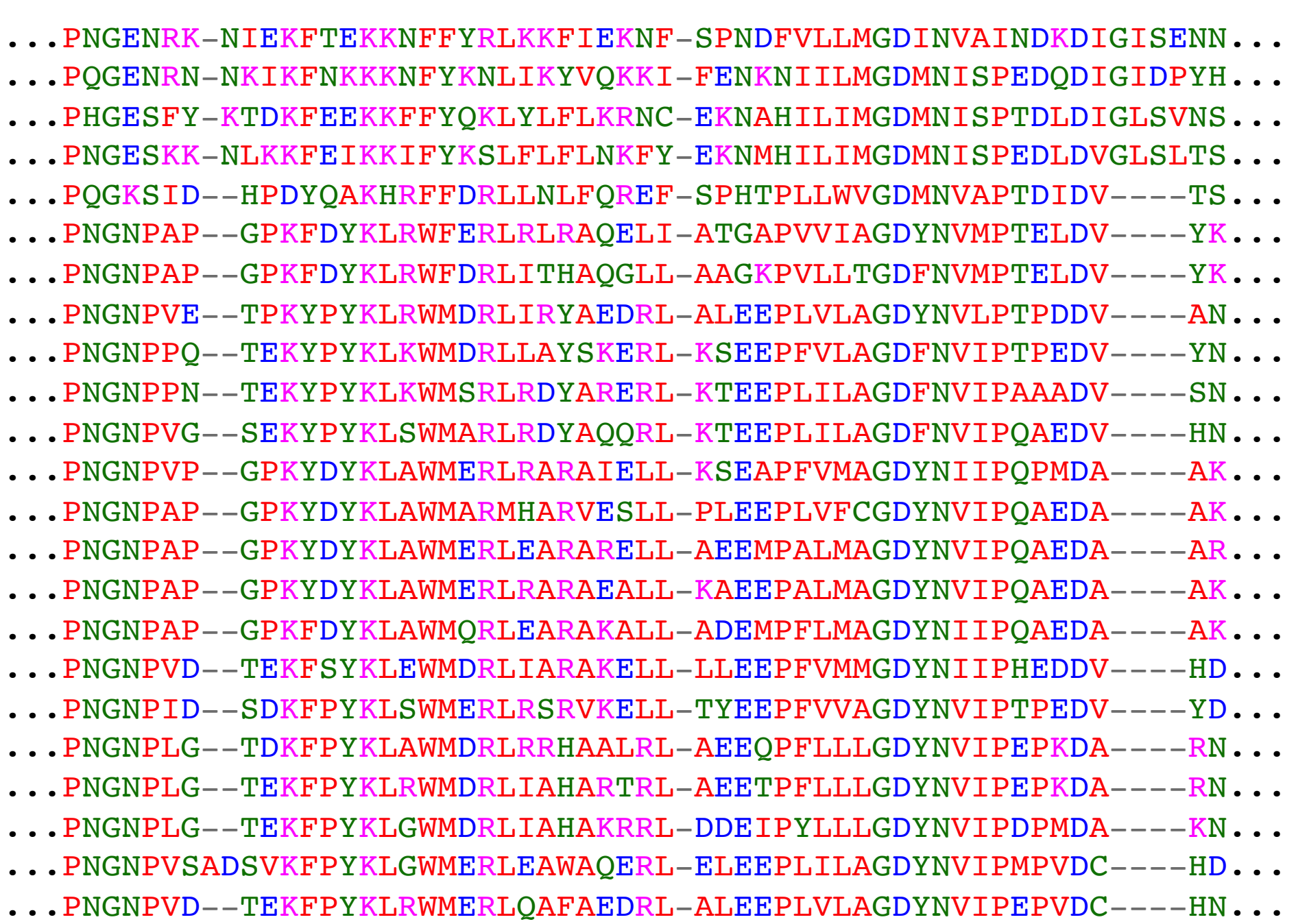
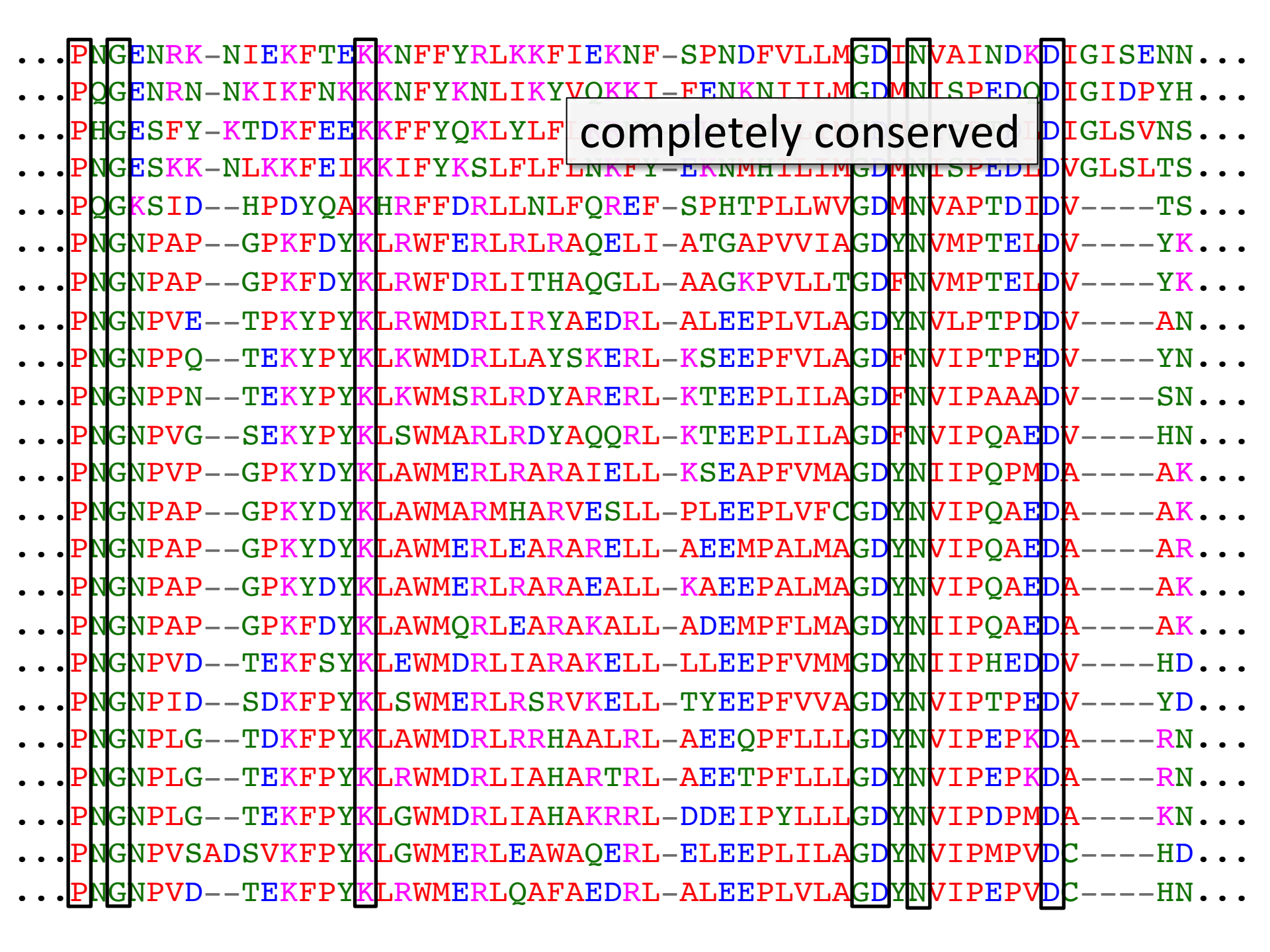
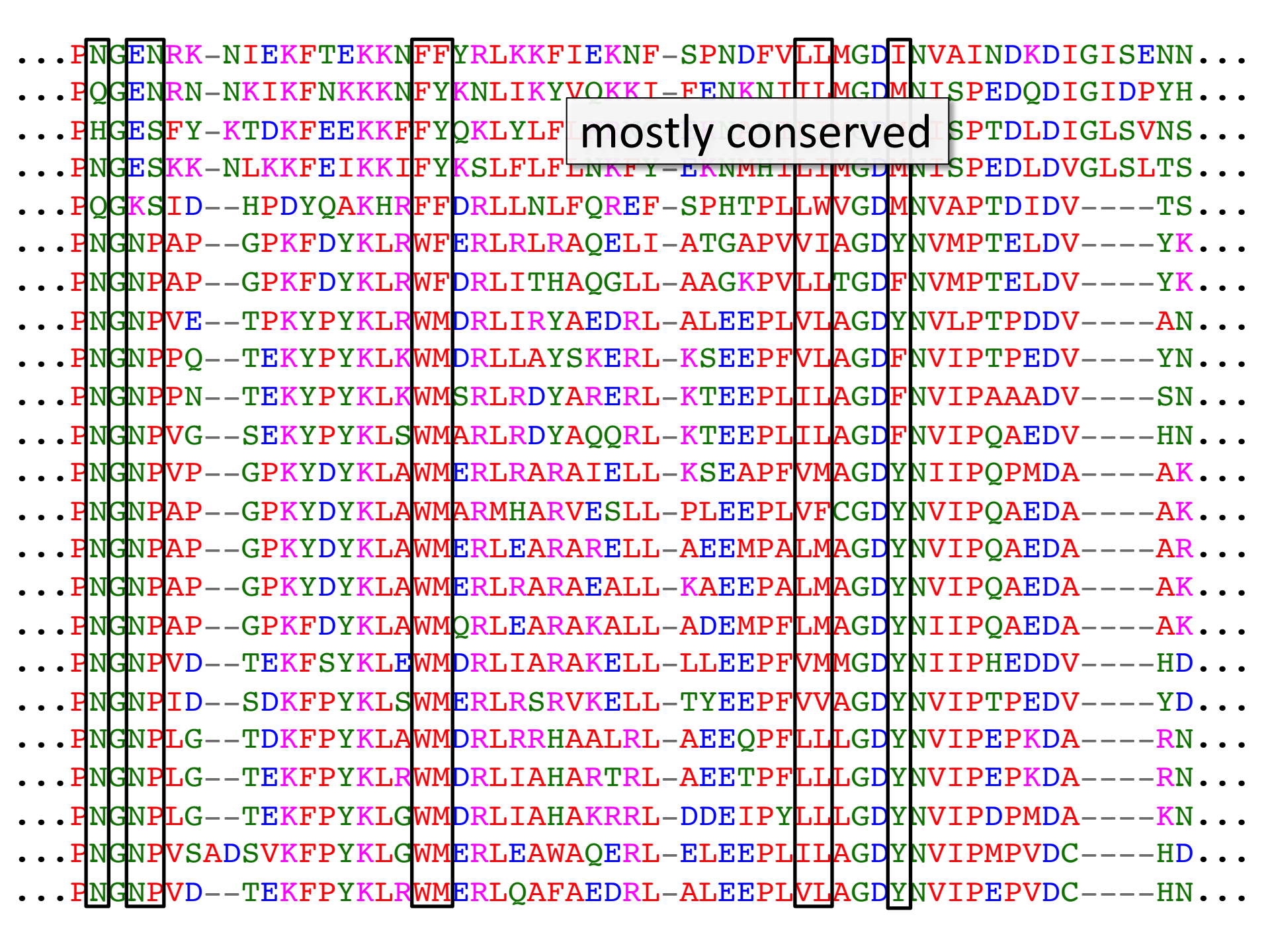
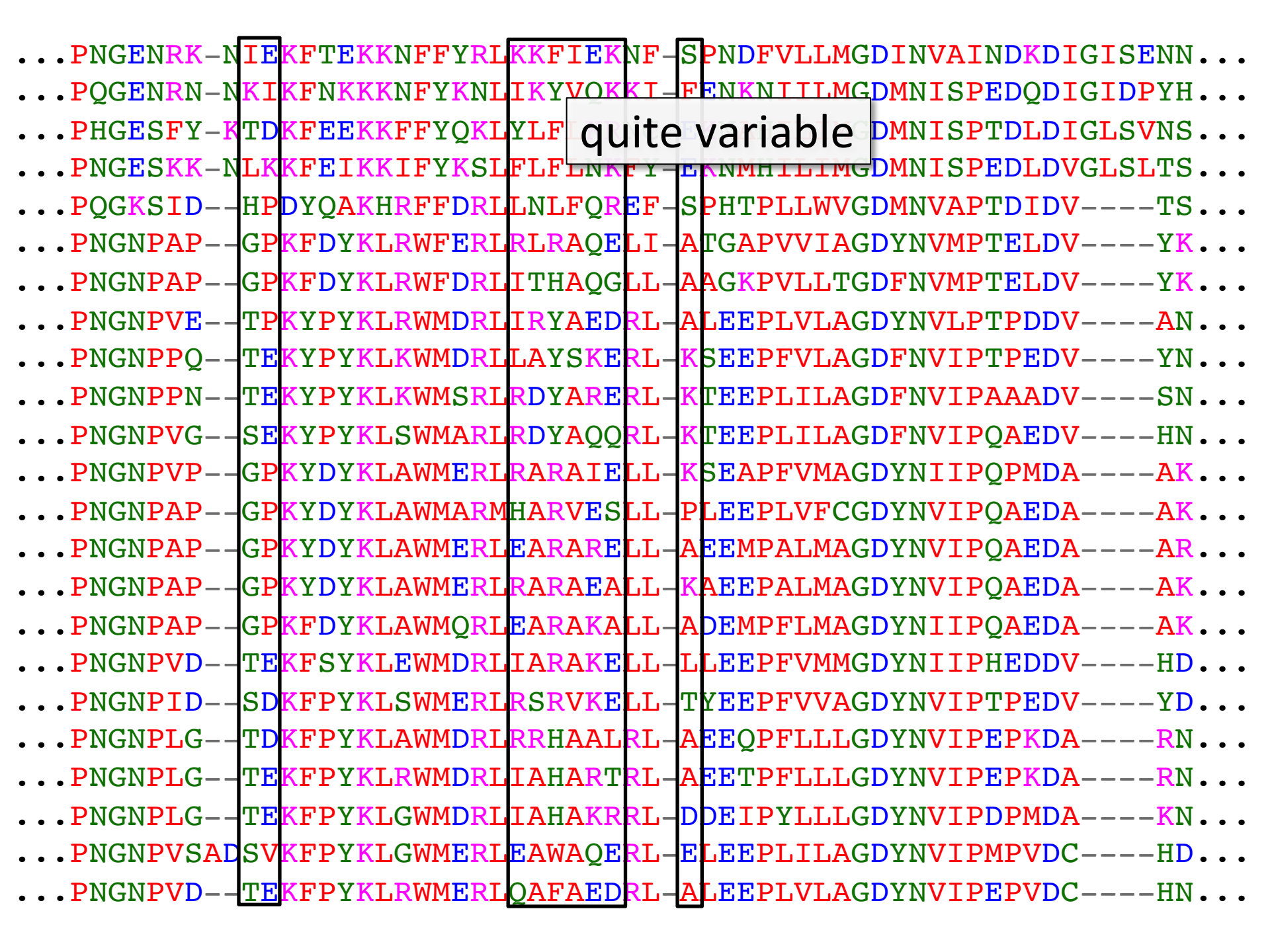
Structural and functional constraints on protein evolution

Claus O. Wilke
The University of Texas at Austin

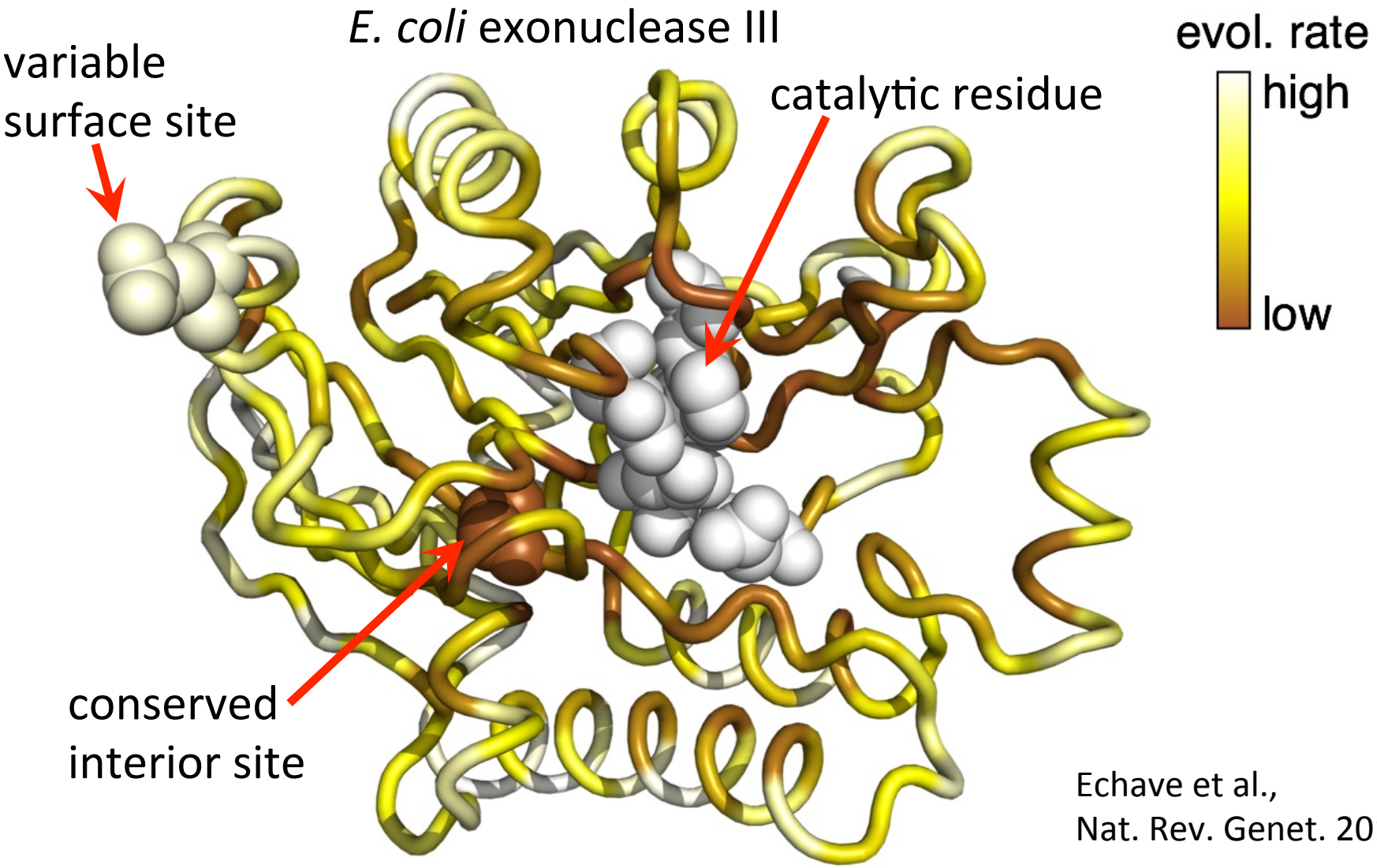

...PNGENRK-NIEKFTEKKNFFYRLKKFIEKNF-SPNDFVLLMGDINVAINDKDIGISENN...
...PQGENRN-NKIKFNKKKNFYKNLIKYVQKKI-FENKNIILMGDMNISPEDQDIGIDPYH...
...PHGESFY-KTDKFEEKKFFYQKLYLFLKRNC-EKNAHILIMGDMNISPTDLDIGLSVNS...
...PNGESKK-NLKKFEIKKIFYKSLFLFLNKFY-EKNMHILIMGDMNISPEDLDVGLSLTS...
...PQGKSID--HPDYQAKHRFFDRLLNLFQREF-SPHTPLLWVGDMNVAPTDIDV----TS...
...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

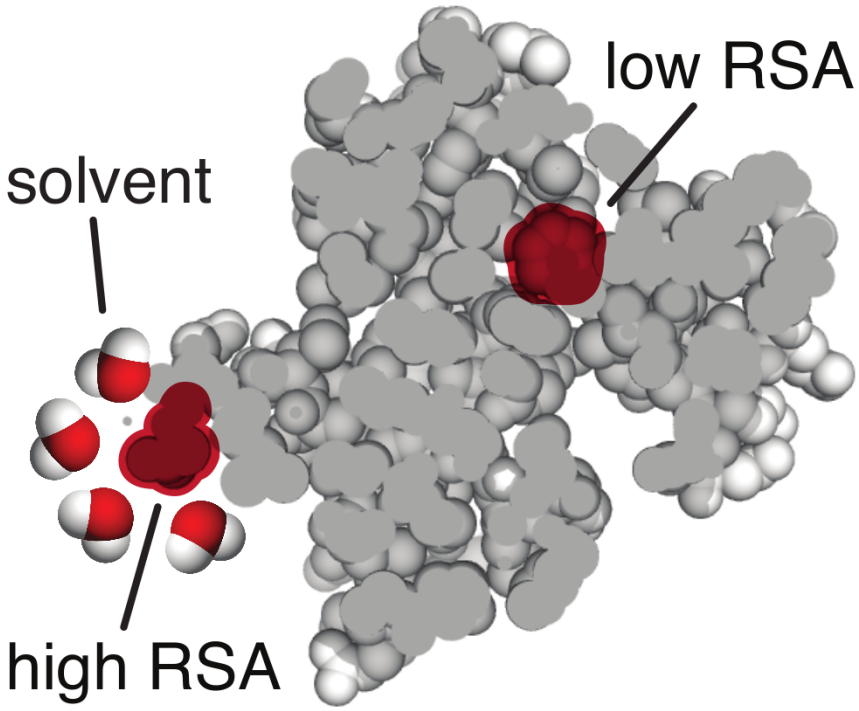
Patterns of sequence variation make more sense in a structural context



