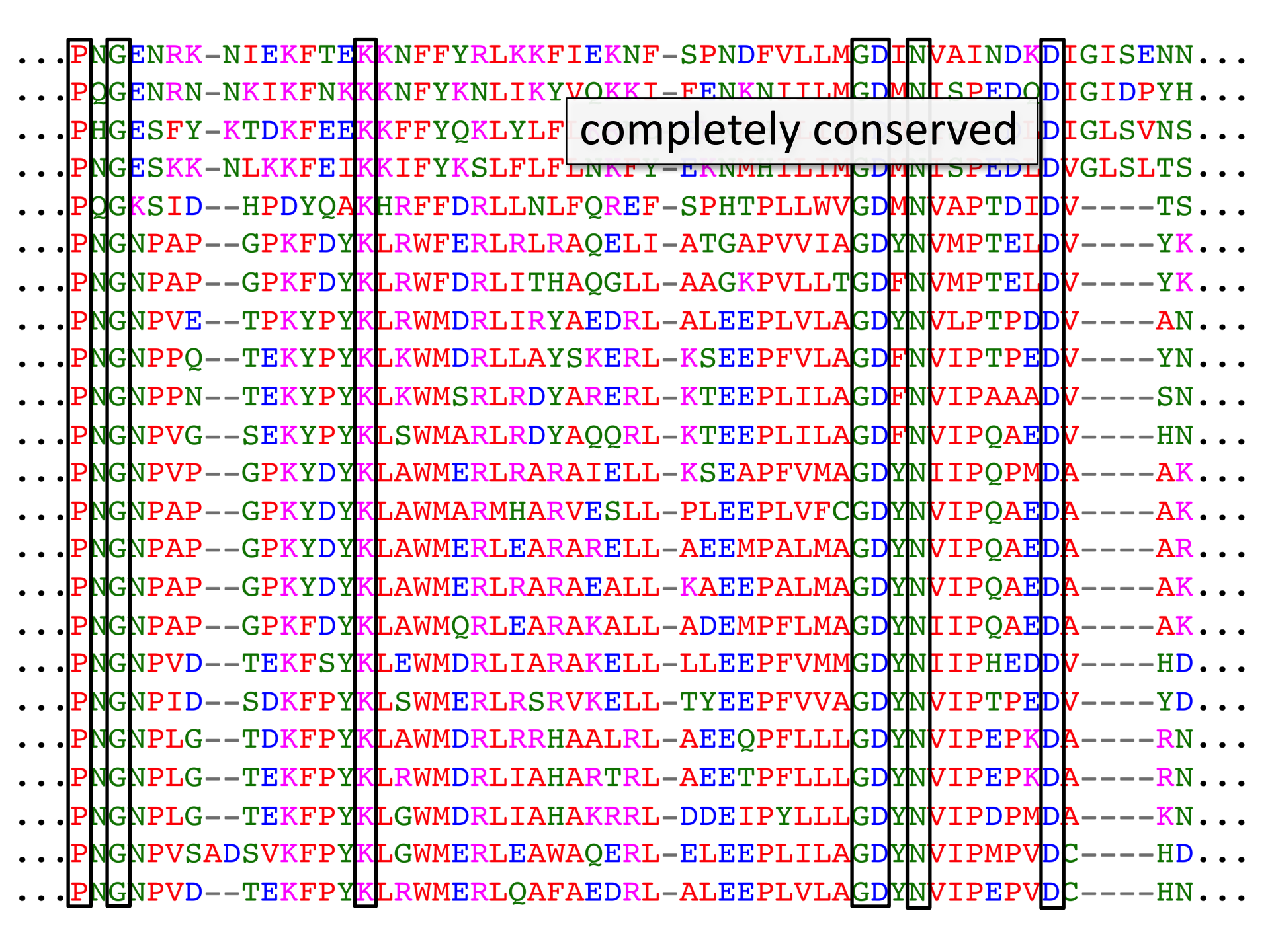
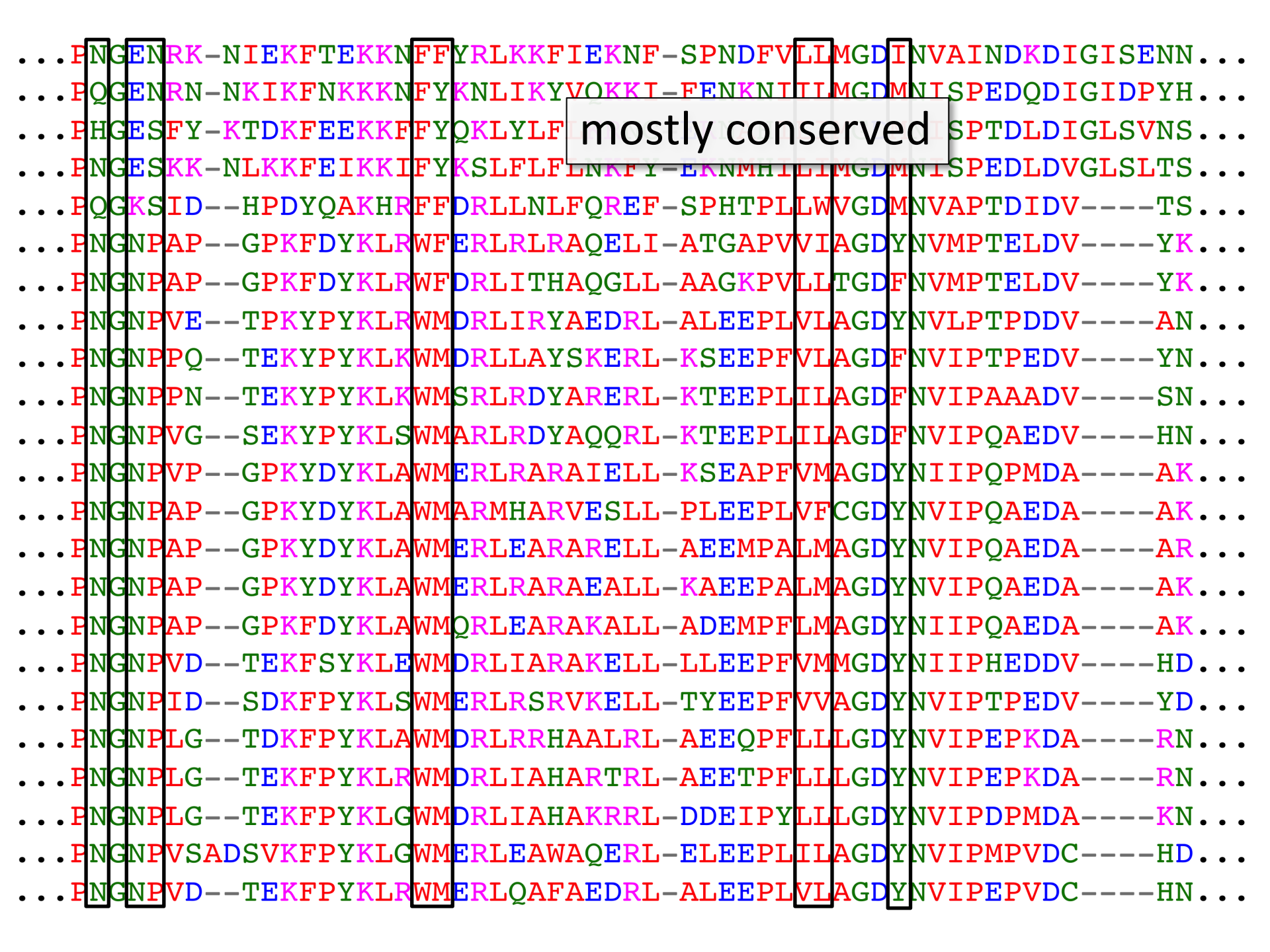
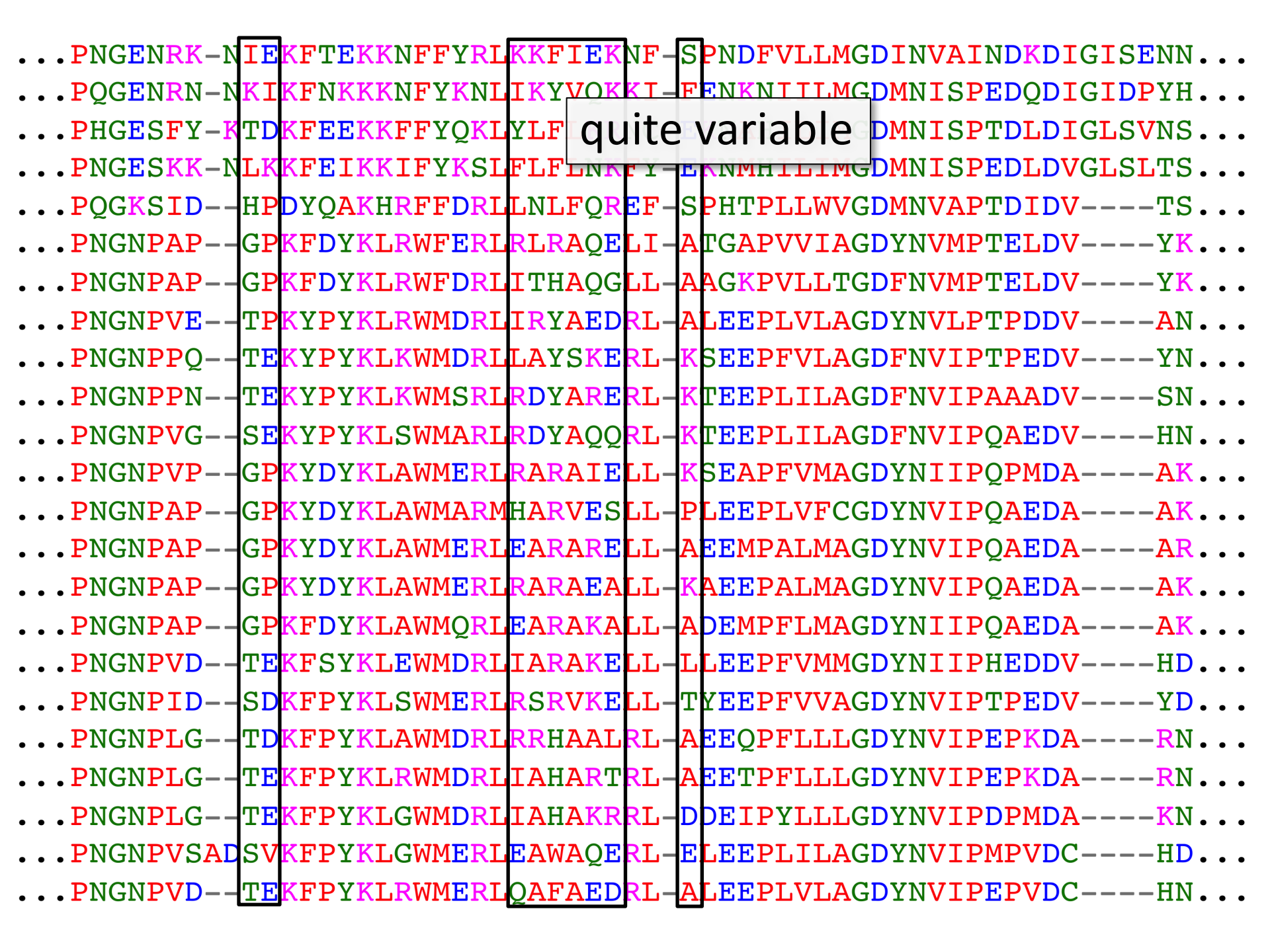
Structural and functional constraints on protein evolution

Claus O. Wilke
The University of Texas at Austin

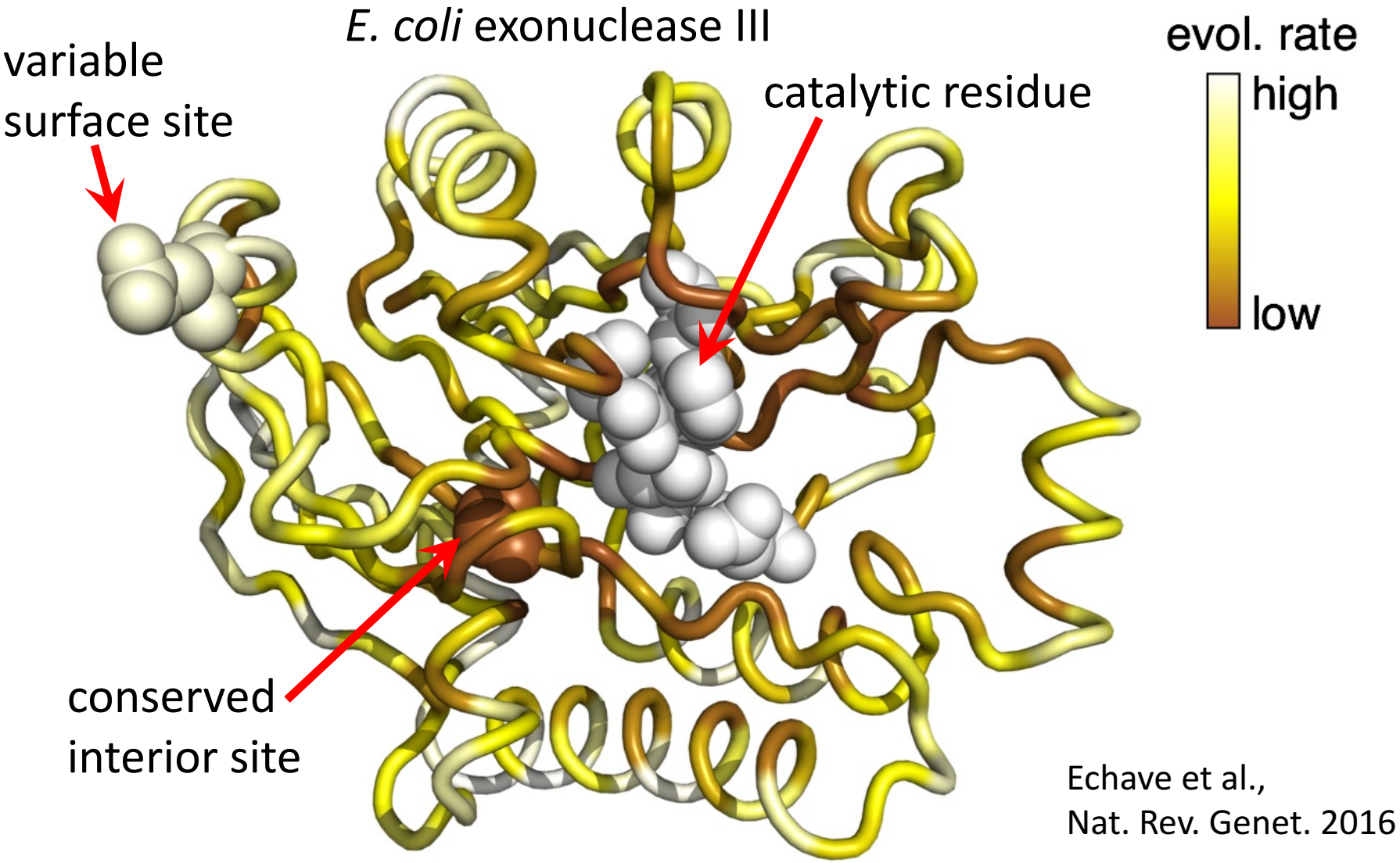

...PNGENRK-NIEKFTEKKNFFYRLKKFIEKNF-SPNDFVLLMGDINVAINDKDIGISENN...
...PQGENRN-NKIKFNKKKNFYKNLIKYVQKKI-FENKNIILMGDMNISPEDQDIGIDPYH...
...PHGESFY-KTDKFEEKKFFYQKLYLFLKRNC-EKNAHILIMGDMNISPTDLDIGLSVNS...
...PNGESKK-NLKKFEIKKIFYKSLFLFLNKFY-EKNMHILIMGDMNISPEDLDVGLSLTS...
...PQGKSID--HPDYQAKHRFFDRLLNLFQREF-SPHTPLLWVGDMNVAPTDIDV----TS...
...PNGNPAP--GPKFDYKLRWFERLRLRAQELI-ATGAPVVIAGDYNVMPTELDV----YK...
...PNGNPAP--GPKFDYKLRWFDRLITHAQGLL-AAGKPVLLTGDFNVMPTELDV----YK...
...PNGNPVE--TPKYPYKLRWMDRLIRYAEDRL-ALEEPLVLAGDYNVLPTPDDV----AN...
...PNGNPPQ--TEKYPYKLKWMDRLLAYSKERL-KSEEPFVLAGDFNVIPTPEDV----YN...
...PNGNPPN--TEKYPYKLKWMSRLRDYARERL-KTEEPLILAGDFNVIPAAADV----SN...
...PNGNPVG--SEKYPYKLSWMARLRDYAQQRL-KTEEPLILAGDFNVIPQAEDV----HN...
...PNGNPVP--GPKYDYKLAWMERLRARAIELL-KSEAPFVMAGDYNIIPQPMDA----AK...
...PNGNPAP--GPKYDYKLAWMARMHARVESLL-PLEEPLVFCGDYNVIPQAEDA----AK...
...PNGNPAP--GPKYDYKLAWMERLEARARELL-AEEMPALMAGDYNVIPQAEDA----AR...
...PNGNPAP--GPKYDYKLAWMERLRARAEALL-KAEEPALMAGDYNVIPQAEDA----AK...
...PNGNPAP--GPKFDYKLAWMQRLEARAKALL-ADEMPFLMAGDYNIIPQAEDA----AK...
...PNGNPVD--TEKFSYKLEWMDRLIARAKELL-LLEEPFVMMGDYNIIPHEDDV----HD...
...PNGNPID--SDKFPYKLSWMERLRSRVKELL-TYEEPFVVAGDYNVIPTPEDV----YD...
...PNGNPLG--TDKFPYKLAWMDRLRRHAALRL-AEEQPFLLLGDYNVIPEPKDA----RN...
...PNGNPLG--TEKFPYKLRWMDRLIAHARTRL-AEETPFLLLGDYNVIPEPKDA----RN...
...PNGNPLG--TEKFPYKLGWMDRLIAHAKRRL-DDEIPYLLLGDYNVIPDPMDA----KN...
...PNGNPVSADSVKFPYKLGWMERLEAWAQERL-ELEEPLILAGDYNVIPMPVDC----HD...
...PNGNPVD--TEKFPYKLRWMERLQAFAEDRL-ALEEPLVLAGDYNVIPEPVDC----HN...


completely conserved


mostly conserved


quite variable

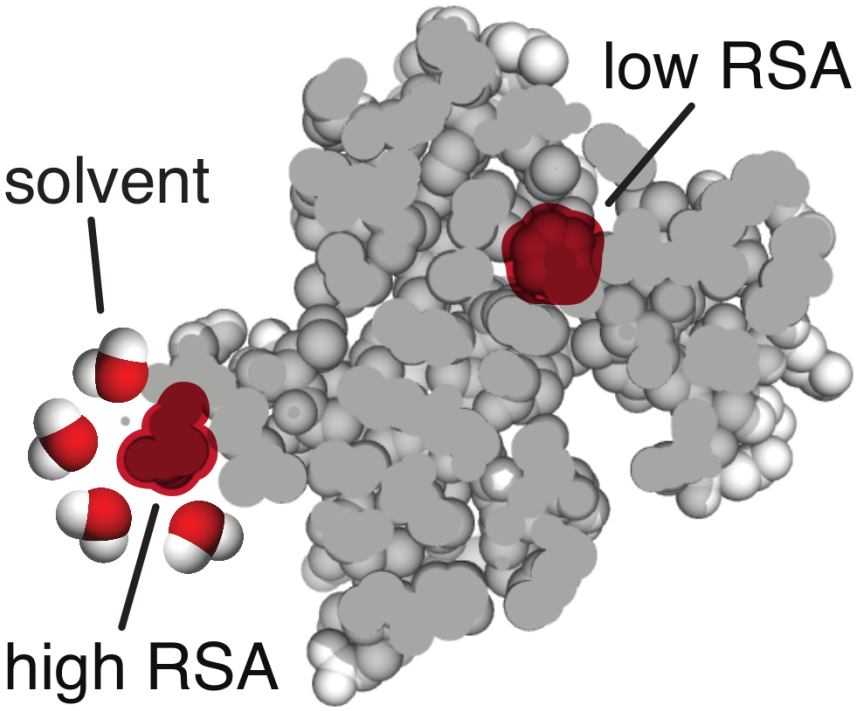
Patterns of sequence variation make more sense in a structural context



