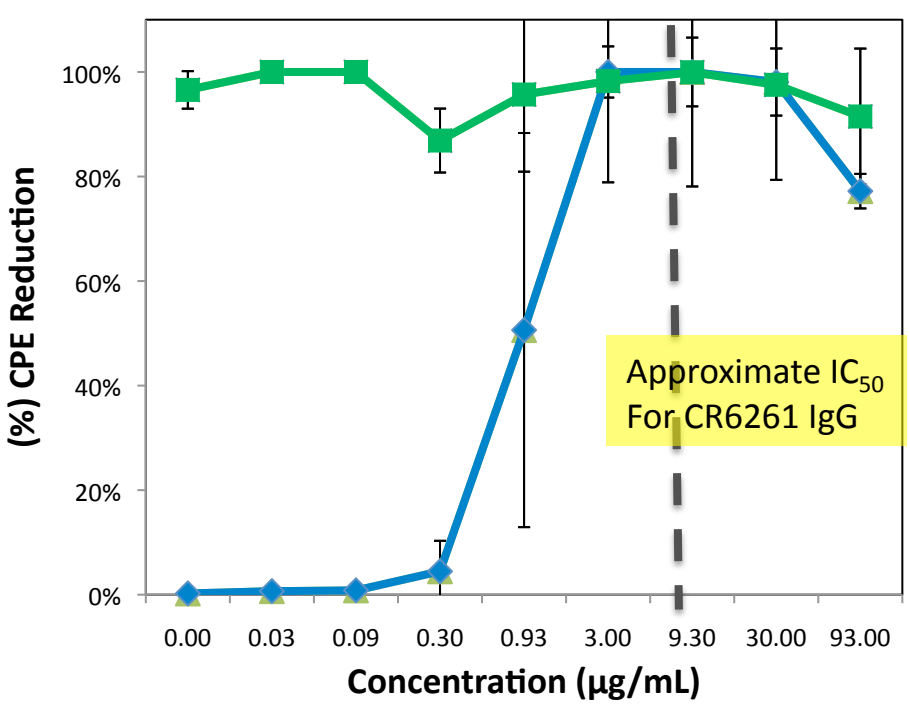
F-HB80.4 has 25-fold improved affinity and broadened specificity compared to its parent

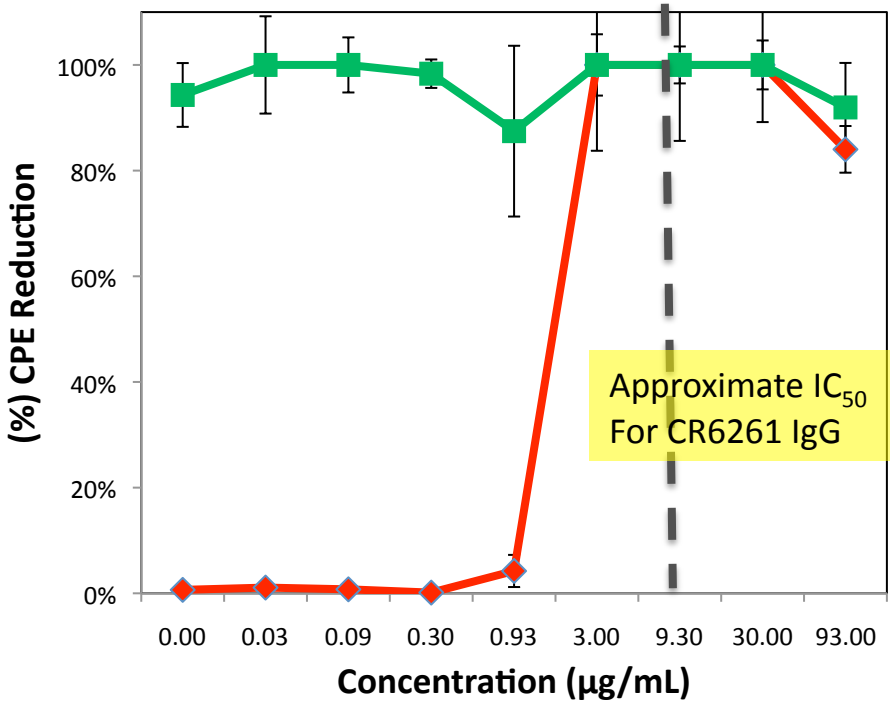
Variant	K_d [nM]	$t_{1/2}$ [s]
F-HB80.3	18	36
F-HB80.4	0.6	550