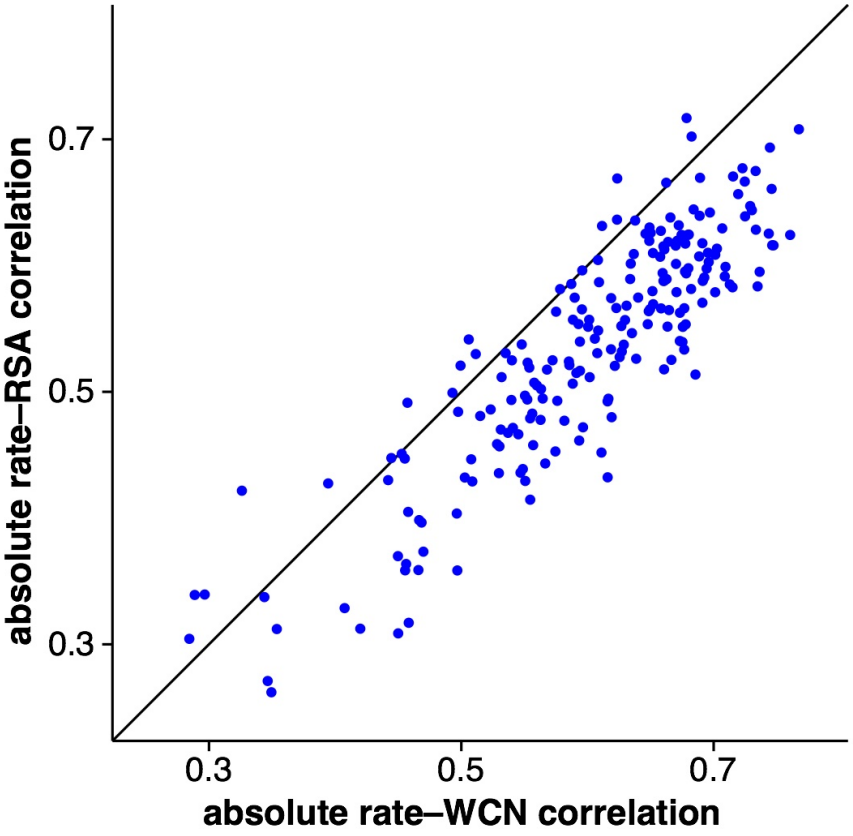
Echave et al.,
Nat. Rev. Genet. 2016

We measure structure with RSA and WCN

Relative Solvent Accessibility

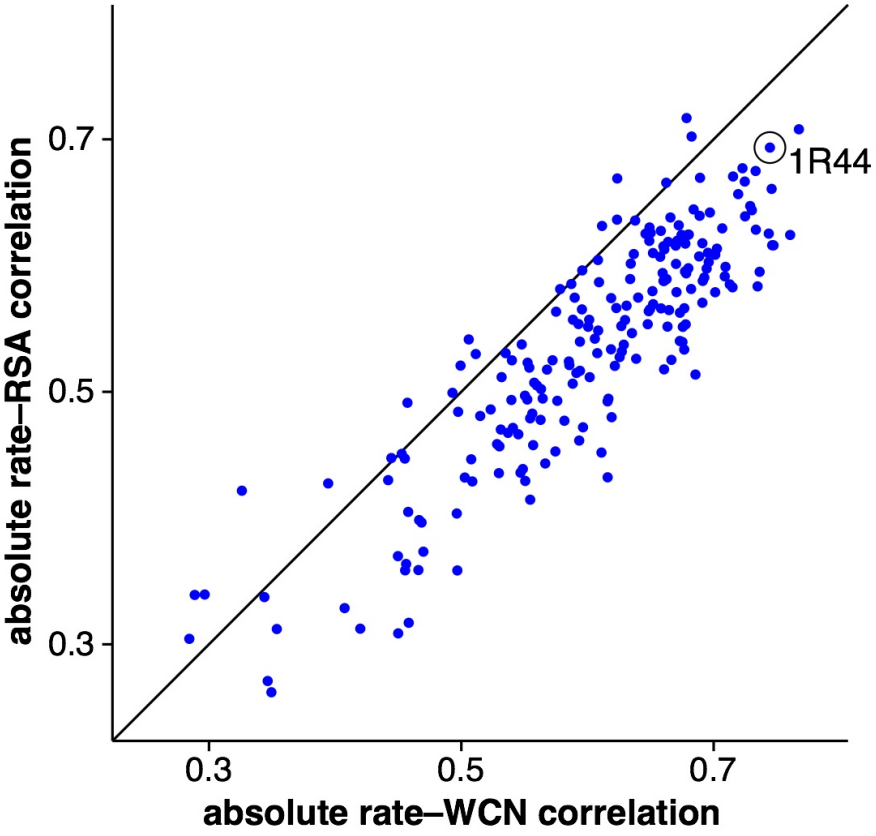


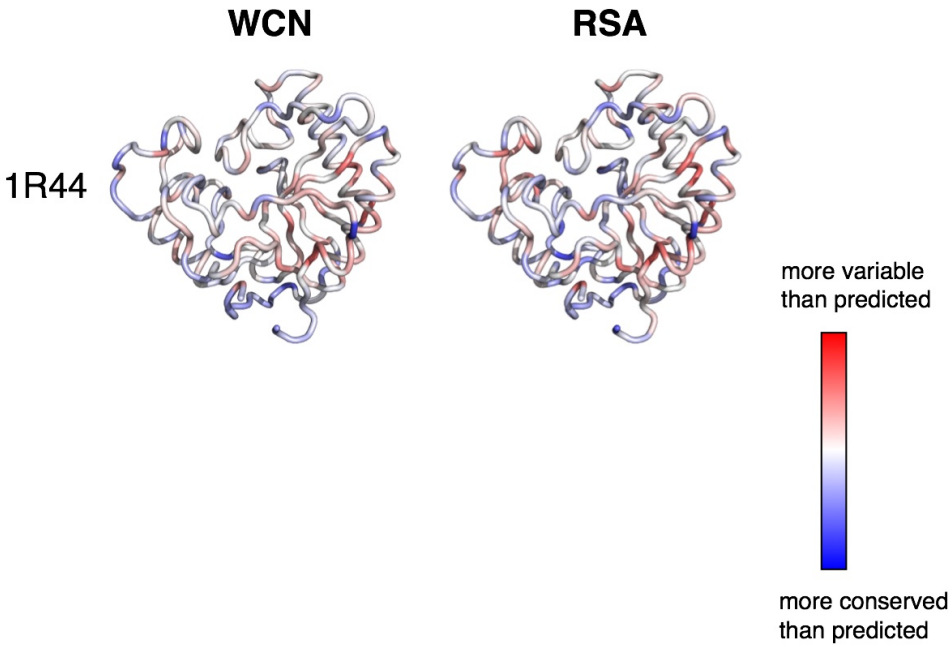
Structure is a good rate predictor for some proteins, but not so much for others



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

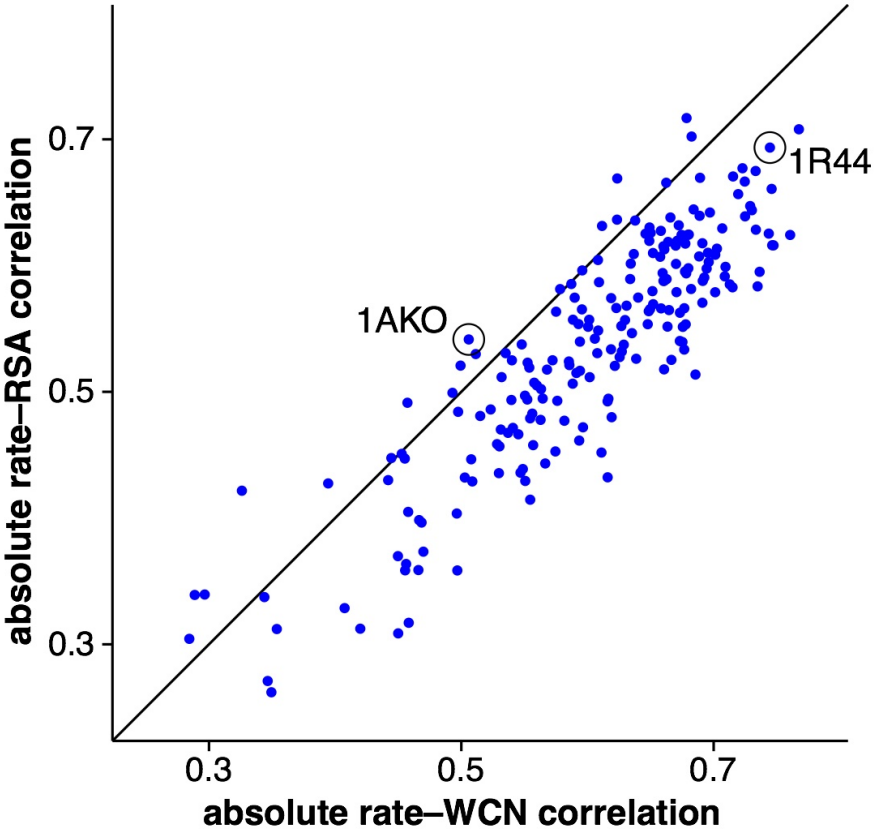
Structure is a good rate predictor for some proteins, but not so much for others

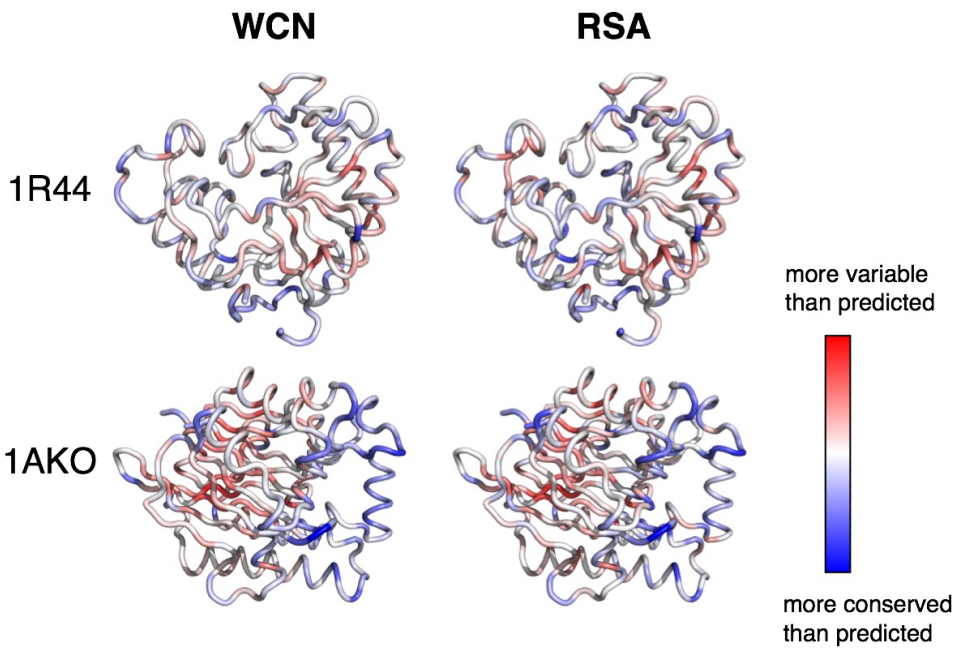




Echave et al., Nat. Rev. Genet. 2016

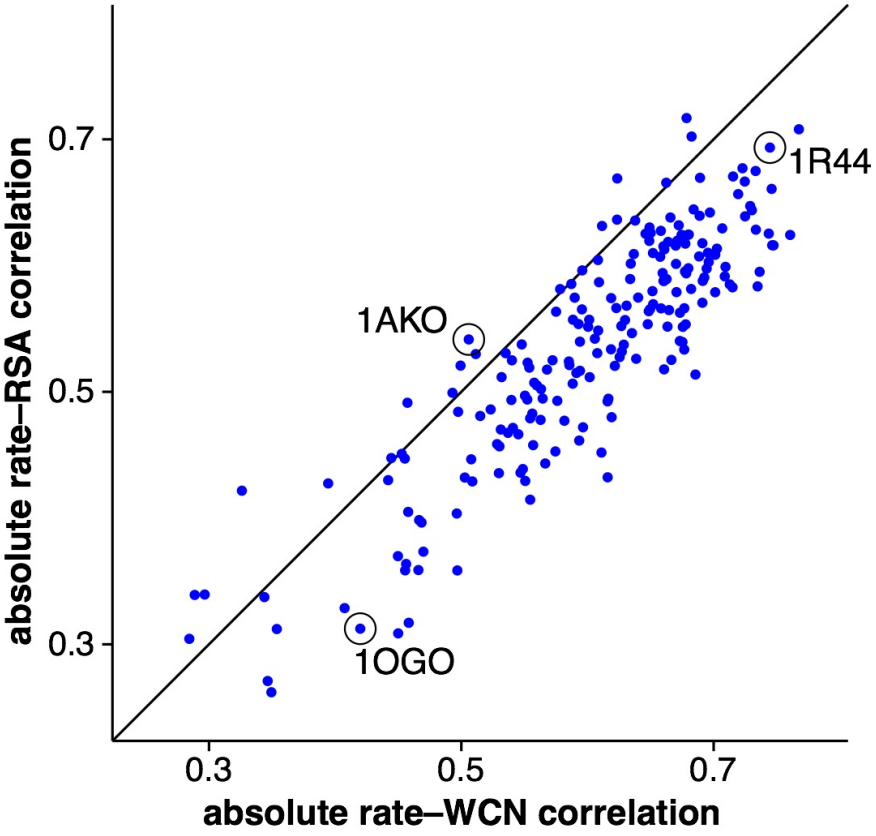
Structure is a good rate predictor for some proteins, but not so much for others