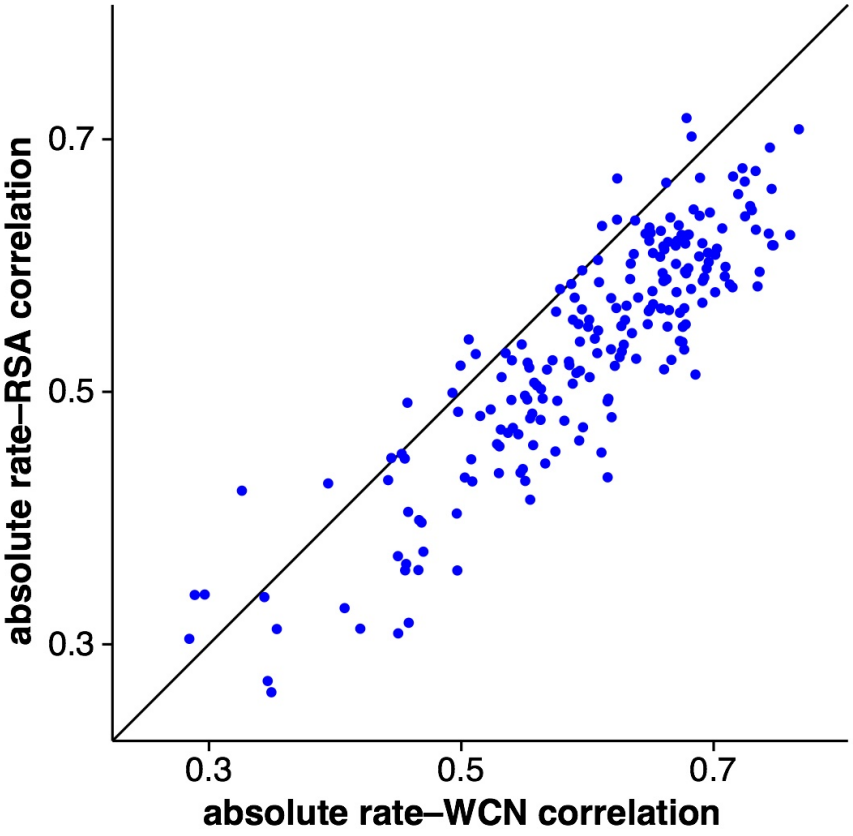
Echave et al.,
Nat. Rev. Genet. 2016

We measure structure with RSA and WCN

Relative Solvent Accessibility

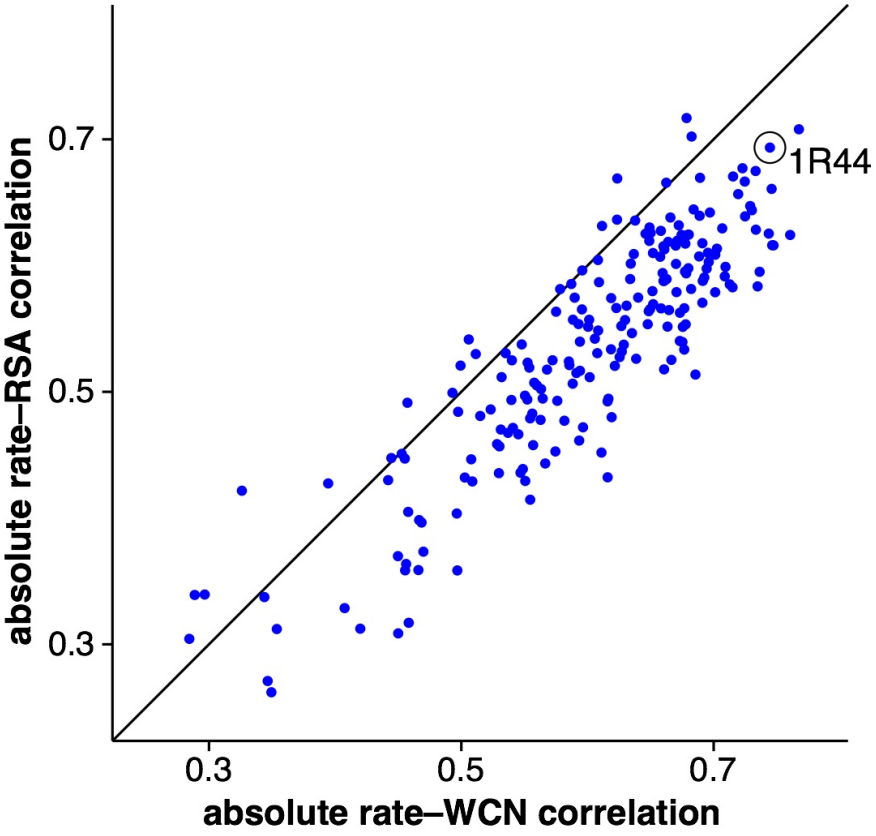


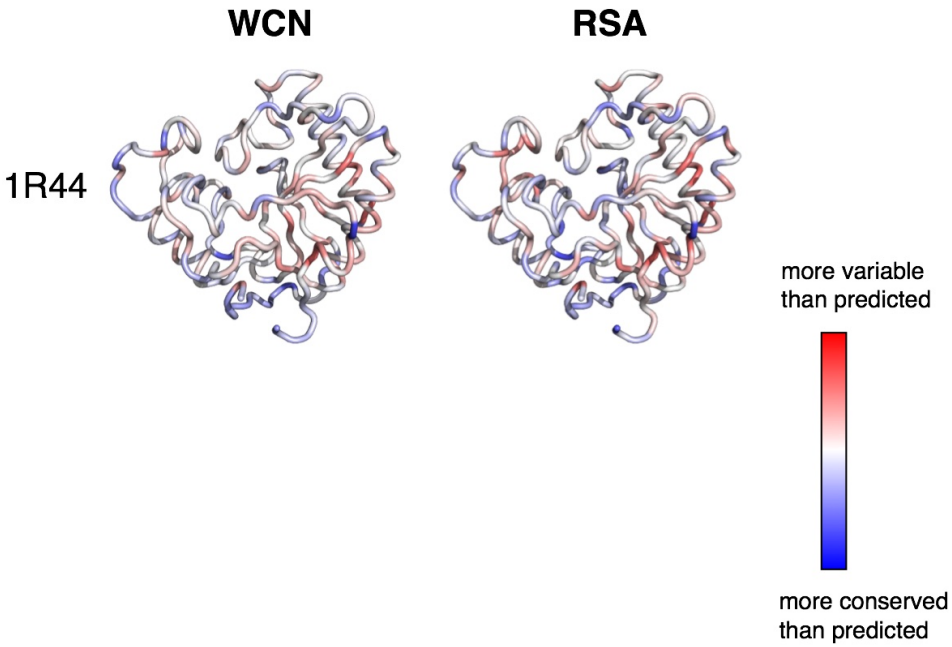
Structure is a good rate predictor for some proteins, but not so much for others



Echave et al., Nat. Rev. Genet. 2016

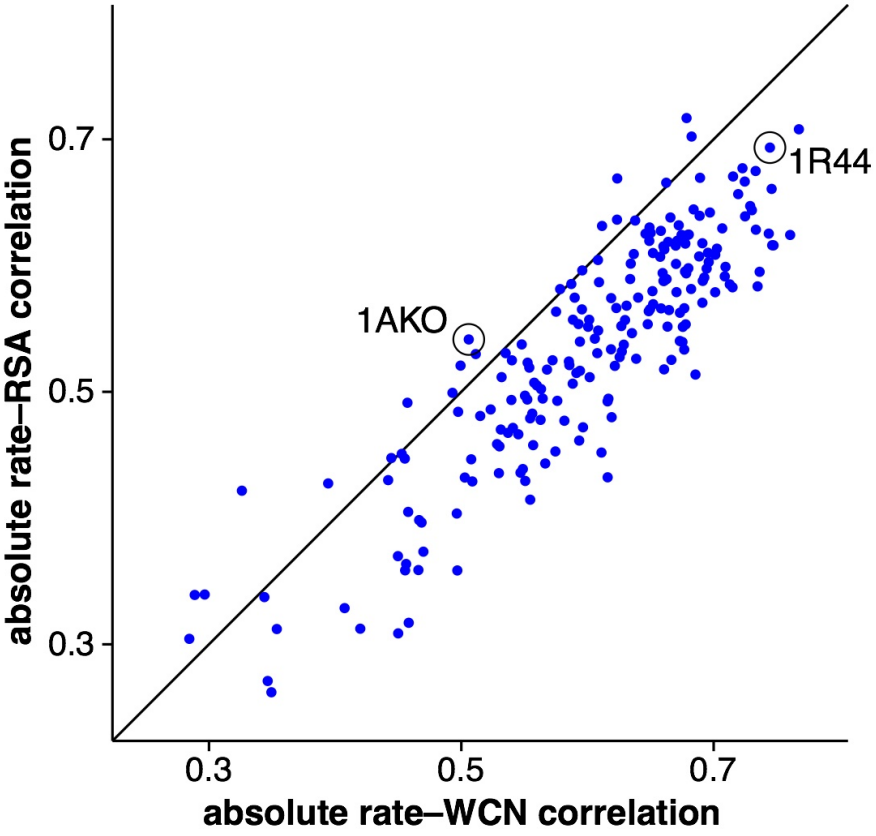
Structure is a good rate predictor for some proteins, but not so much for others

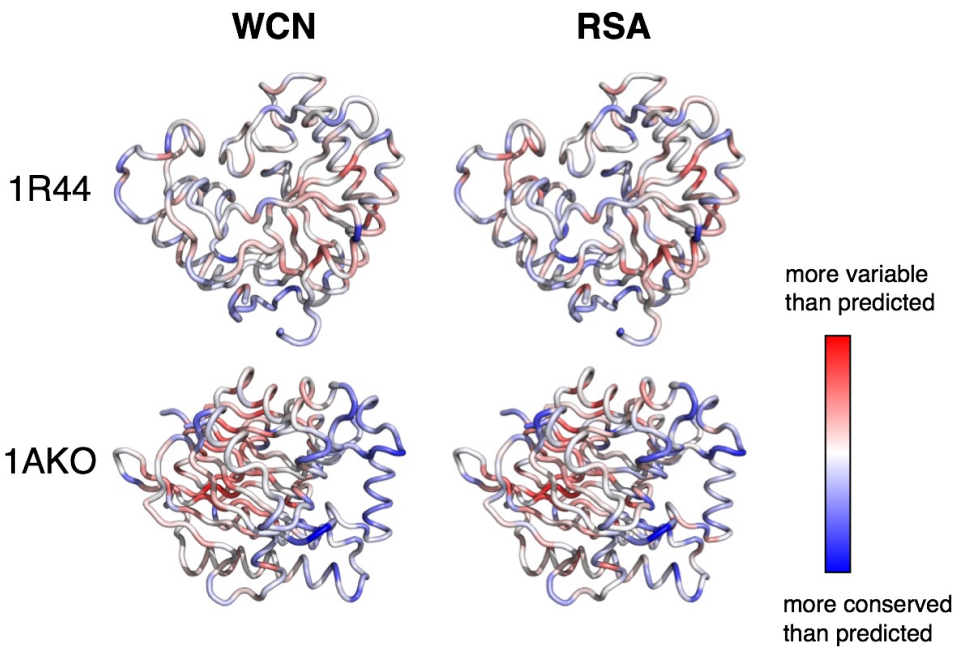




Echave et al., Nat. Rev. Genet. 2016

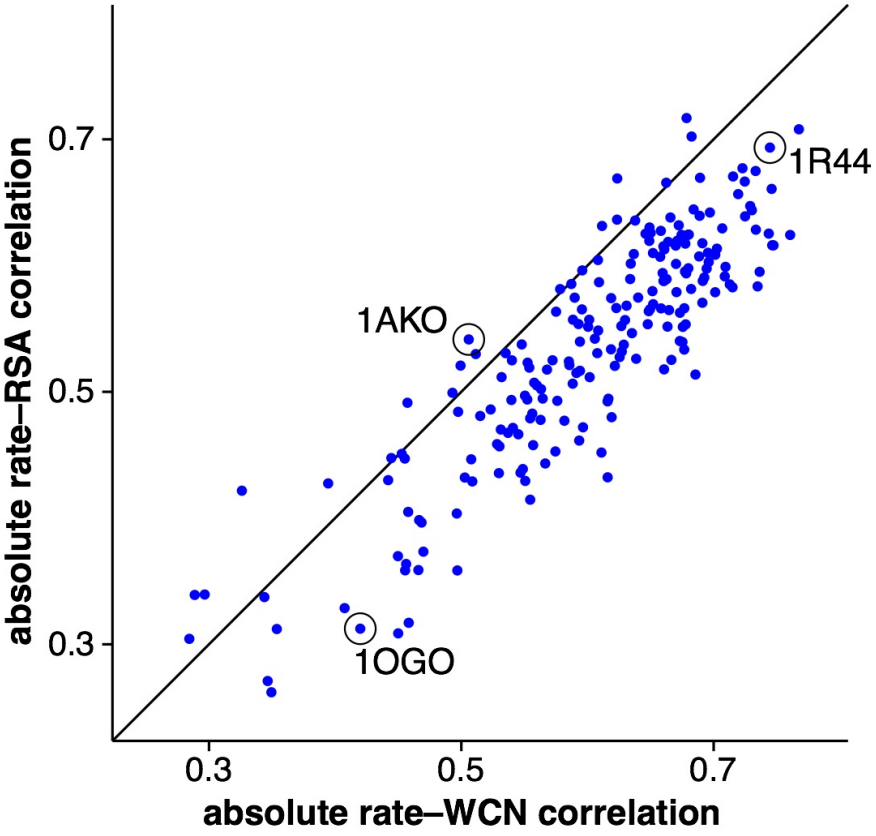
Structure is a good rate predictor for some proteins, but not so much for others

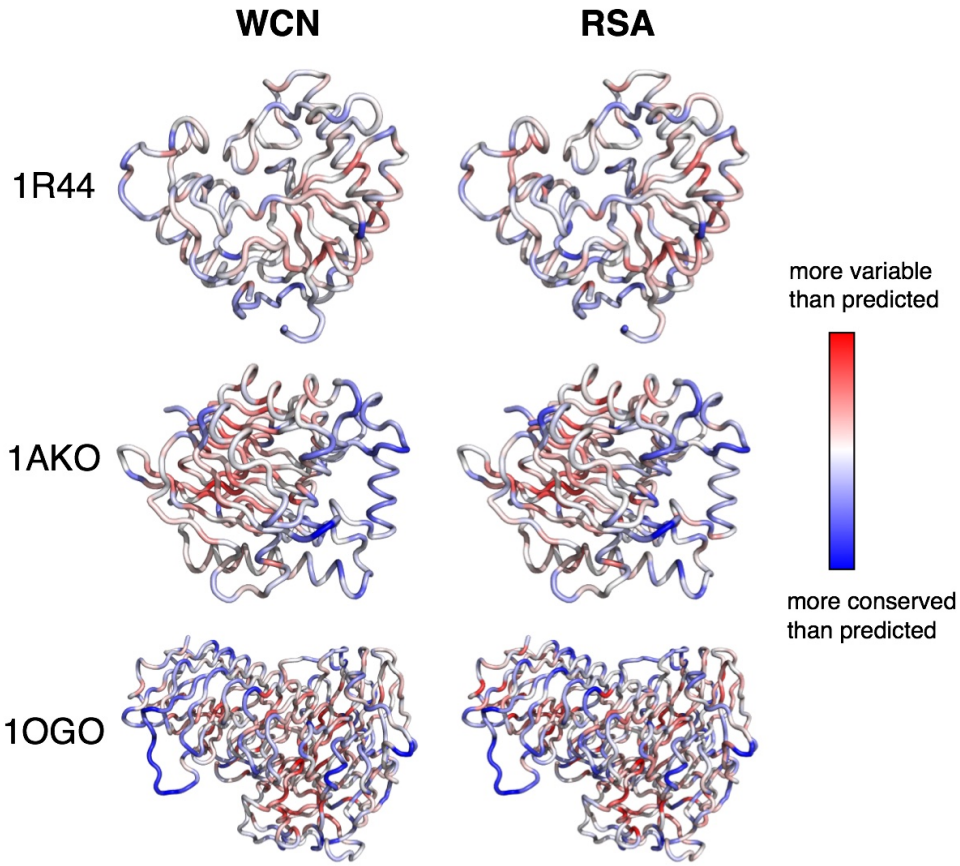




Echave et al., Nat. Rev. Genet. 2016

Structure is a good rate predictor for some proteins, but not so much for others





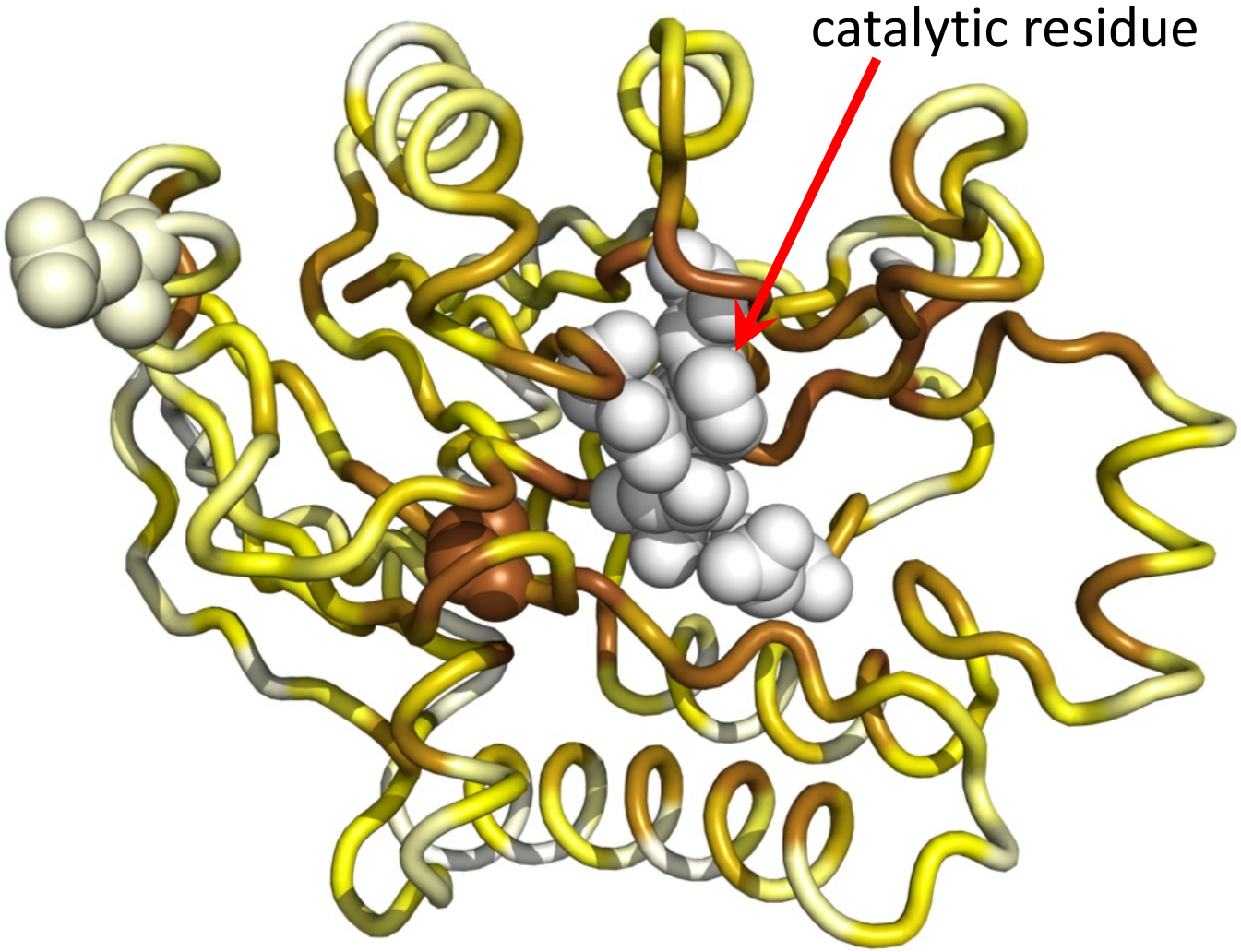
Echave et al., Nat. Rev. Genet. 2016

Part I: How does protein function constrain site-specific evolution?

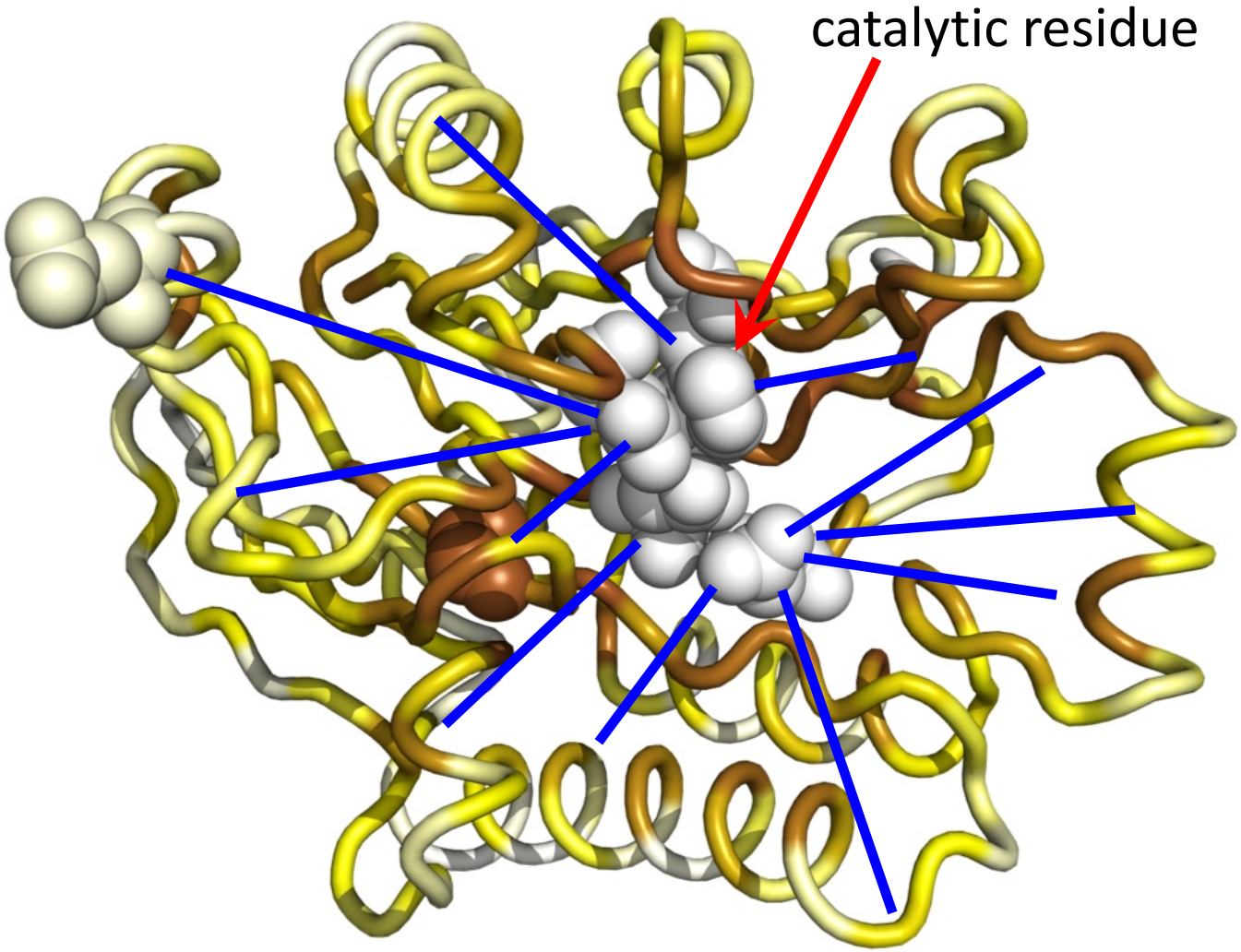
Work by graduate student Ben Jack

In collaboration with Julian Echave

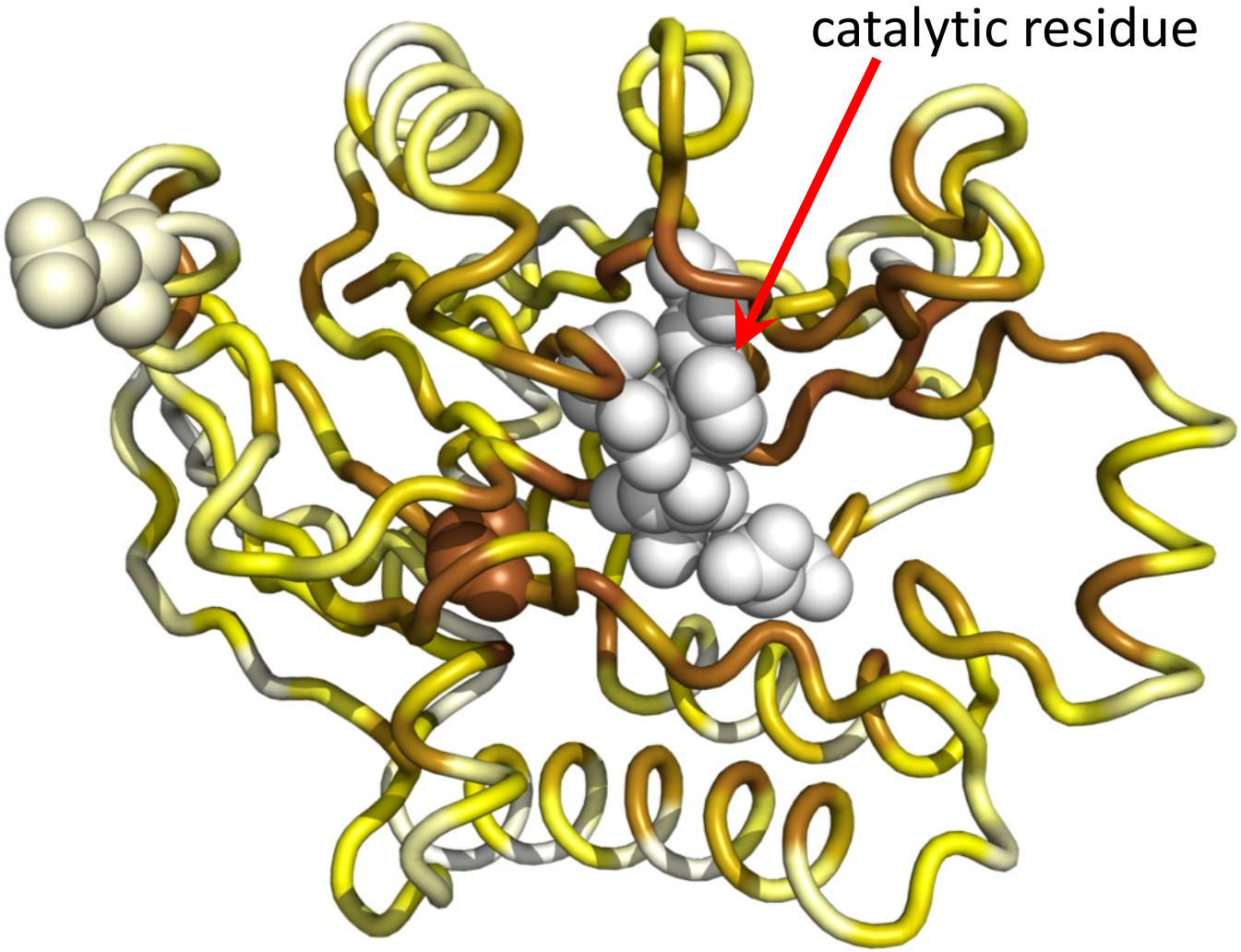
We measure function with distance to the nearest catalytic residue



