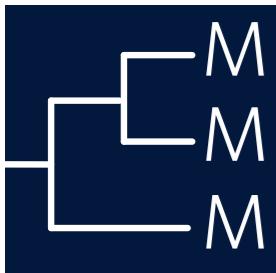
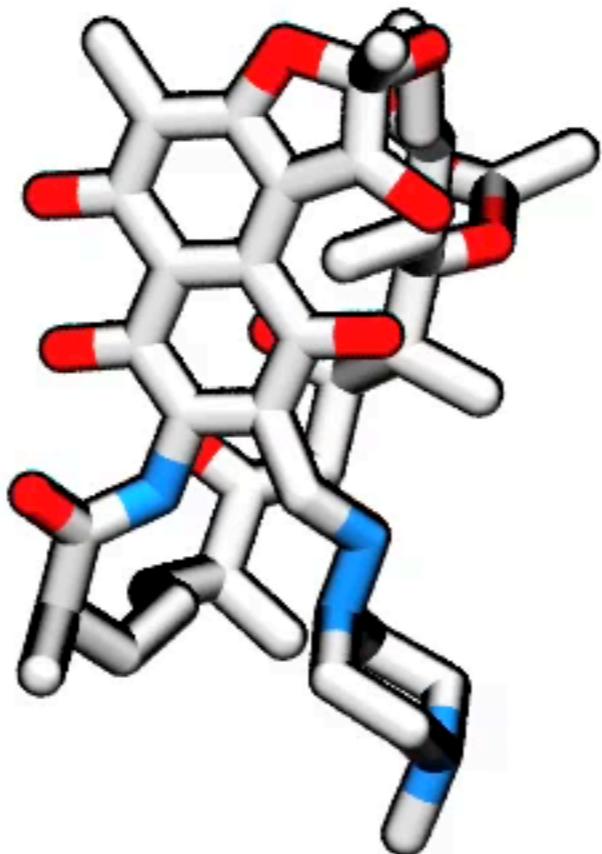
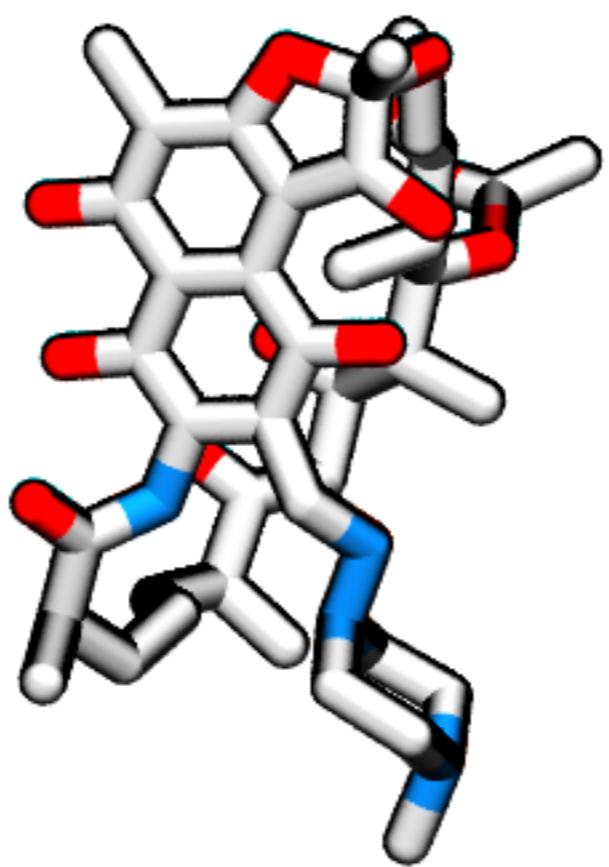