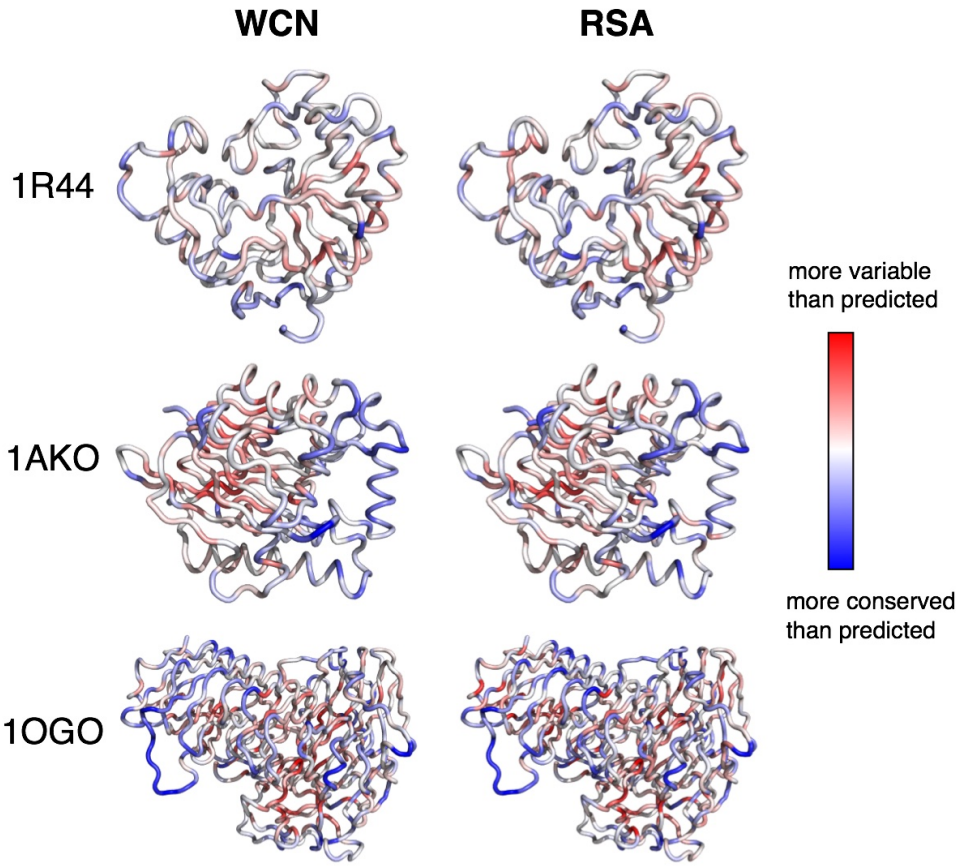




Echave et al., Nat. Rev. Genet. 2016

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





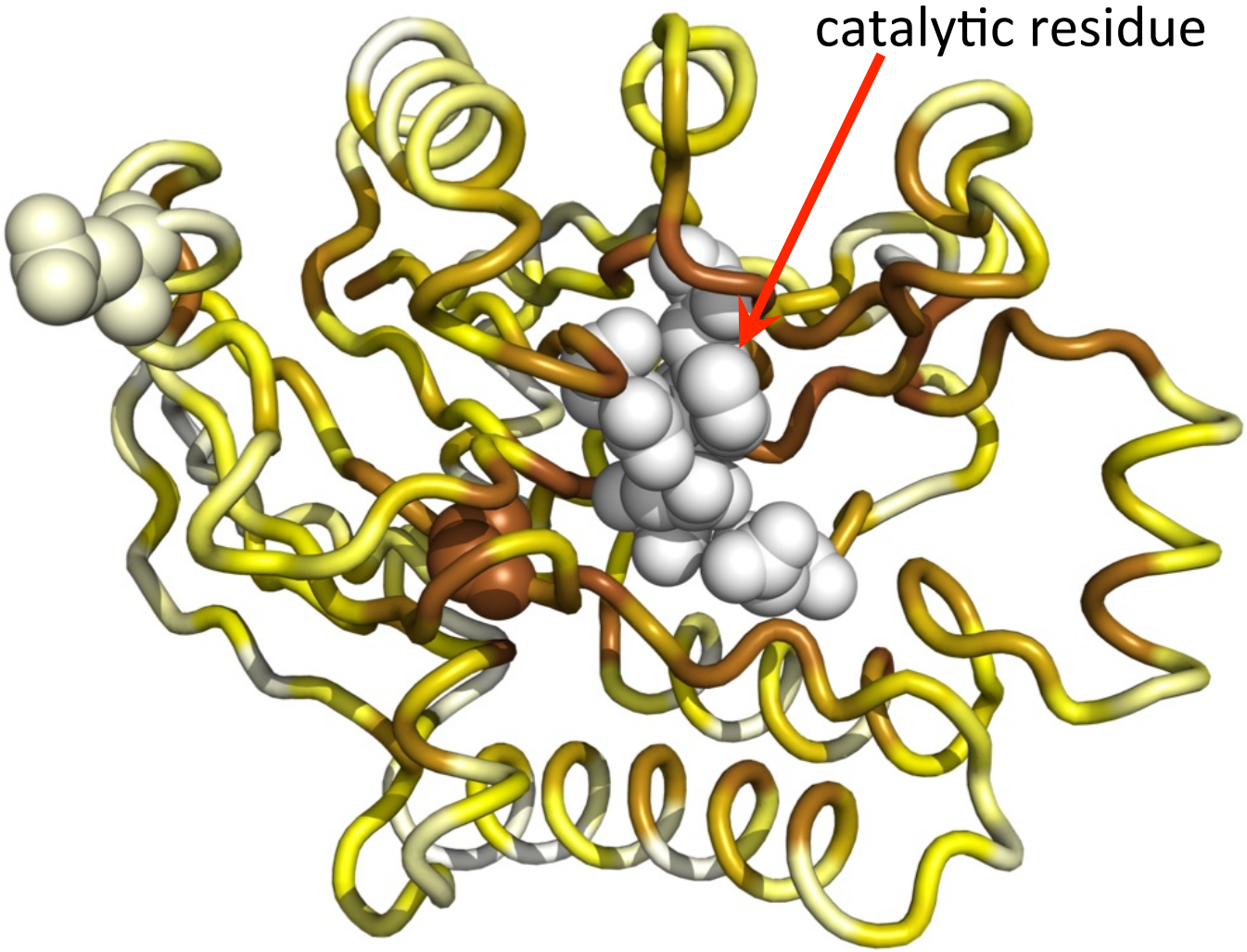
Echave et al., Nat. Rev. Genet. 2016

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

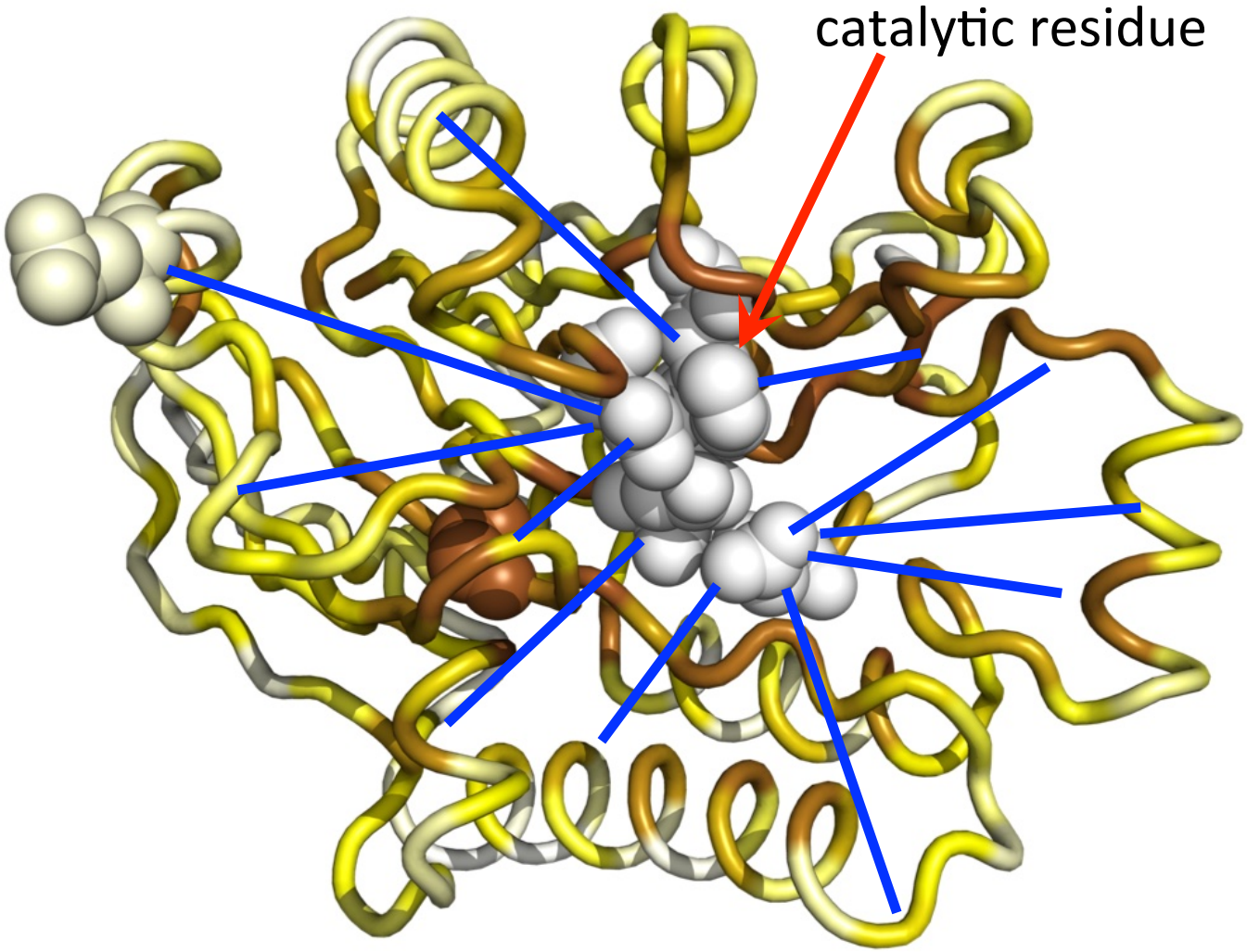
Work by graduate student Ben Jack

In collaboration with Julian Echave

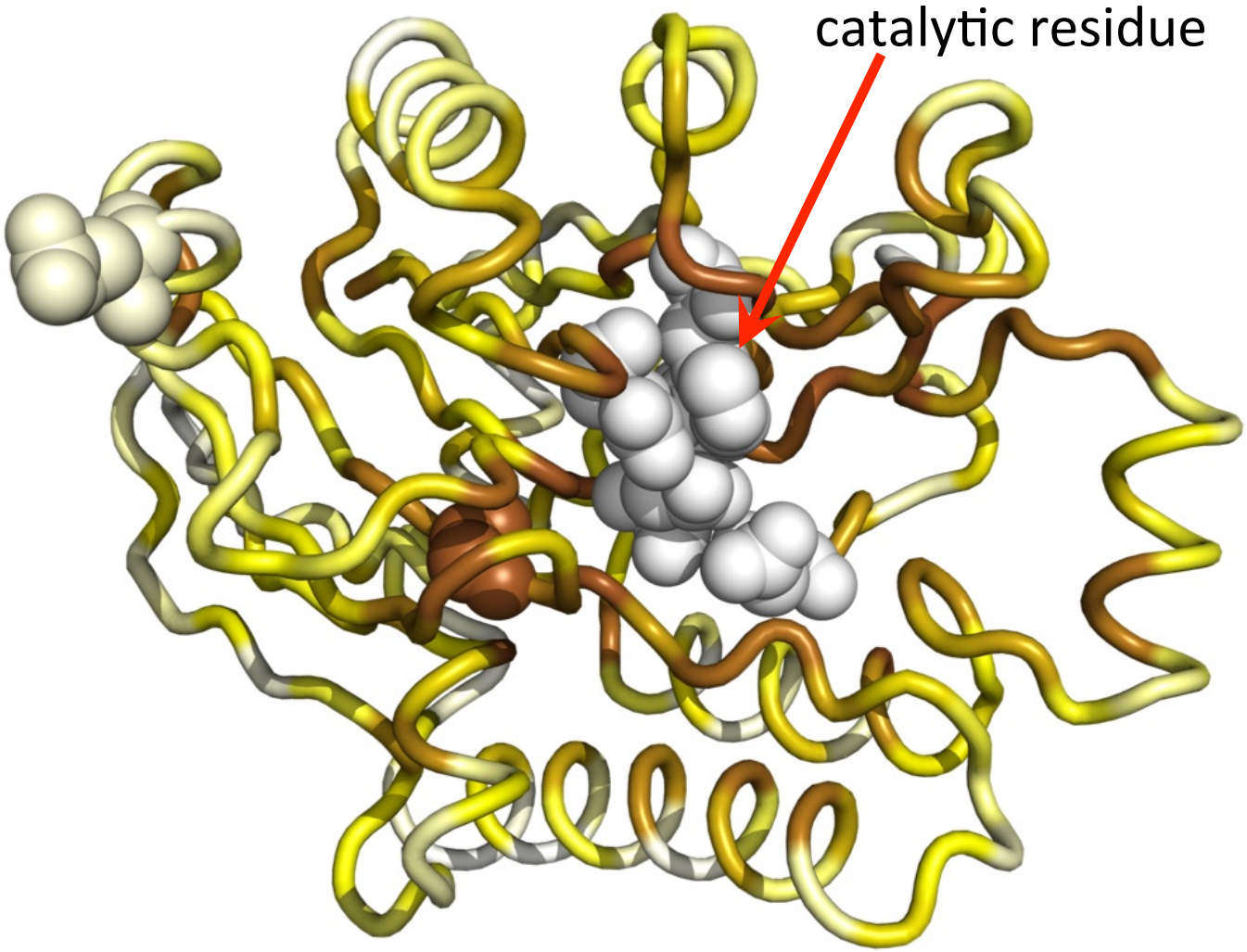
We measure function with distance to the nearest catalytic residue



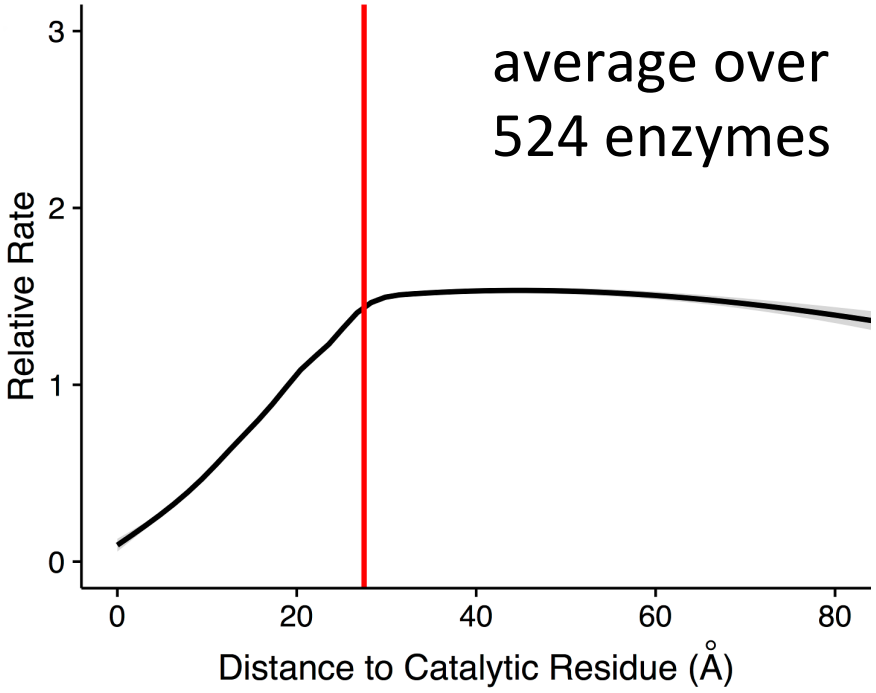
We measure function with distance to the nearest catalytic residue