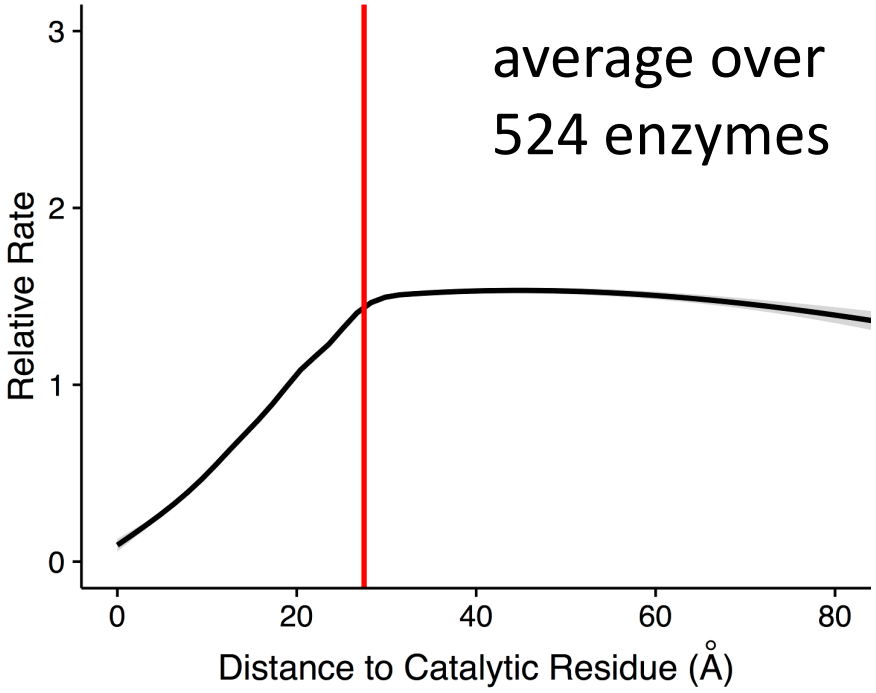
We measure function with distance to the nearest catalytic residue



We measure function with distance to the nearest catalytic residue

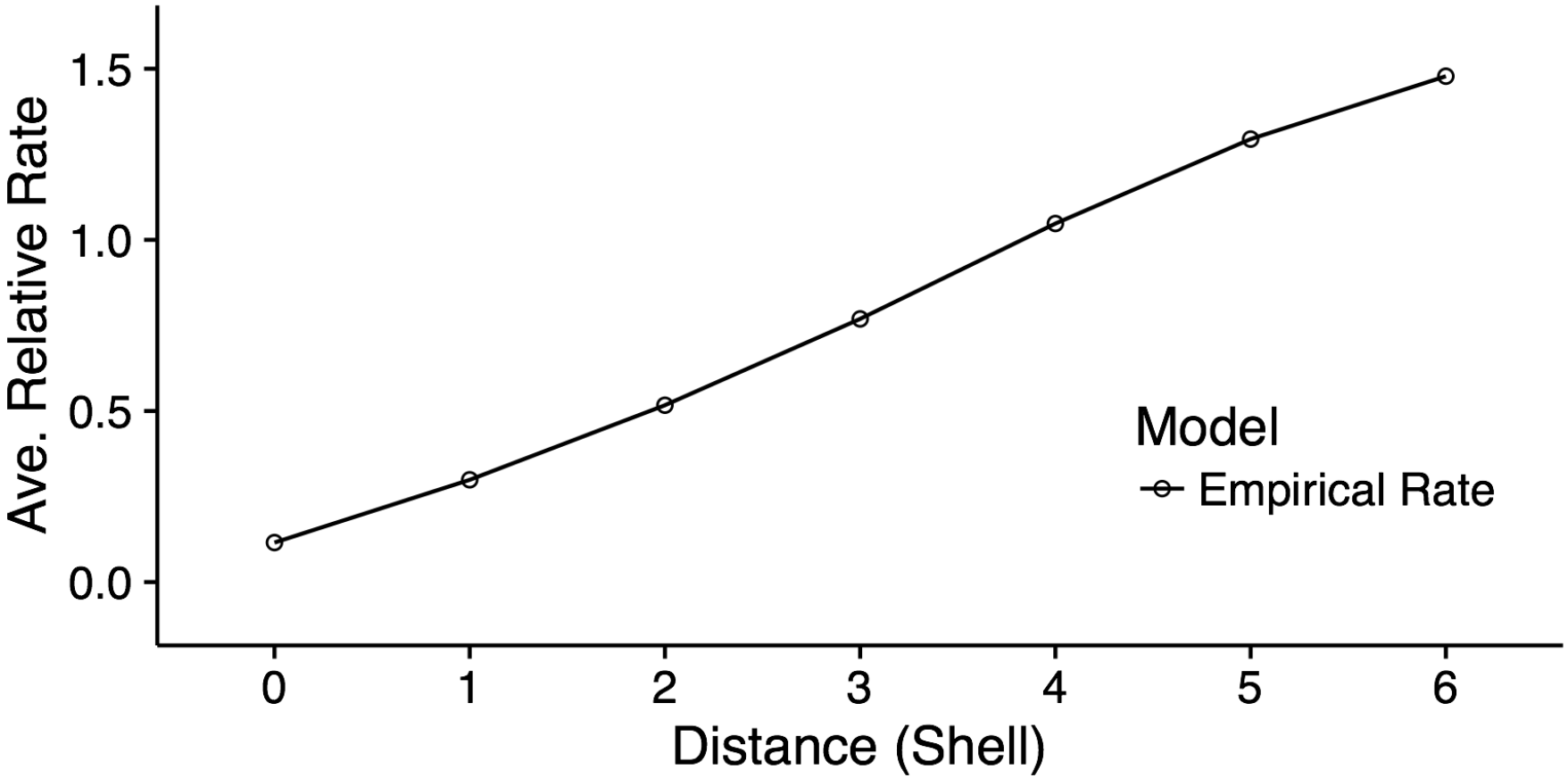


Catalytic residues impose long-range conservation gradients



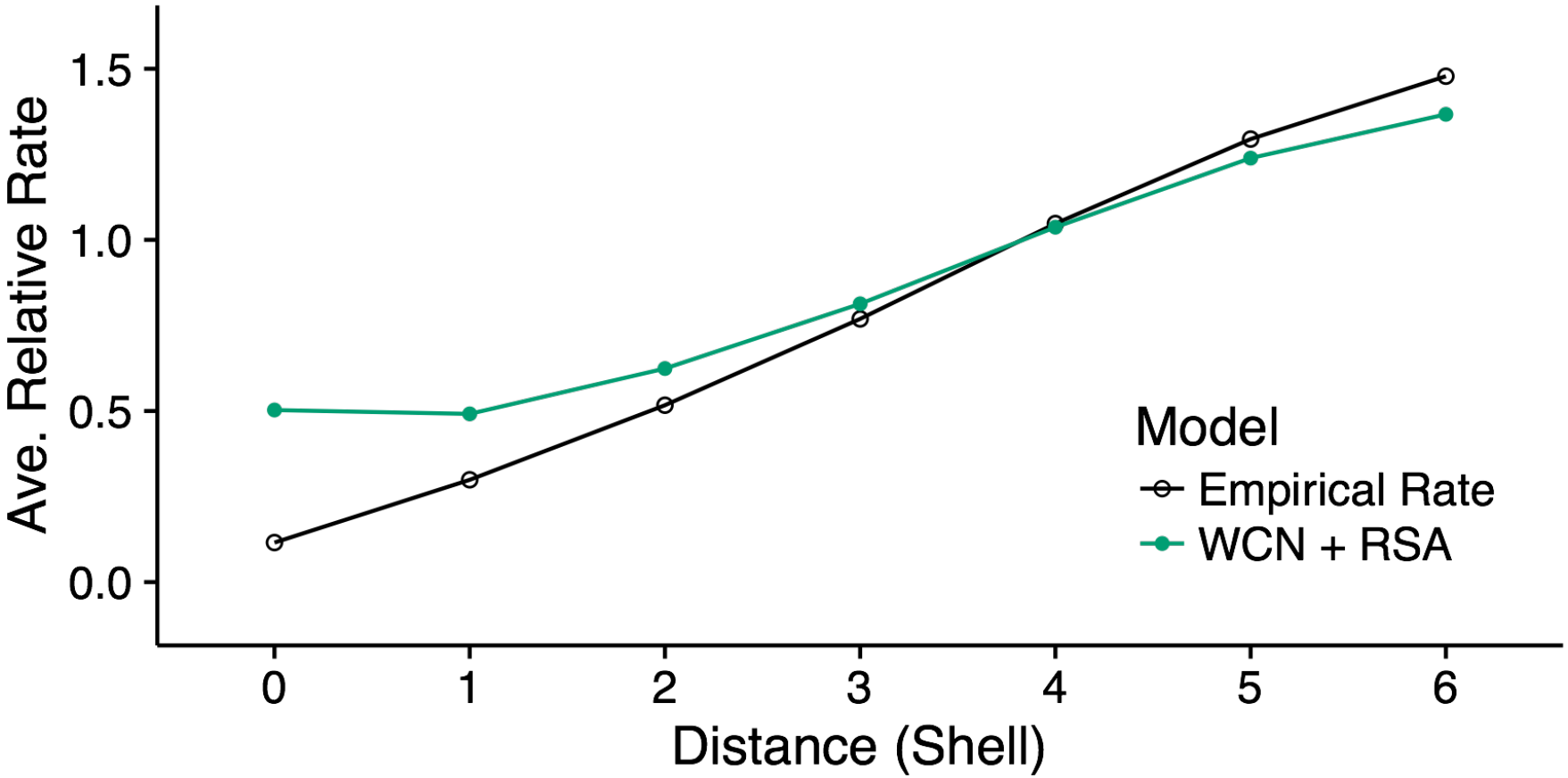
Jack et al., PLOS Biol. 2016

Distance is critical to accurately predict average rate near catalytic residues



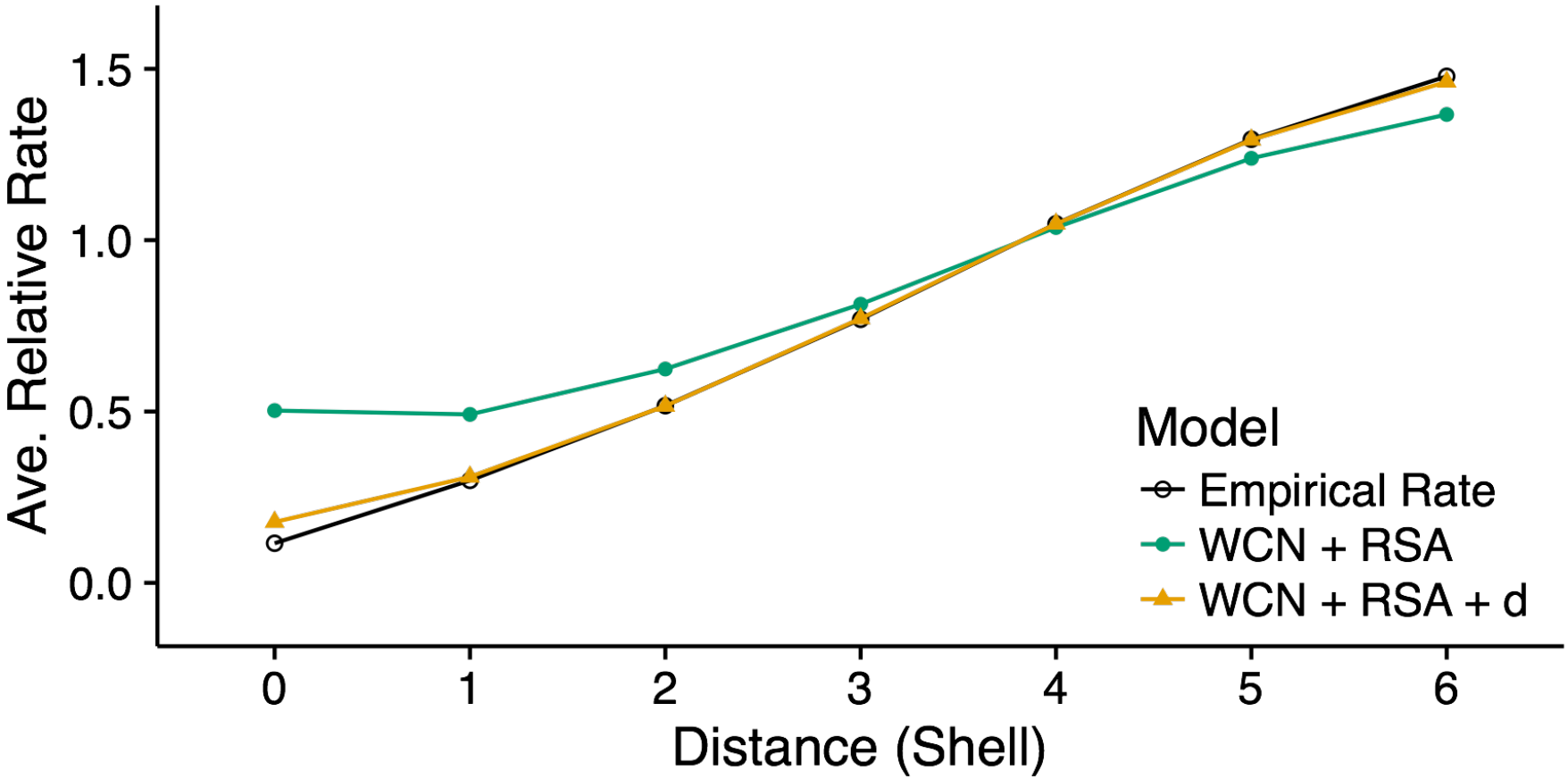
Jack et al., PLOS Biol. 2016

Distance is critical to accurately predict average rate near catalytic residues



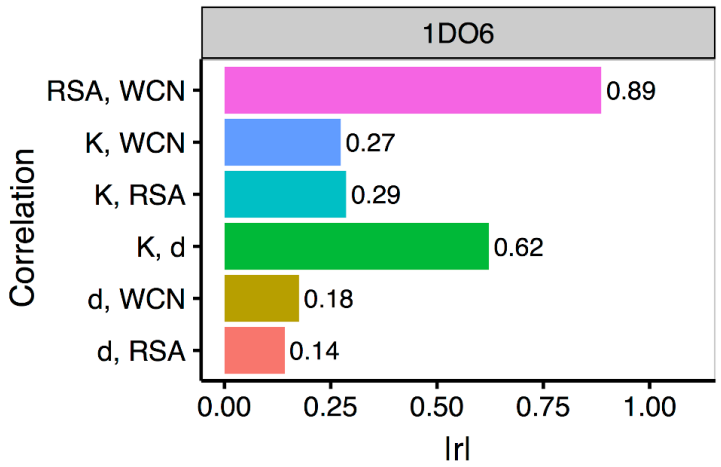
Jack et al., PLOS Biol. 2016

Distance is critical to accurately predict average rate near catalytic residues

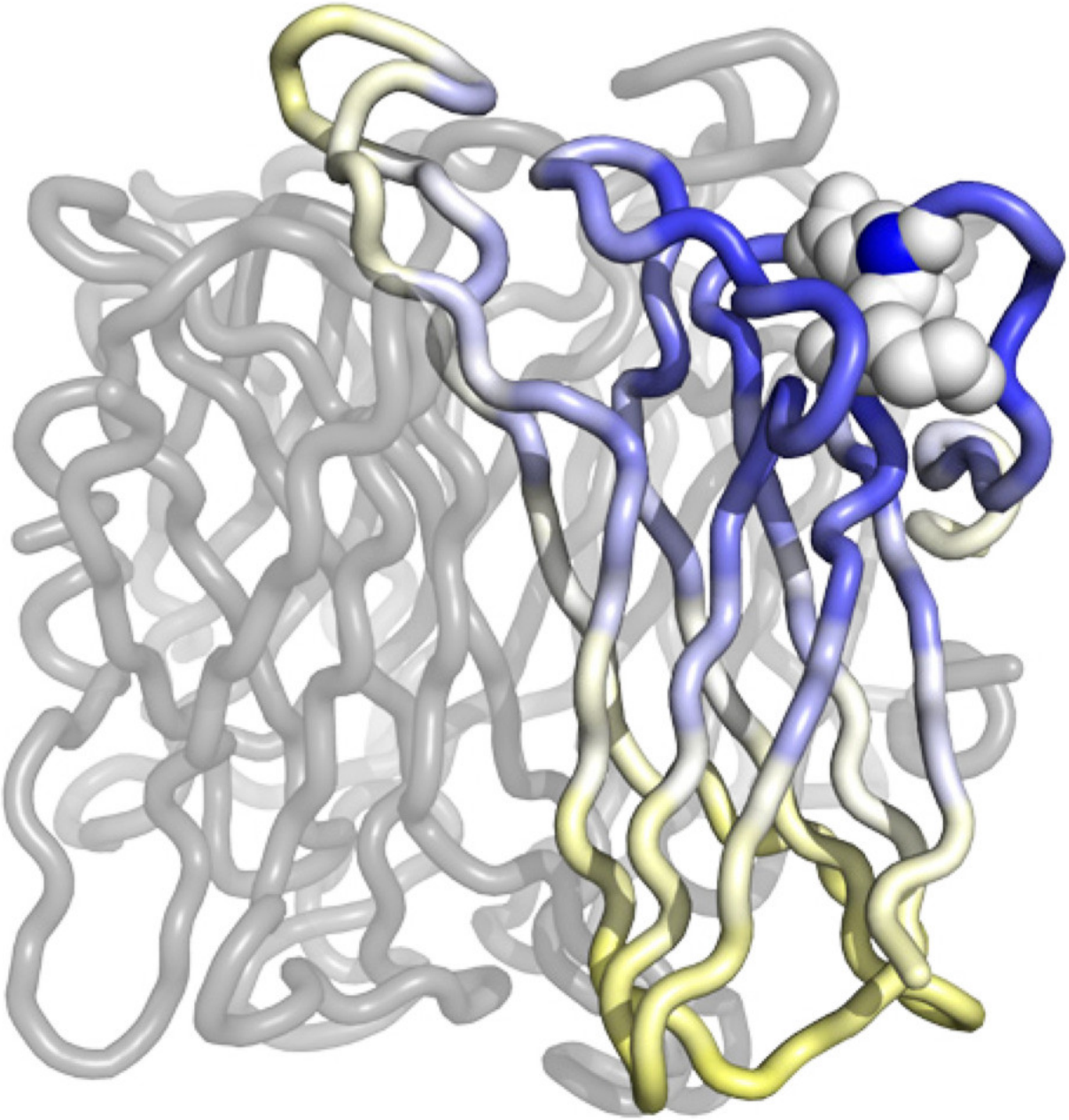


Jack et al., PLOS Biol. 2016

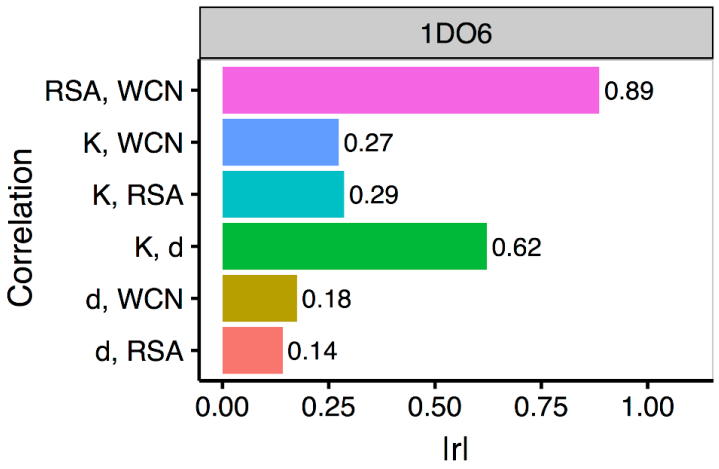
Effect is strongest for active sites on the protein surface