Lock it or Lose it: The story of rifampicin resistance



Philip Fowler
7 June 2018

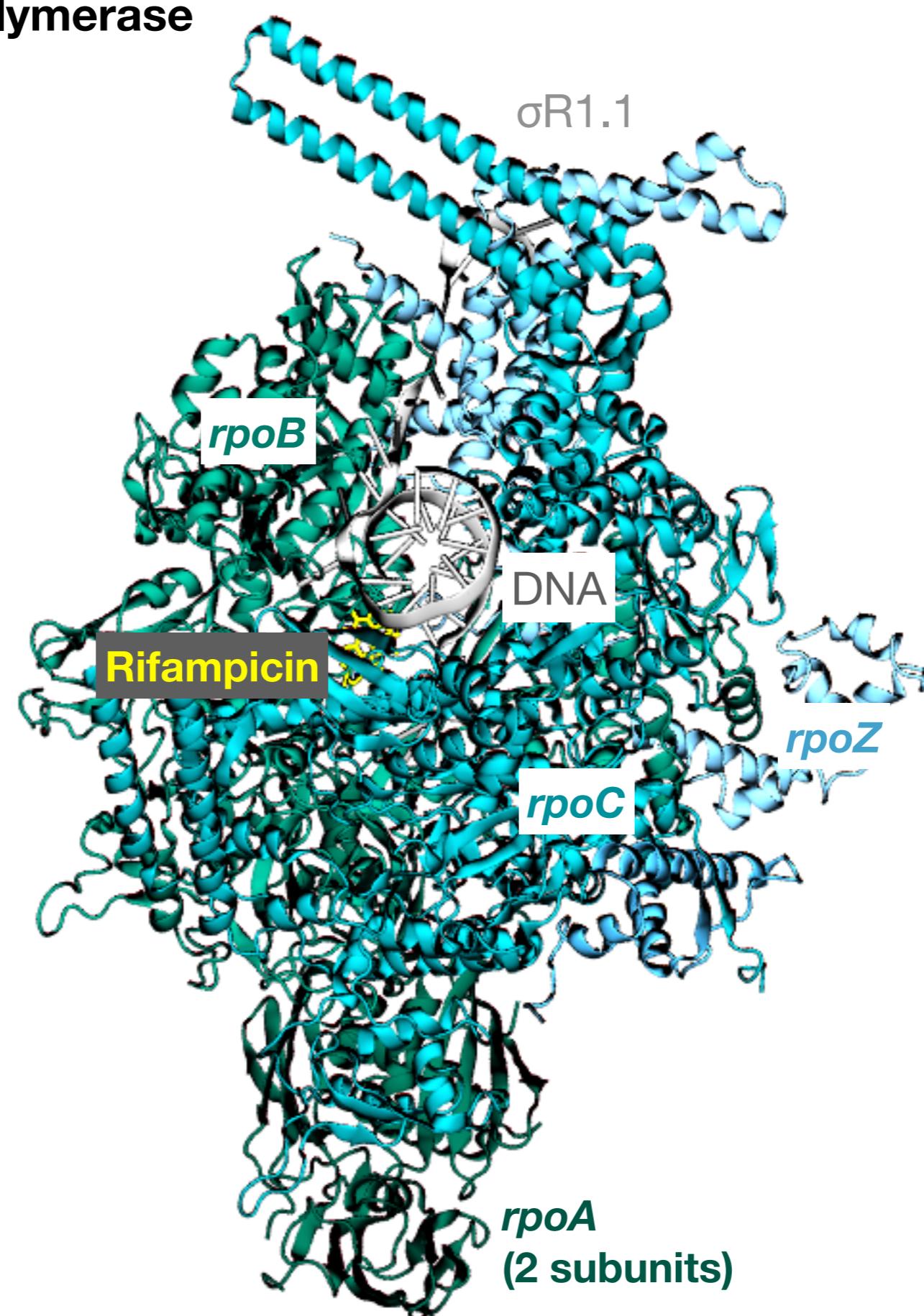




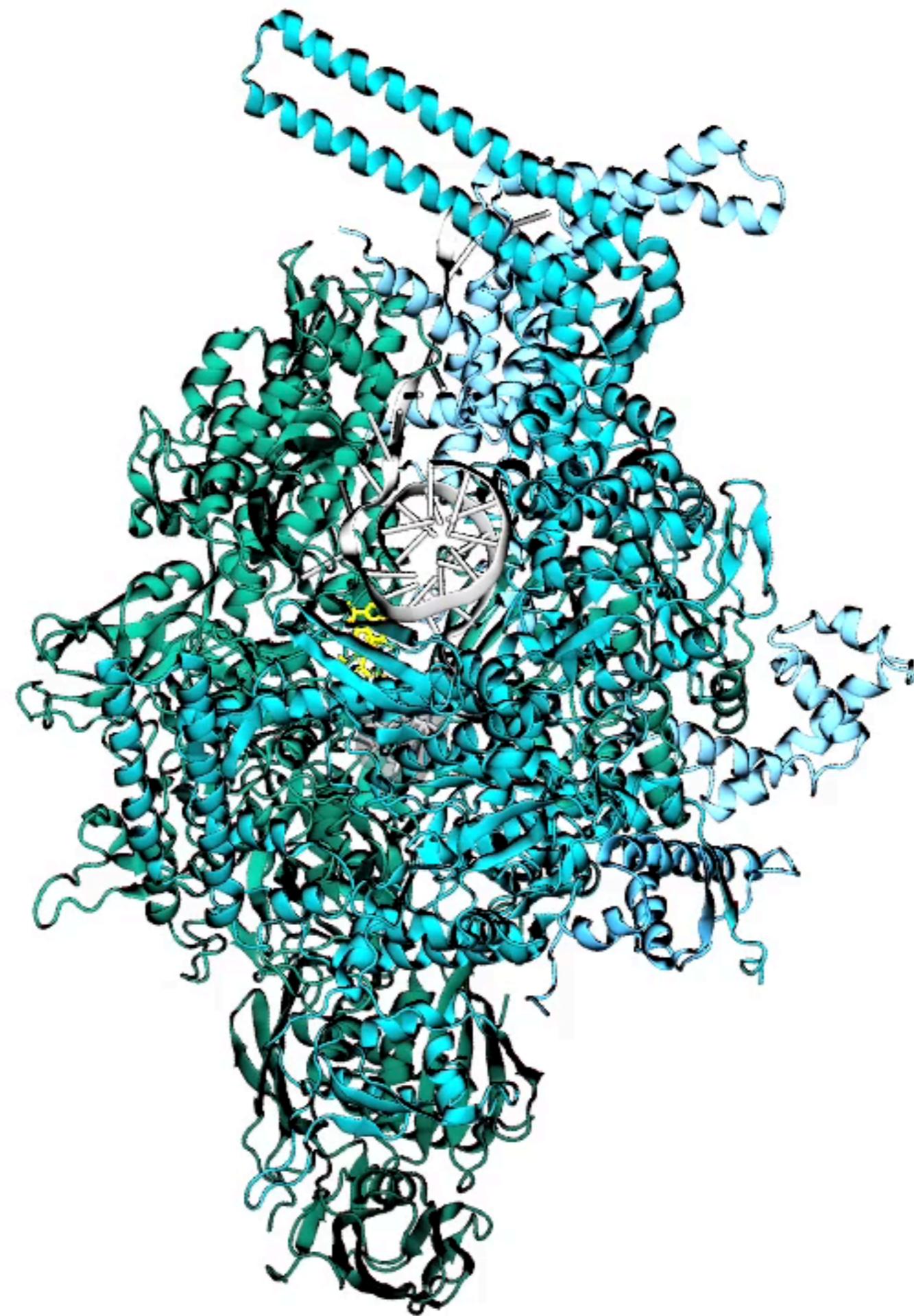


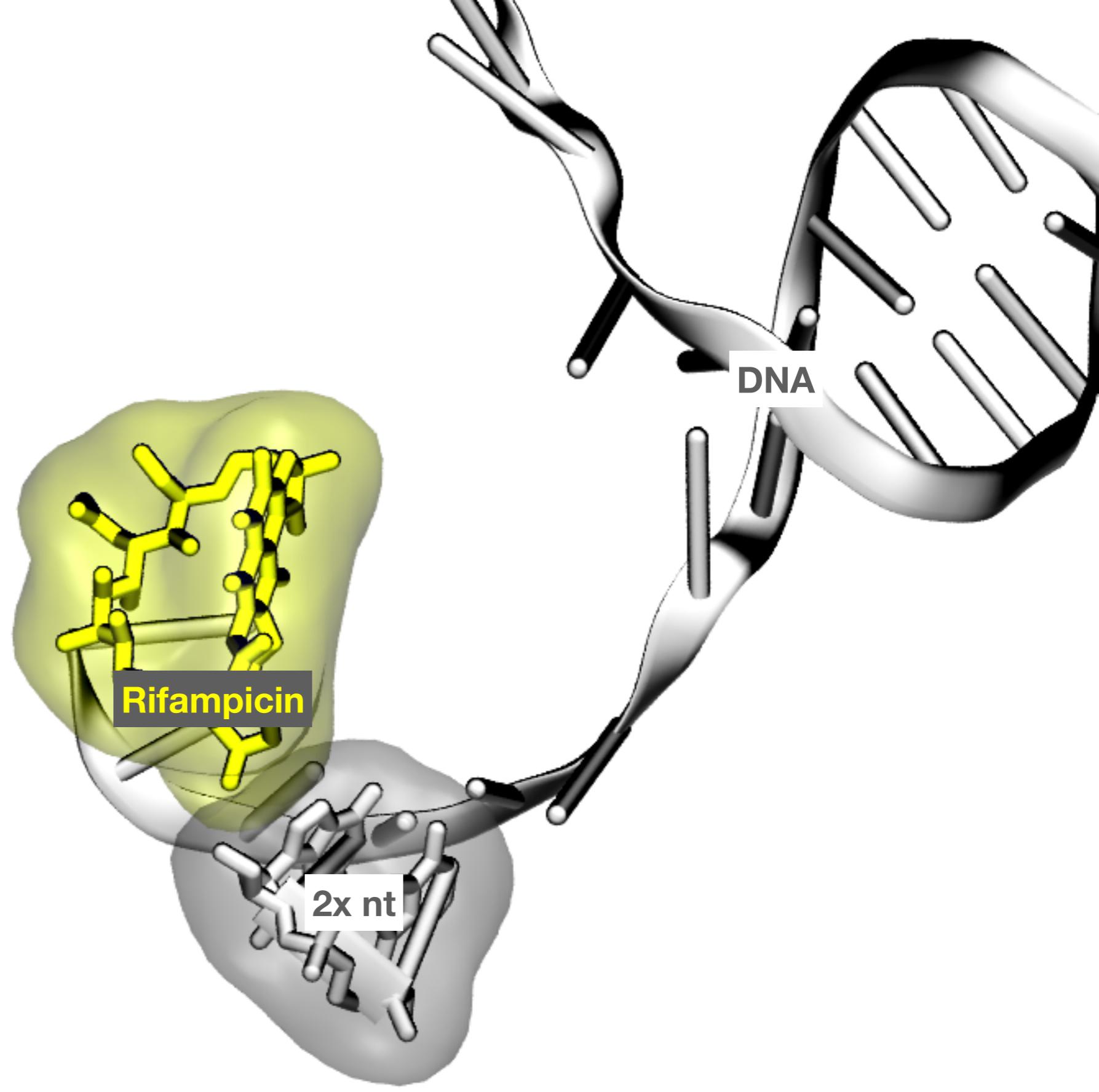