We measure function with distance to the nearest catalytic residue

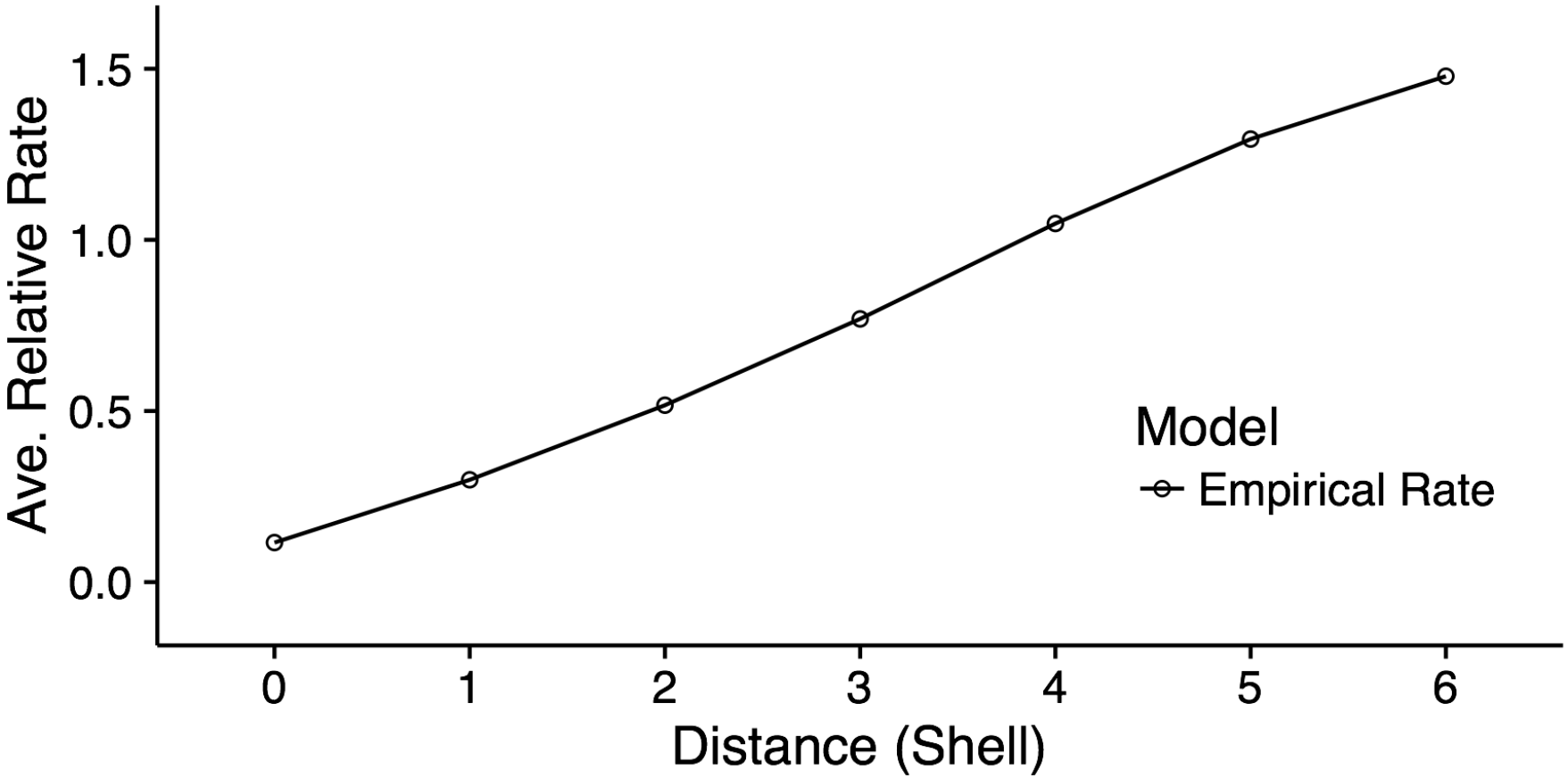


Catalytic residues impose long-range conservation gradients



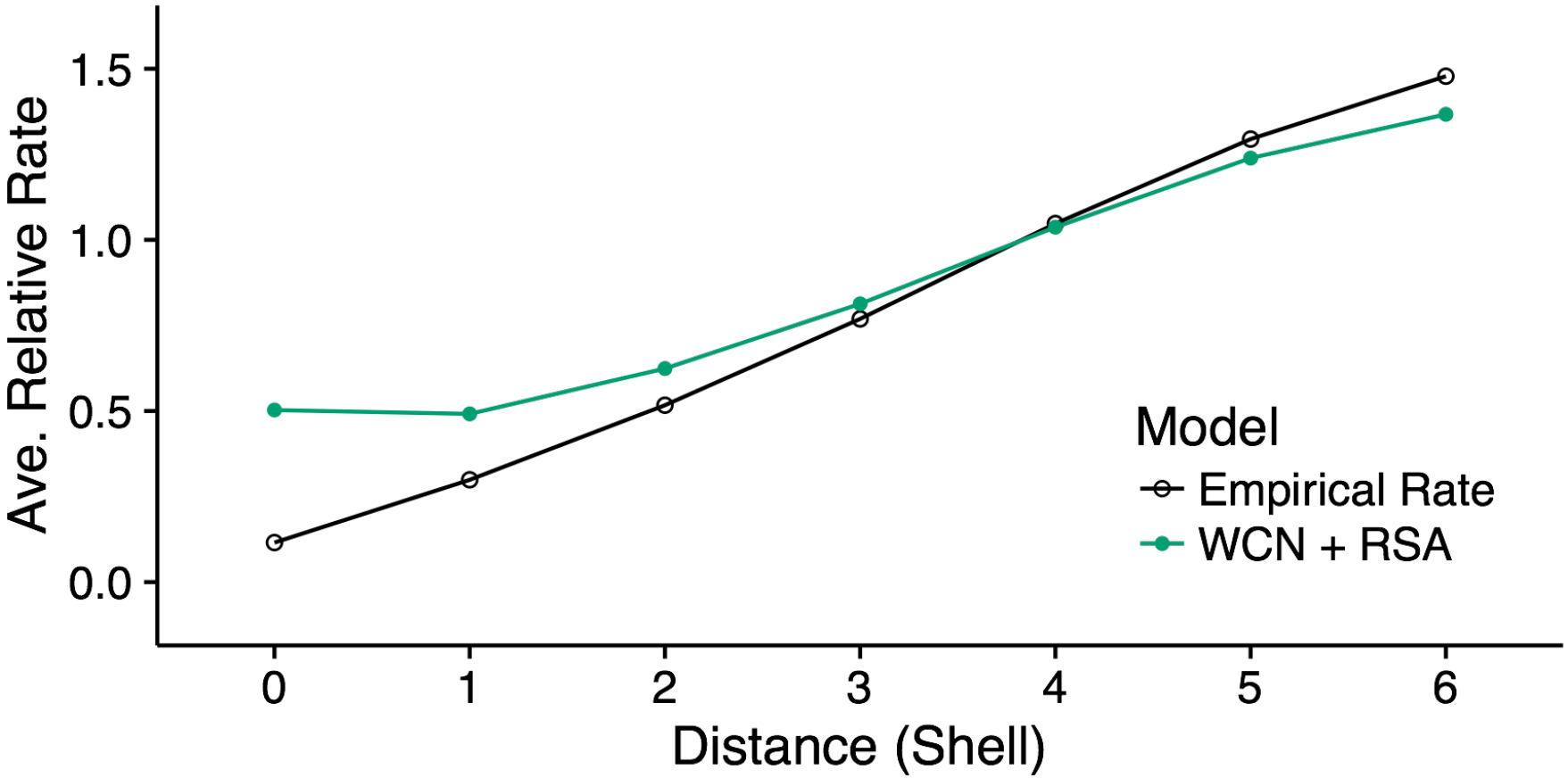
Jack et al., PLOS Biol. 2016

Distance is critical to accurately predict average rate near catalytic residues



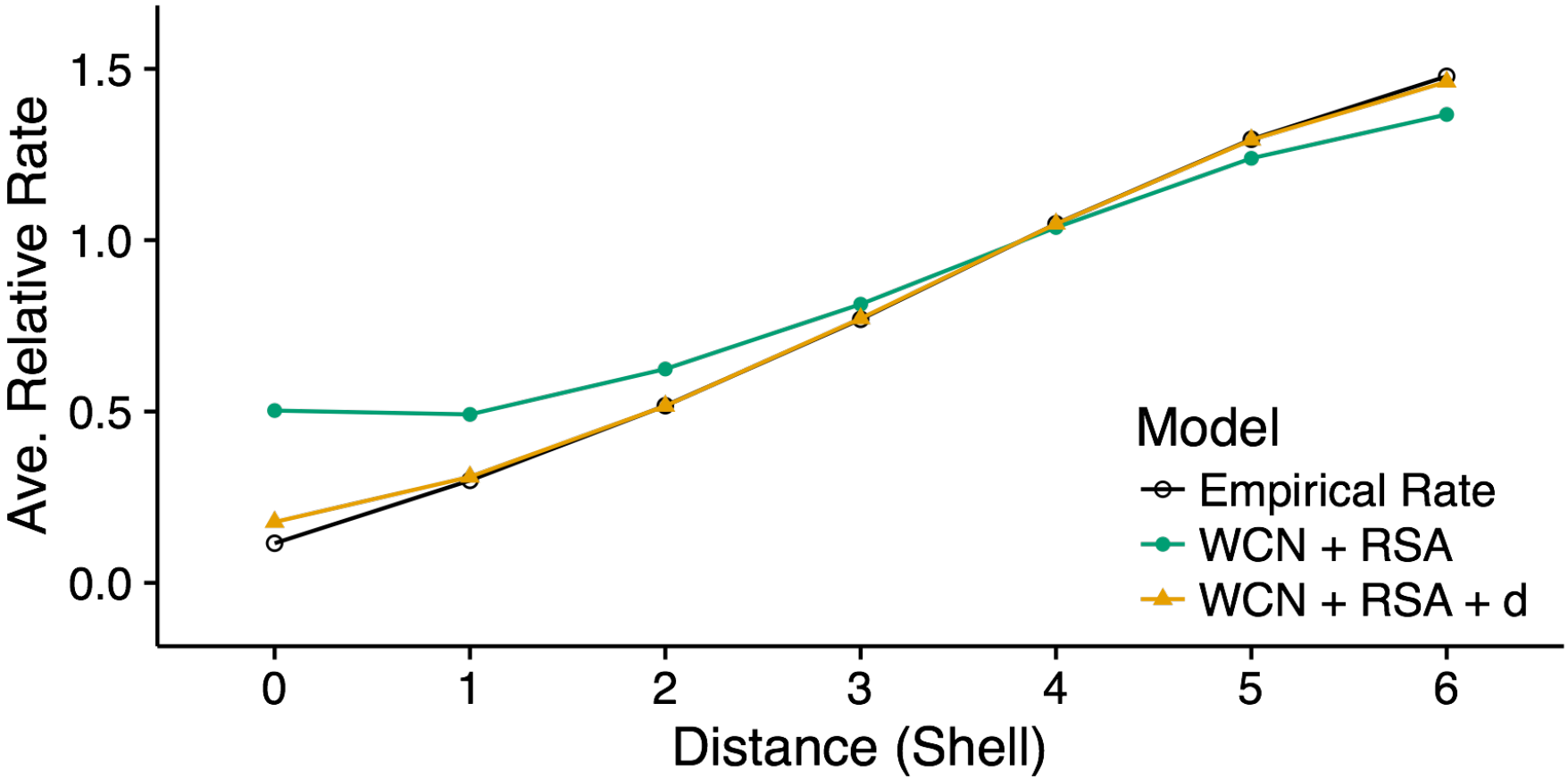
Jack et al., PLOS Biol. 2016

Distance is critical to accurately predict average rate near catalytic residues



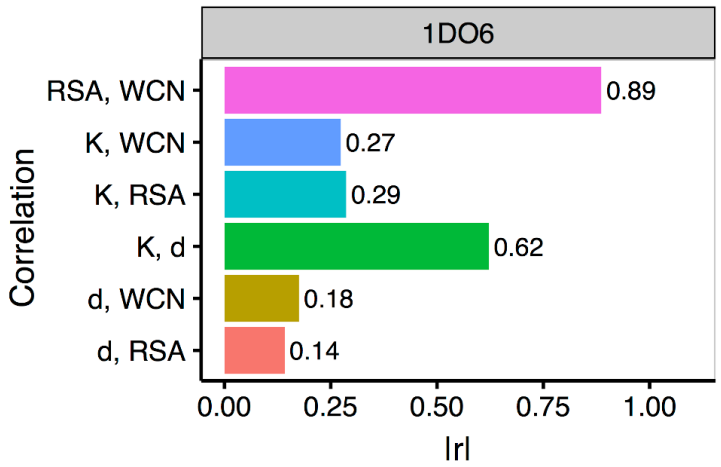
Jack et al., PLOS Biol. 2016

Distance is critical to accurately predict average rate near catalytic residues

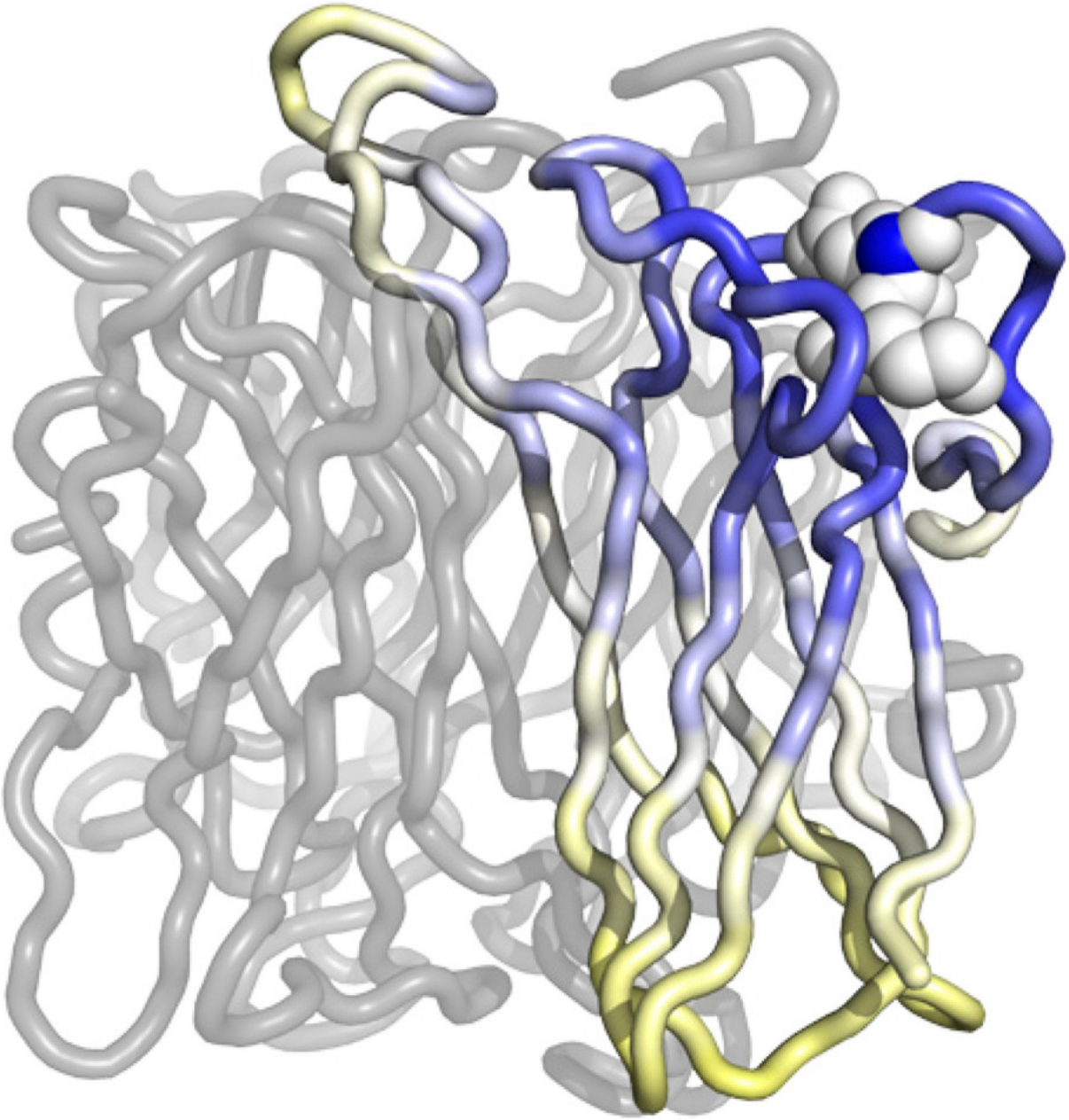


Jack et al., PLOS Biol. 2016

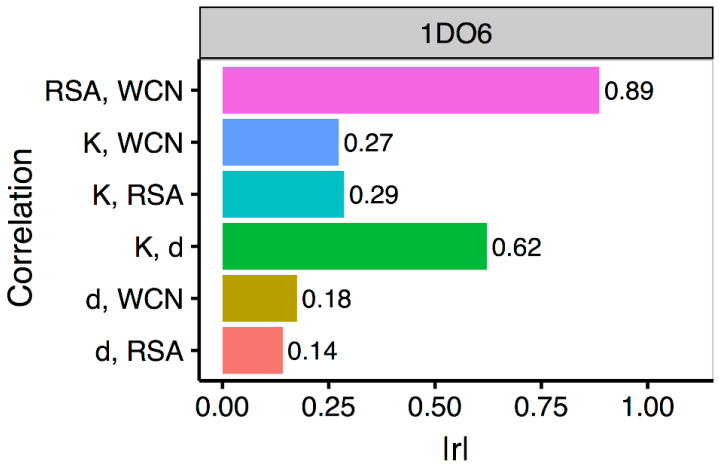
Effect is strongest for active sites on the protein surface

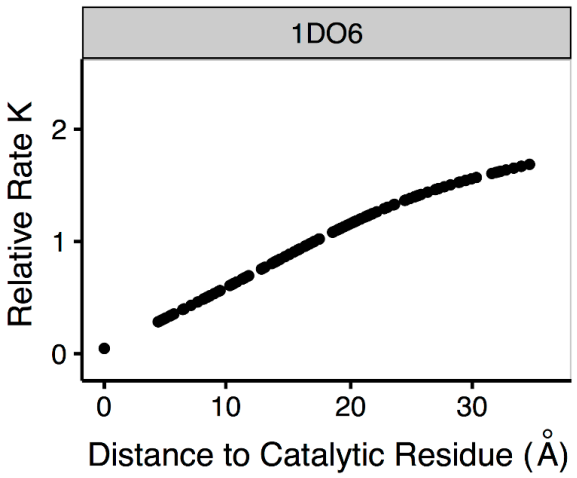


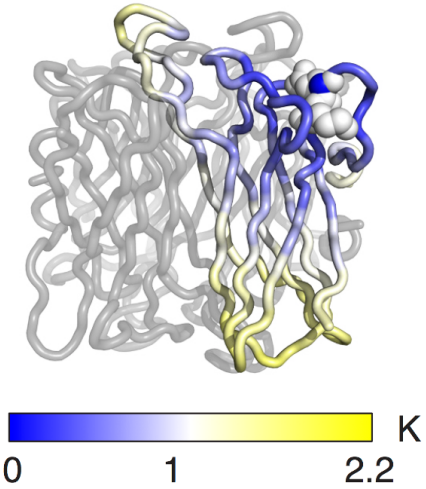
Jack et al., PLOS Biol. 2016

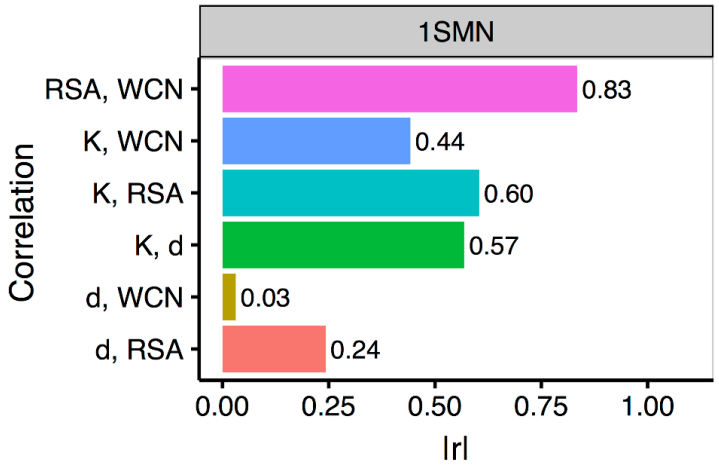


Effect is strongest for active sites on the protein surface

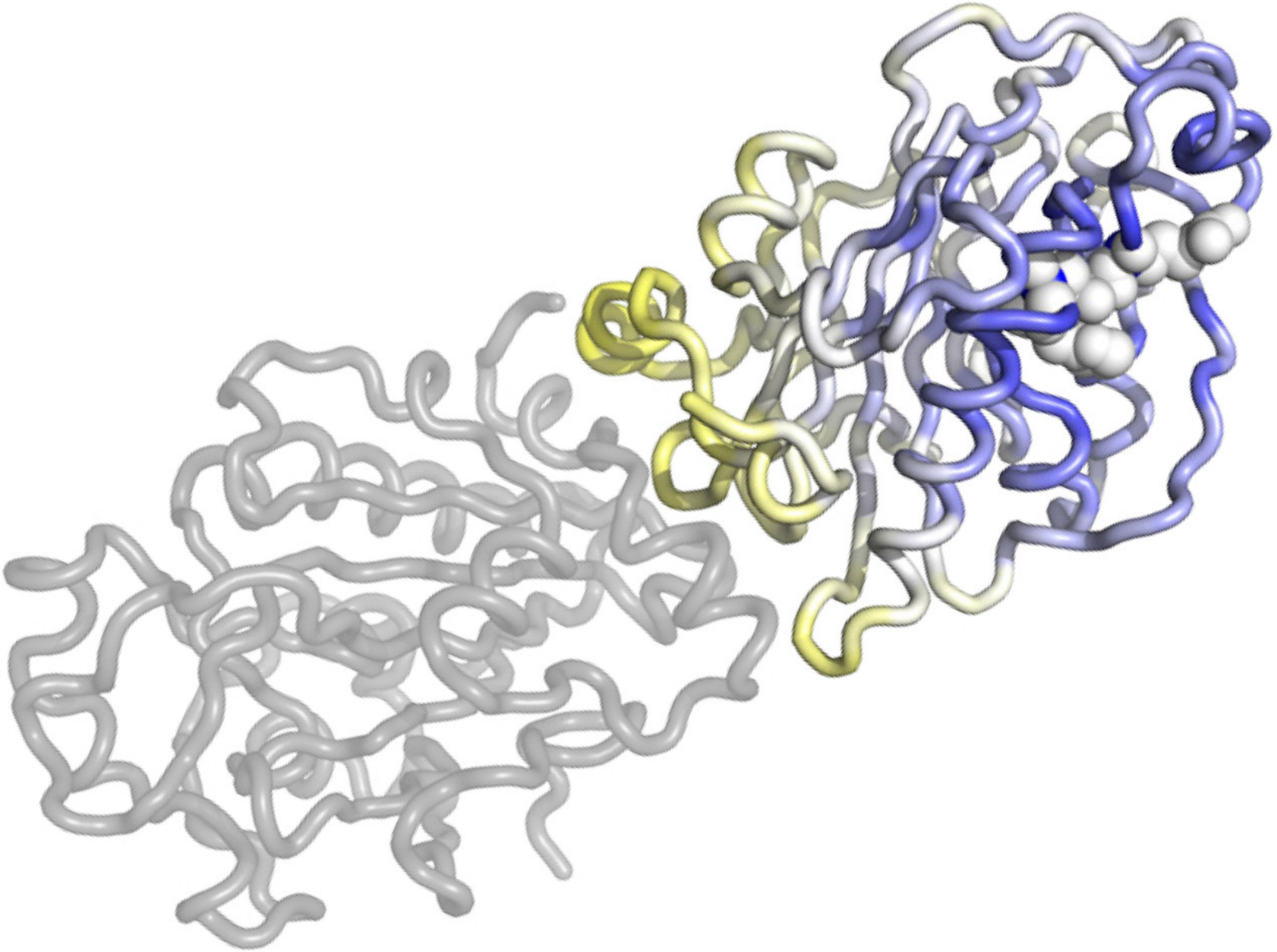




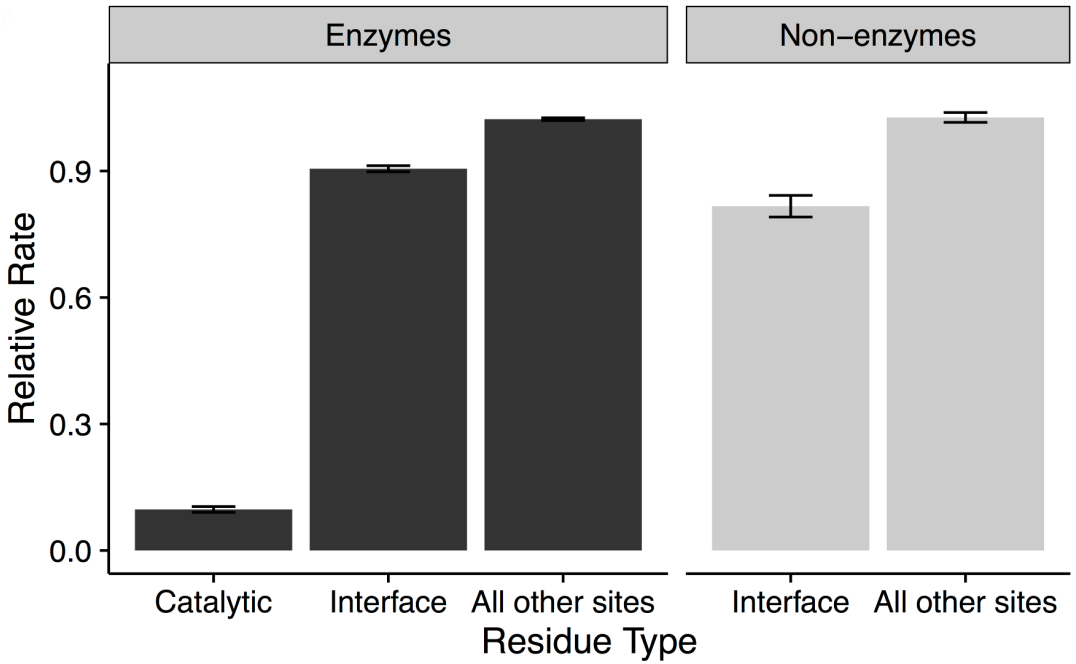




Jack et al., PLOS Biol. 2016



Catalytic residues impose much stronger constraints than interface residues



Jack et al., PLOS Biol. 2016

Take-home message

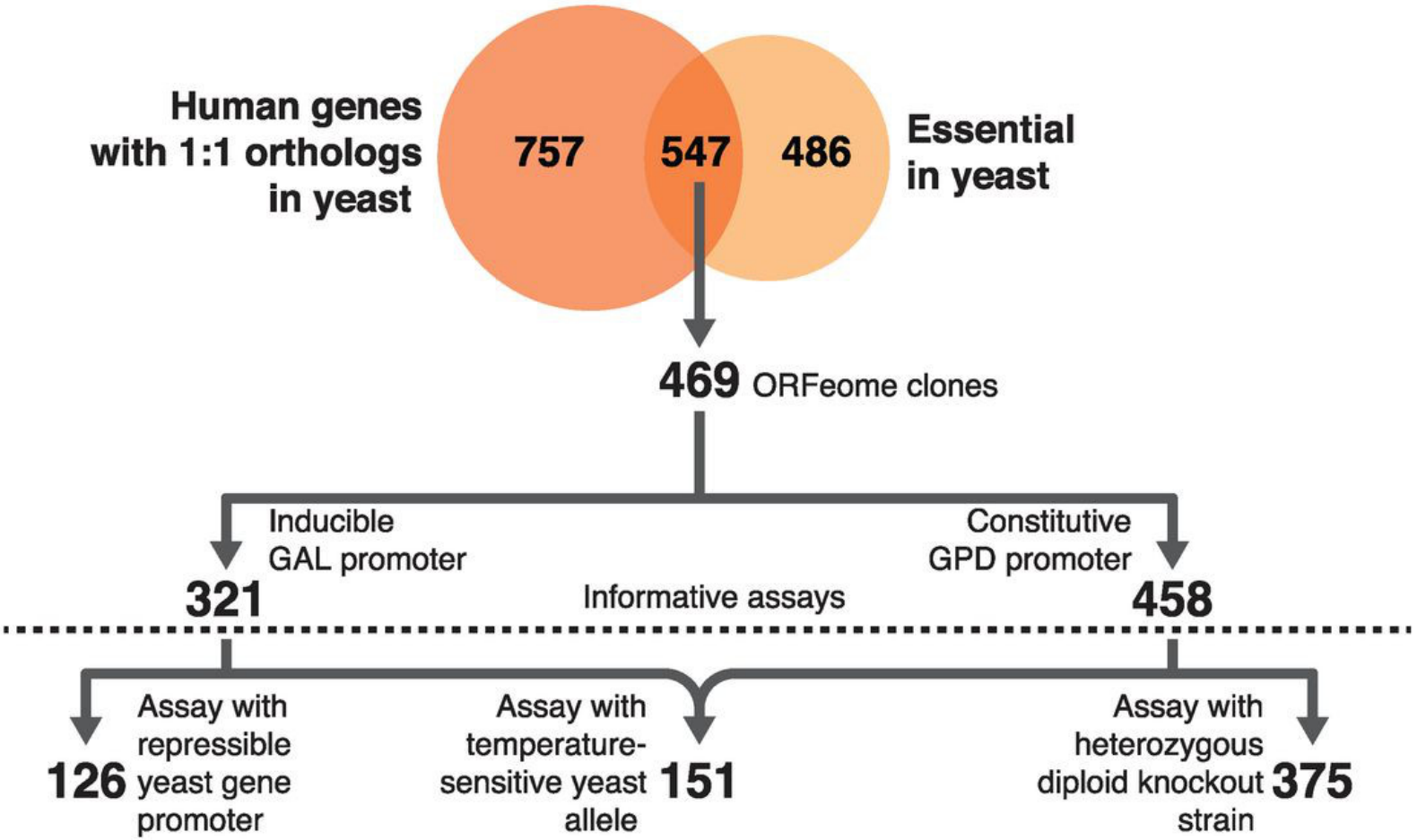
Catalytic residues generate long-range selection gradients covering most of a typical enzyme structure.

Part II: When can proteins interact with ancestral partners?

In collaboration with the Marcotte lab

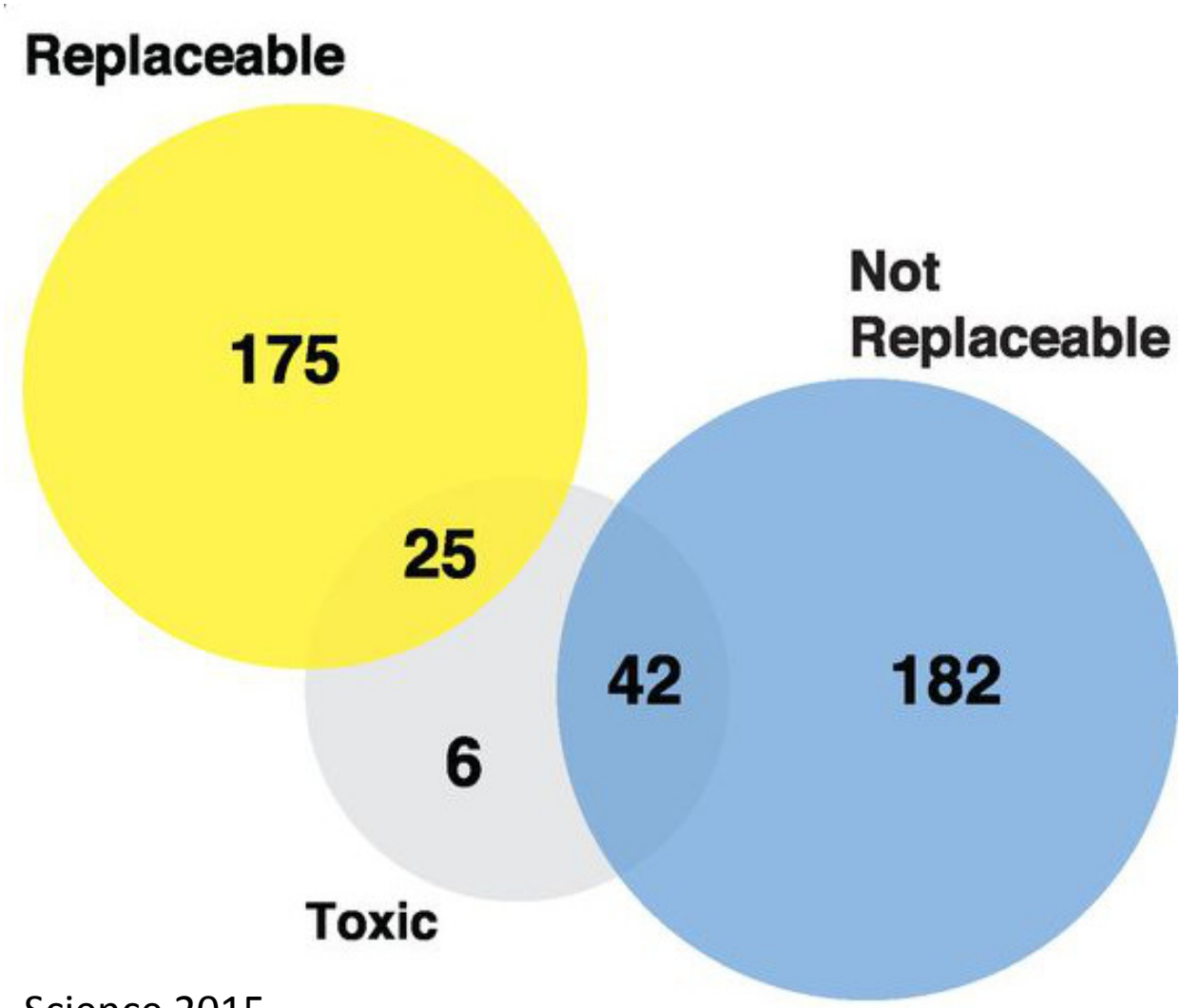
Protein evolution simulations by Austin Meyer

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



Kachroo et al., Science 2015

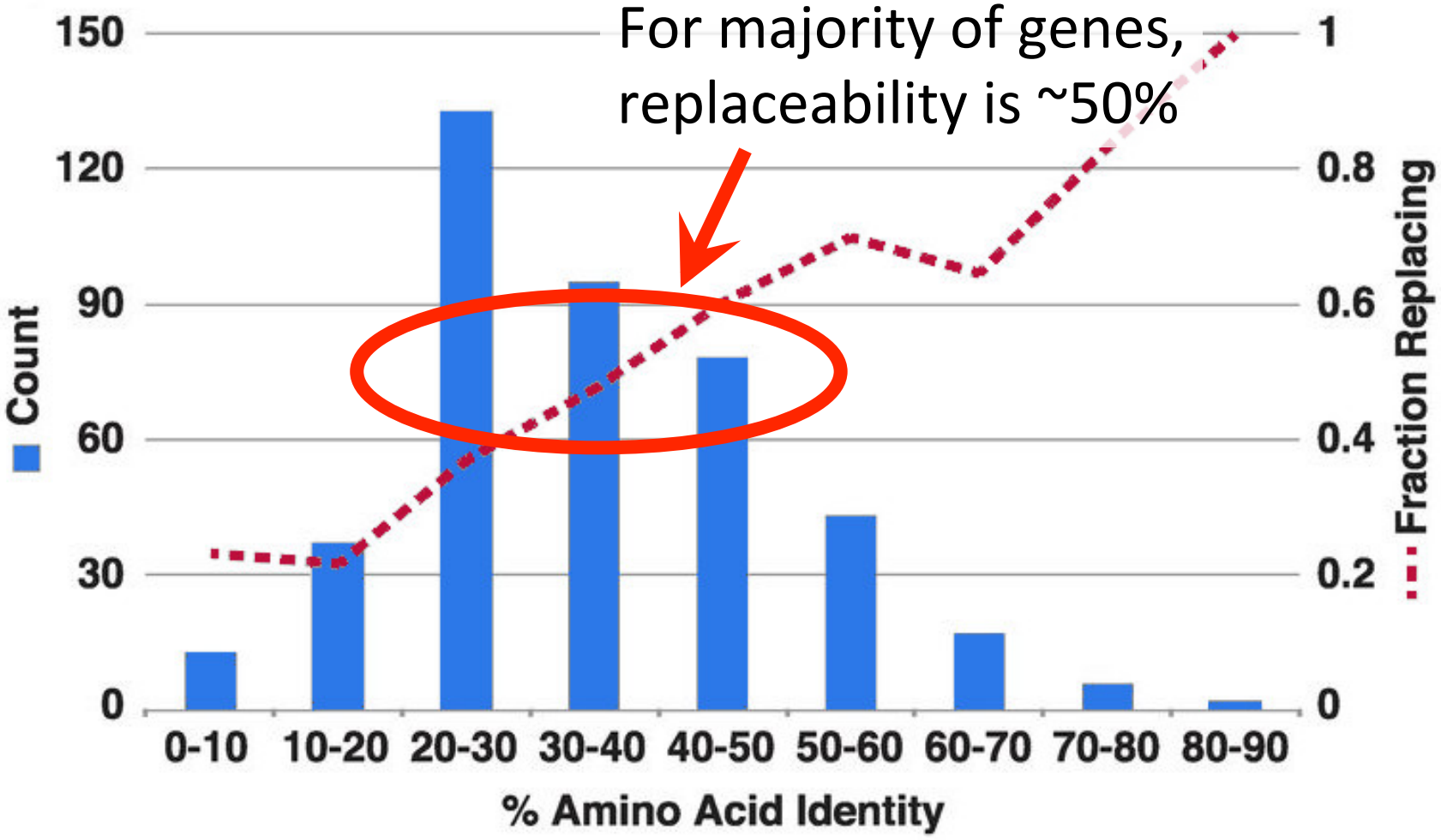
Yes!