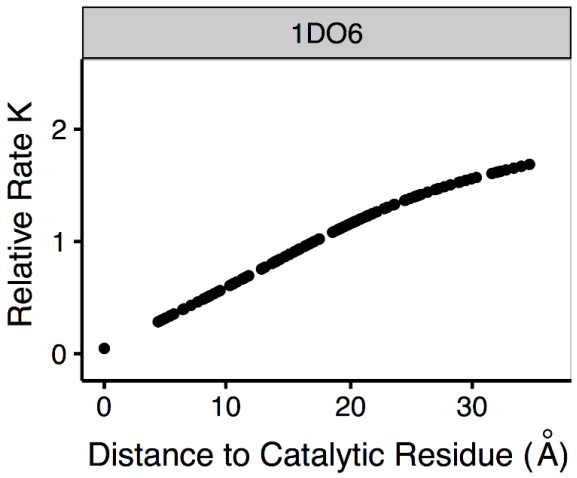


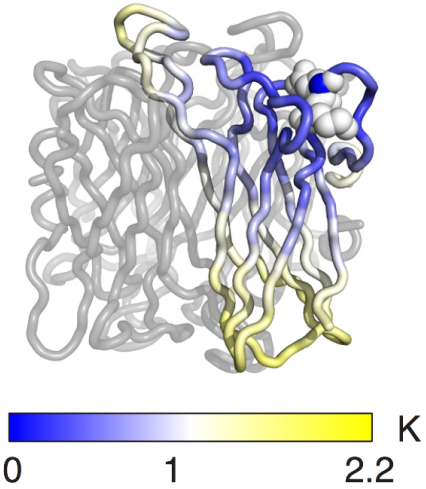
Jack et al., PLOS Biol. 2016

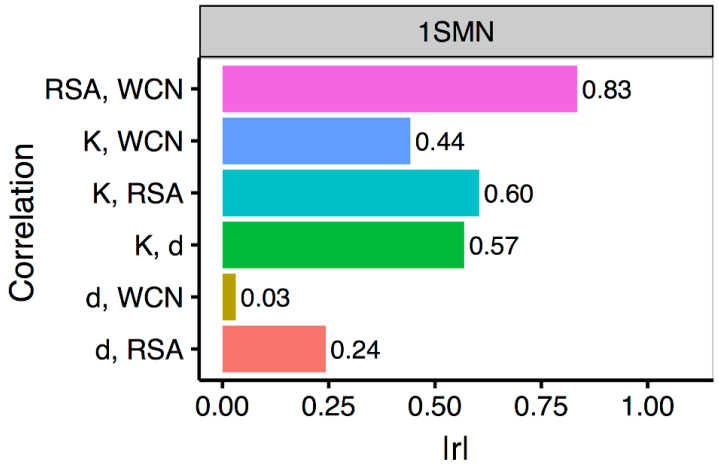


Effect is strongest for active sites on the protein surface

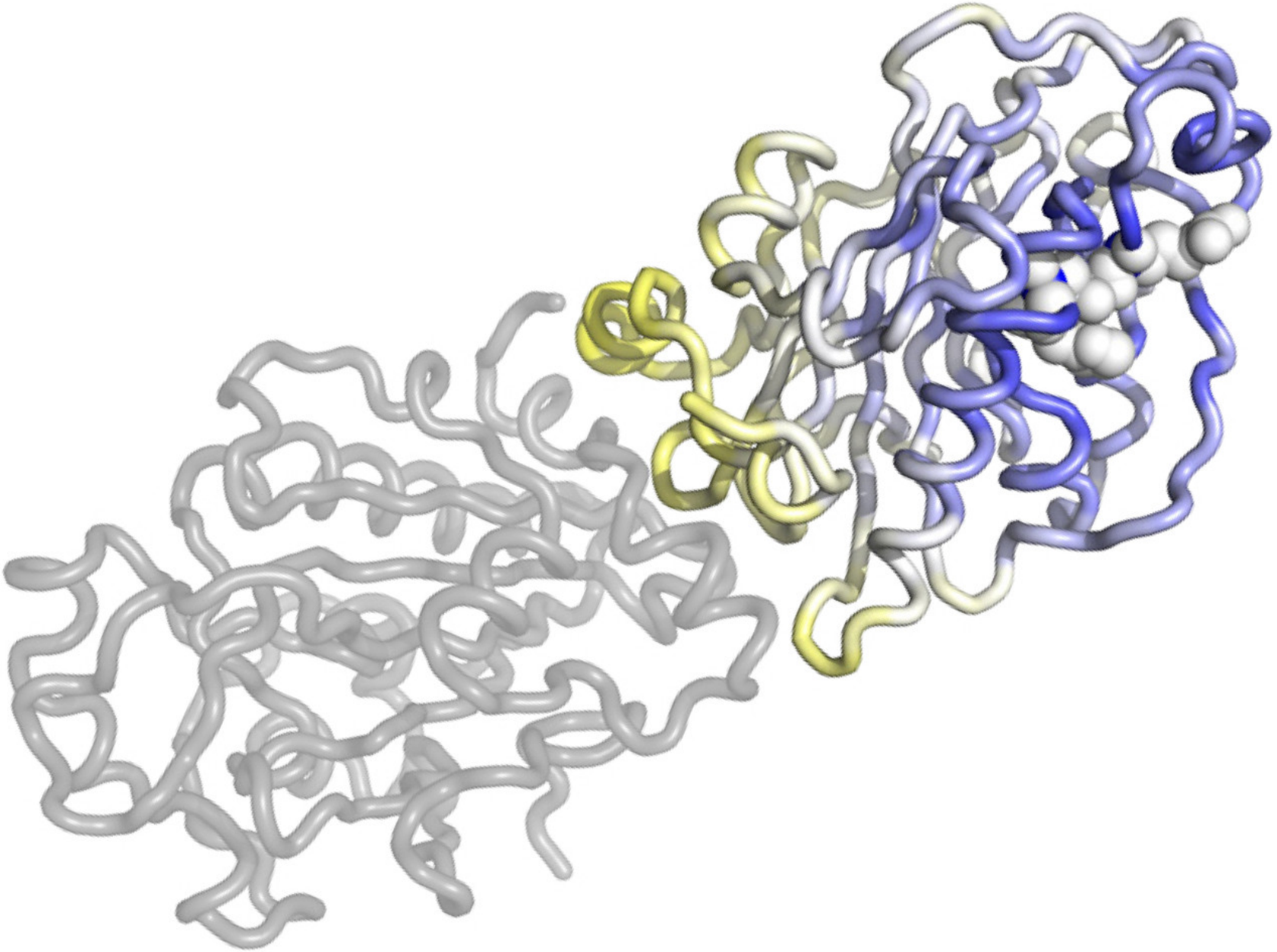




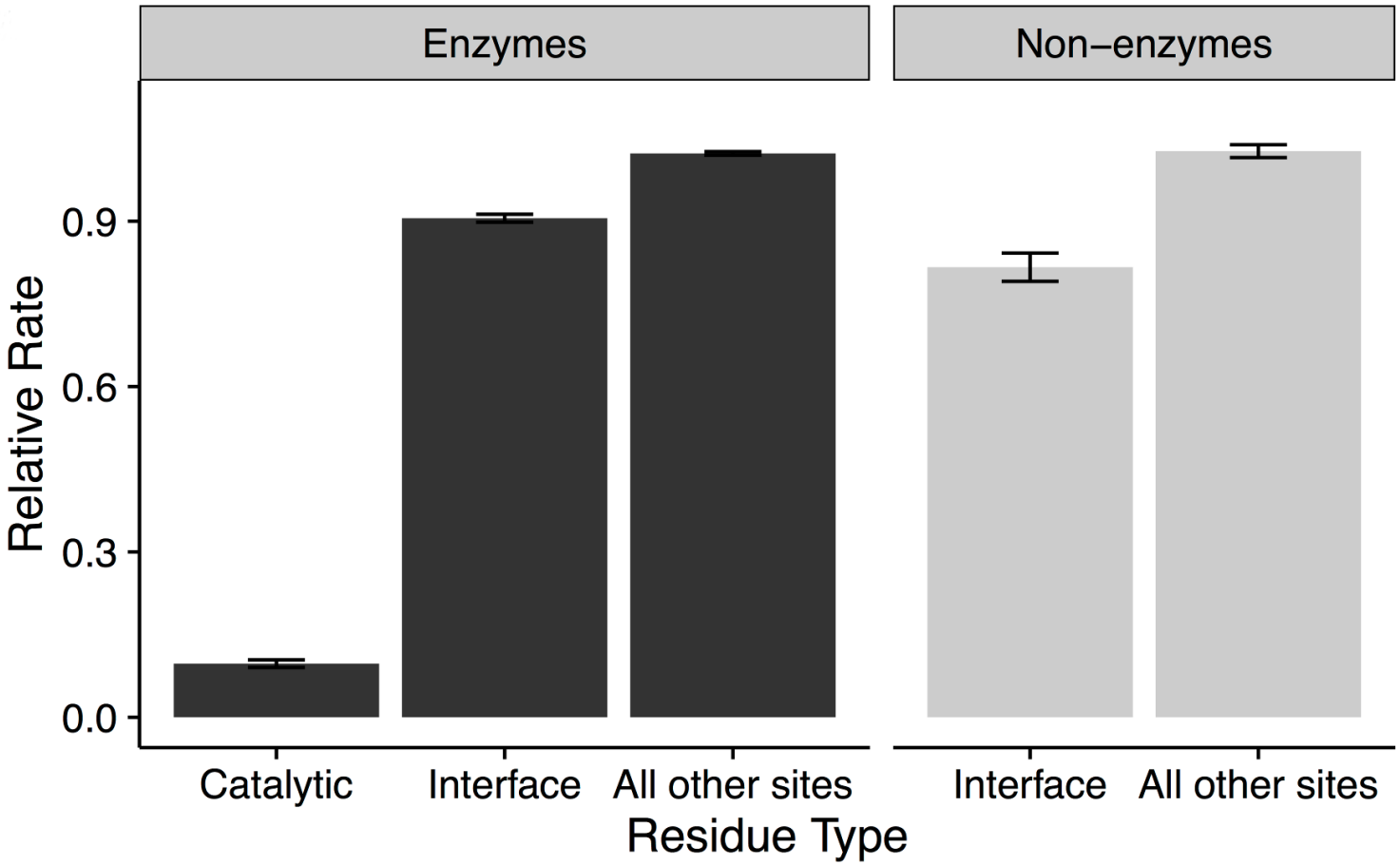




Jack et al., PLOS Biol. 2016



Catalytic residues impose much stronger constraints than interface residues



Jack et al., PLOS Biol. 2016

Take-home message

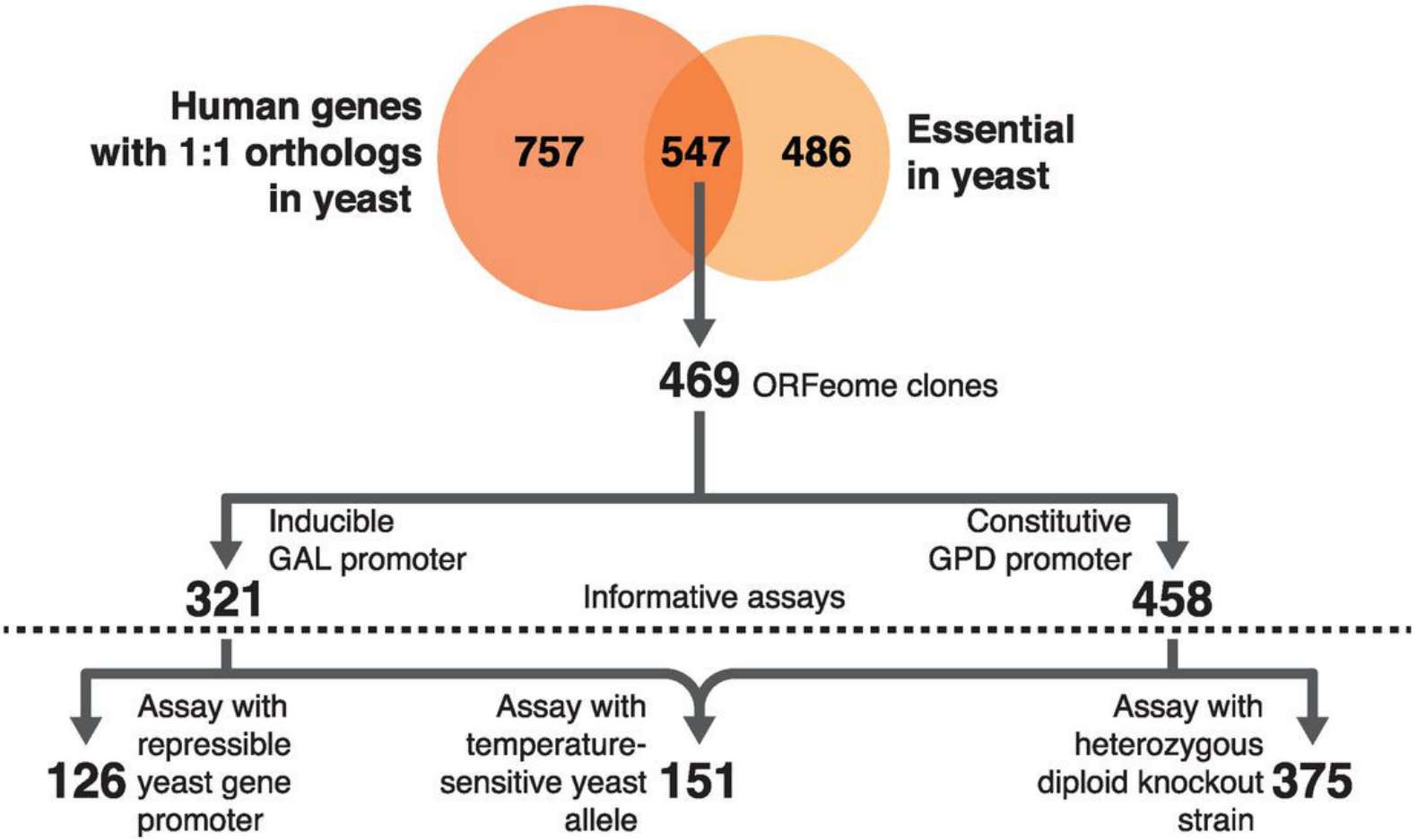
- Catalytic residues generate long-range selection gradients covering most of a typical enzyme structure

Part II: How does evolution constrain protein–protein interfaces?

In collaboration with the Marcotte lab

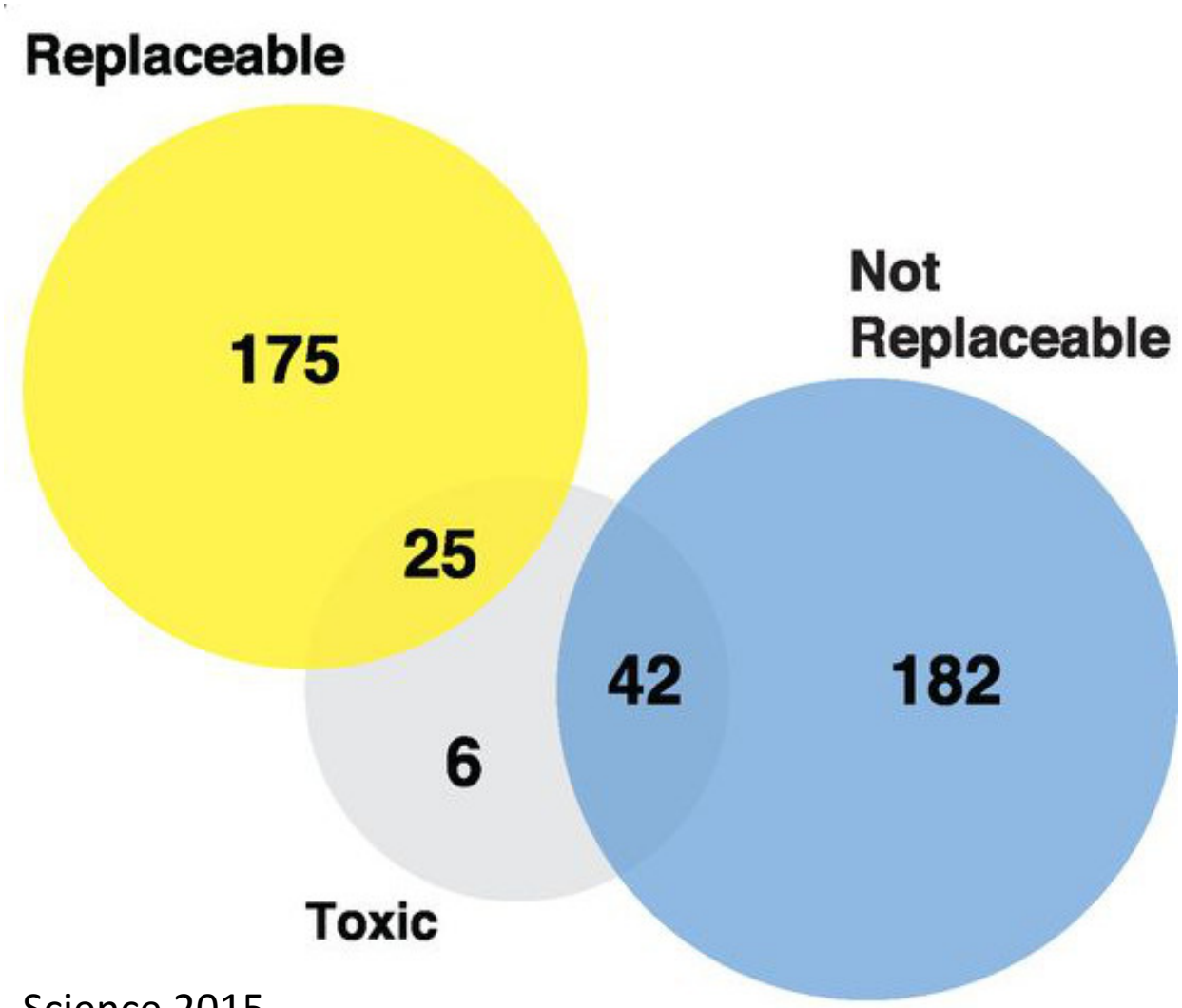
Protein evolution simulations by Austin Meyer

Can human genes function in a distant relative (baker's yeast)?



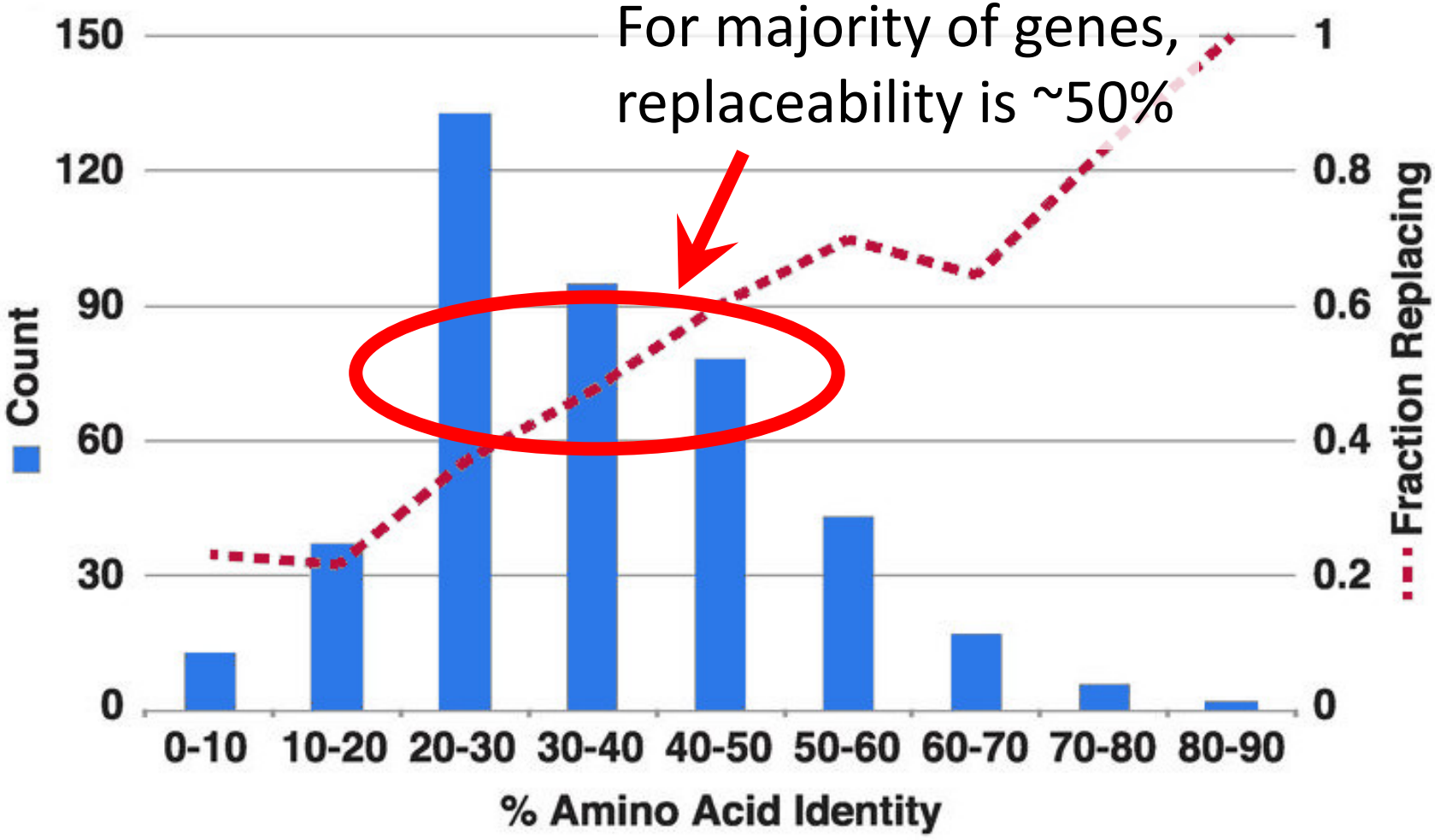
Kachroo et al., Science 2015

Yes!