M. tuberculosis RNA polymerase

PDB: 5uh6



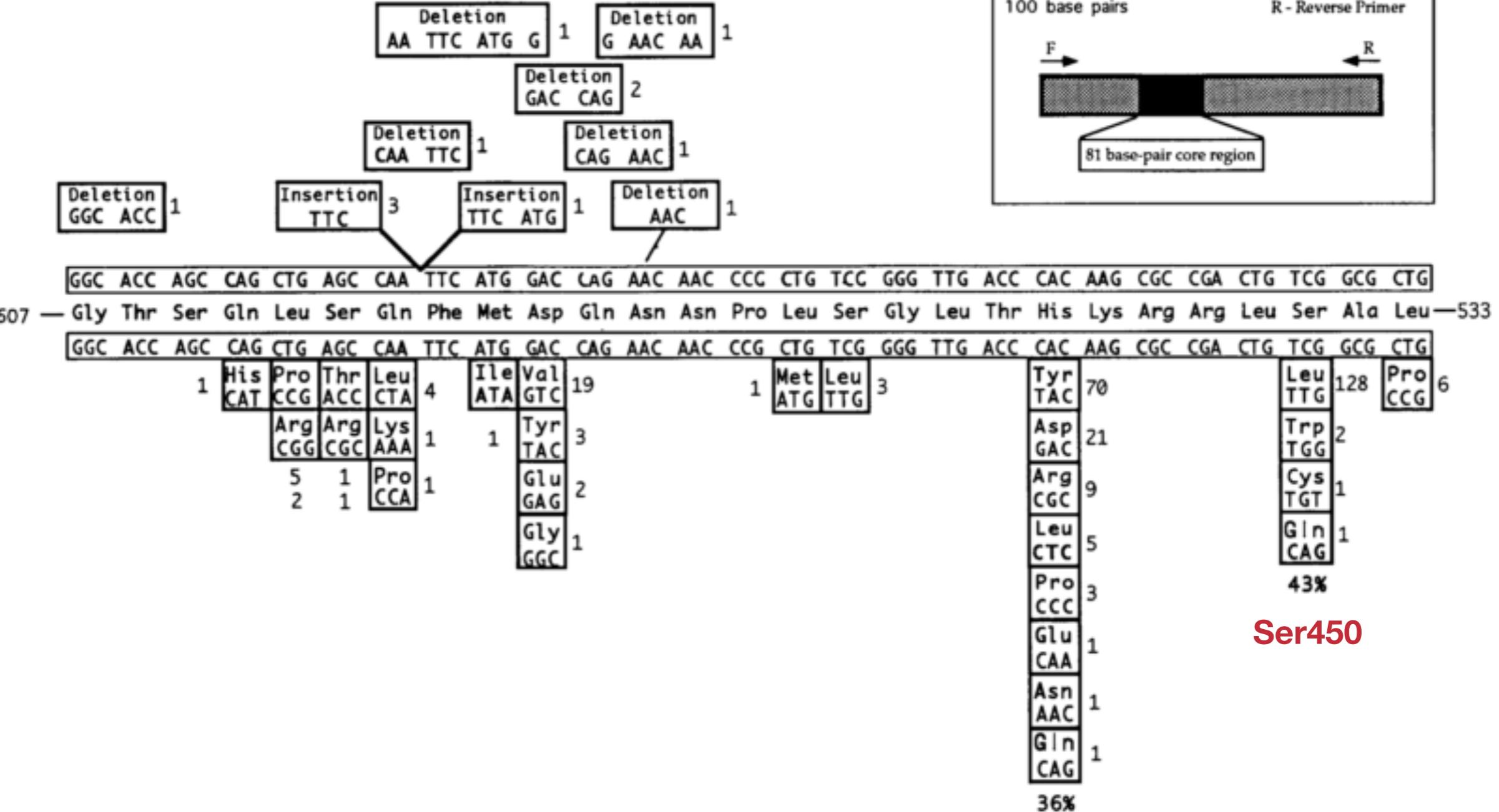


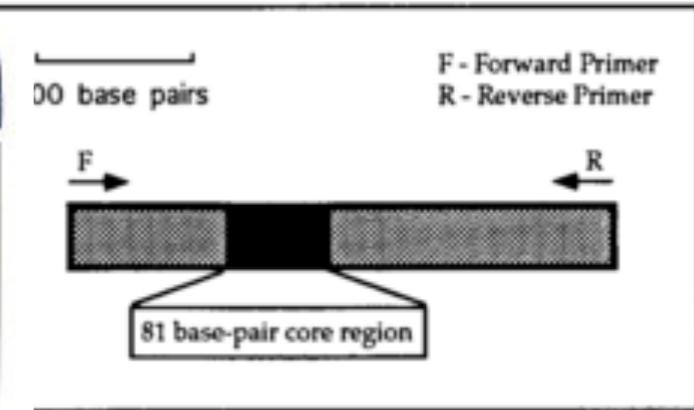
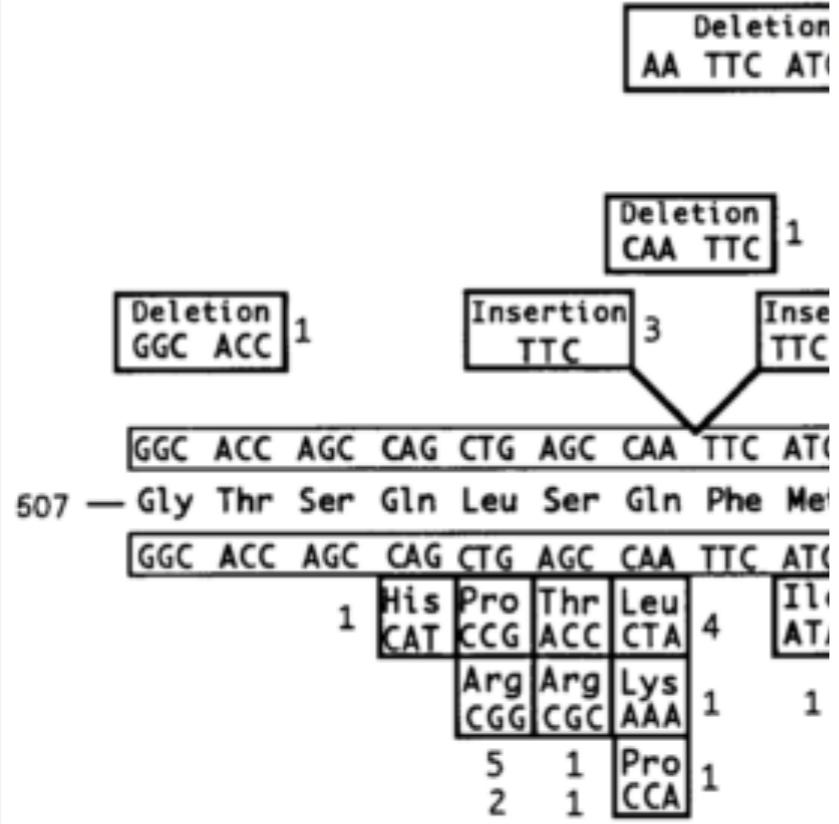