Kachroo et al., Science 2015

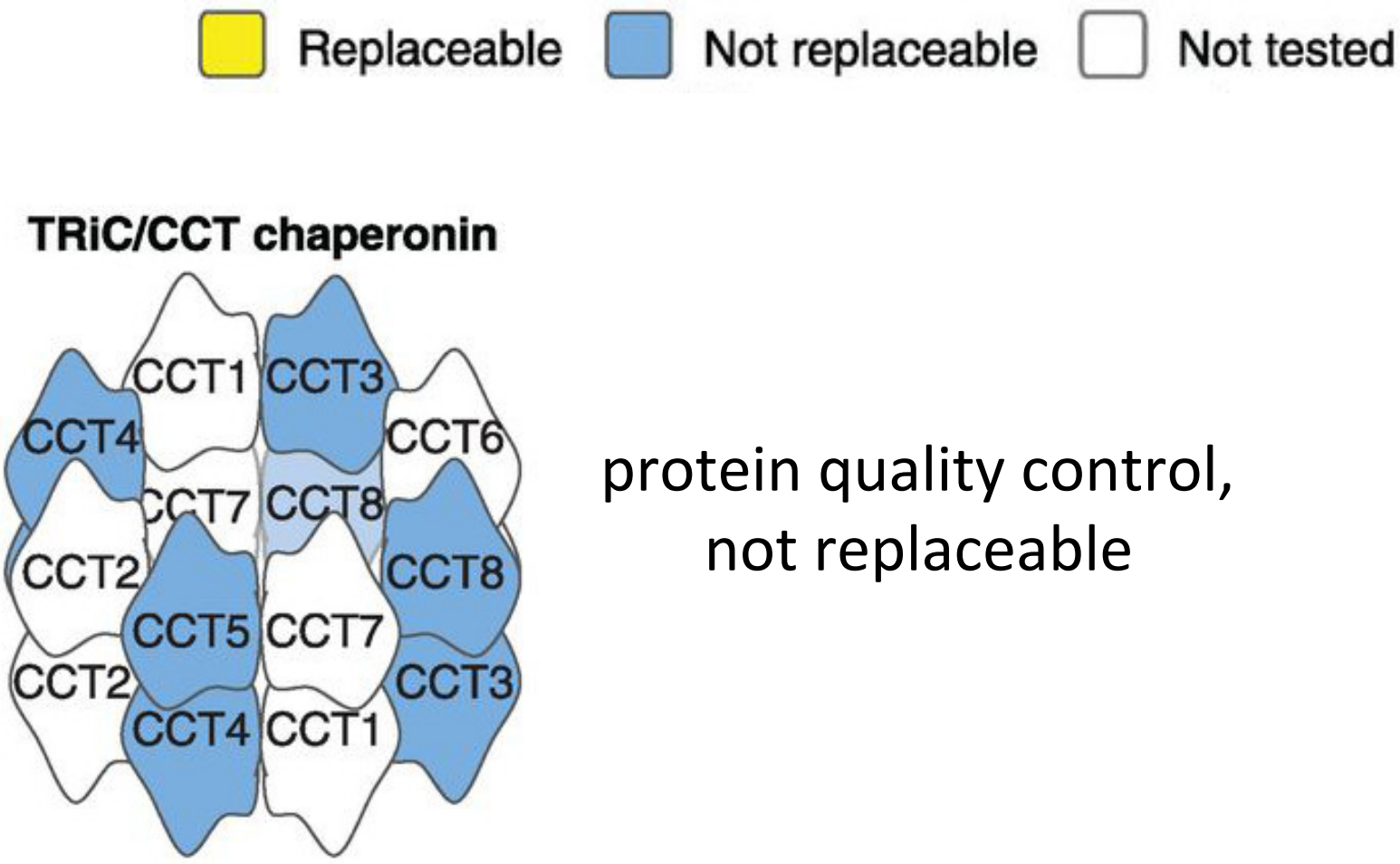
What predicts whether a gene will be replaceable?

Sequence divergence is not a good predictor of replaceability



Kachroo et al., Science 2015

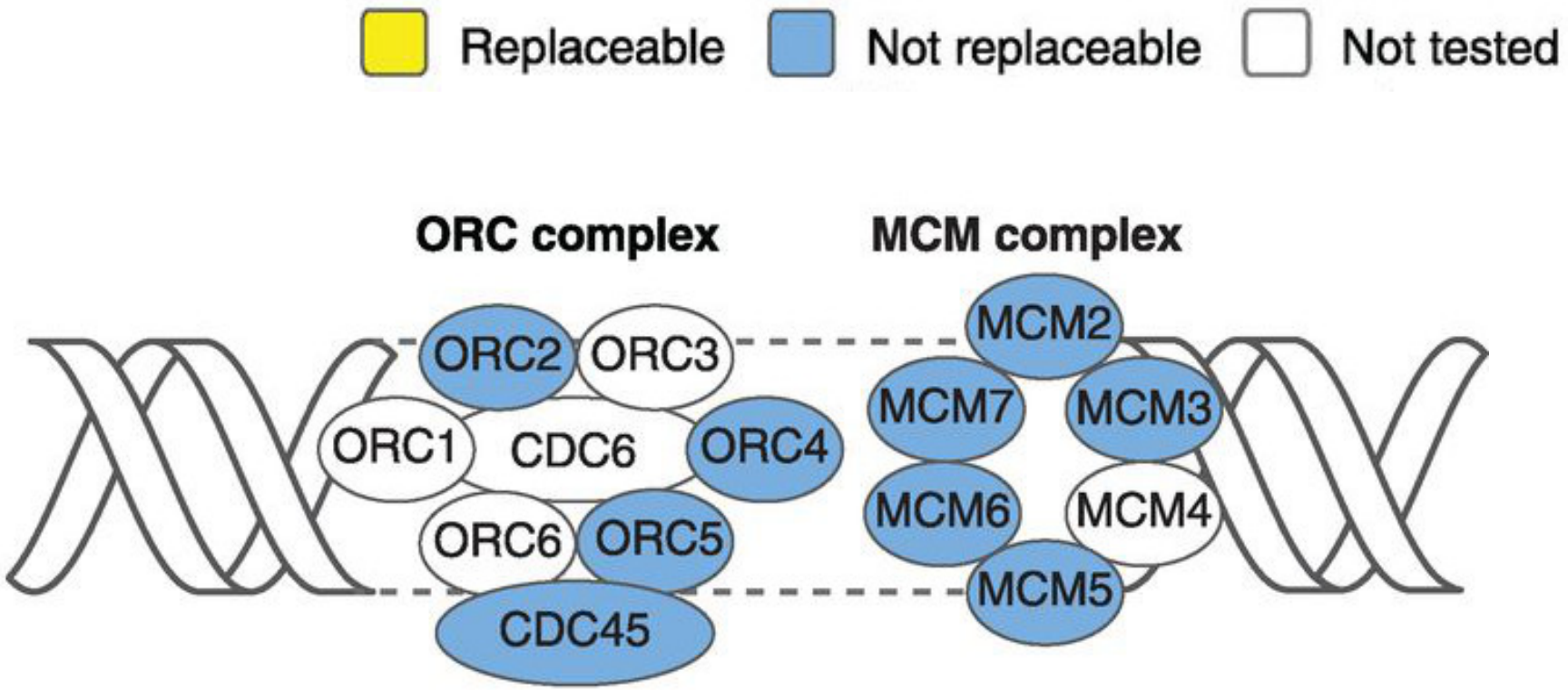
Genes tend to be replaceable in modules



protein quality control,
not replaceable

Kachroo et al., Science 2015

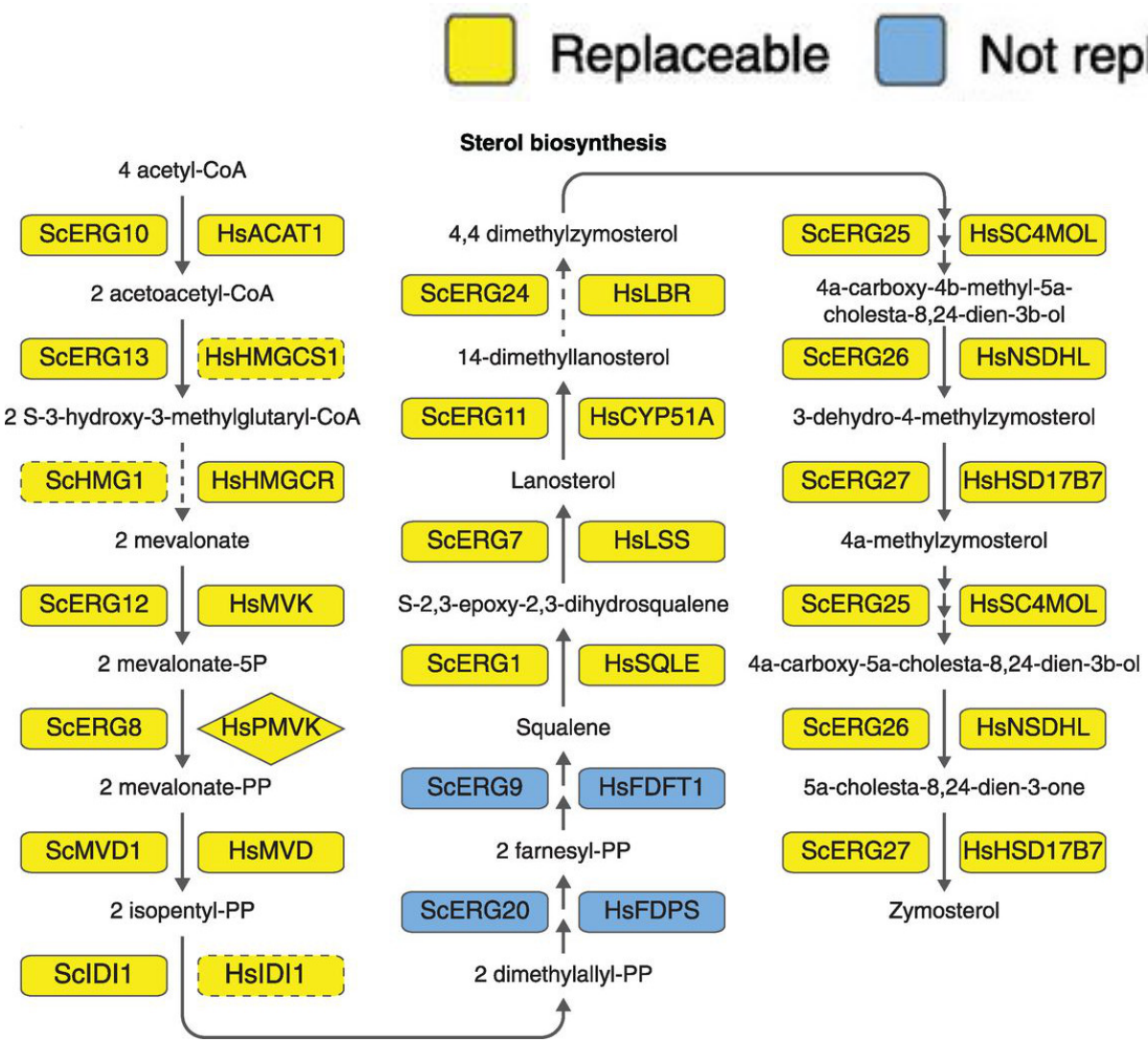
Genes tend to be replaceable in modules



genome replication,
not replaceable

Kachroo et al., Science 2015

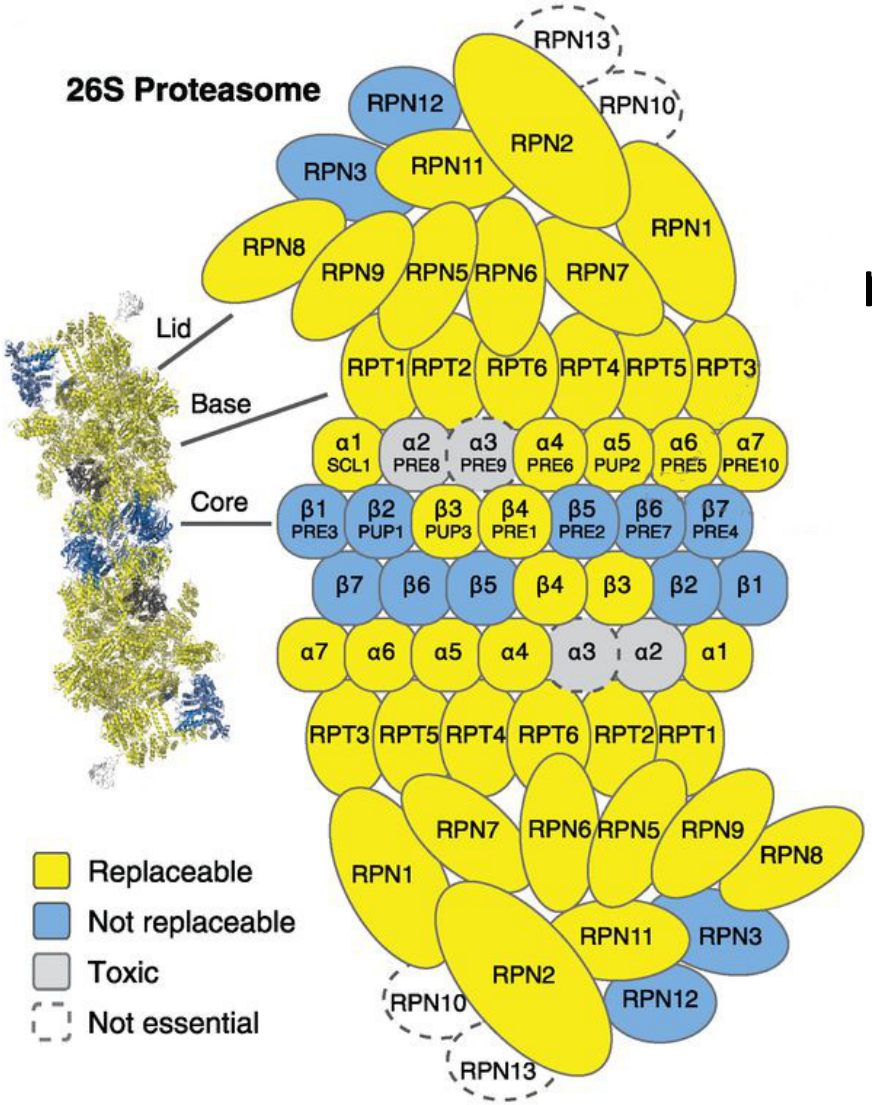
Genes tend to be replaceable in modules



sterol biosynthesis, replaceable

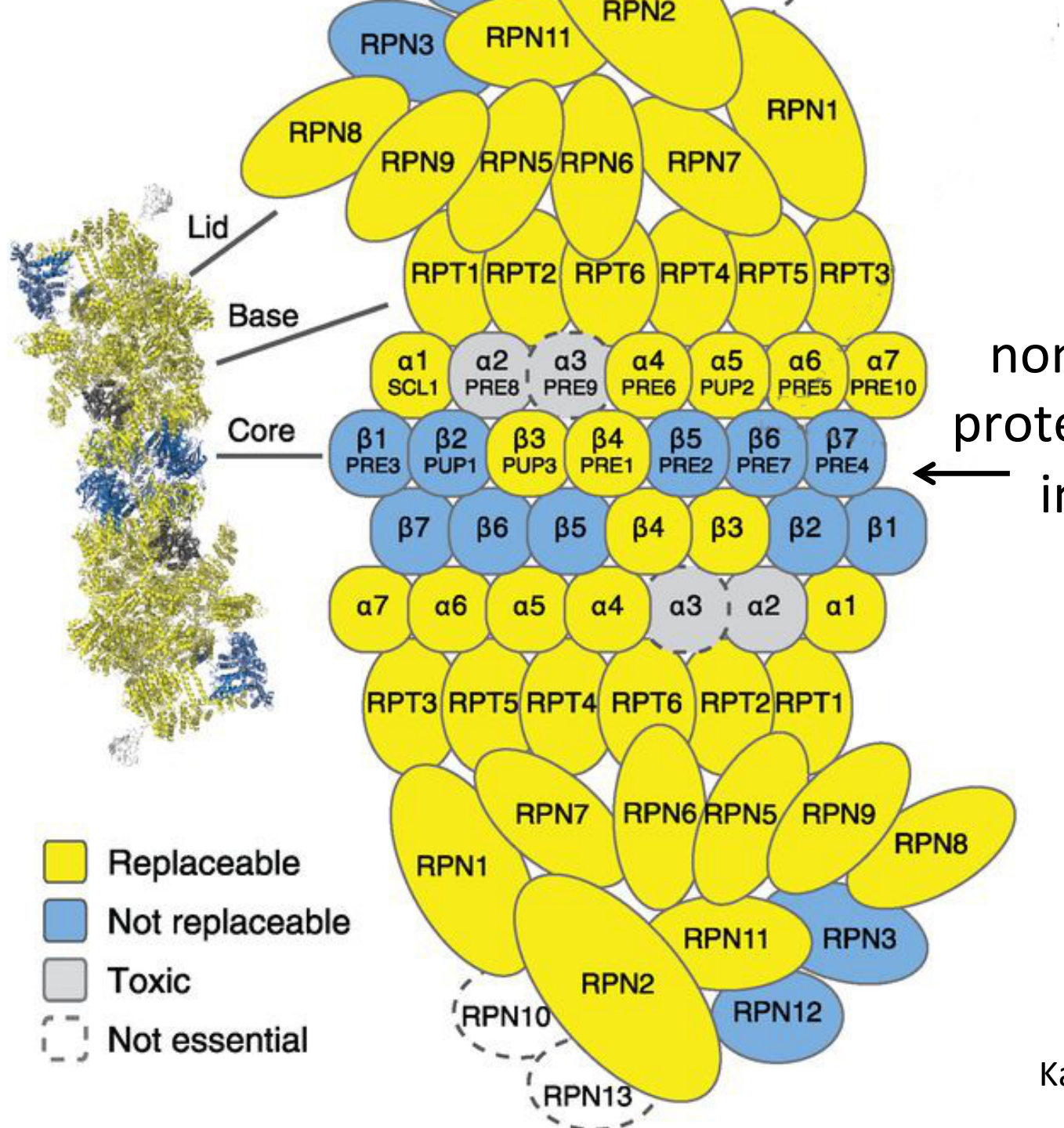
Kachroo et al., Science 2015

Genes tend to be replaceable in modules



proteasome,
mostly replaceable

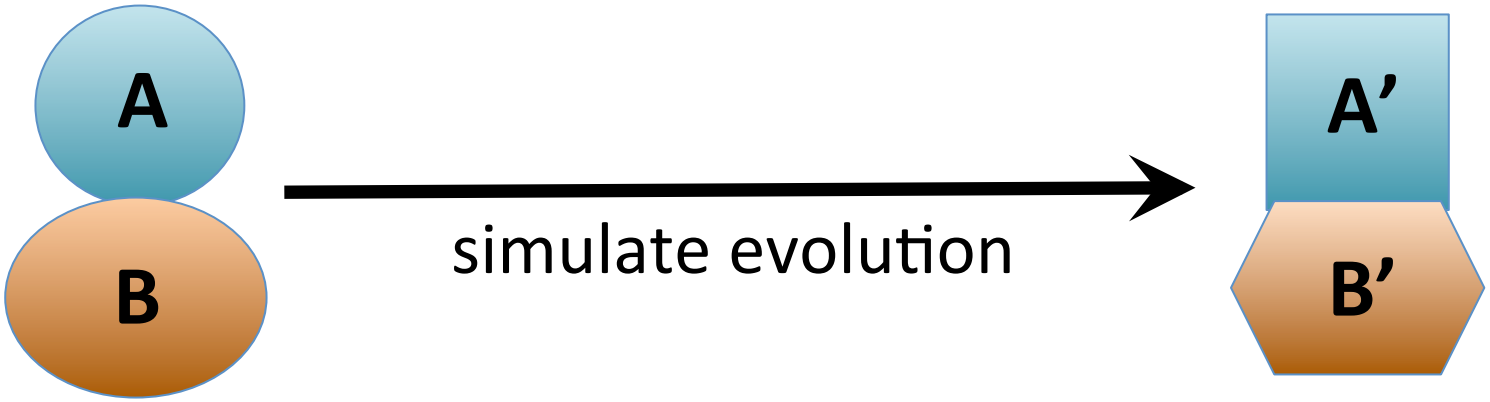
Kachroo et al., Science 2015

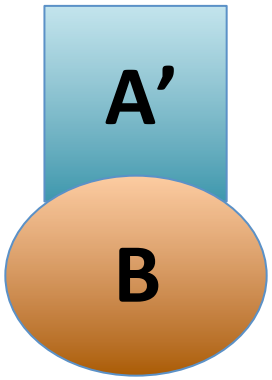


non-replaceable proteins are mostly in the β ring

Kachroo et al., Science 2015

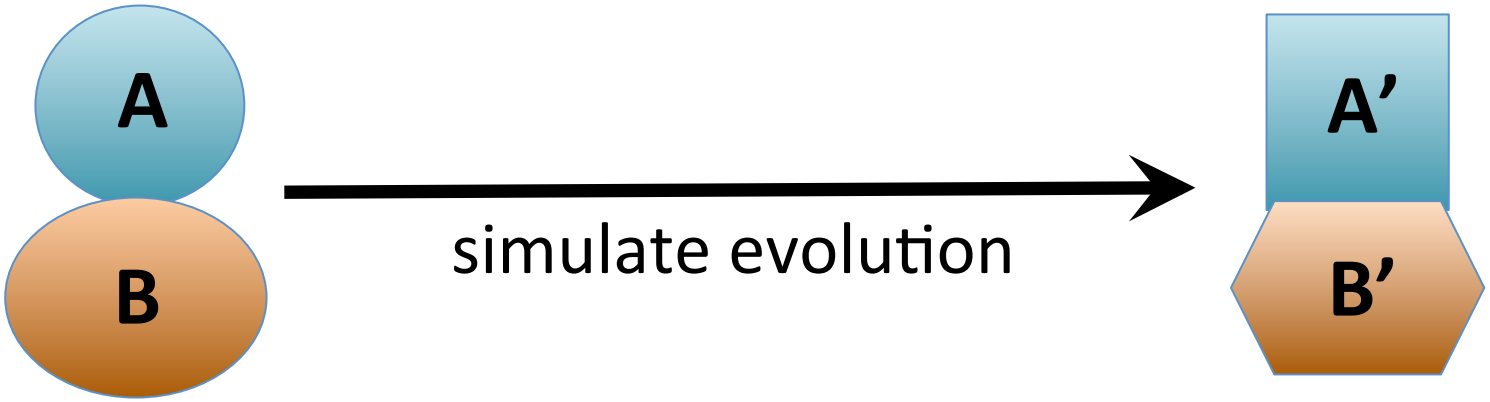
Hypothesis: Conserved protein–protein interactions cause replaceability