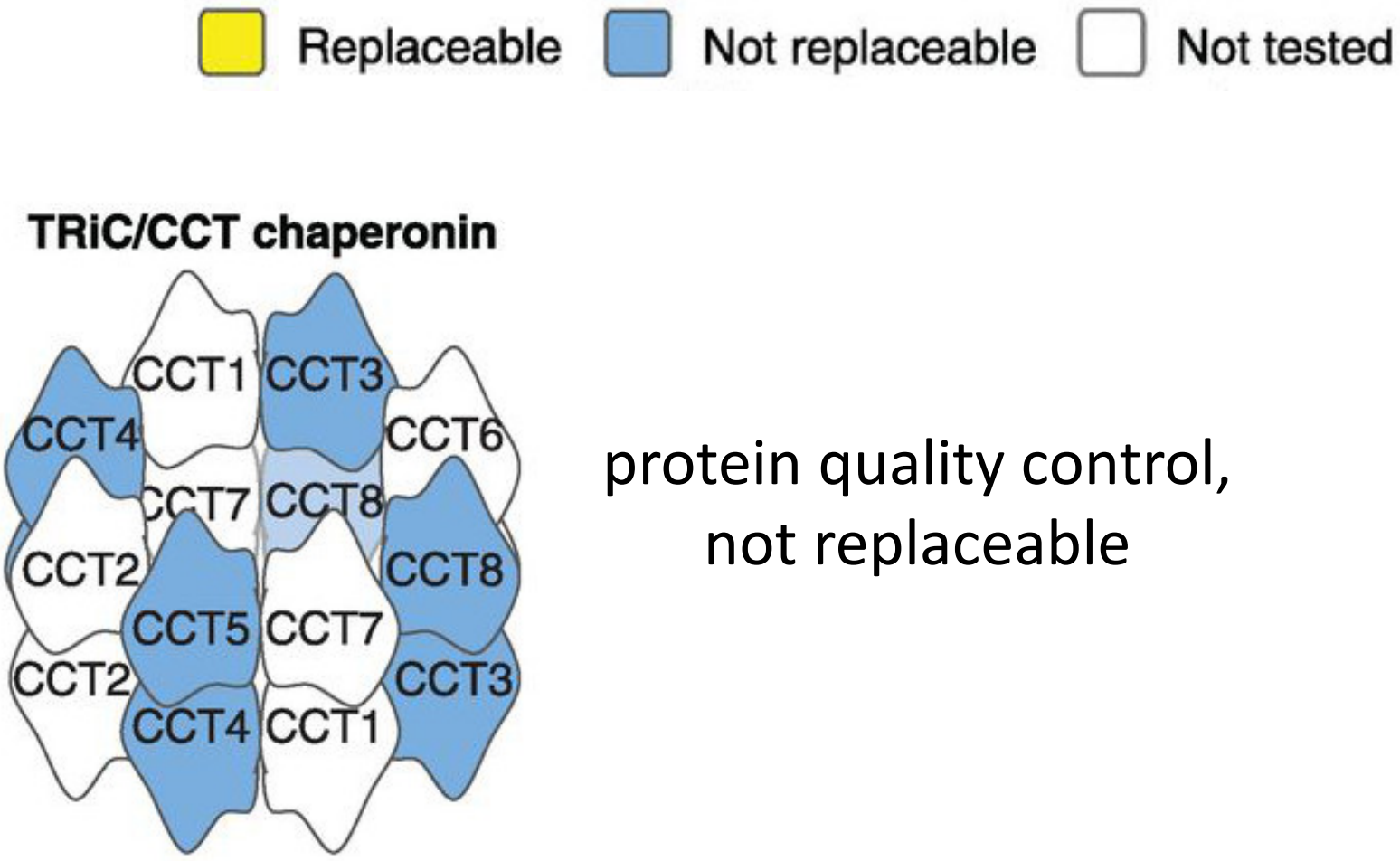
Kachroo et al., Science 2015

Sequence divergence is not a good predictor of replaceability



Kachroo et al., Science 2015

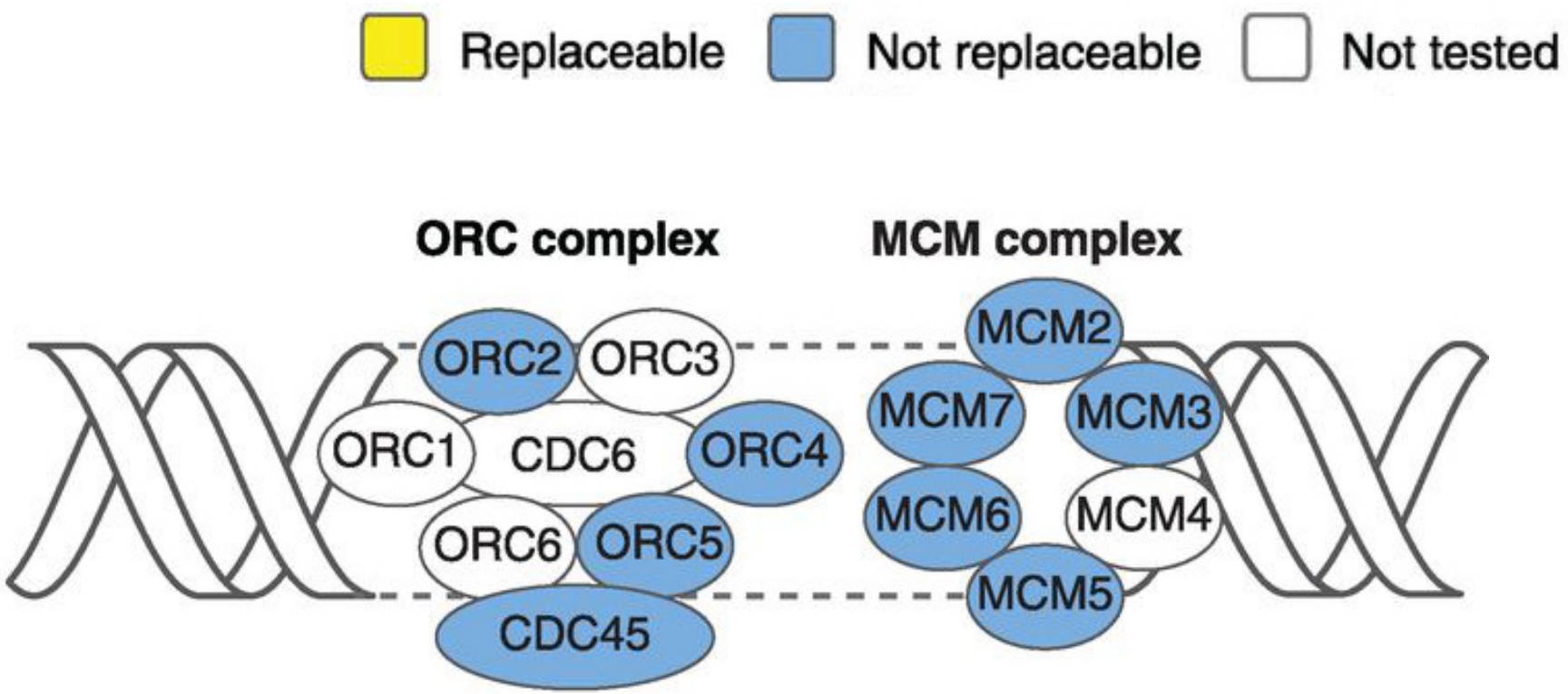
Genes tend to be replaceable in modules



protein quality control,
not replaceable

Kachroo et al., Science 2015

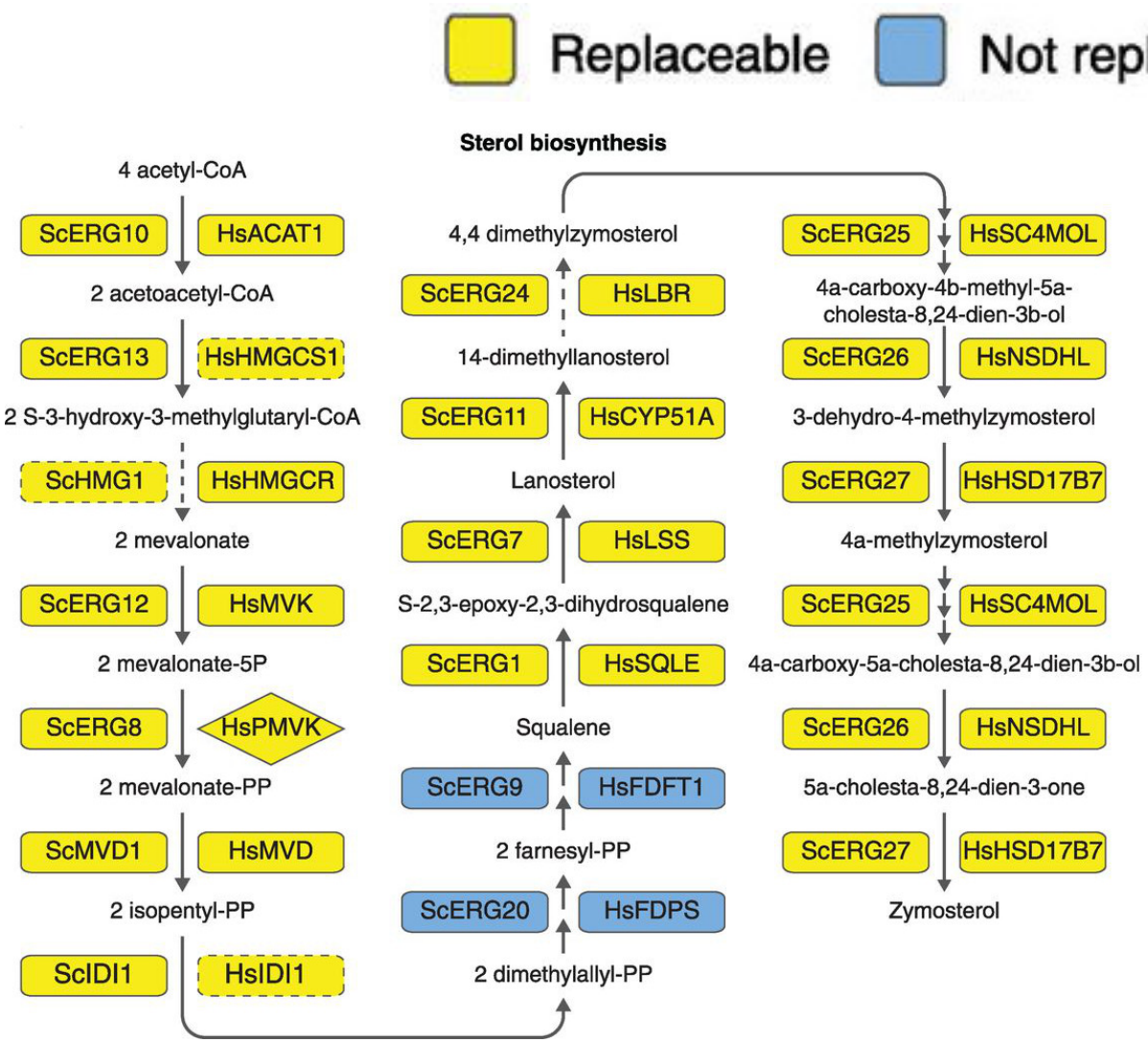
Genes tend to be replaceable in modules



genome replication,
not replaceable

Kachroo et al., Science 2015

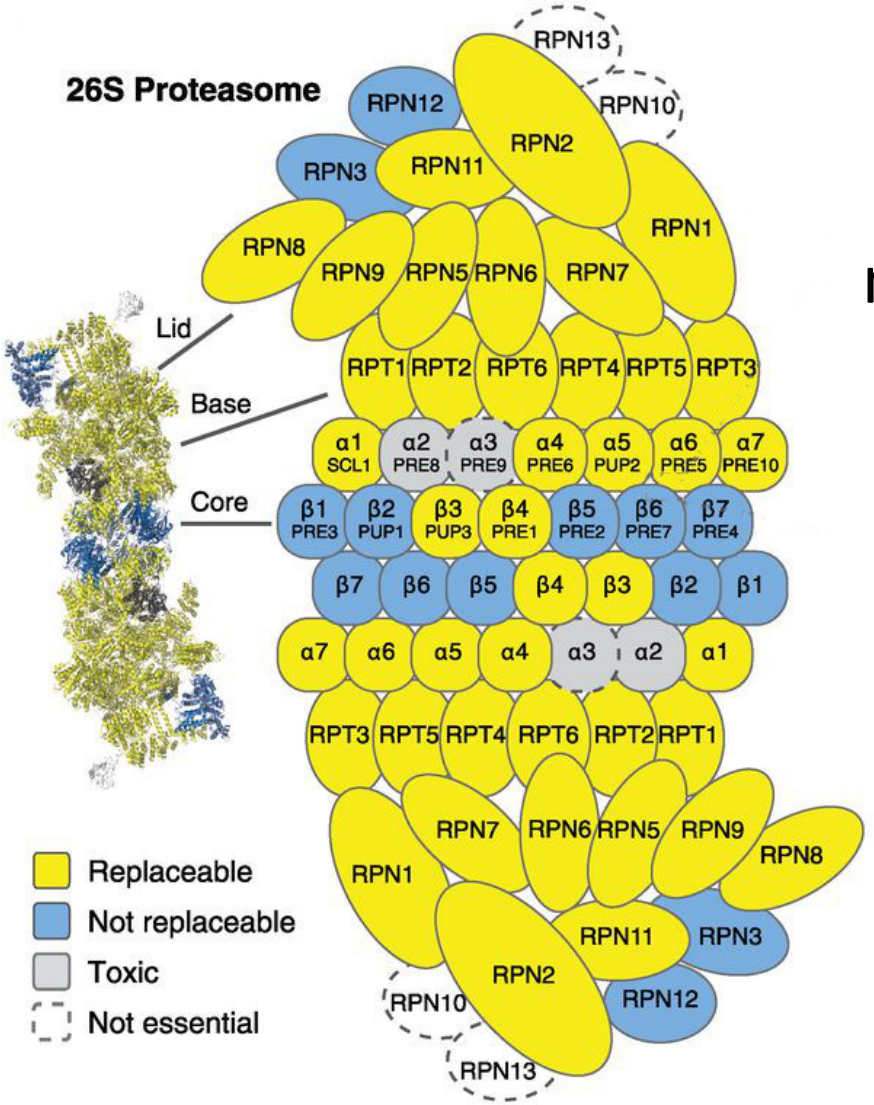
Genes tend to be replaceable in modules



sterol biosynthesis, replaceable

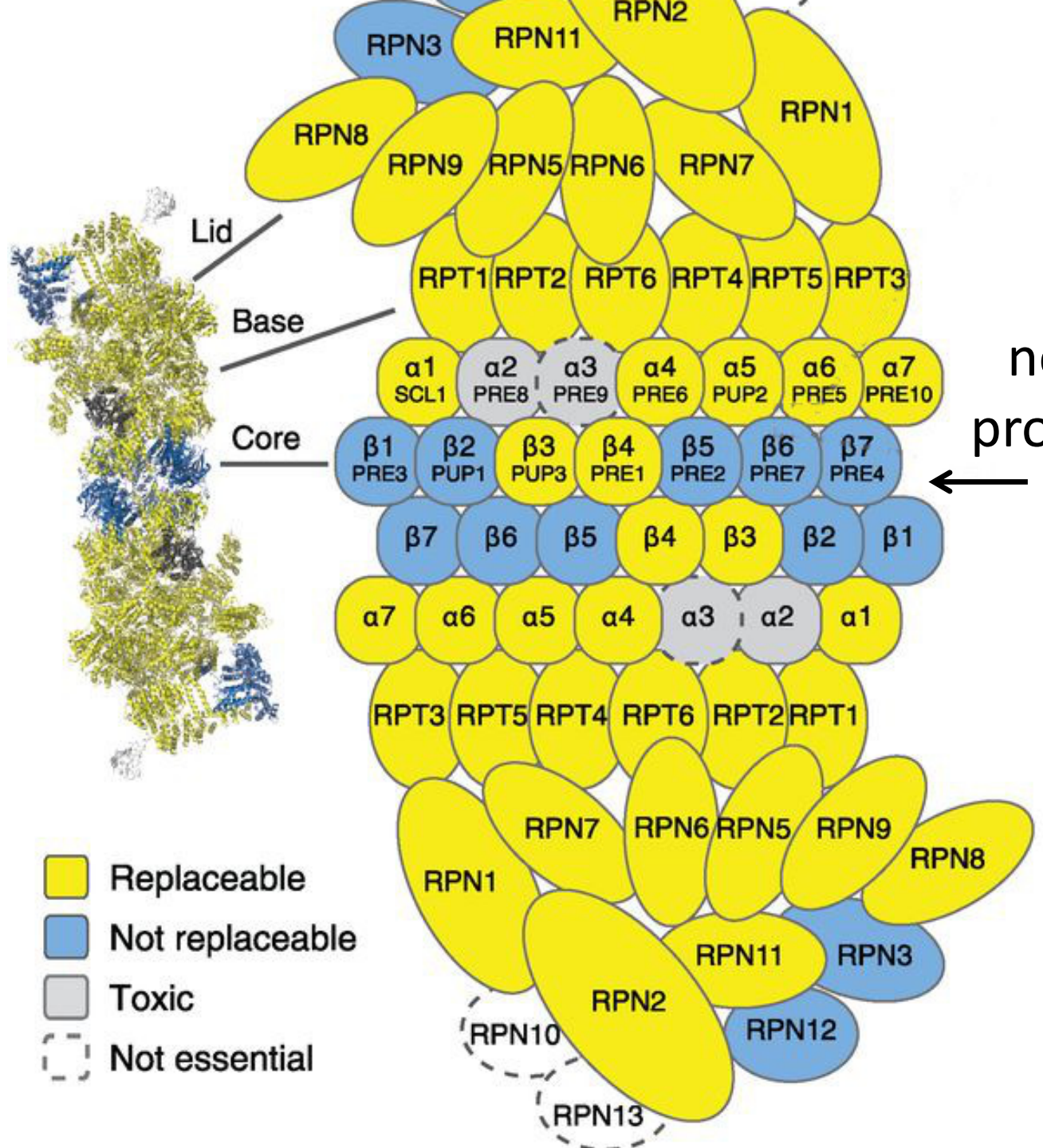
Kachroo et al., Science 2015

Genes tend to be replaceable in modules



proteasome,
mostly replaceable

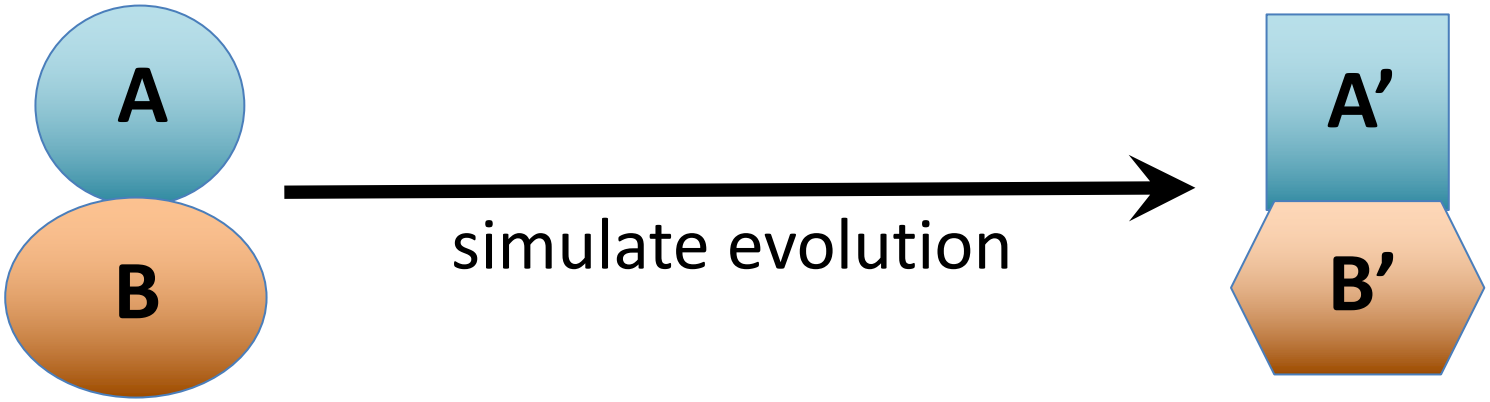
Kachroo et al., Science 2015

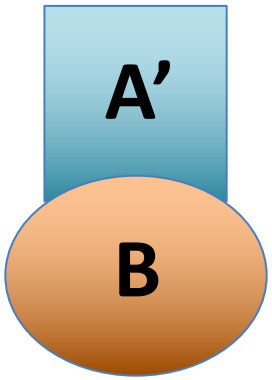


non-replaceable proteins are mostly in the β ring

Kachroo et al., Science 2015

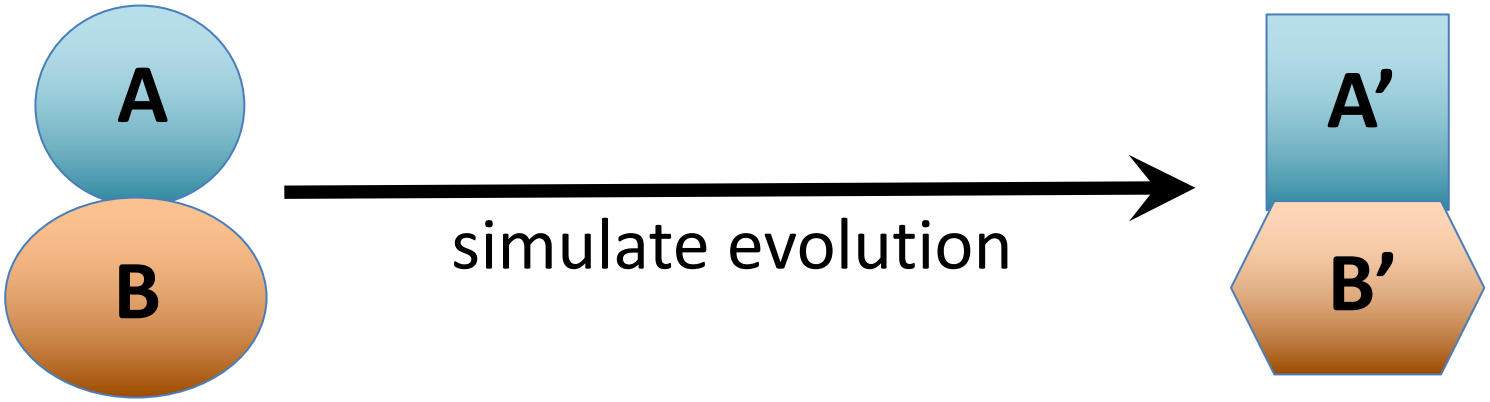
Hypothesis: Conserved protein–protein interactions cause replaceability

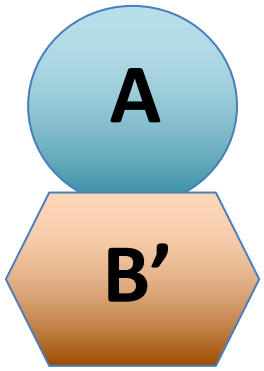




Does evolved A’ bind to ancestral B?

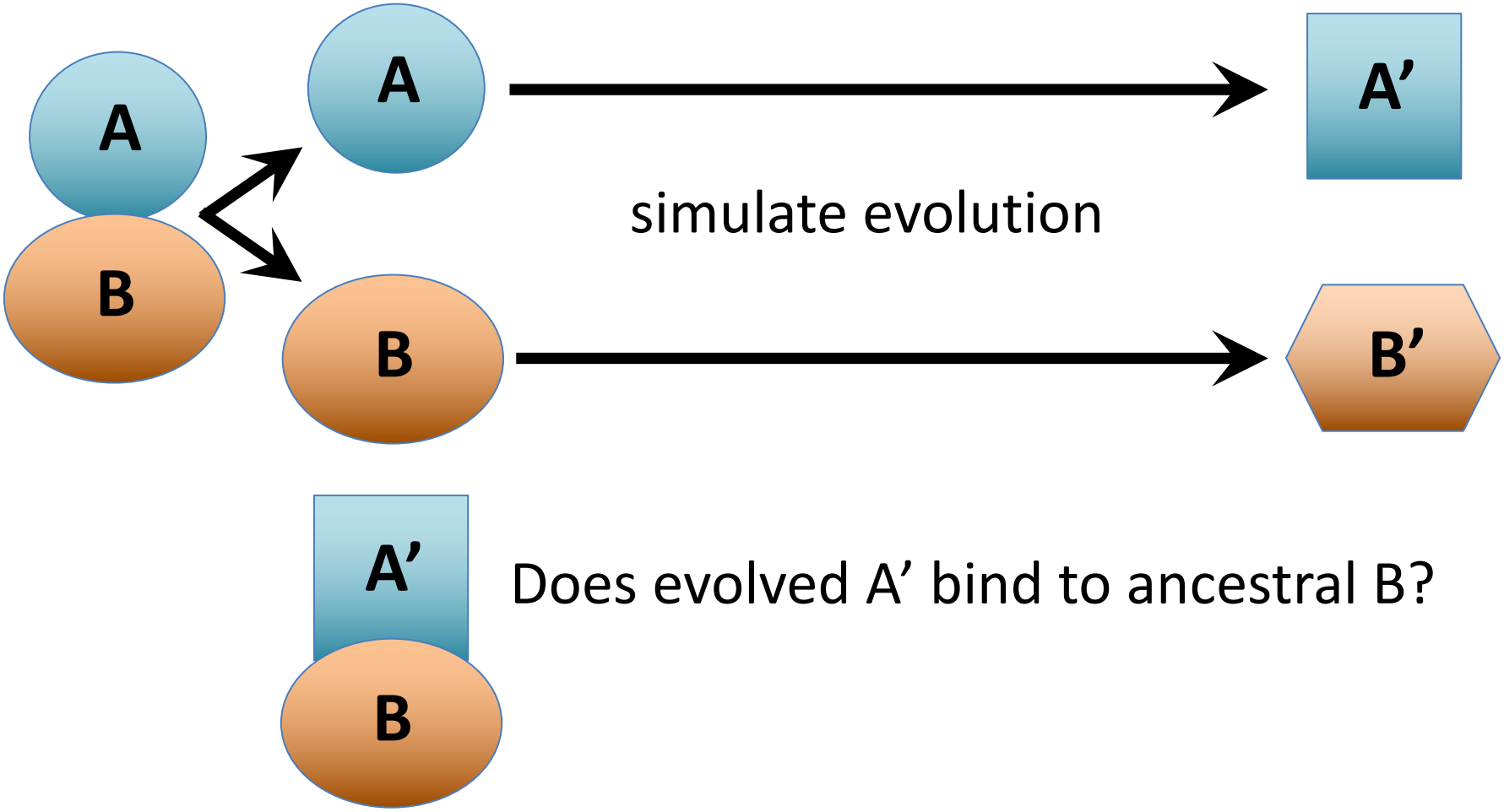
Hypothesis: Conserved protein–protein interactions cause replaceability