Rifampicin binds to the RNA polymerase, preventing the extension of RNA

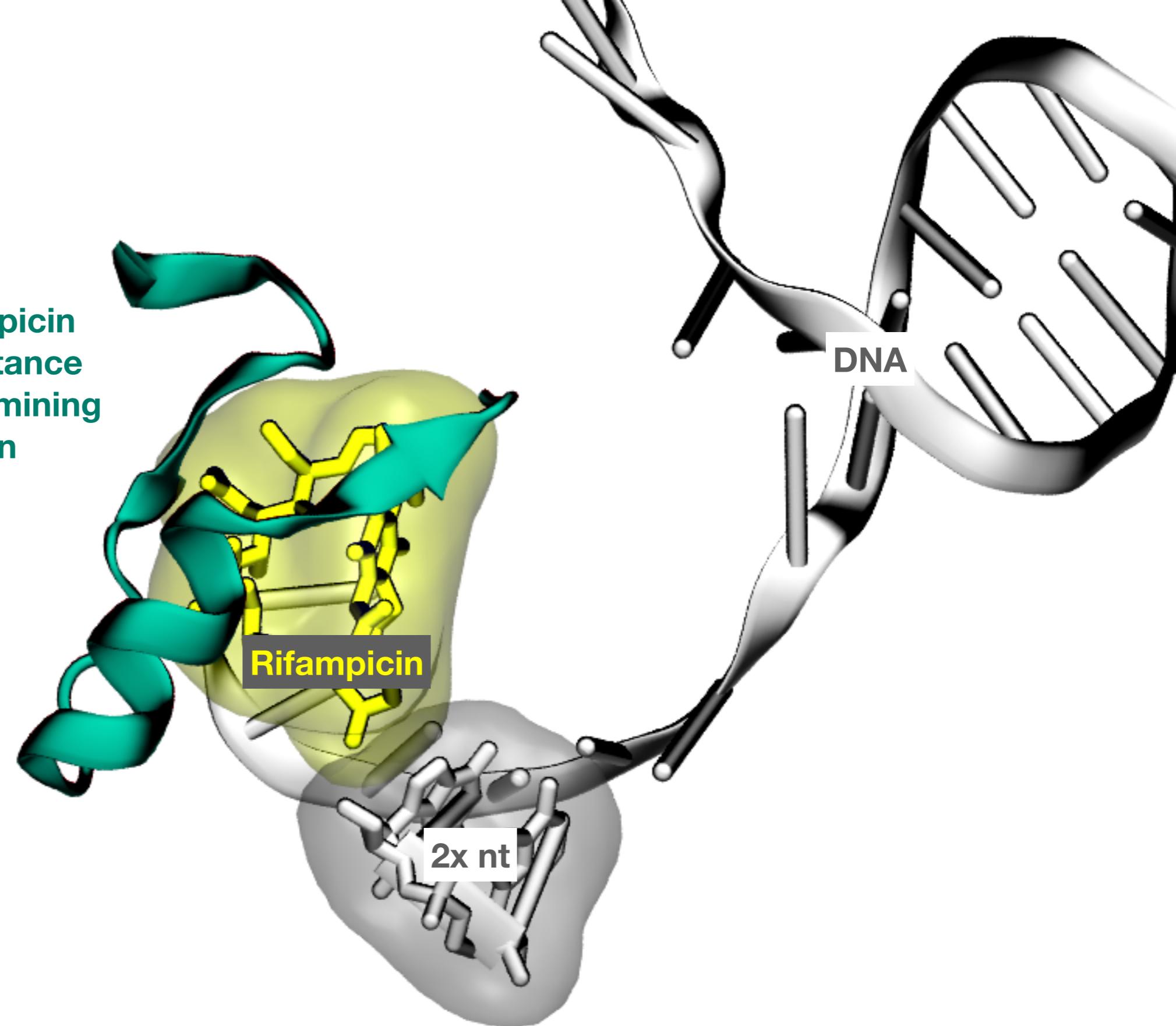






CAC AAG CGC CGA CTG TCG GCG CTG			
· His Lys Arg Arg Leu Ser Ala Leu — 533			
CAC AAG CGC CGA CTG TCG GCG CTG			
Tyr 70	Leu 128	Pro 6	
CAT TAC	TTG CCG	CCG	
Asp 21	Trp 2		
GAC TGG			
Arg 9	Cys 1		
CGC TGT			
Leu 5	Gln 1		
CTC CAG	CAG		
Pro 3	43%		
CCC CCC			
Glu 1			
CAA AAC			
Asn 1			
AAC CAG			
Gln 1			
CAG CAG			
36%			