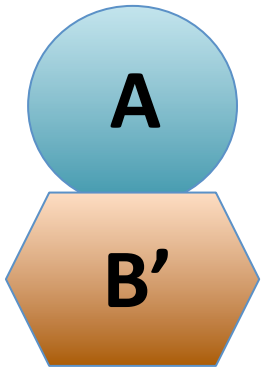




Does evolved A’ bind to ancestral B?

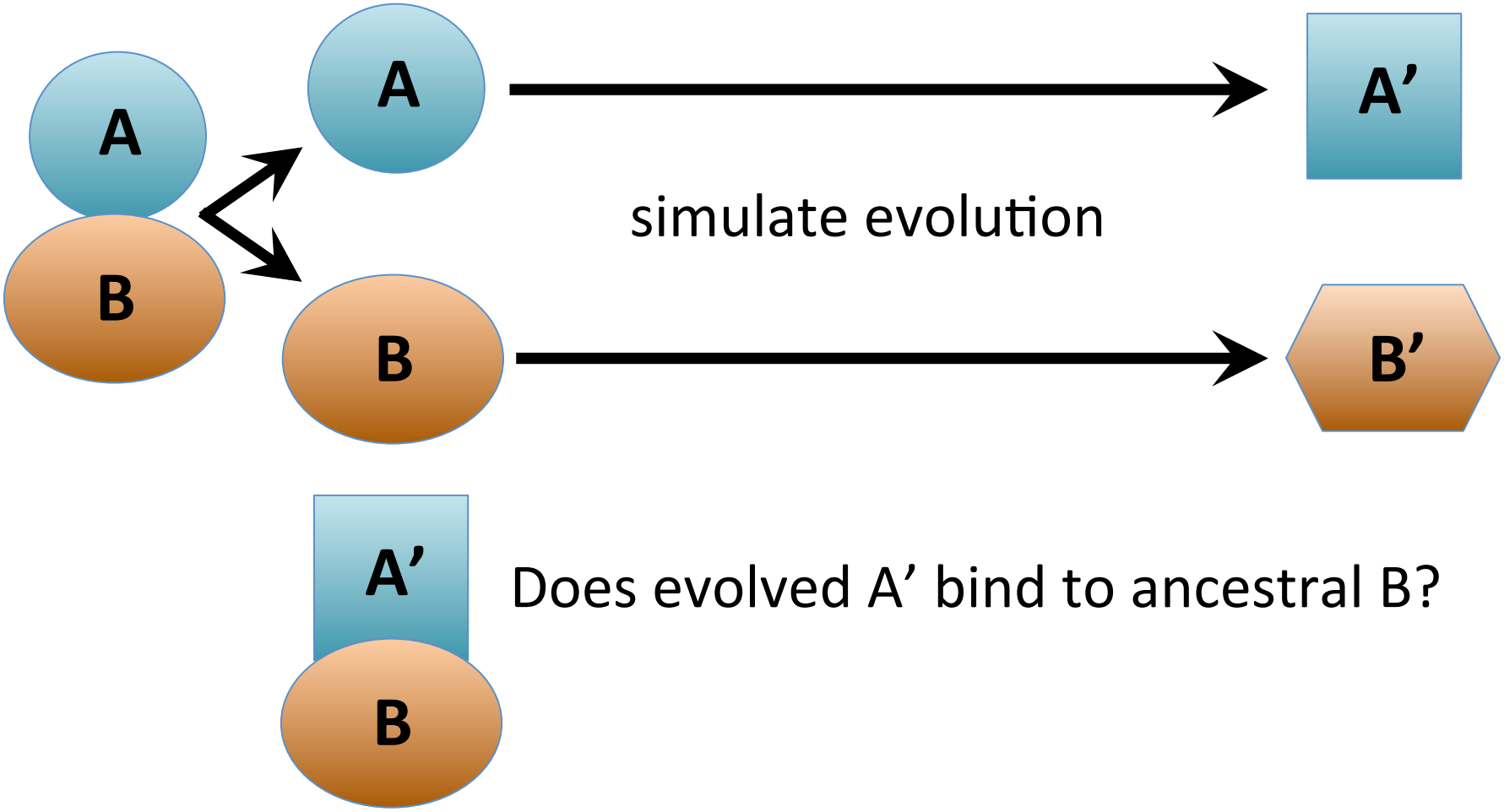
Hypothesis: Conserved protein–protein interactions cause replaceability



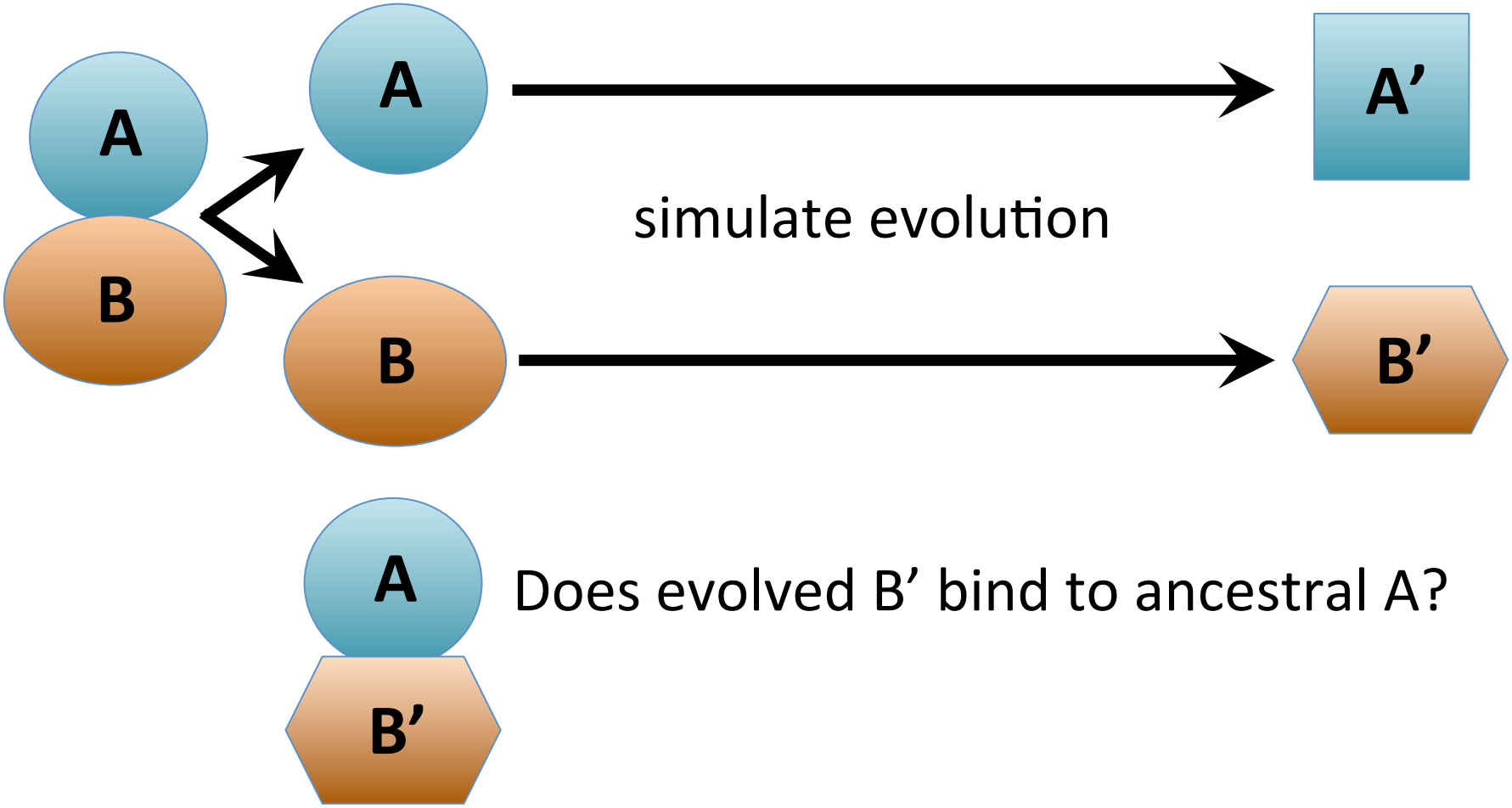


Does evolved B’ bind to ancestral A?

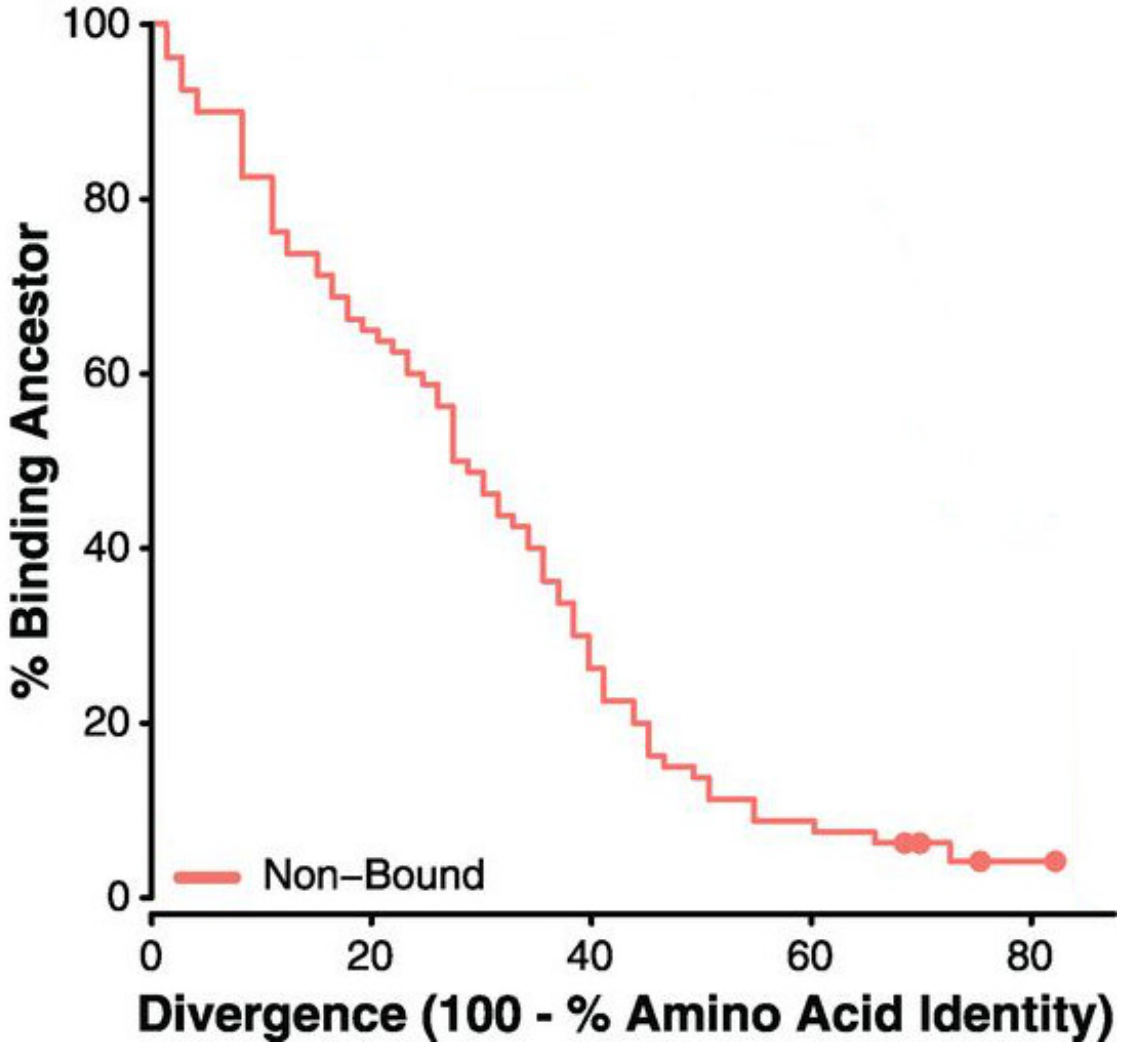
Control simulation: Evolve without binding



Control simulation: Evolve without binding

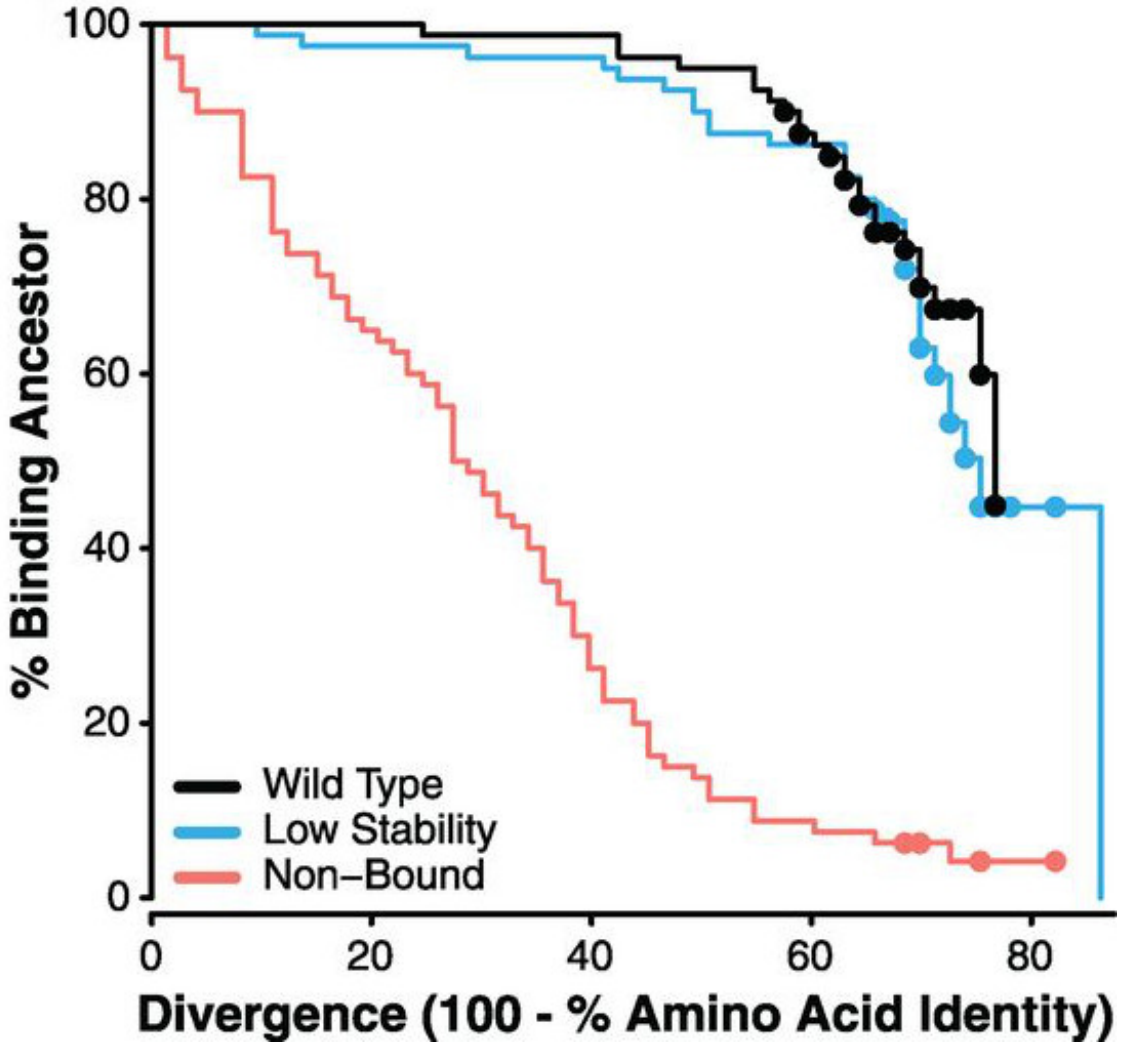


Binding to ancestor declines rapidly in the control scenario



Kachroo et al., Science 2015

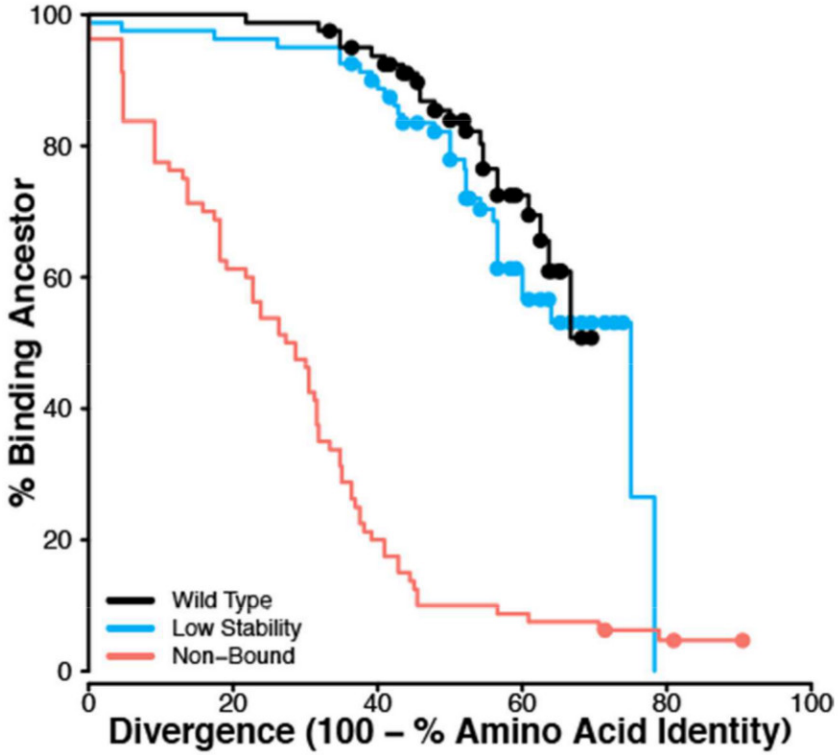
Binding to ancestor declines rapidly in the control scenario, but not in WT scenario



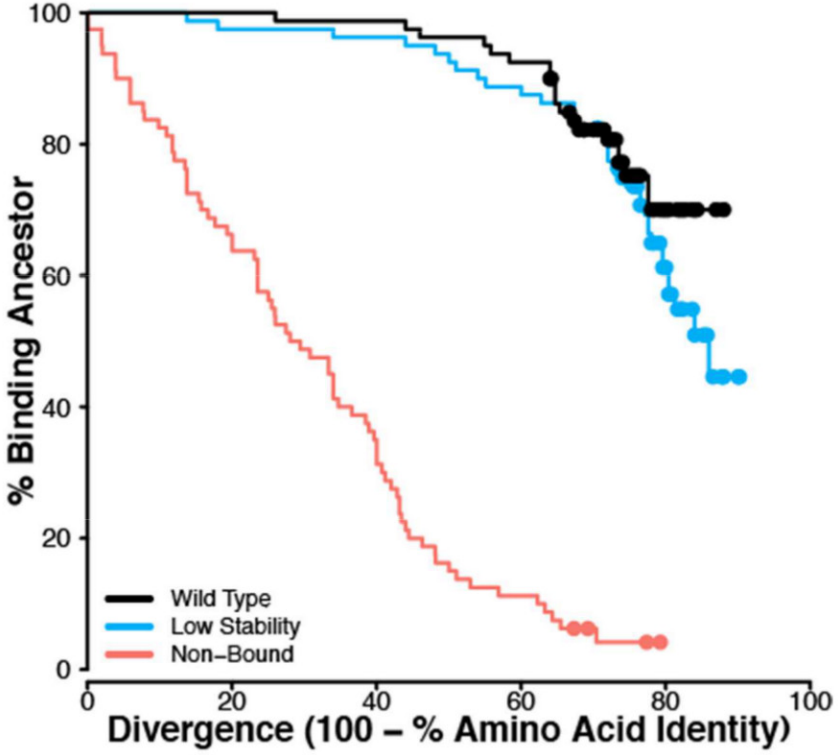
Kachroo et al., Science 2015

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.