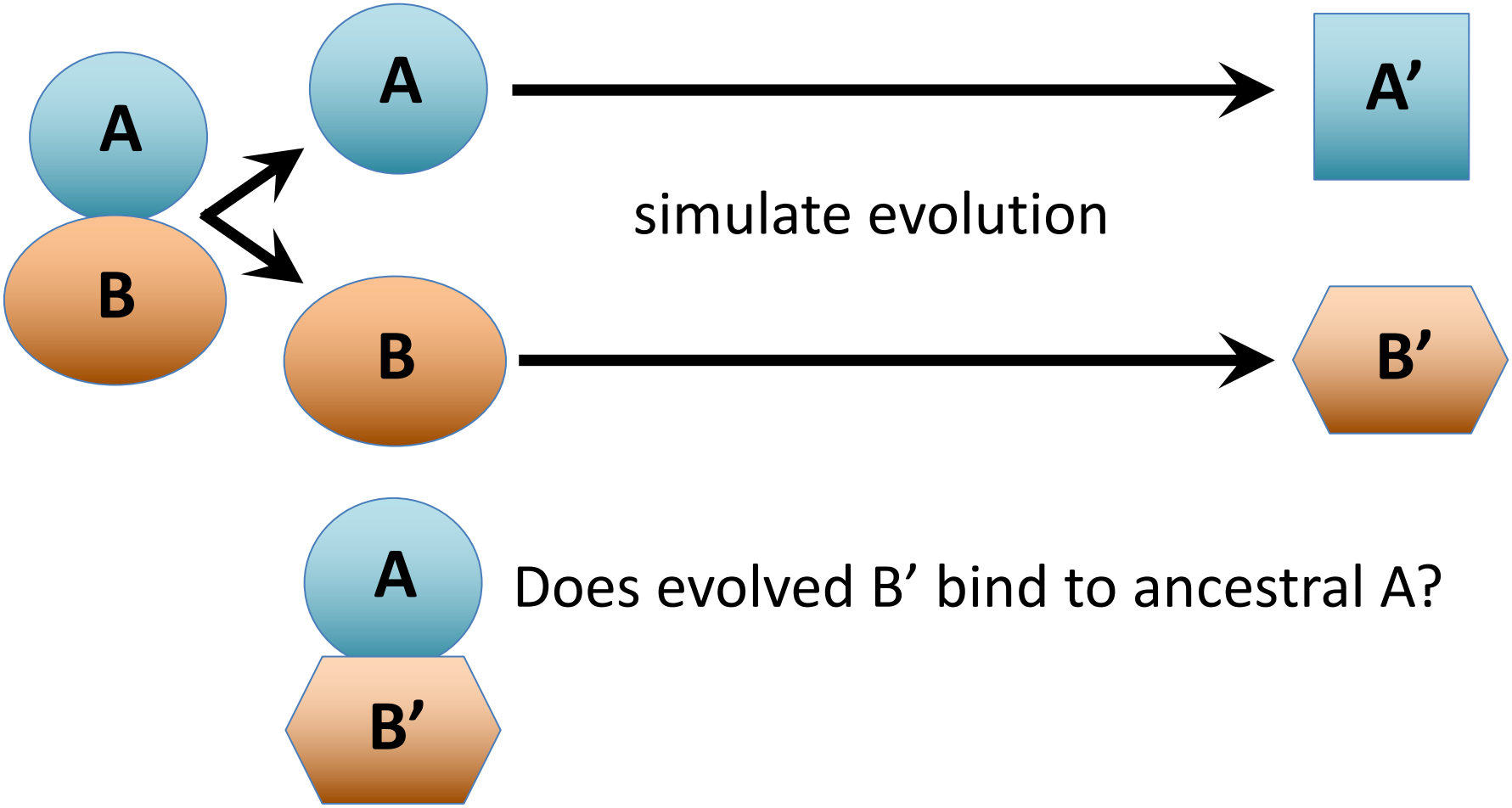


Does evolved B’ bind to ancestral A?

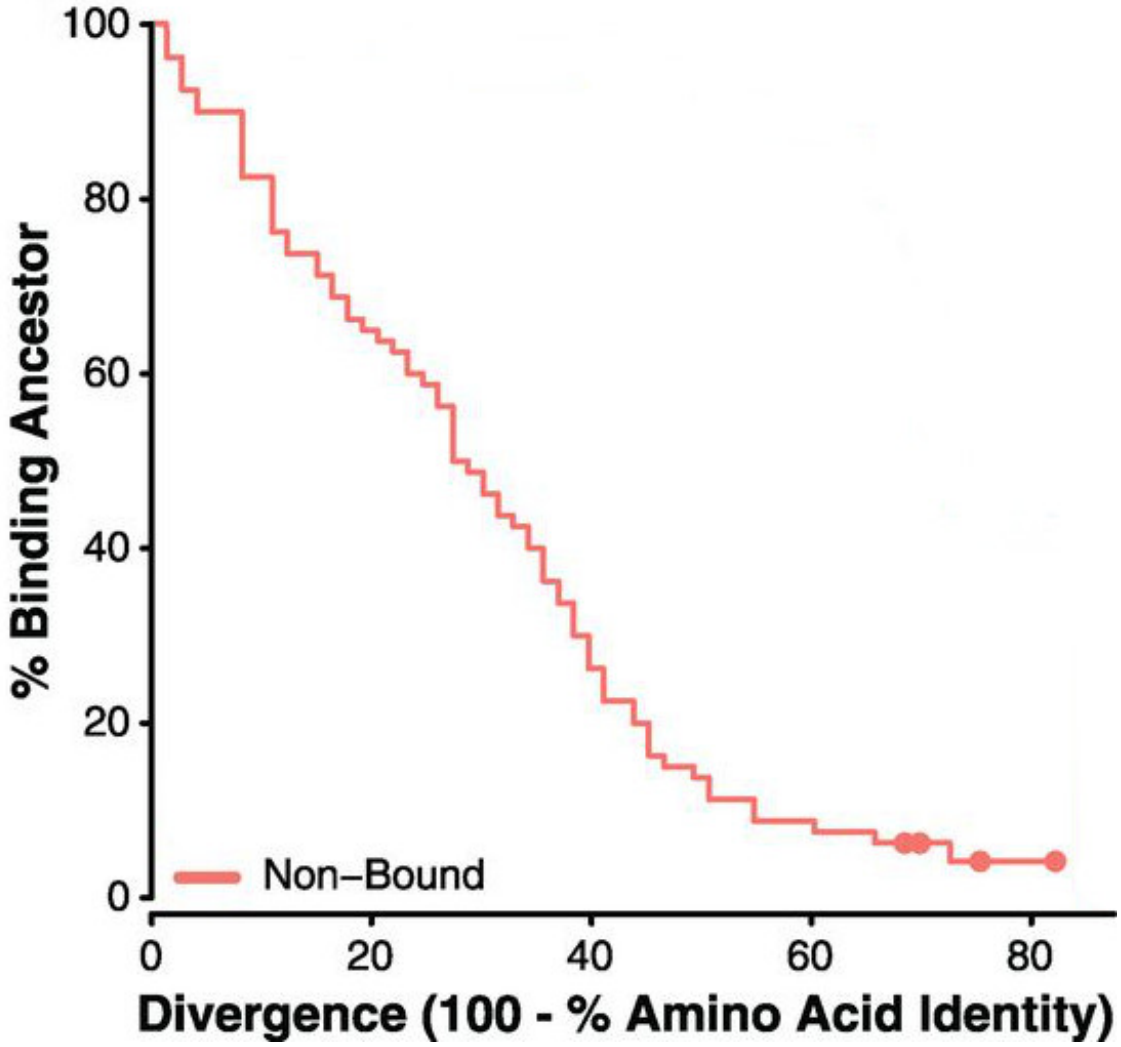
Control simulation: Evolve without binding



Control simulation: Evolve without binding

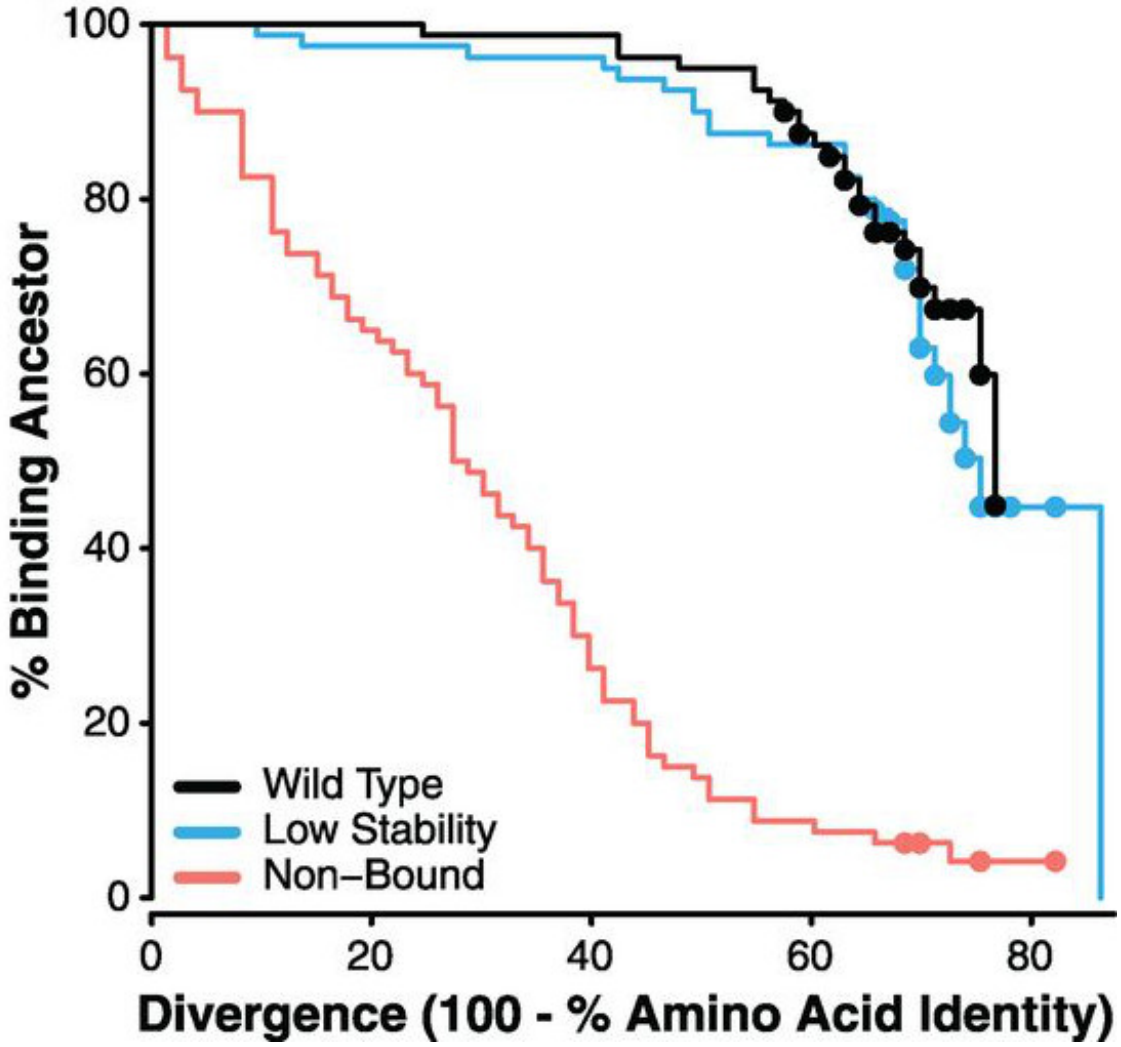


Binding to ancestor declines rapidly in the control scenario



Kachroo et al., Science 2015

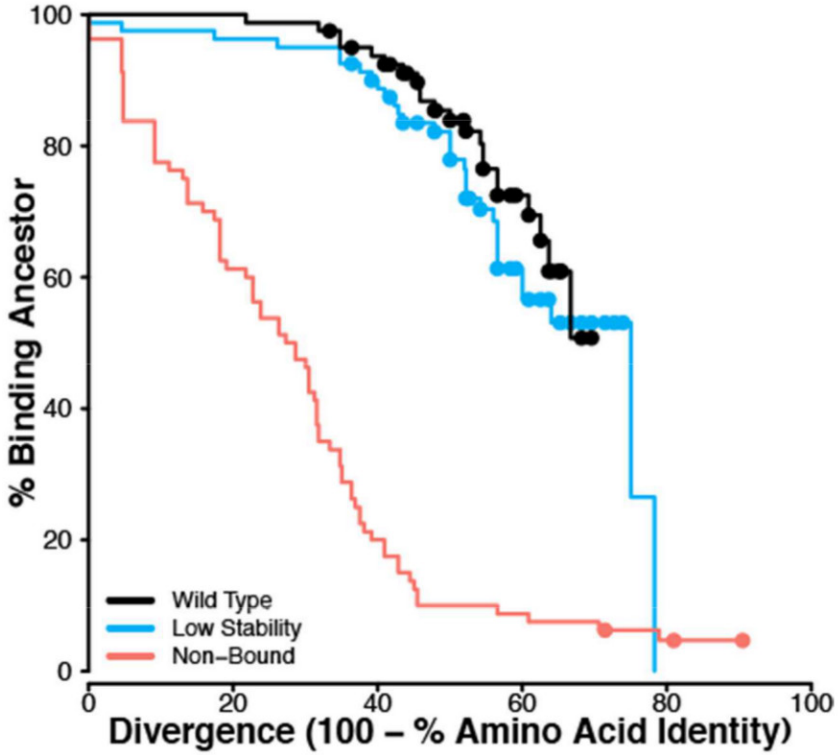
Binding to ancestor declines rapidly in the control scenario, but not in WT scenario



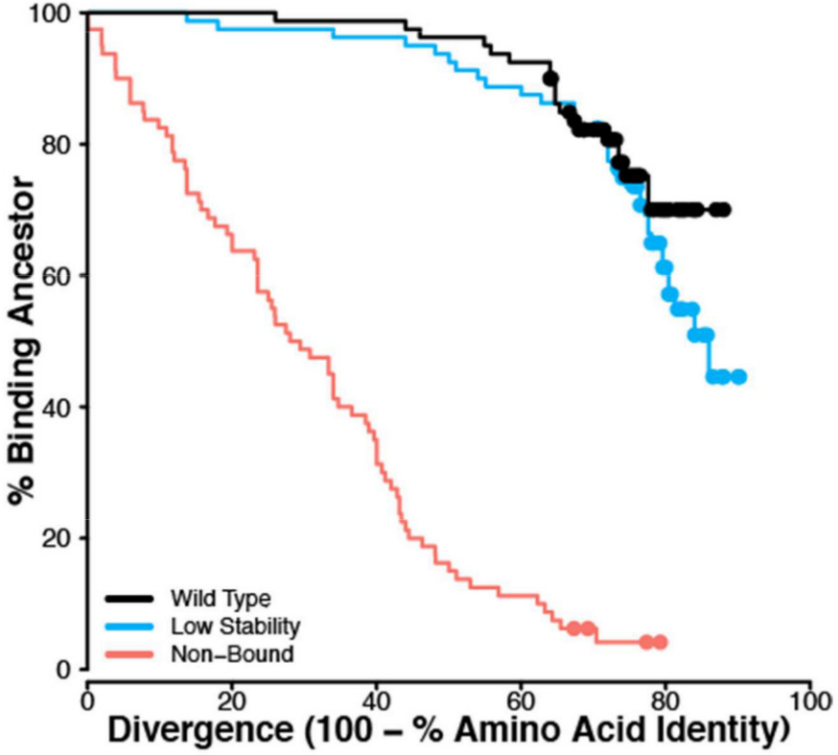
Kachroo et al., Science 2015

Even with substantial divergence in interface sites, evolved proteins bind ancestors

divergence in interface sites



divergence in non-interface sites



Kachroo et al., Science 2015

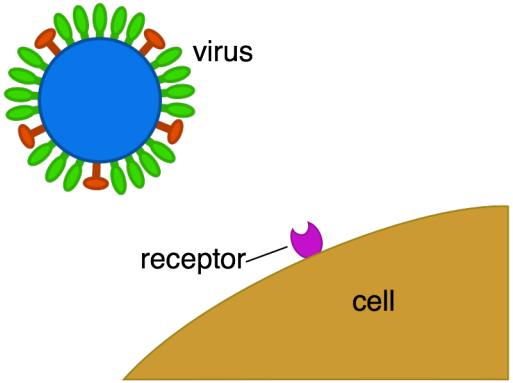
Take-home message

- As protein–protein interfaces diverge, they seem to maintain biochemical similarity to their ancestral state

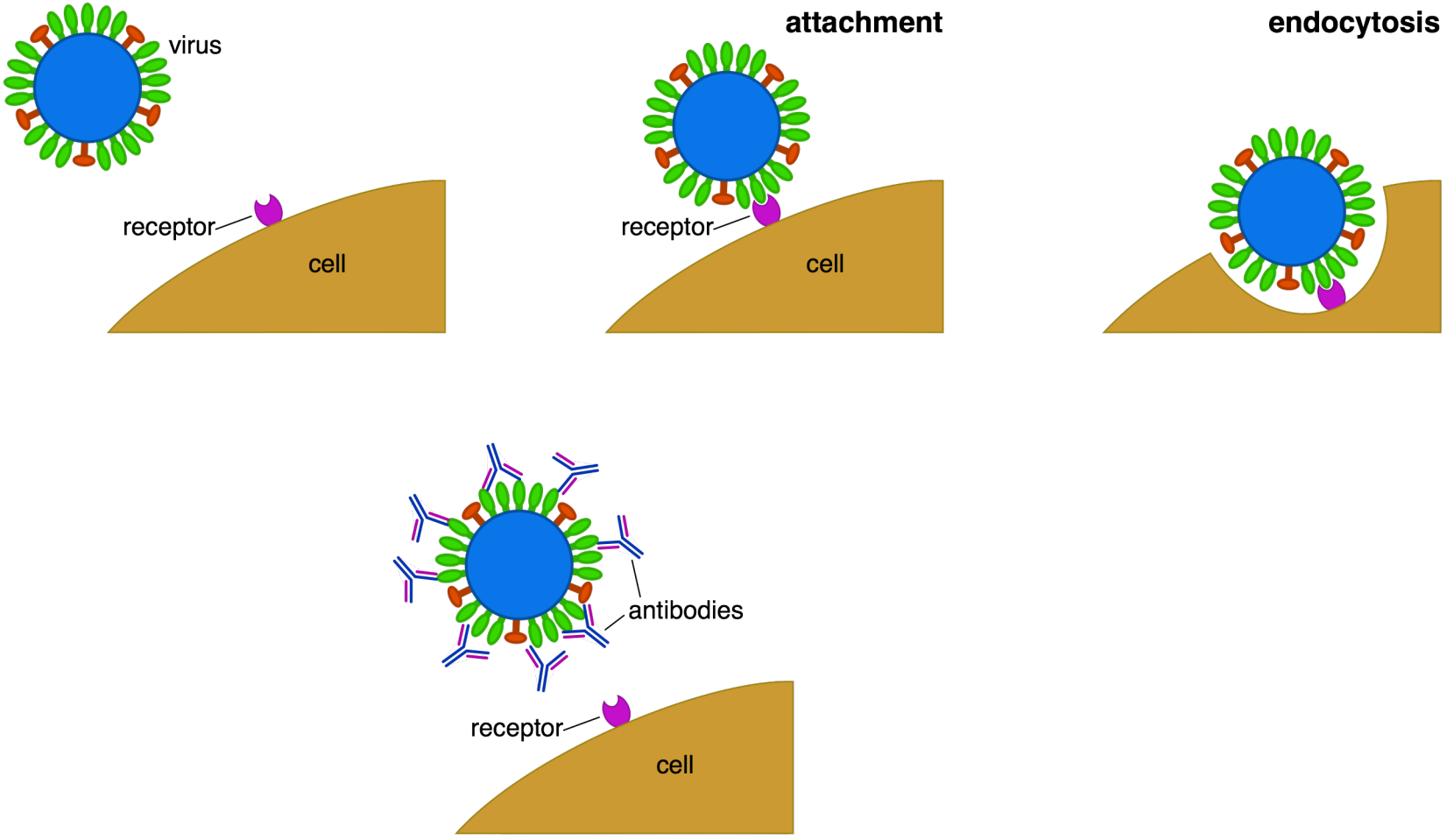
Part III: Can we identify sites under positive selection in influenza HA?