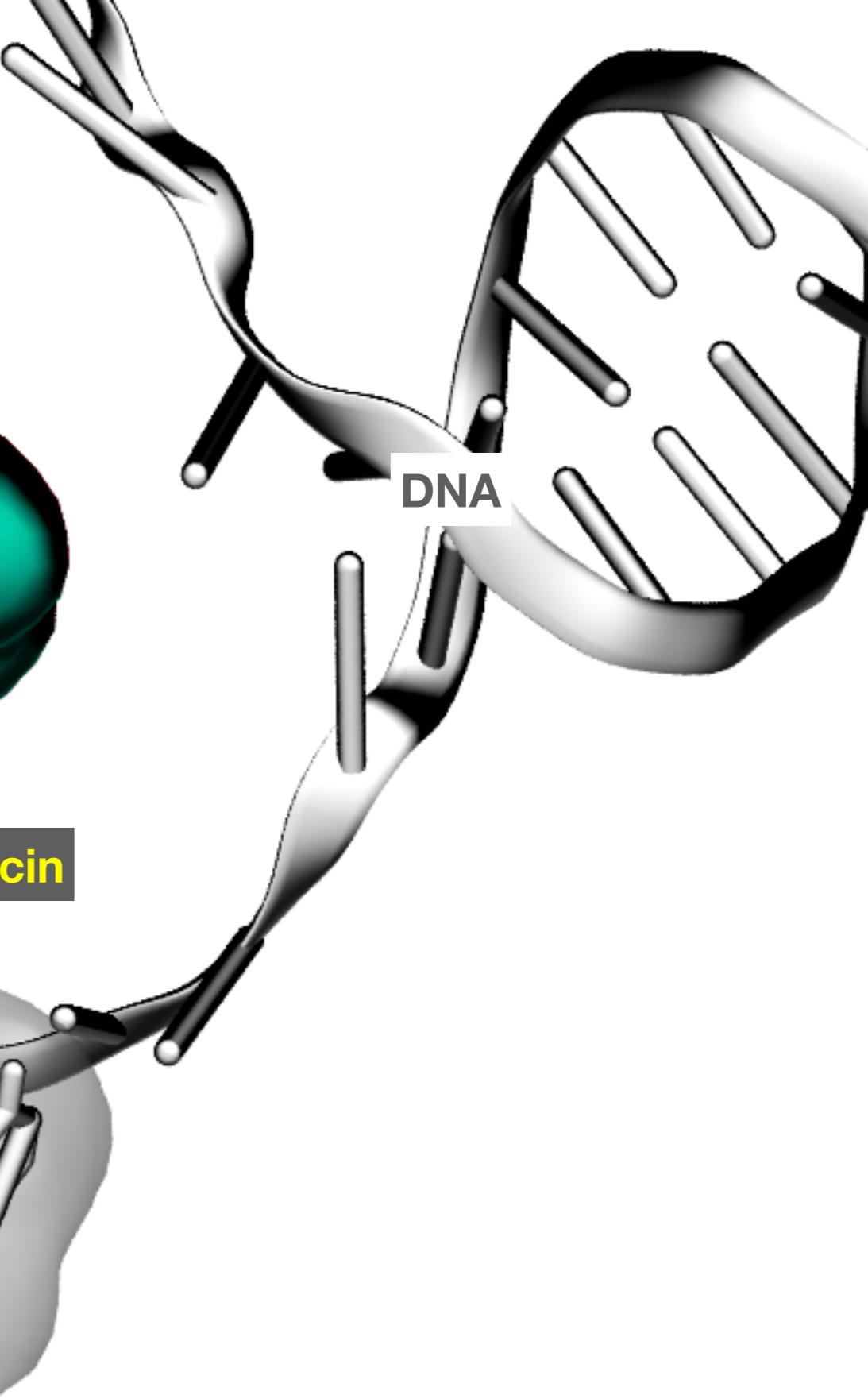
**Rifampicin
Resistance
Determining
Region**