	Strains	K_D (nM)	
		F-HB80.4	CR6261
Group 1	A/South Carolina/1/1918 (H1N1)	4.9	1.3
	A/Japan/305/1957 (H2N2)	16.7	n.d.
	A/Adachi/2/1957 (H2N2)	10.5	n.d.
	A/Vietnam/1203/2004 (H5N1)	1.4	1.2
	A/Indonesia/05/2005 (H5N1)	0.7	n.d.
	A/turkey/Massachusetts/3740/1965 (H6N2)	9.3	1.8
	A/turkey/Wisconsin/1/1966 (H9N2)	7.5	9
	A/duck/Alberta/60/1976 (H12N5)	430	-
	A/gull/Maryland/704/1977 (H13N6)	55	36
	A/black-headed gull/Sweden/4/1999 (H16N3)	540	52
Group 2	A/Hong Kong/1/1968 (H3N2)	-	-
	A/duck/Czechoslovakia/1956 (H4N6)	-	-
	A/Netherlands/219/2003 (H7N7)	-	-
	A/chicken/Germany/N/1949 (H10N7)	-	-
	A/mallard duck/Astrakhan/263/1982 (H14N5)	-	-
	A/shearwater/West Australia/2576/1979 (H15N8)	-	-

F-HB80.4 neutralizes Influenza virus at doses 10-fold less than the best-in-class antibody!





Cecilia de Mattos, NHRC

Outlook

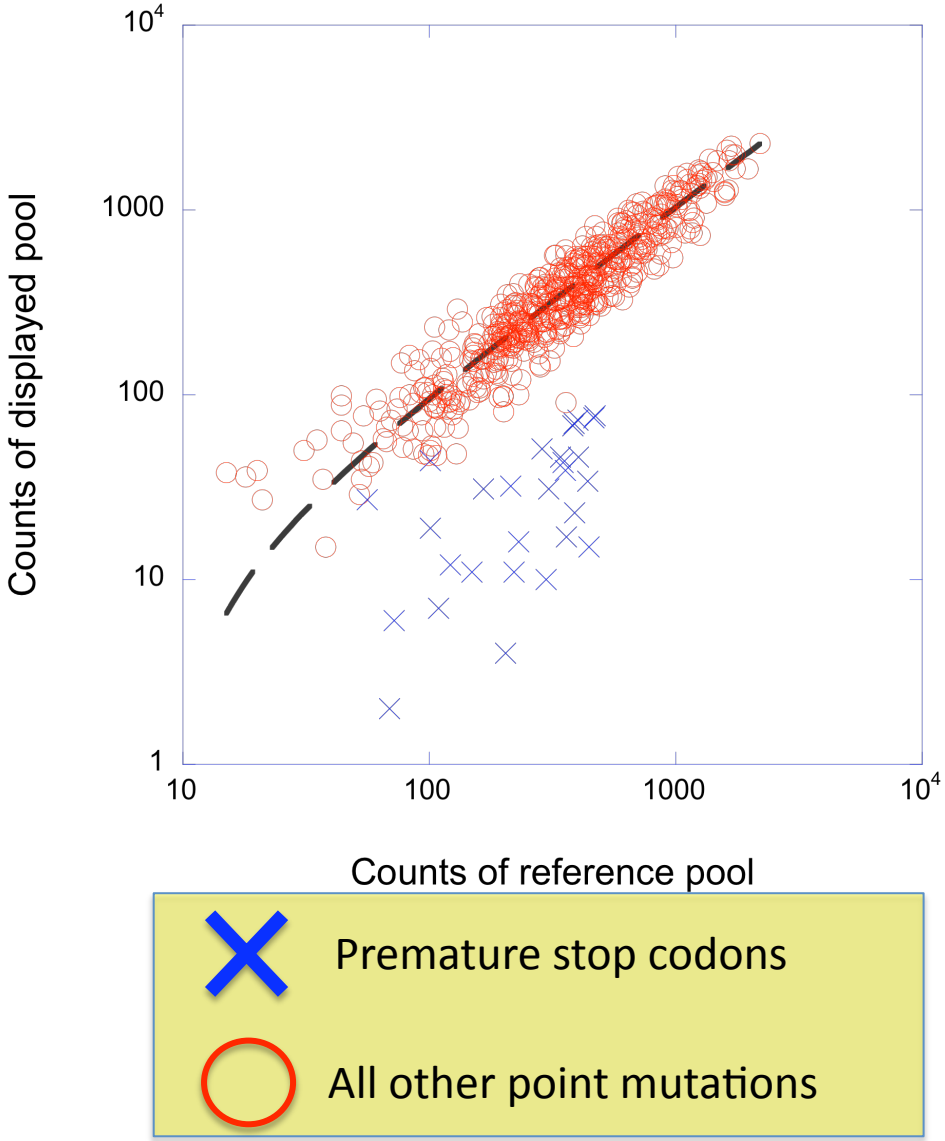
From conception of a small protein that binds a specific conserved epitope on Influenza and neutralizes the virus to a lead molecule took:

2 years, 2 postdocs, 2 grad students, 2 PI's

In 20 years, can we do the same in 2 weeks?

Surprise/Problem #1

- "yeast polypeptide fusion surface display predicts thermal stability"[1]
- "proteins are subject to quality control in the ER, wherein chaperones and foldases ensure the fidelity of the secreted protein species"[2]



1. J Mol Biol 292, 949 (1999); 2. Handbook of therapeutic monoclonal antibodies" (2009)