Work by Claire McWhite (Marcotte lab), Austin Meyer

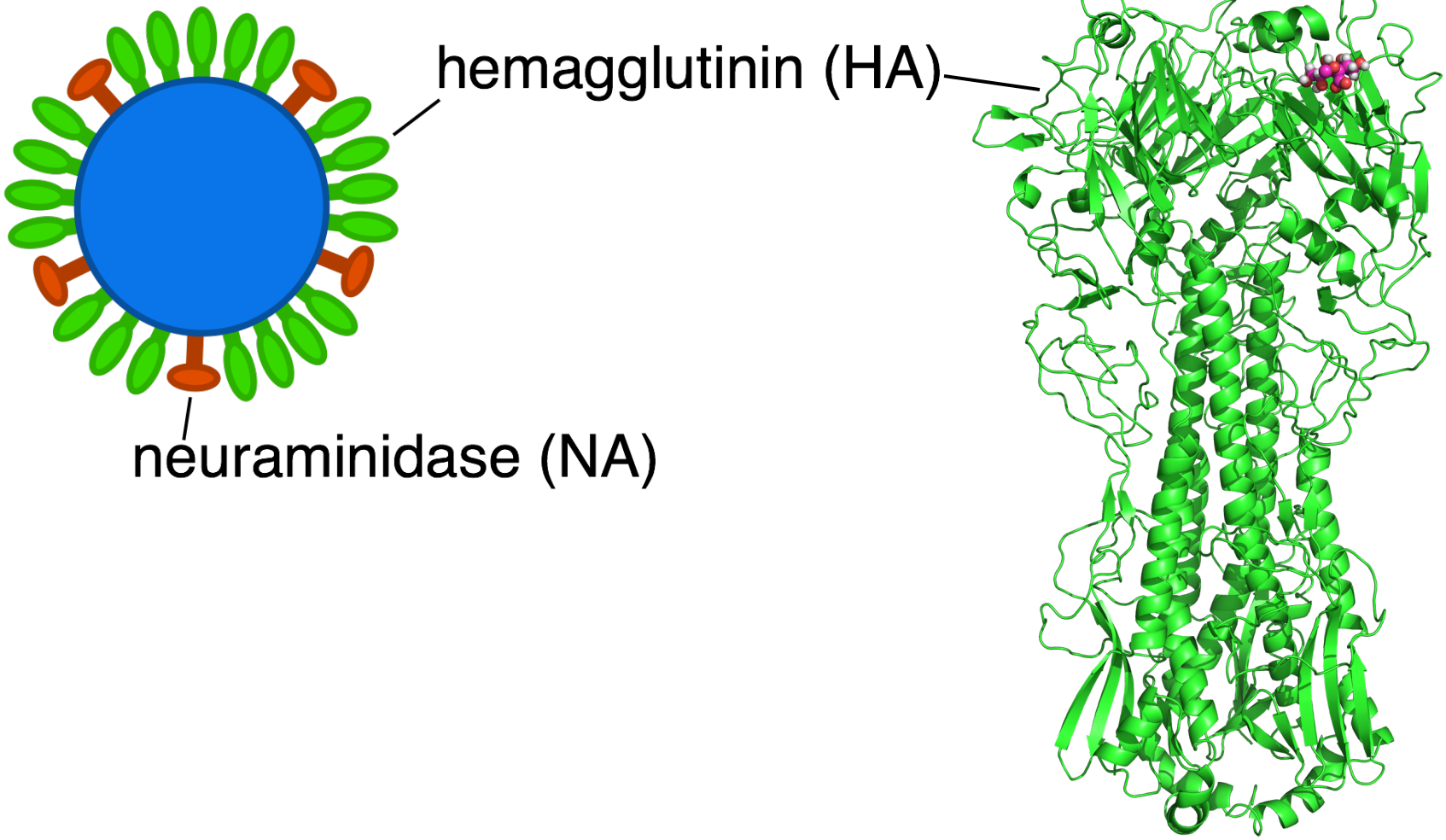
Influenza enters cells via attachment and endocytosis



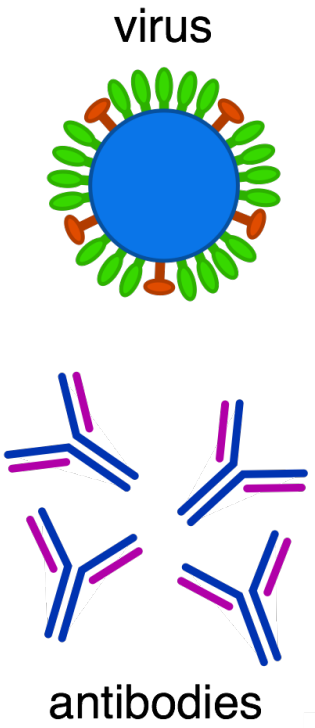
Antibodies bind to viral surface proteins and prevent attachment



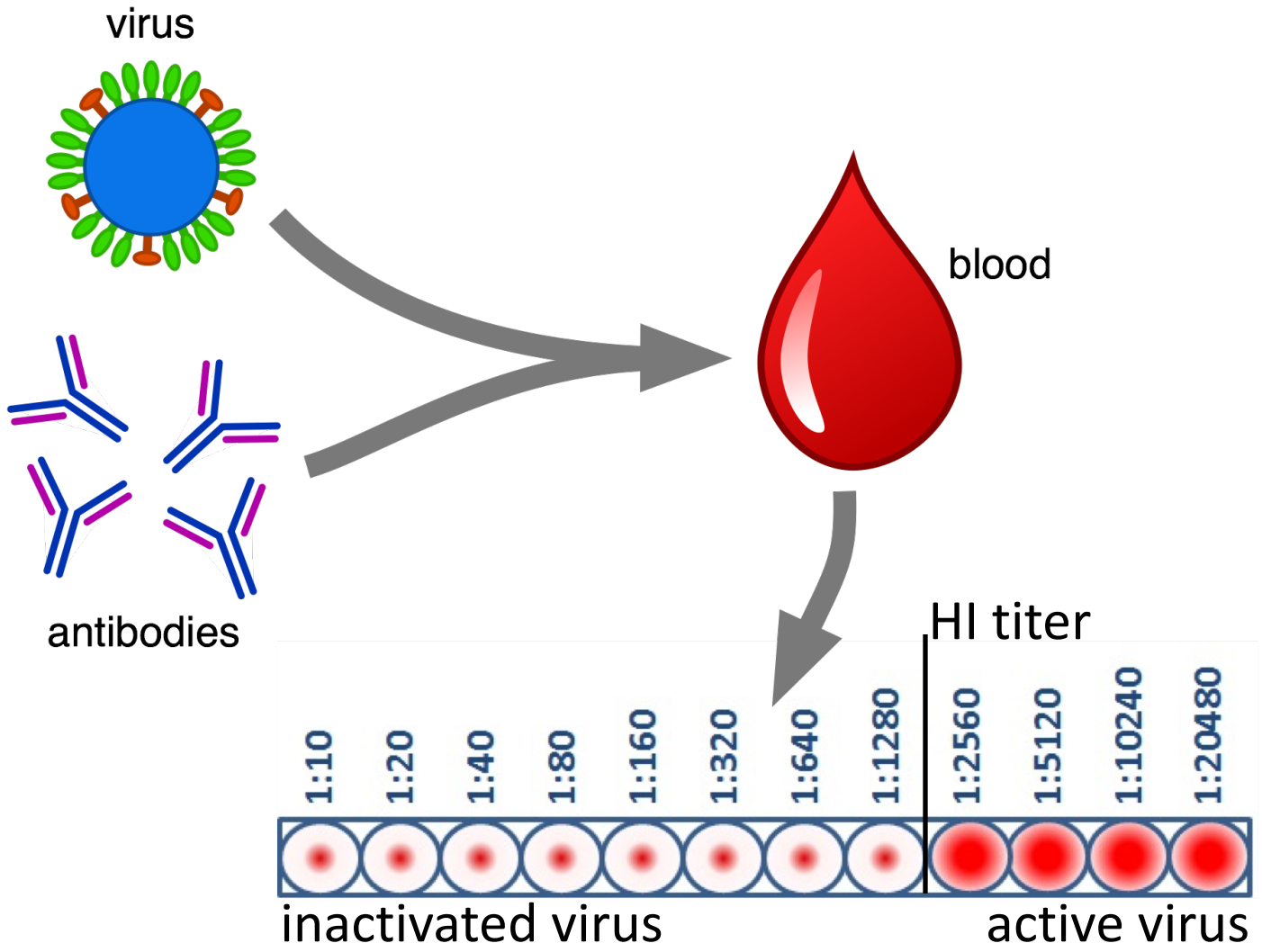
The primary antigenic target is the hemagglutinin protein (HA)



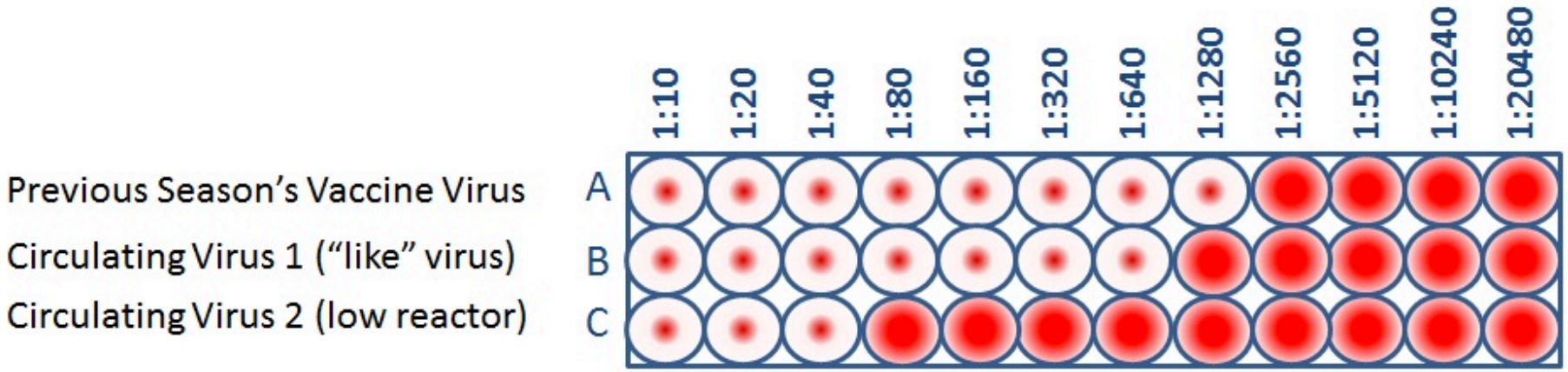
The hemagglutinin inhibition assay characterizes a virus's antigenic profile



The hemagglutinin inhibition assay characterizes a virus's antigenic profile

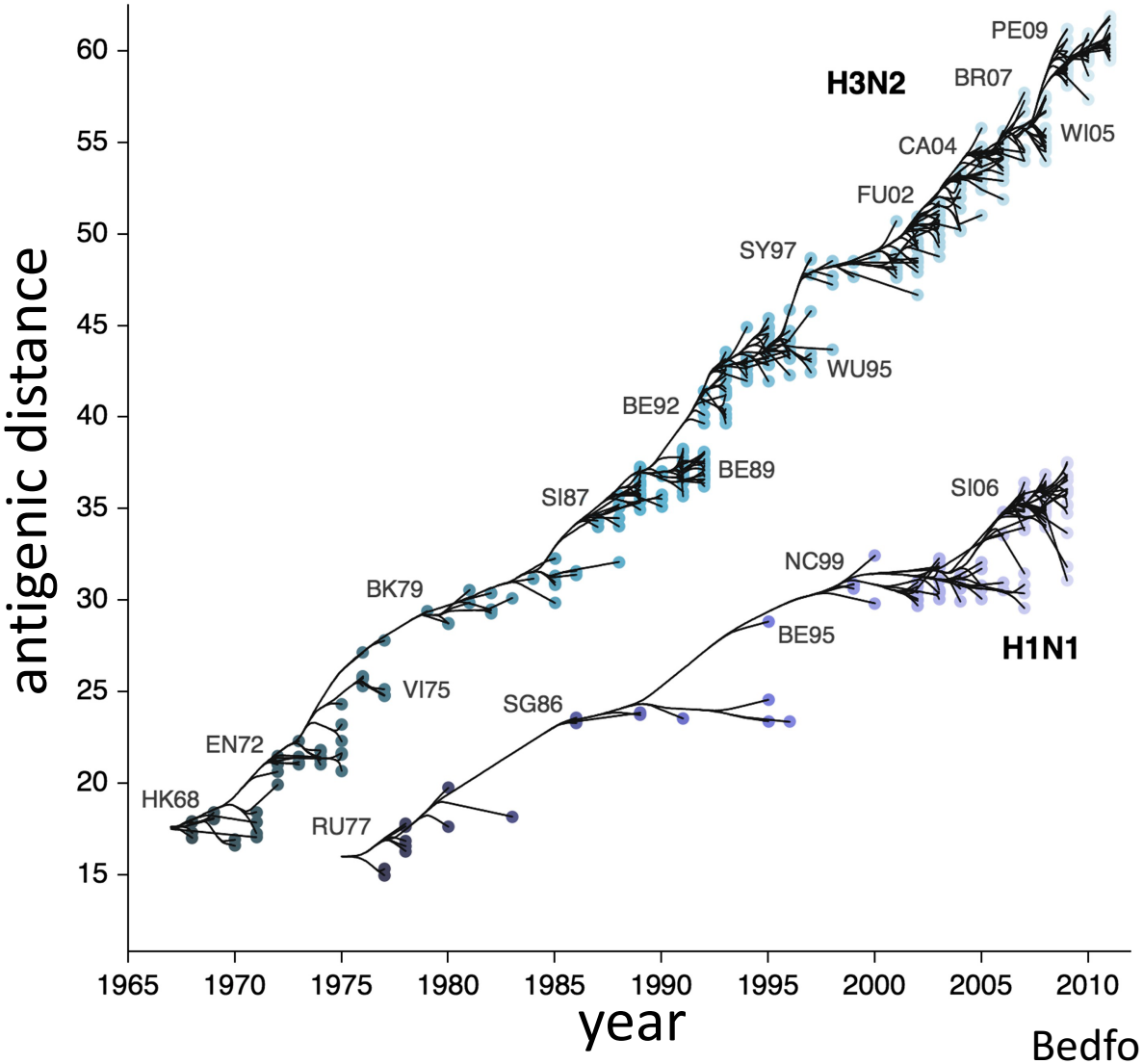


The hemagglutinin inhibition assay characterizes a virus's antigenic profile



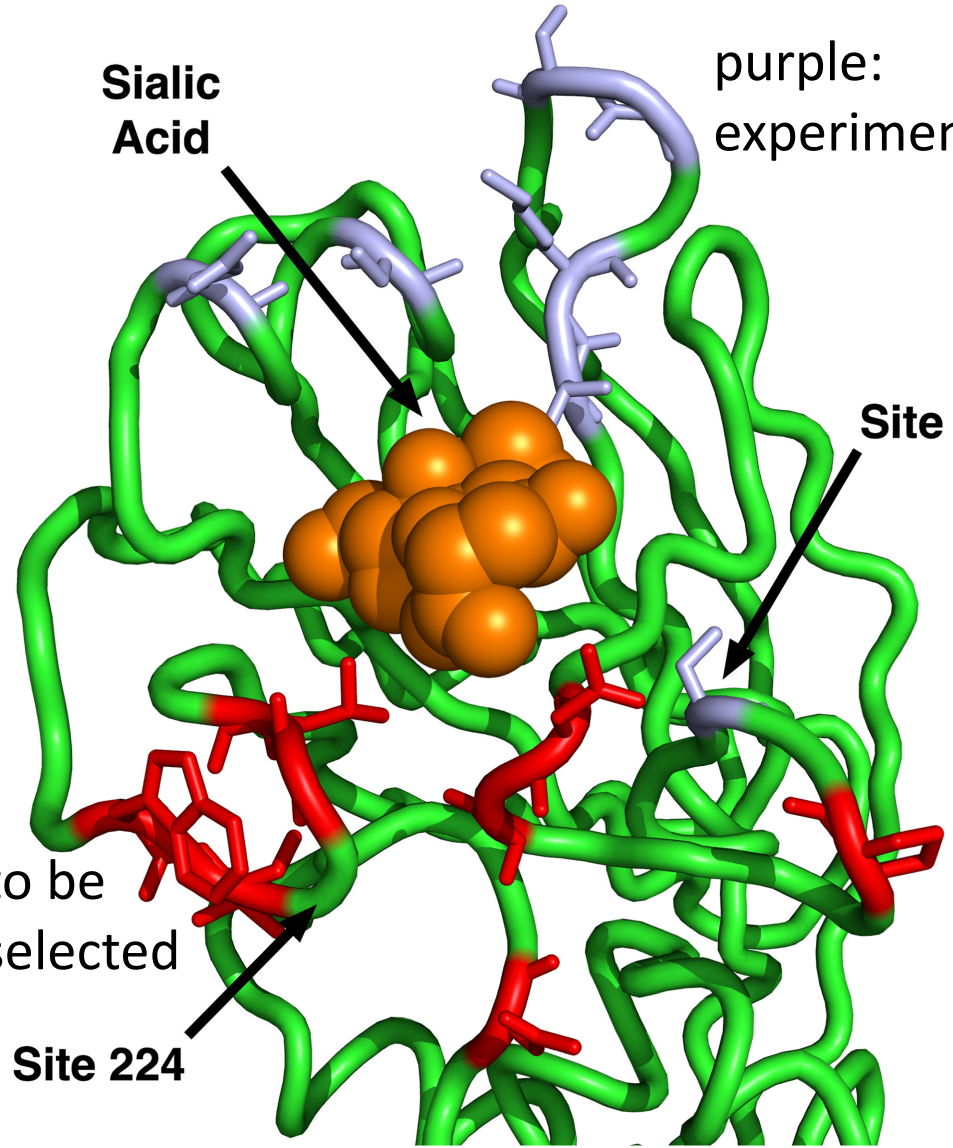
Source: CDC

Influenza viruses rapidly evolve in response to the human immune system



Bedford et al., eLife 2014

Positively selected sites disagree with experimentally verified antigenic sites

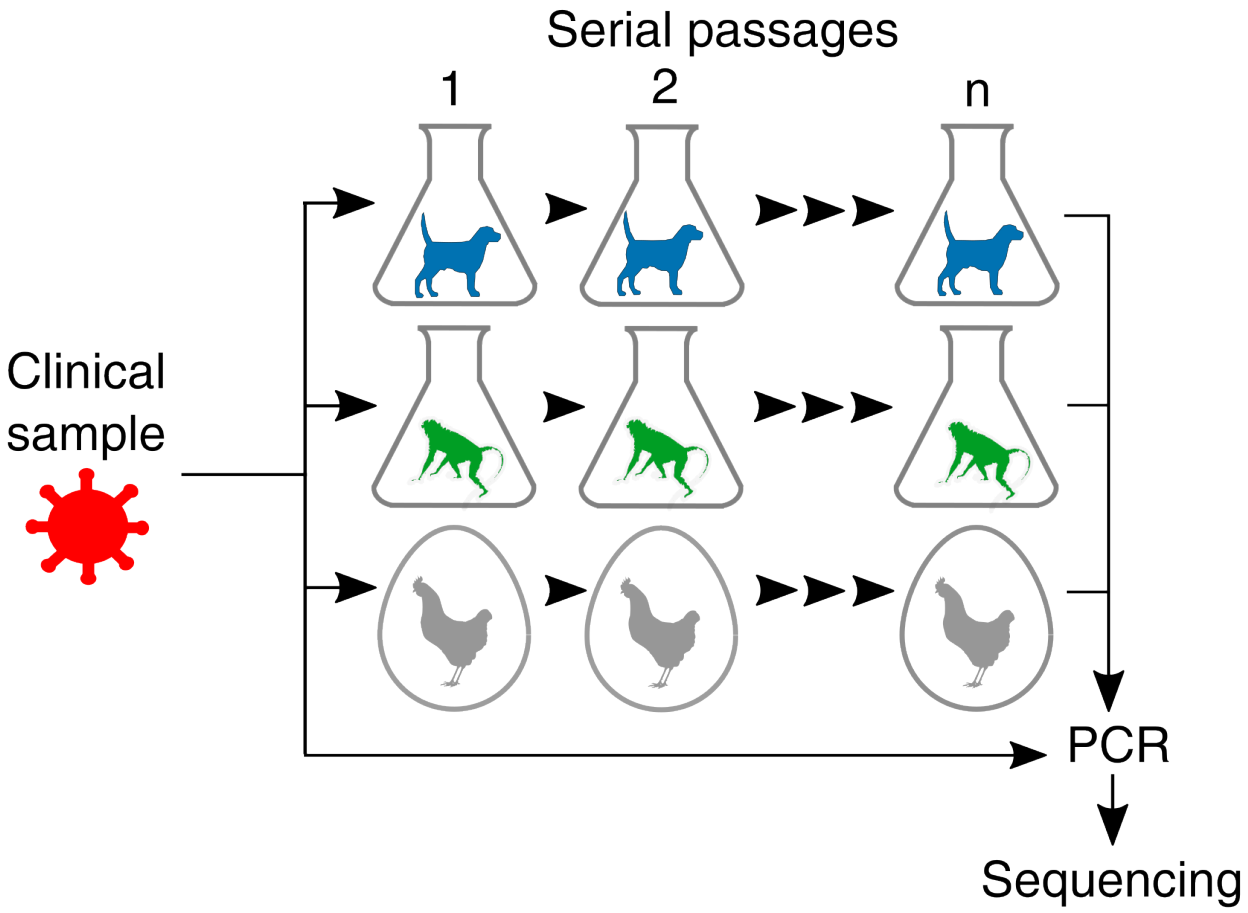


purple:
experimentally verified

red:
predicted to be
positively selected

Meyer & Wilke,
PLOS Pathogens 2015

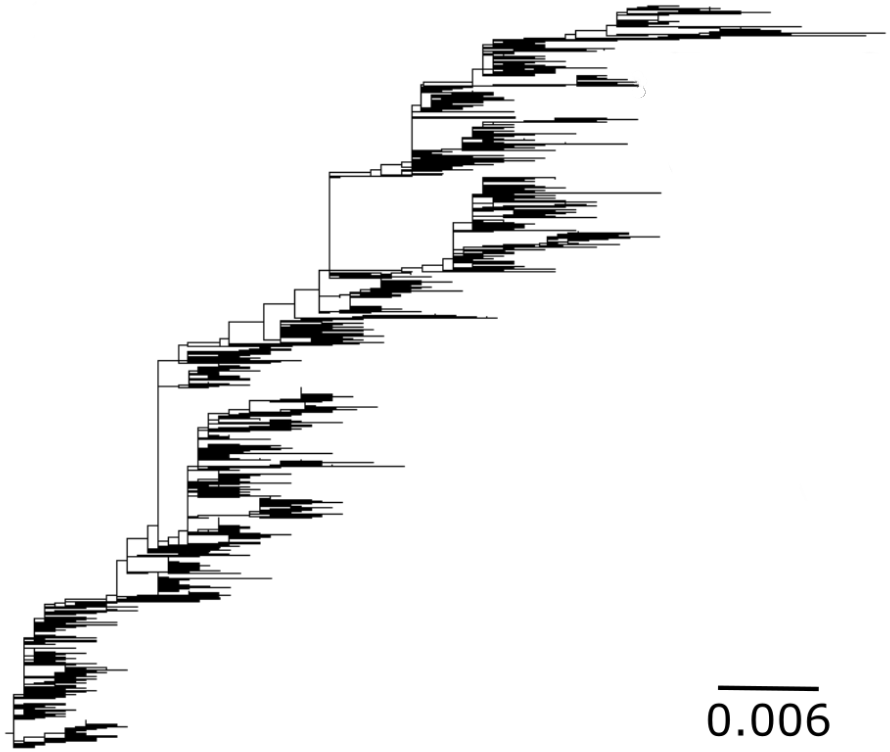
Influenza virus is commonly passaged before analysis, to obtain high viral titers