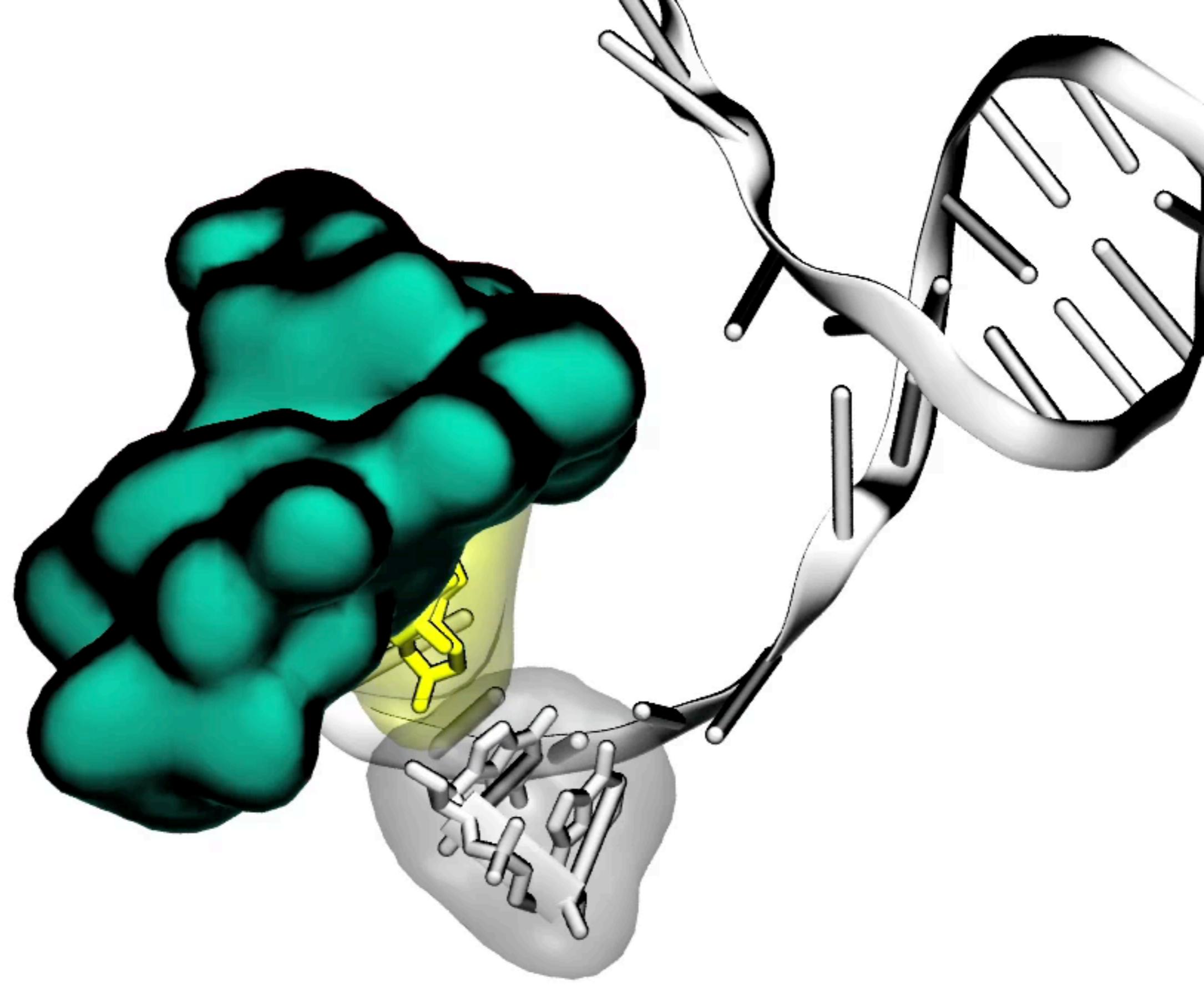


**Rifampicin
Resistance
Determining
Region**

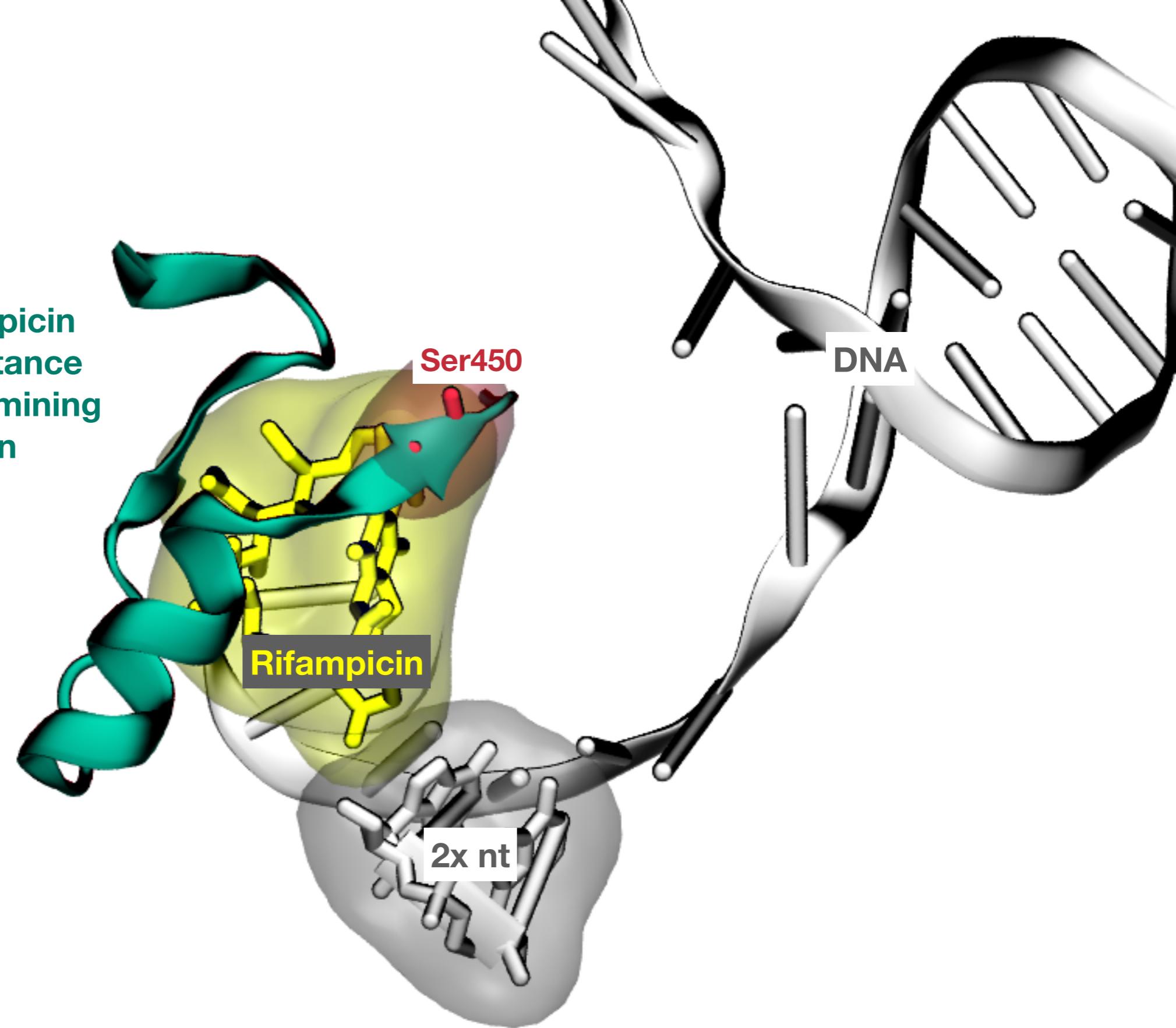
Rifampicin