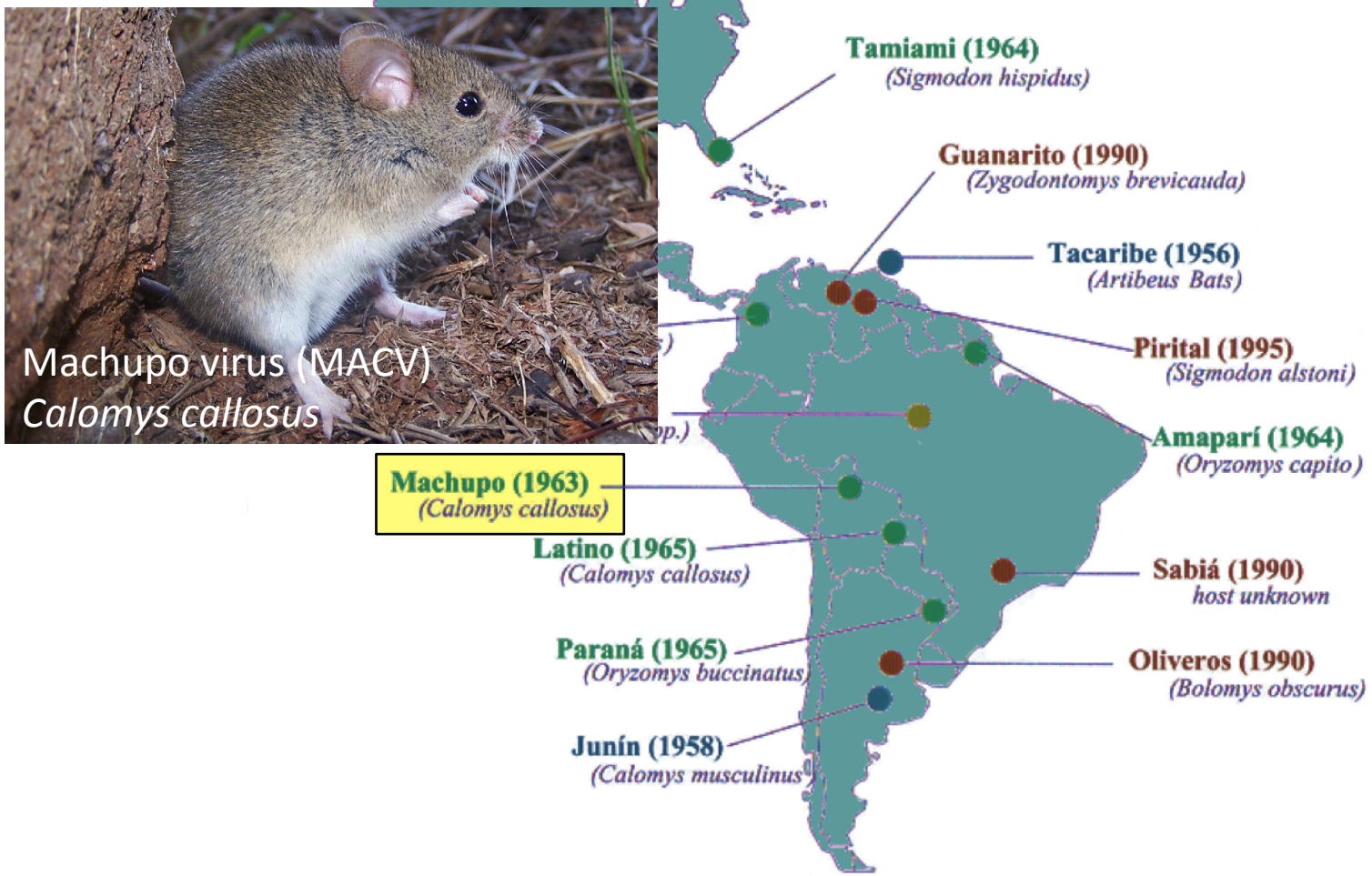
Part III: Can we predict host specificity in arenaviruses?

In collaboration with Sawyer, Ellington, and Georgiou labs

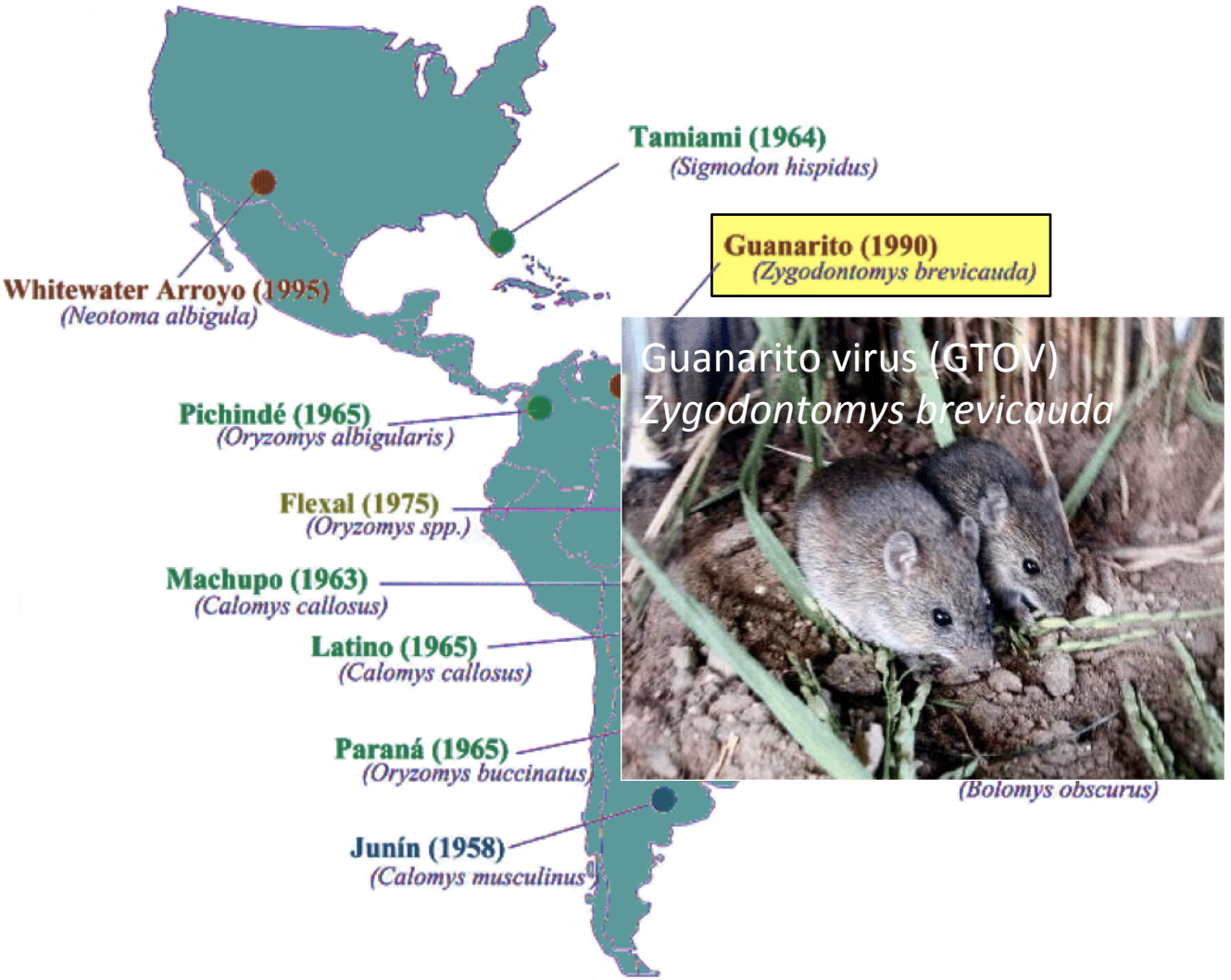
Work by Eleisha Jackson, Austin Meyer, Oana Lungu

New-World arenaviruses are rodent/bat viruses found all over the Americas

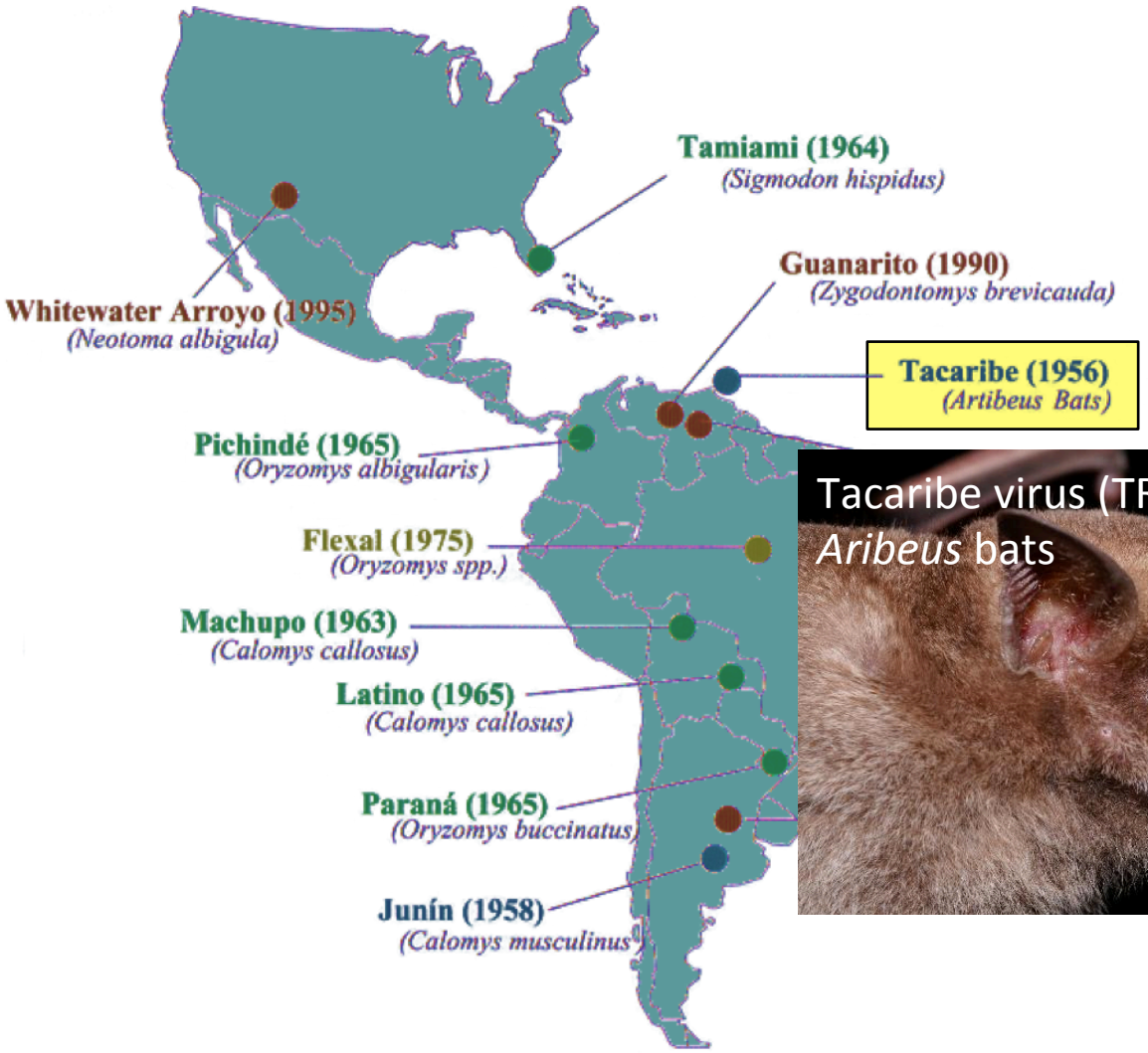
New-World arenaviruses are rodent/bat viruses found all over the Americas

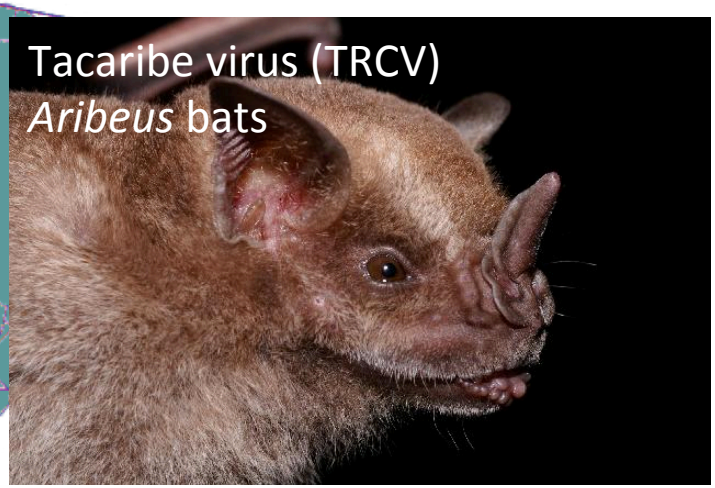


New-World arenaviruses are rodent/bat viruses found all over the Americas



New-World arenaviruses are rodent/bat viruses found all over the Americas





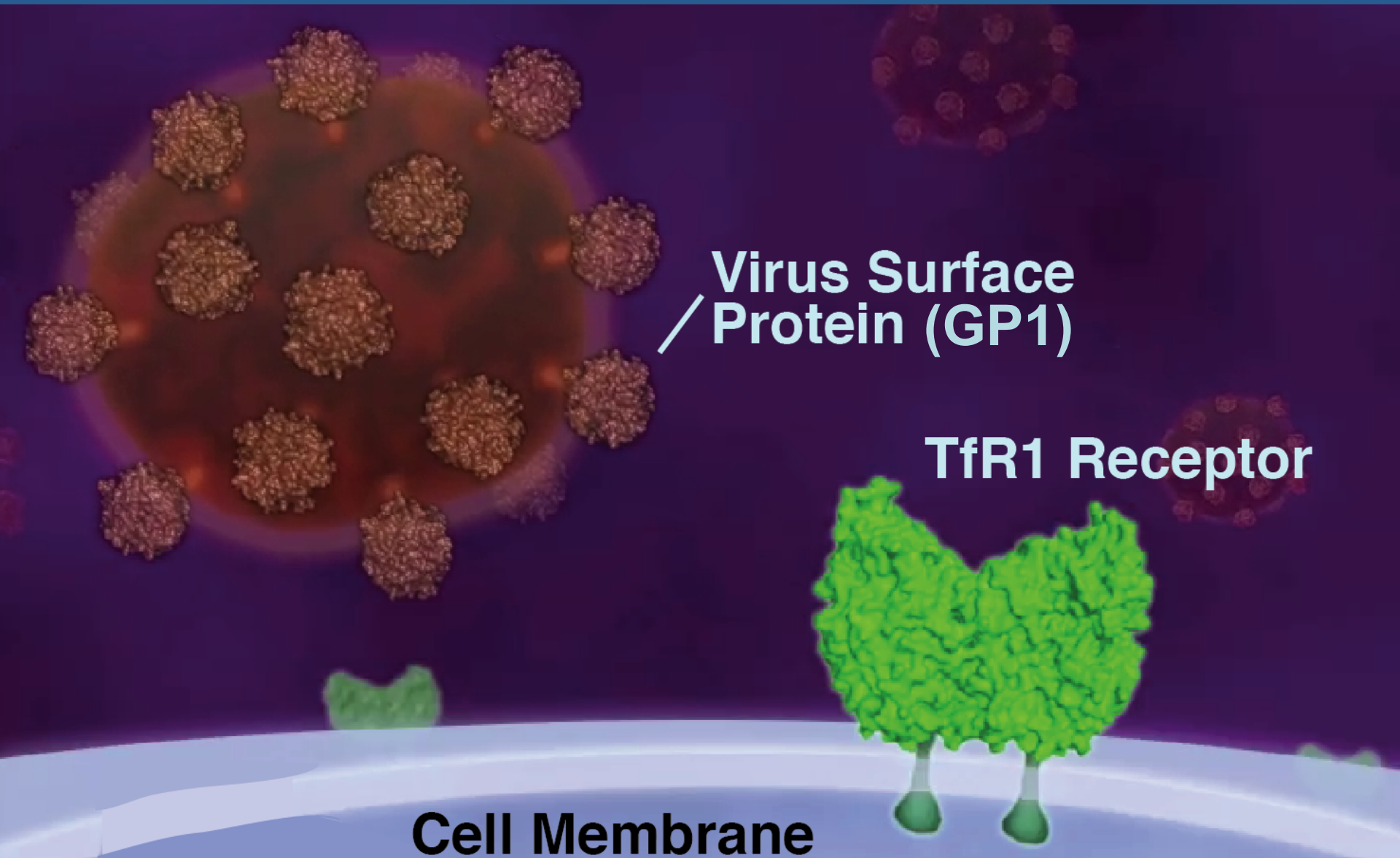
These viruses cause severe hemorrhagic fever and high fatality in humans

Erythematous rash in Bolivian Hemorrhagic Fever (Machupo virus)

Subconjunctival hemorrhage in Bolivian Hemorrhagic Fever

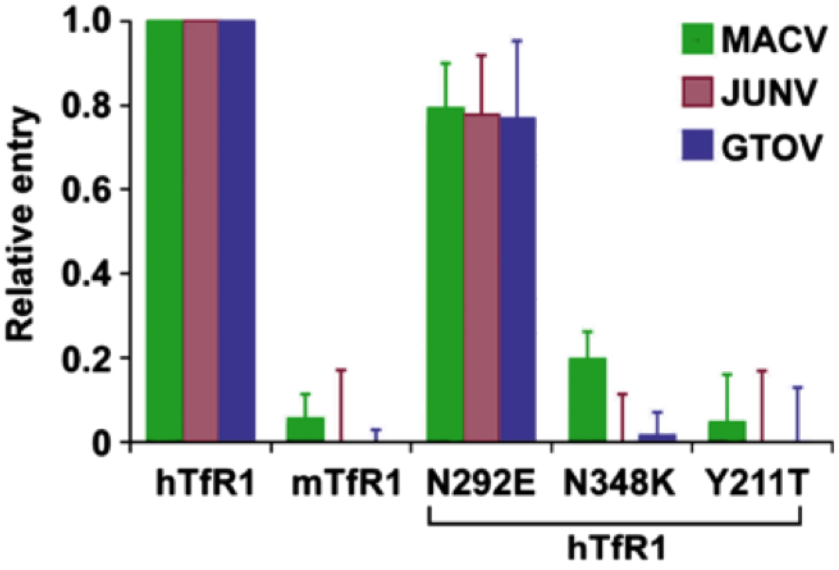
Borio et al., JAMA 2002

Arenaviruses infect cells via the transferrin receptor (TfR1)