McWhite et al., Virus Evol. 2016

Passaging treatments are intermingled in the phylogeny

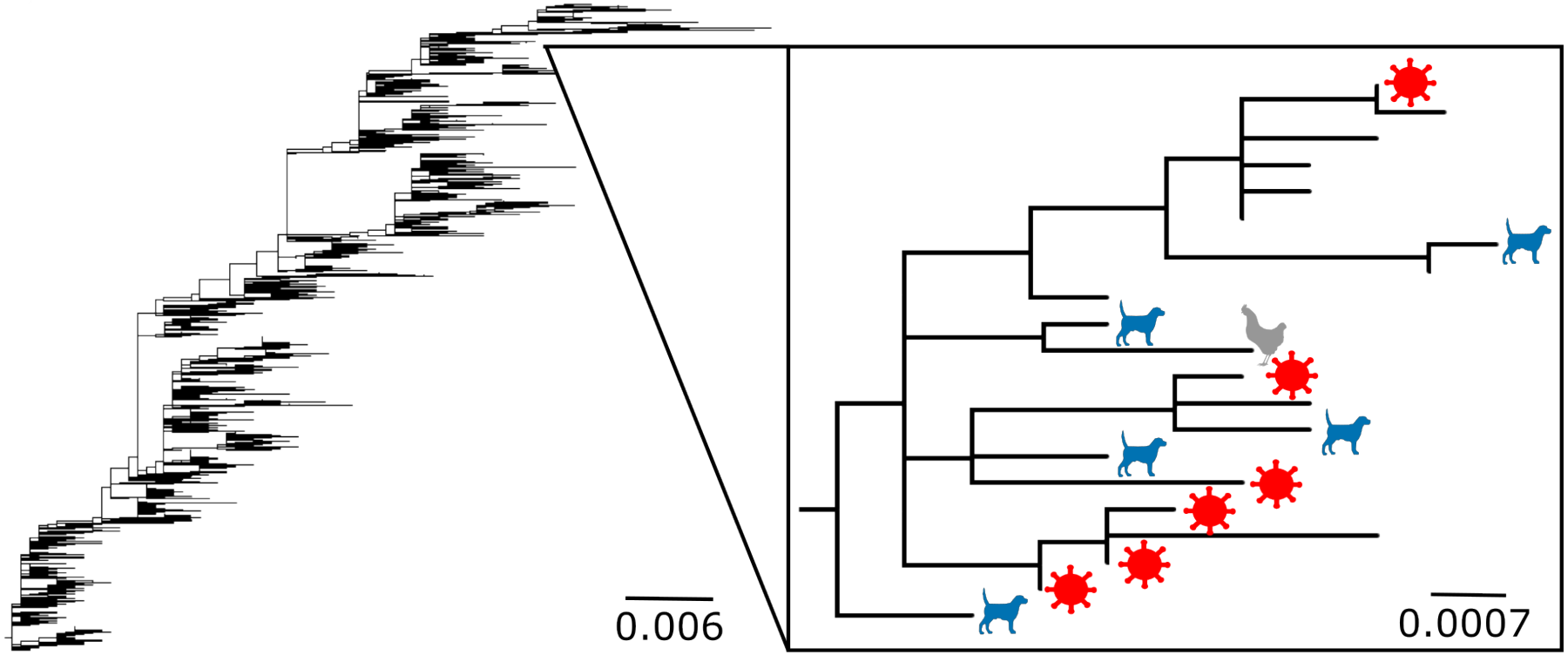
Influenza H3N2 phylogeny, 2005–2015



McWhite et al., bioRxiv 2016

Passaging treatments are intermingled in the phylogeny

Influenza H3N2 phylogeny, 2005–2015



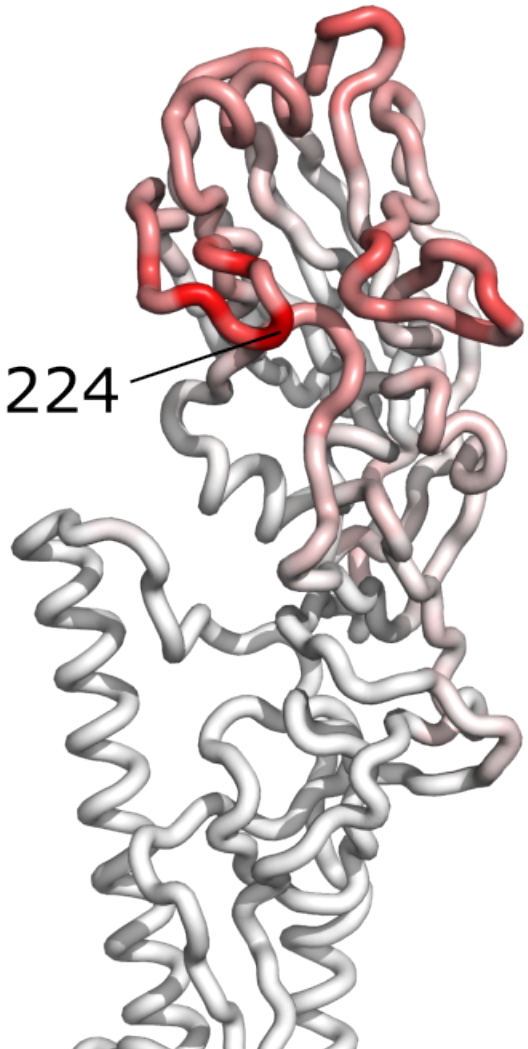
McWhite et al., Virus Evol. 2016

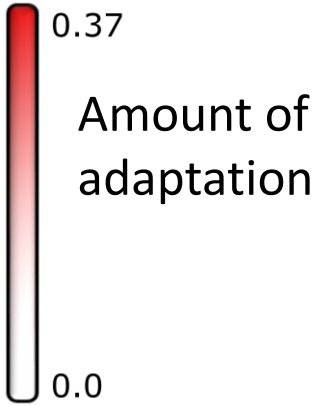
Unfortunately, passaging annotations are a mess

cell	egg	monkey	unpassaged
XMDCK2	E5_E1	PRHMK_2	ORIGINAL
C_3__1	AM1	R_MIX_1_RHMK_1	LUNG_1
MX_S2	E4_E1	PRHMK_3	ISOLATED_DIRECTLY_FRO
CX_S2	EX_E1	RMK_1ST_PASSAGE	M_HOST__NO_PASSA
C_2__1	SPFCK2E3	RHMK_5	GE
MDCK_1	PASSAGE_DETAILS__CK3_	RMK1	P0
M_1_C_2___SIAT1	_E6	RHMK1	OR
M1M2_C1	E6	R1	
MDCK1_SIAT1_SIAT1	E4E8_E1	RII	
PASSAGE_DETAILS__ND_	E4_E2_E1	PRHMK_1	
MDCK	SPFCK3E3	PASSAGE_DETAILS__PMK	
S3	EGG	01	
X1_S2	EGG_PASSAGE	RHMK_1	
. . .	. . .	. . .	

McWhite et al., Virus Evol. 2016

Positive selection in H3N2 HA near site 224



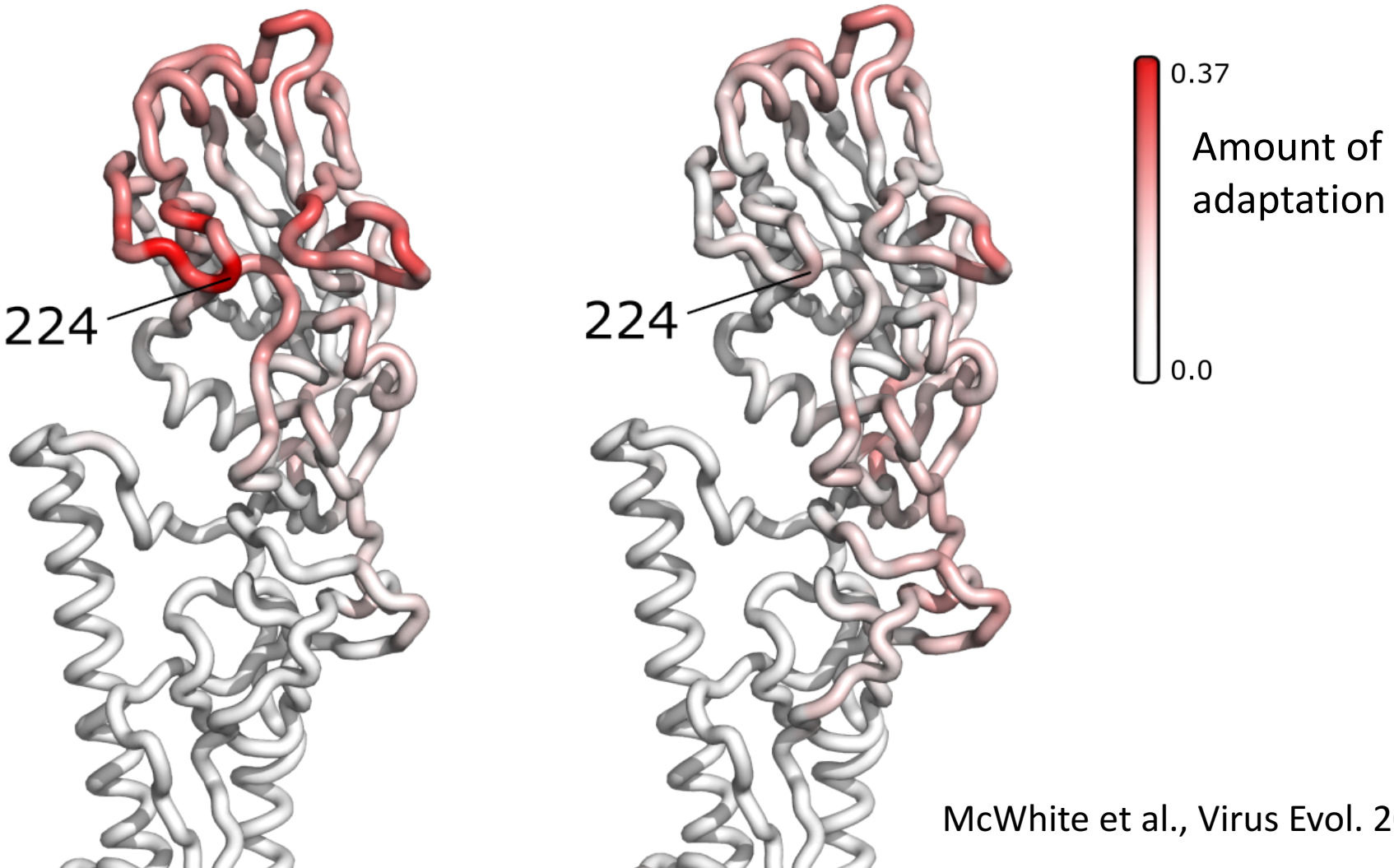


Meyer and Wilke, PLOS Path. 2015

Positive selection in H3N2 HA near site 224 is caused by adaptation to MDCK cells

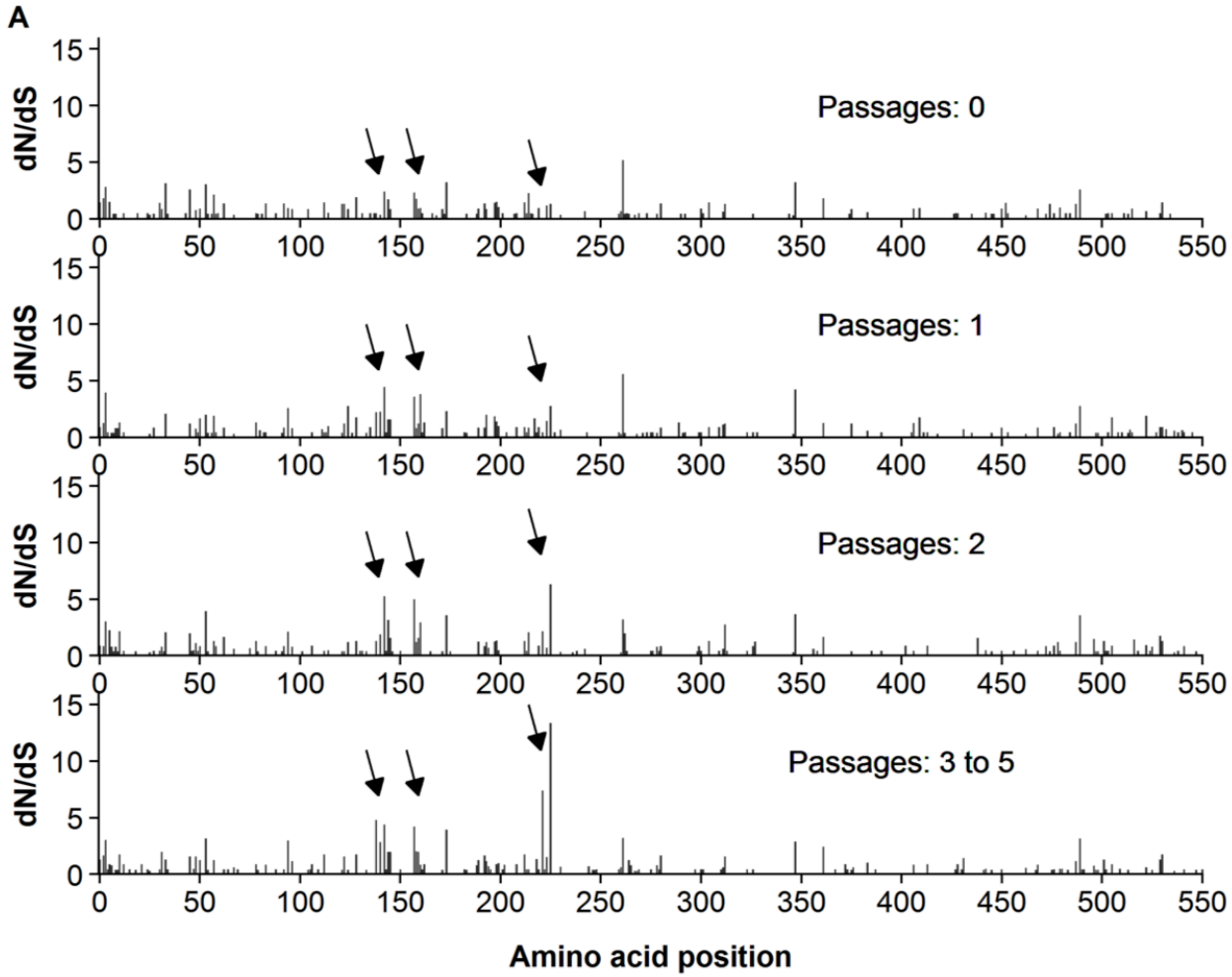
MDCK passaged virus

unpassaged virus

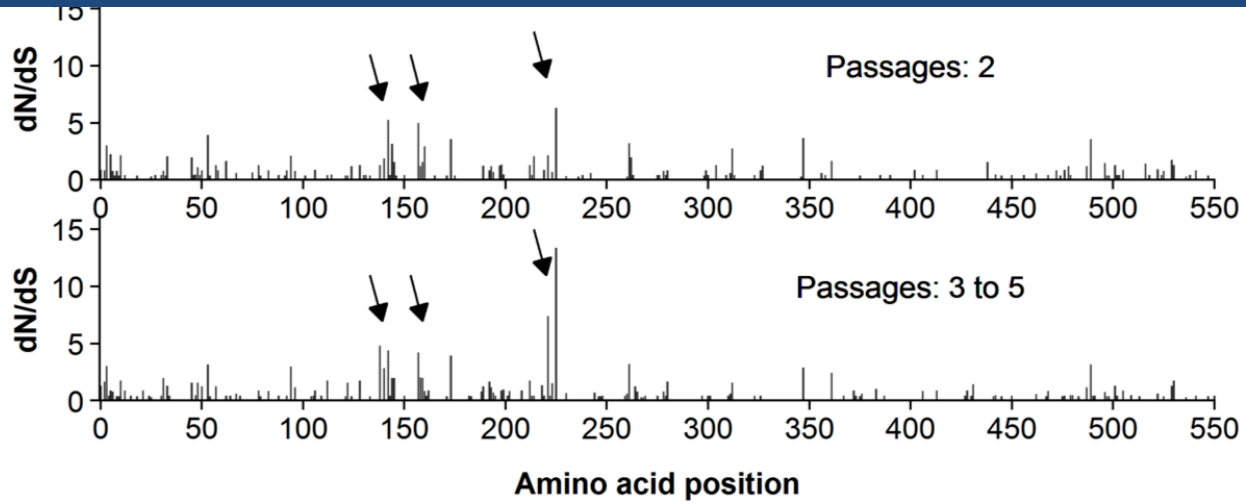


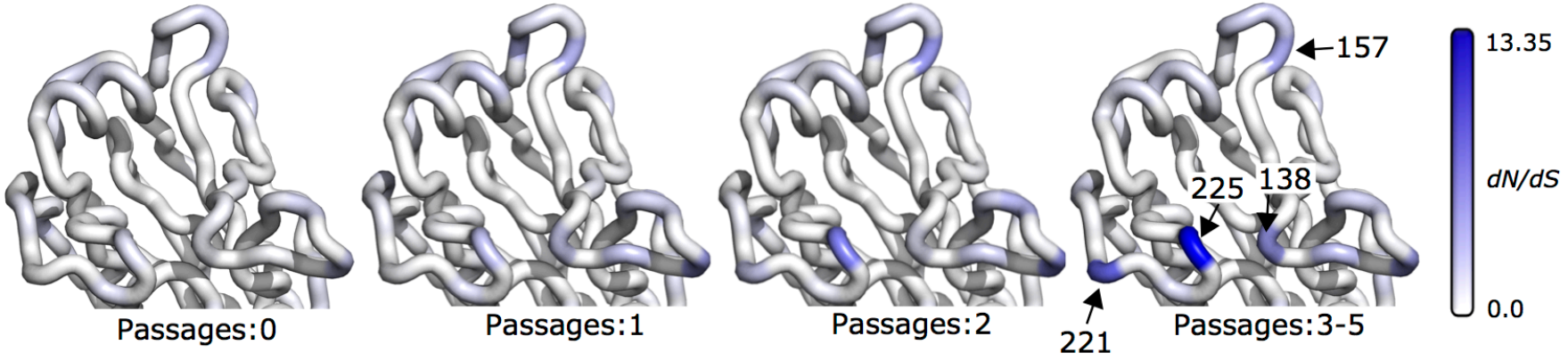
McWhite et al., Virus Evol. 2016

Signal increases with the number of passages

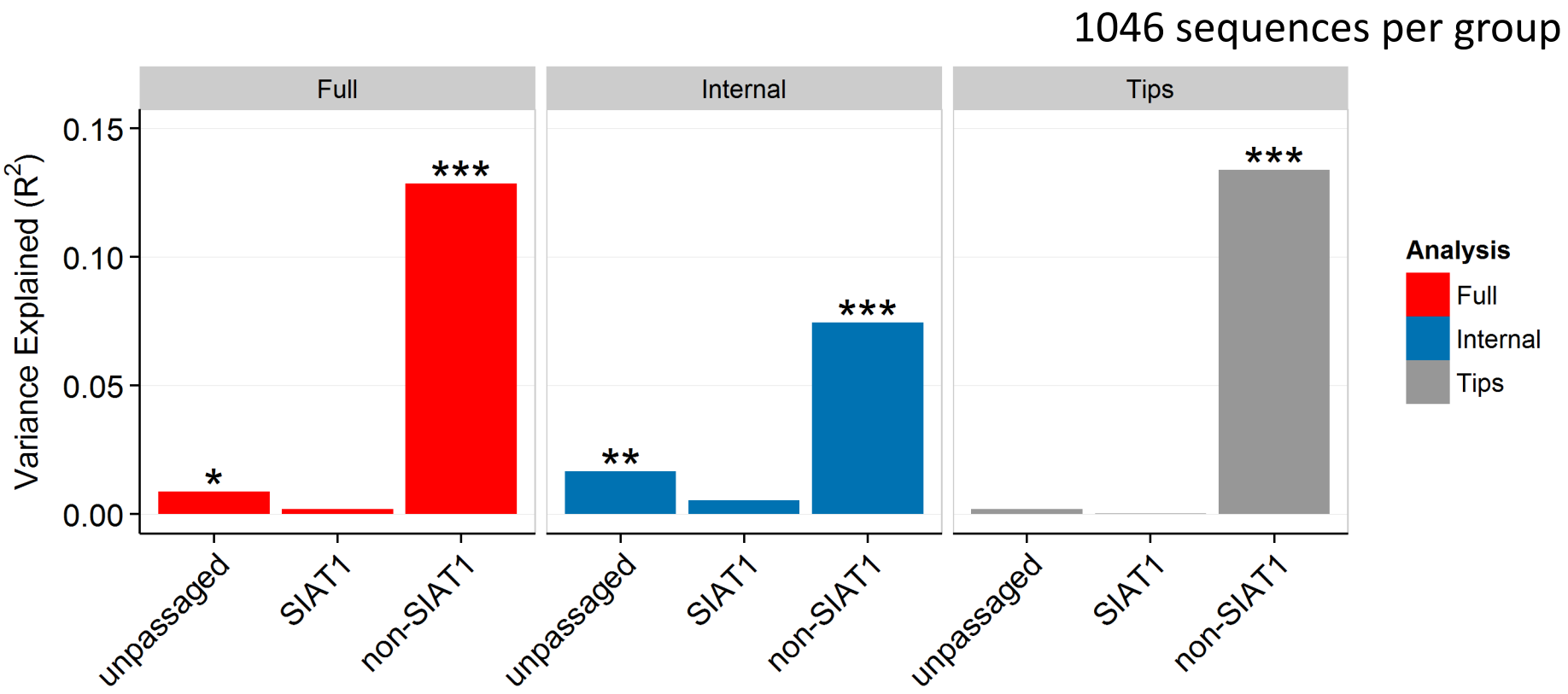


Signal increases with the number of passages





Virus passaged in MDCK-SIAT1 cells looks like unpassaged virus

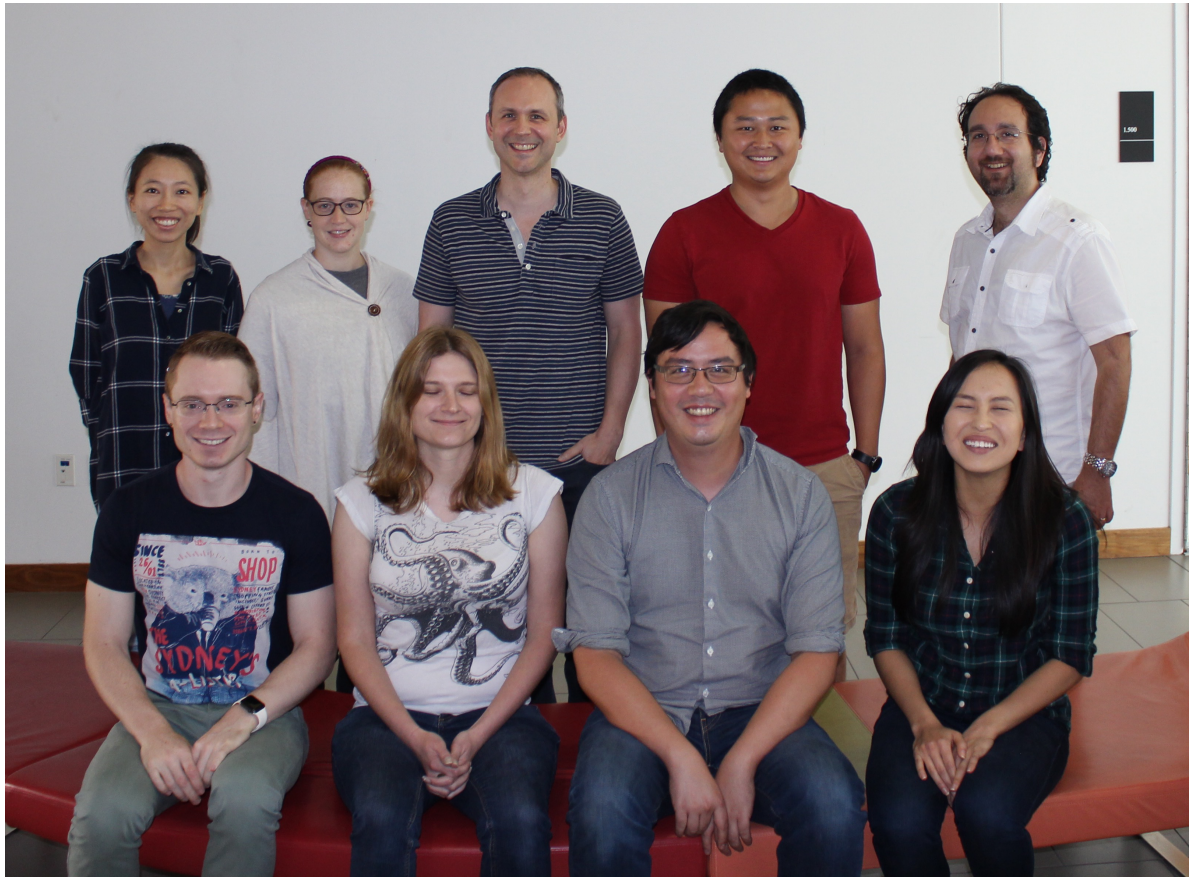


McWhite et al., bioRxiv 2016

Take-home messages

- Influenza virus is regularly adapted to cell culture, with the purpose of measuring its antigenic profile
- As a result, analyzed viral strains are often not representative of the circulating virus

Acknowledgments



Collaborators
- Julian Echave
- Ed Marcotte
- Claire McWhite (Marcotte lab)
- Austin Meyer

Funding
- NIH/NIGMS
- NIH/NIAID
- Army Research Office
- NSF BEACON Center