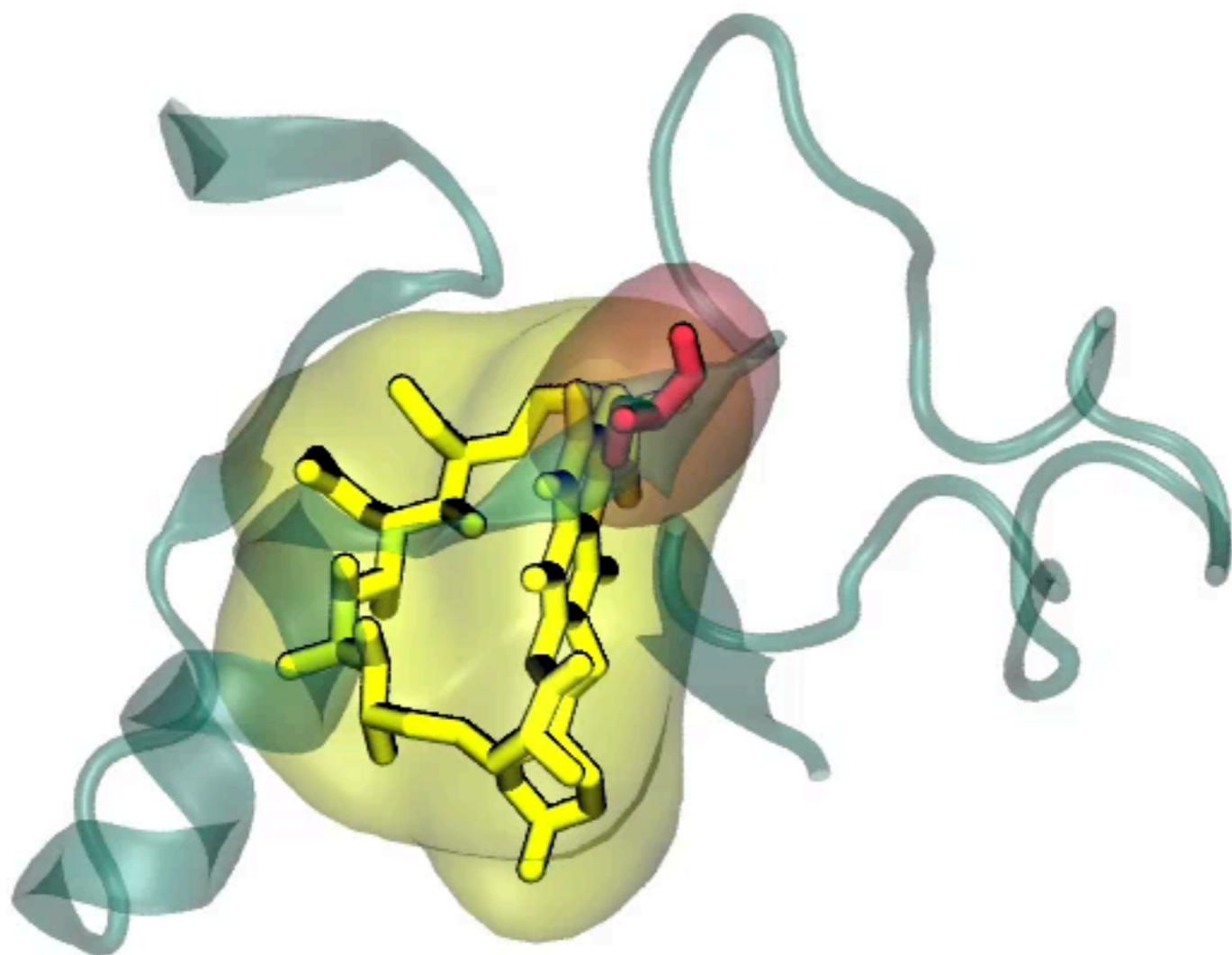
2x nt

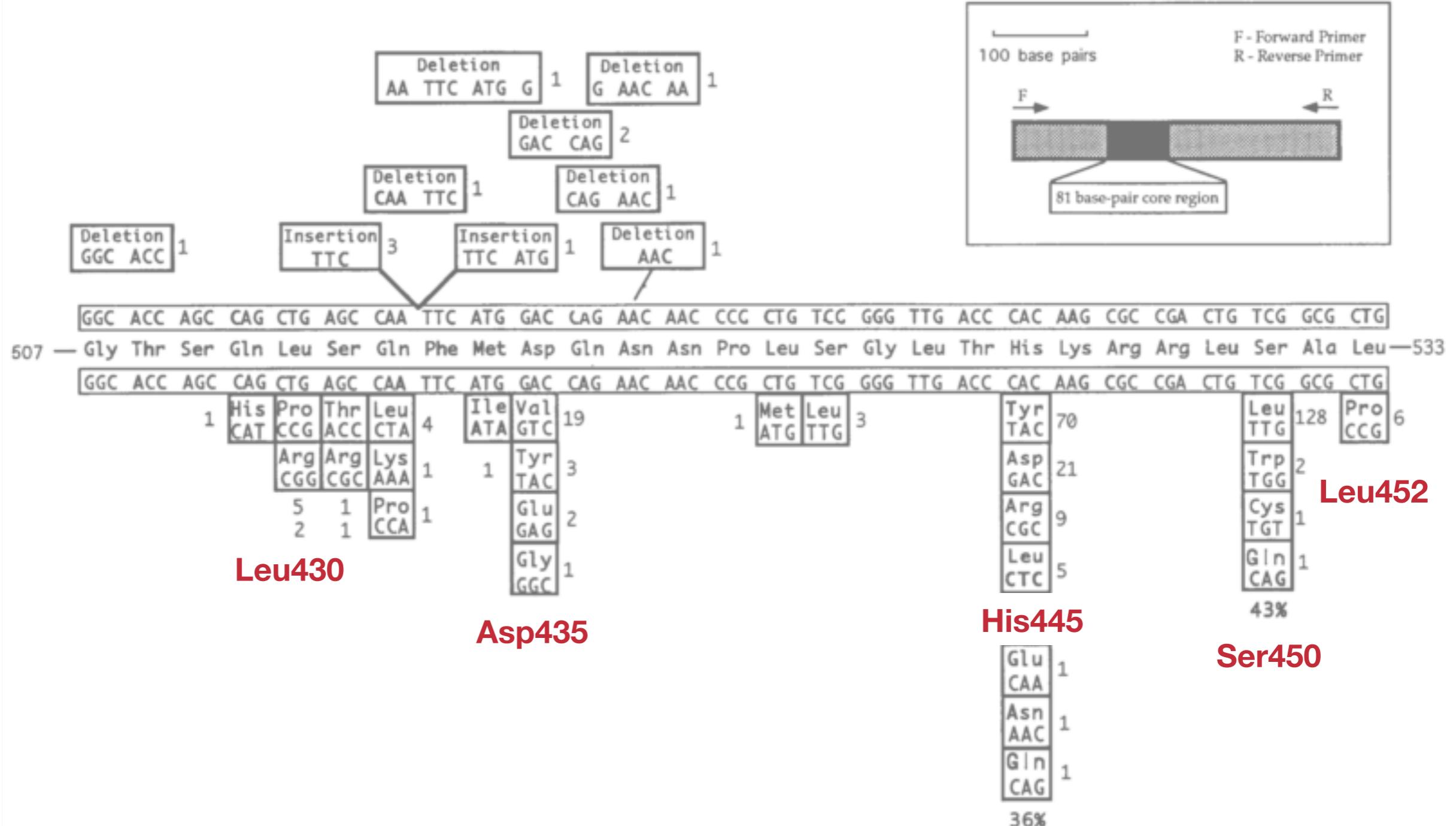




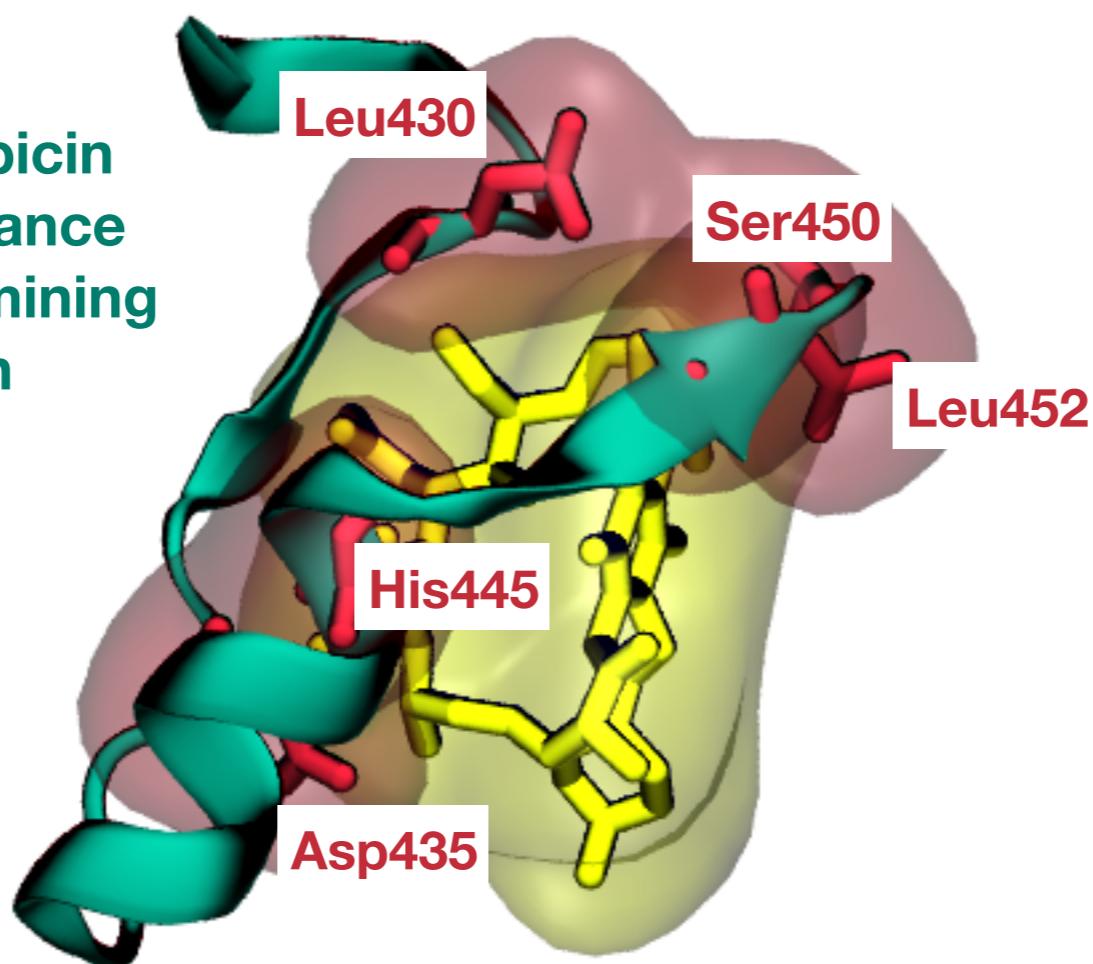
**Rifampicin
Resistance
Determining
Region**