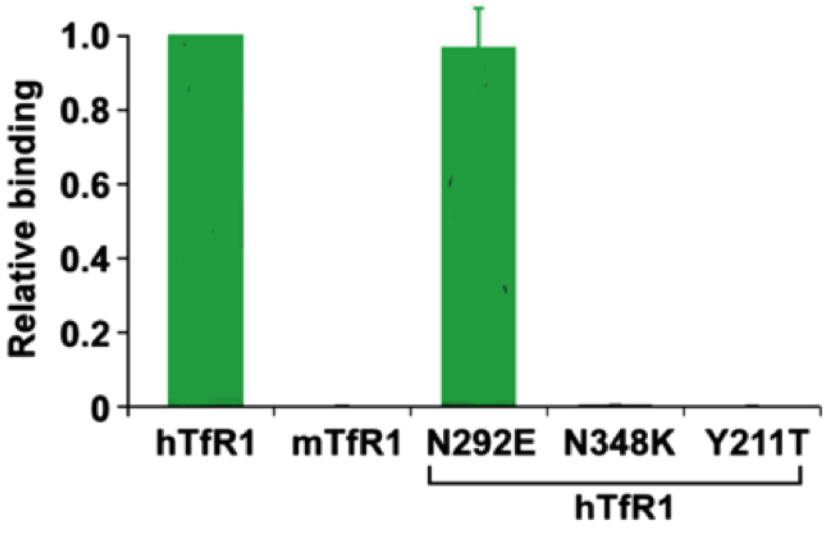


GP1 – TfR1 binding determines viral entry

Viral entry



GP1–TfR1 binding

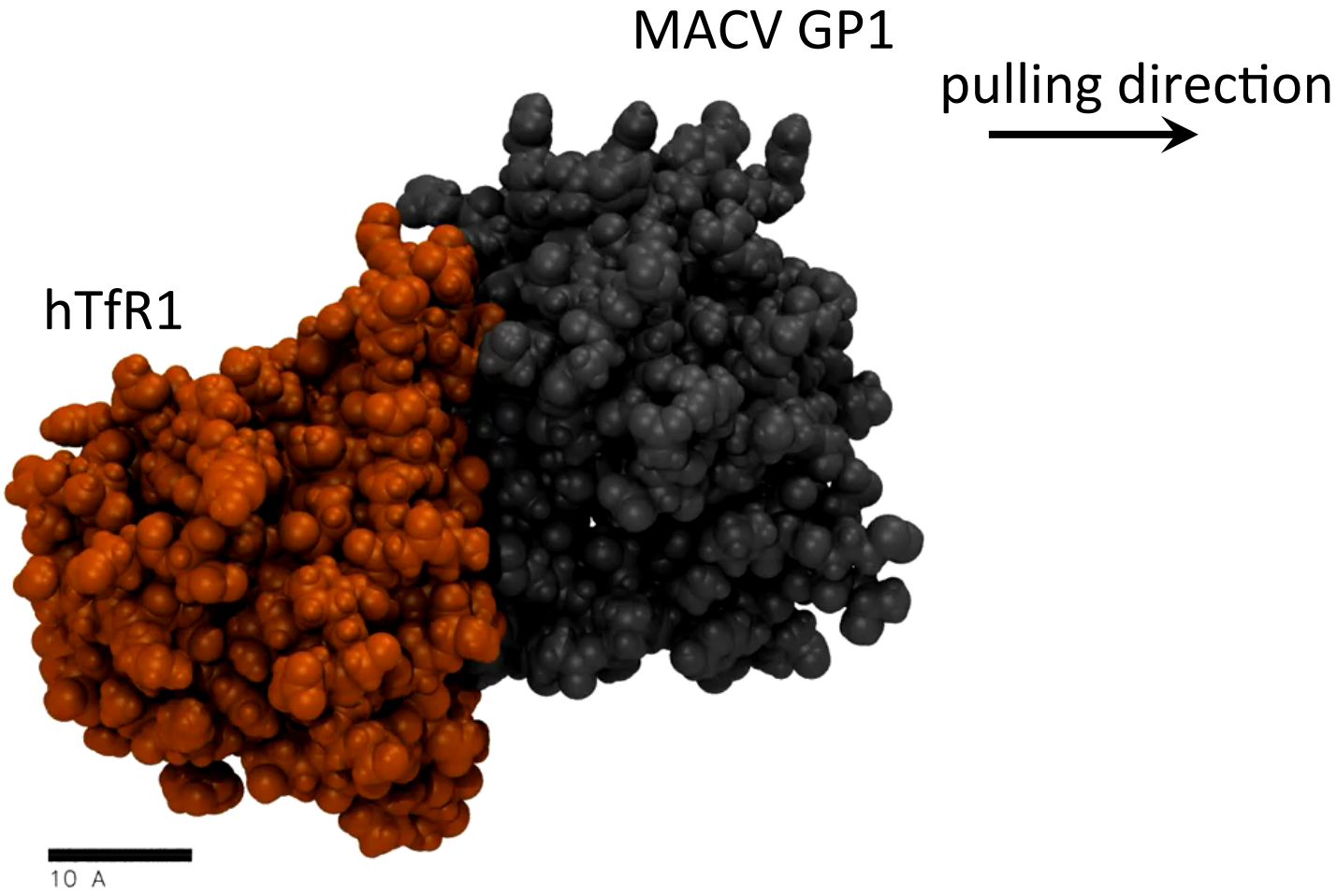


N348K and Y211T
in hTfR1 abolish binding
and entry

Radoshitzky et al., PNAS 2008

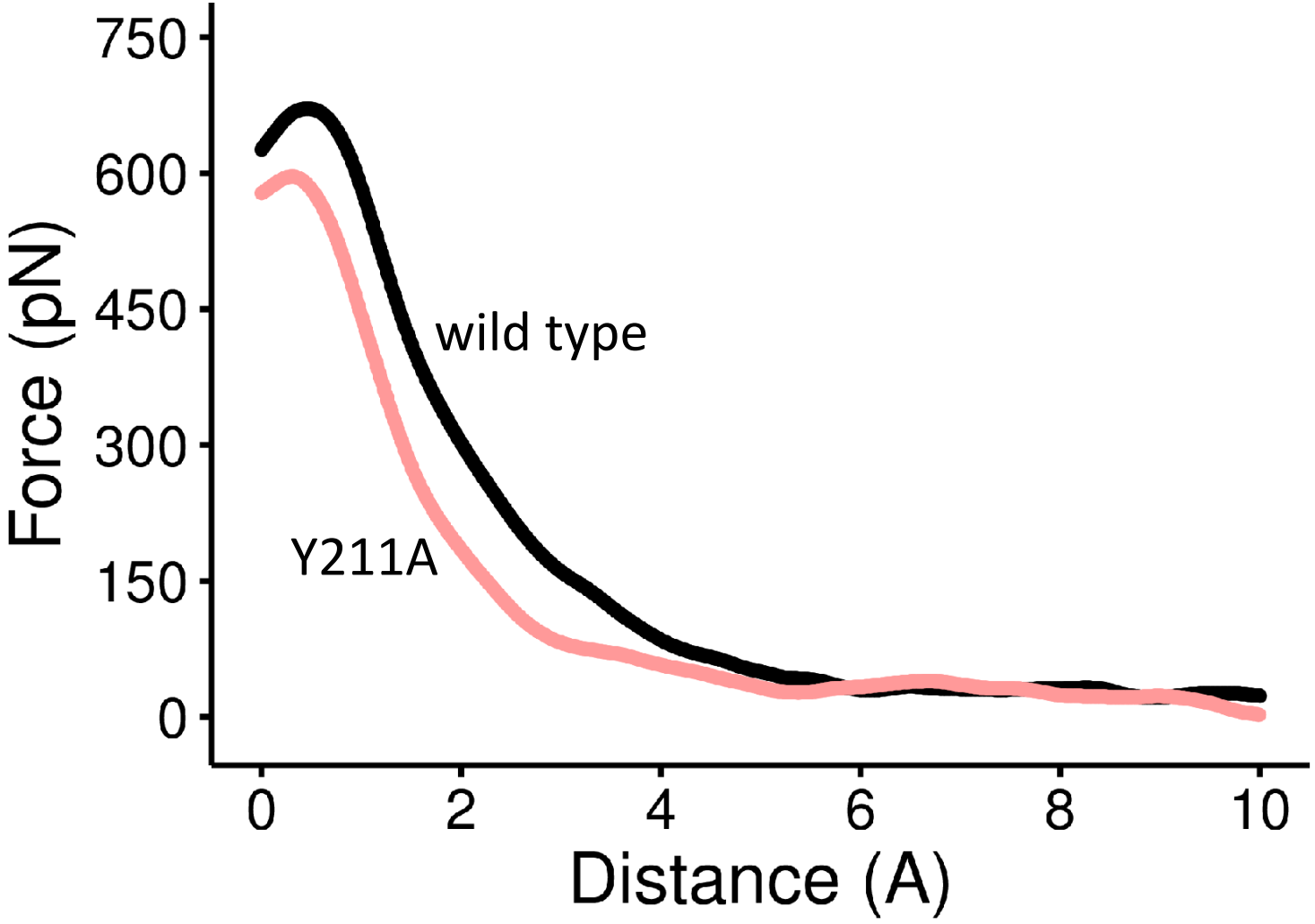
Can we computationally predict mutations that determine host range?

Approach 1: Steered Molecular Dynamics



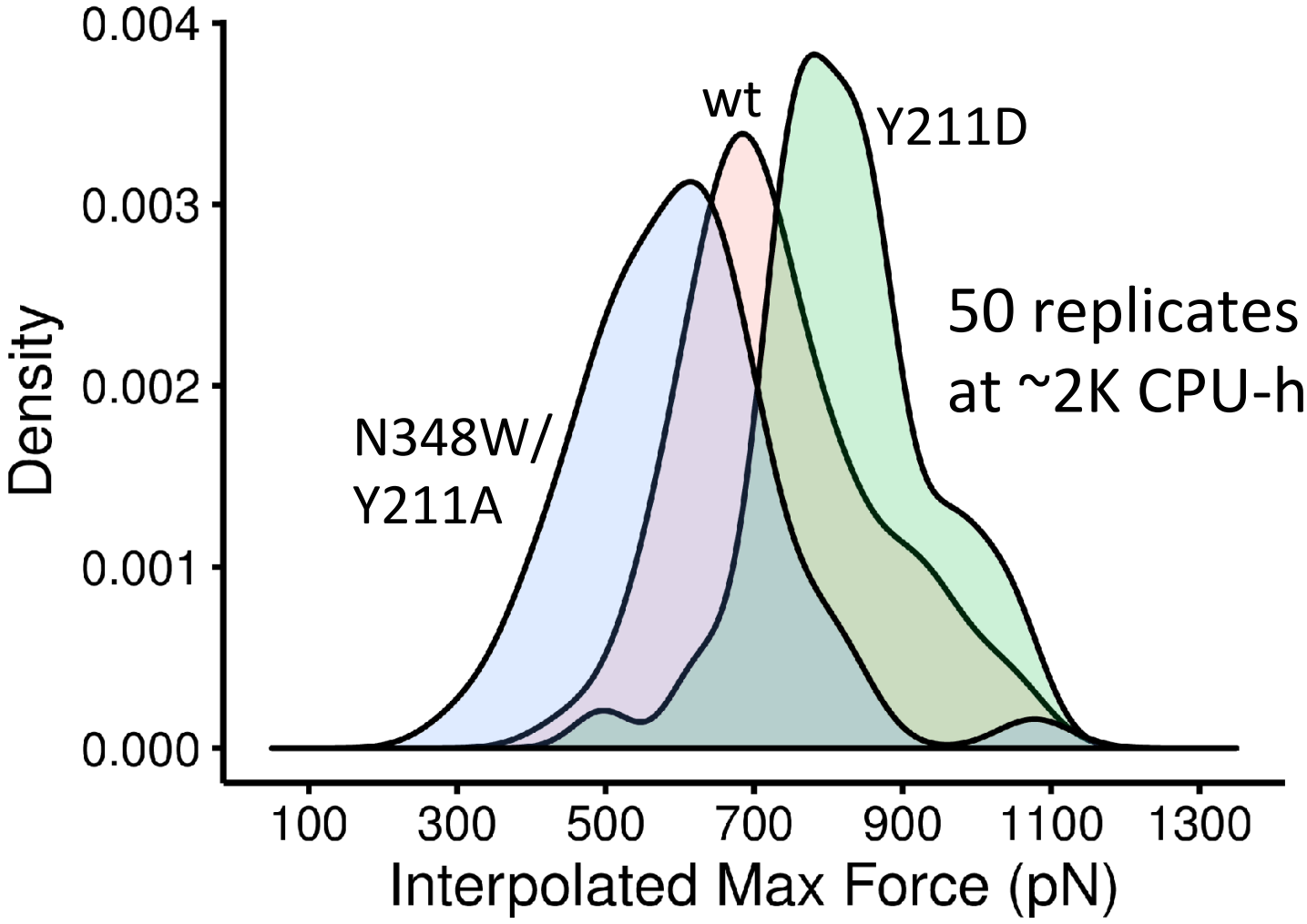
Meyer et al., PeerJ 2013

Force-distance curves differ between wild type and mutants



Meyer et al., PeerJ 2013

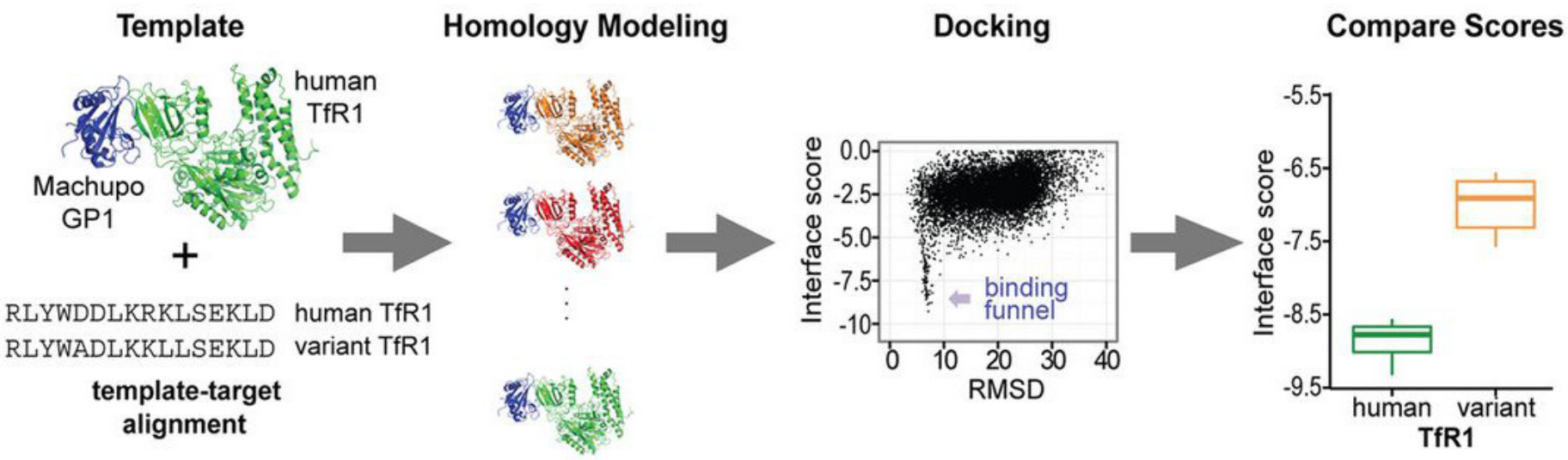
However, the method is noisy and requires millions of CPU-hours



50 replicates per mutant, at ~2K CPU-hours each

Meyer et al., PeerJ 2013

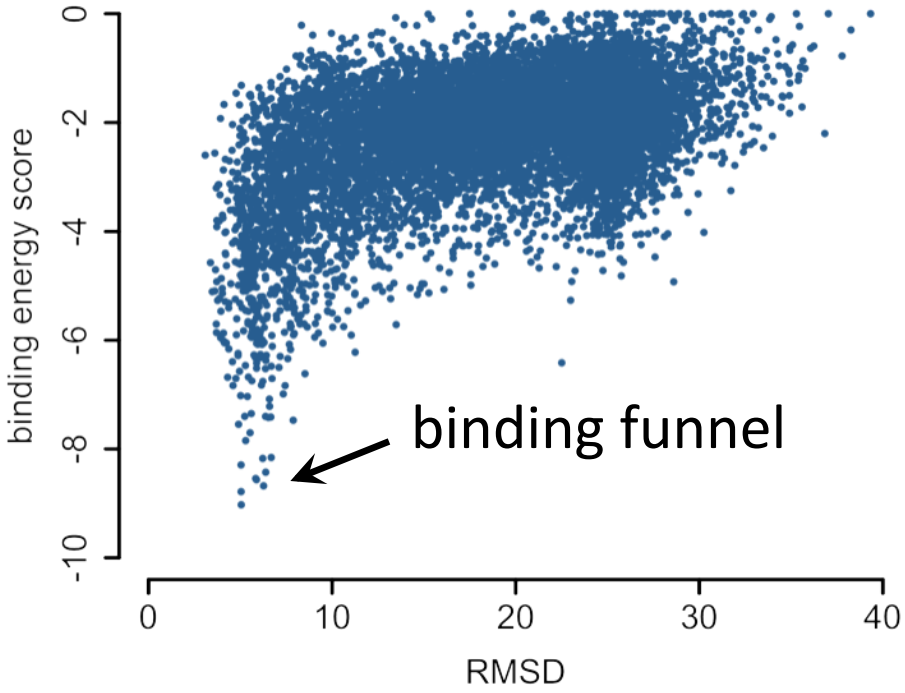
Alternative approach: Homology modeling and docking



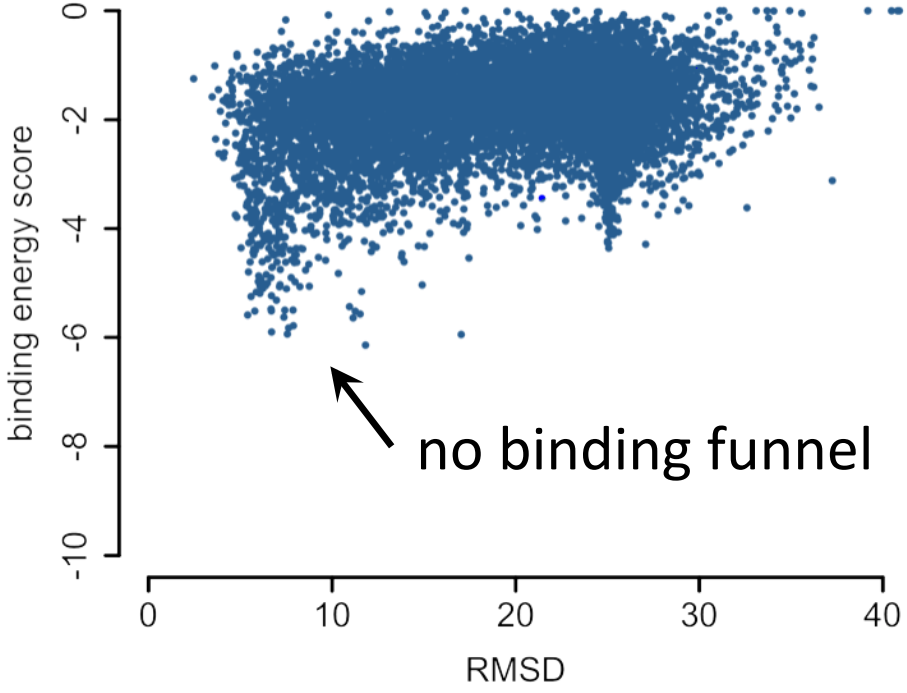
Kerr et al., J. Virol. 2015

Docking results agree with experimentally observed infection patterns

C. callosus
native host to MACV

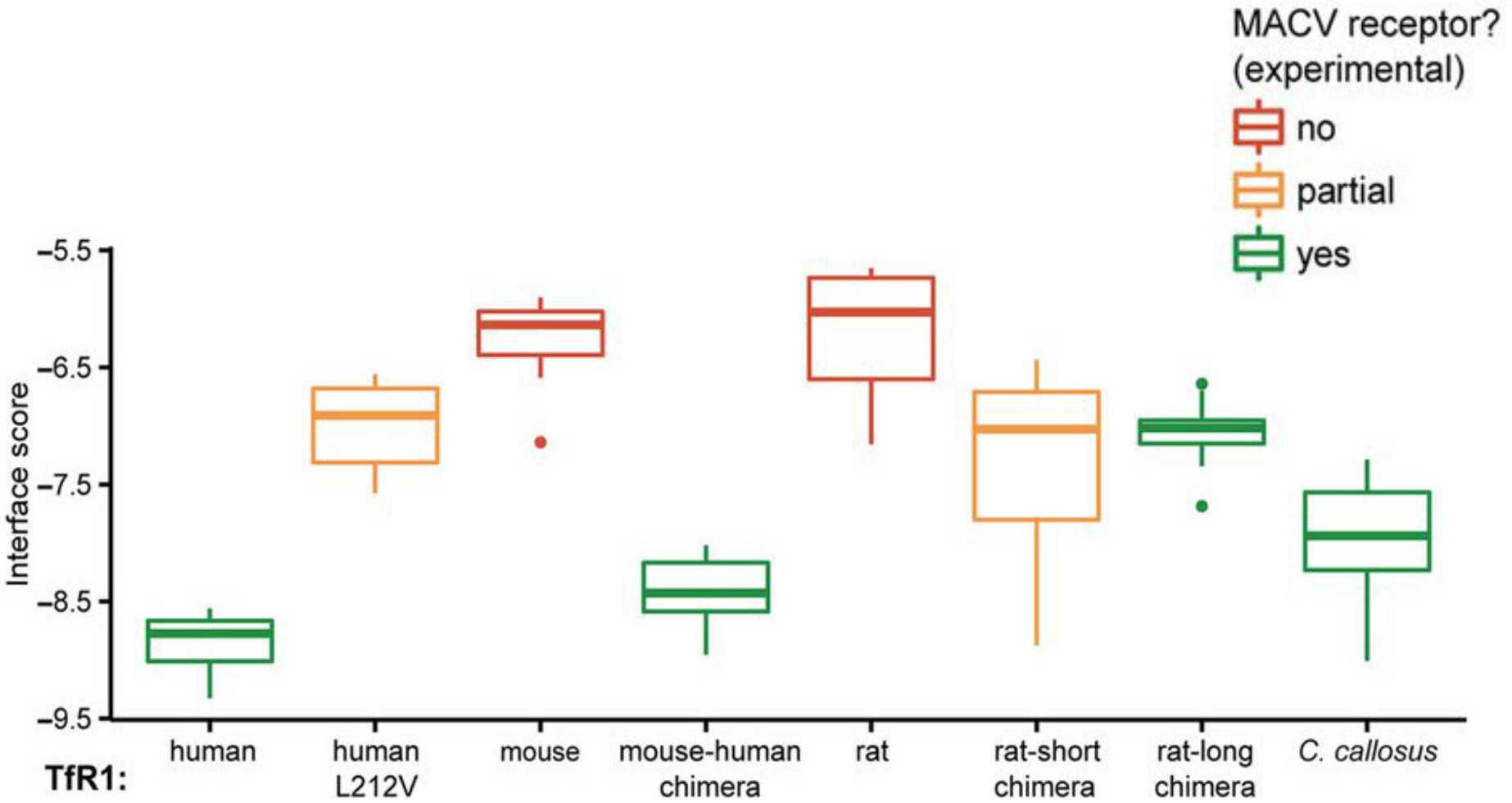


mouse
not infected by MACV



Kerr et al., J. Virol. 2015

Docking results agree with experimentally observed infection patterns



Kerr et al., J. Virol. 2015

Take-home message

Computational prediction of viral host range is becoming feasible with state-of-the-art protein design.

Acknowledgments

Wilke lab

Postdocs
- Umut Caglar
- Bart Smith
- Ashley Teufel

Graduate Students
- Dakota Derryberry
- Benjamin Jack
- Eleisha Jackson
- Stephanie Spielman
- Dariya Sydykova

Former members
- Austin Meyer (now @ Texas Tech)

Collaborators
- Julian Echave
- Andy Ellington
- Aashiq Kachroo (Marcotte lab)
- Jon Laurent (Marcotte lab)
- Oana Lungu (Ellington lab)
- Edward Marcotte
- Sara Sawyer

Funding
- NSF BEACON center
- NIH
- DTRA
- ARO