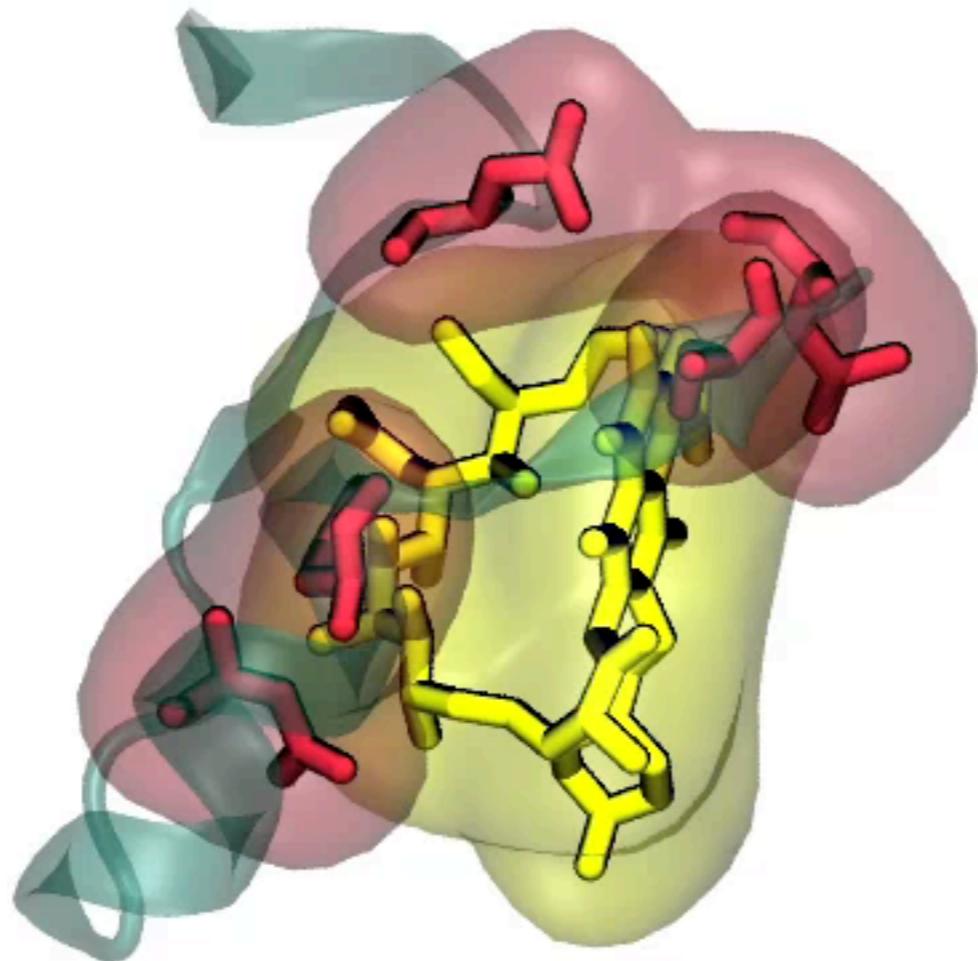


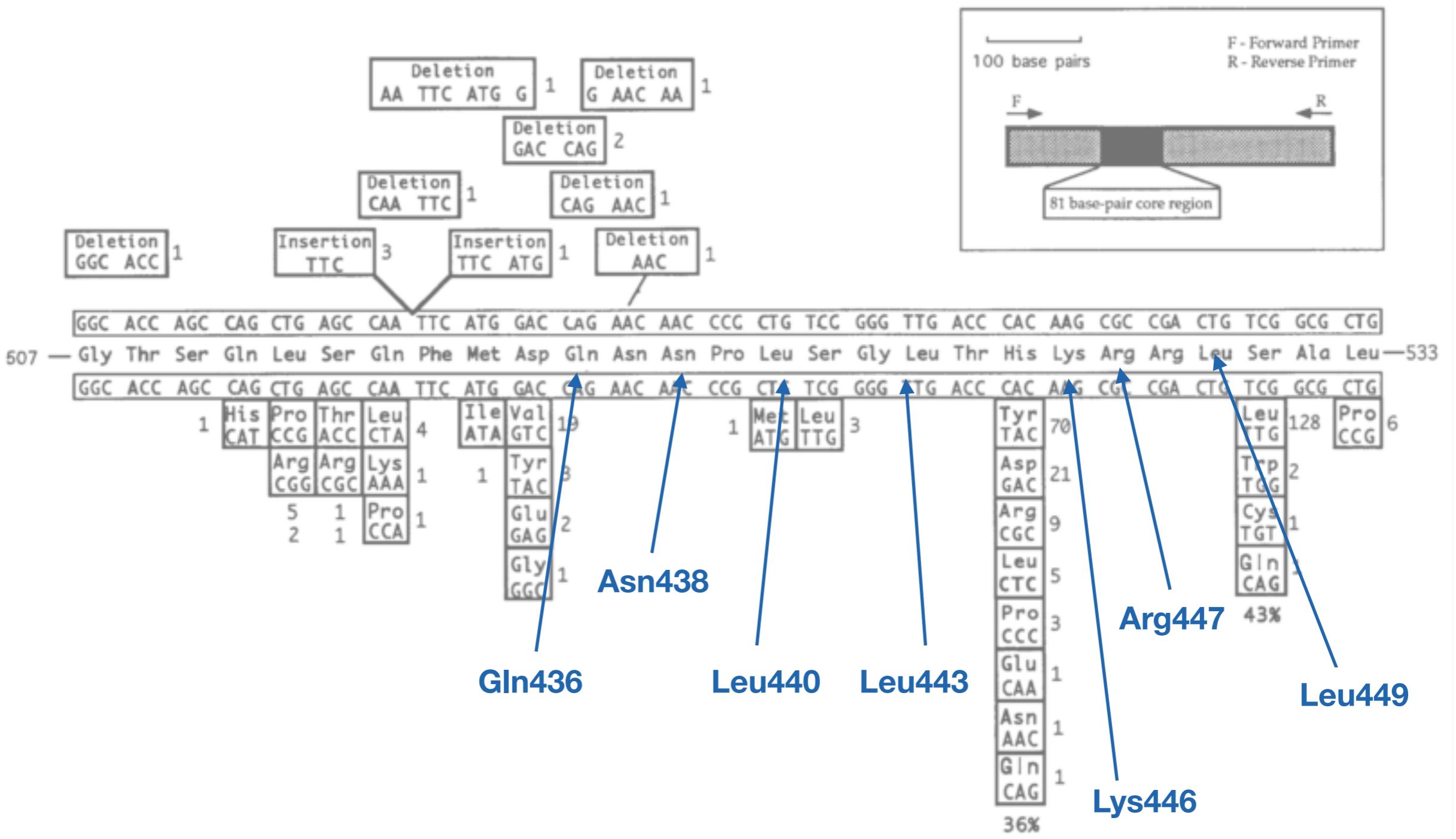


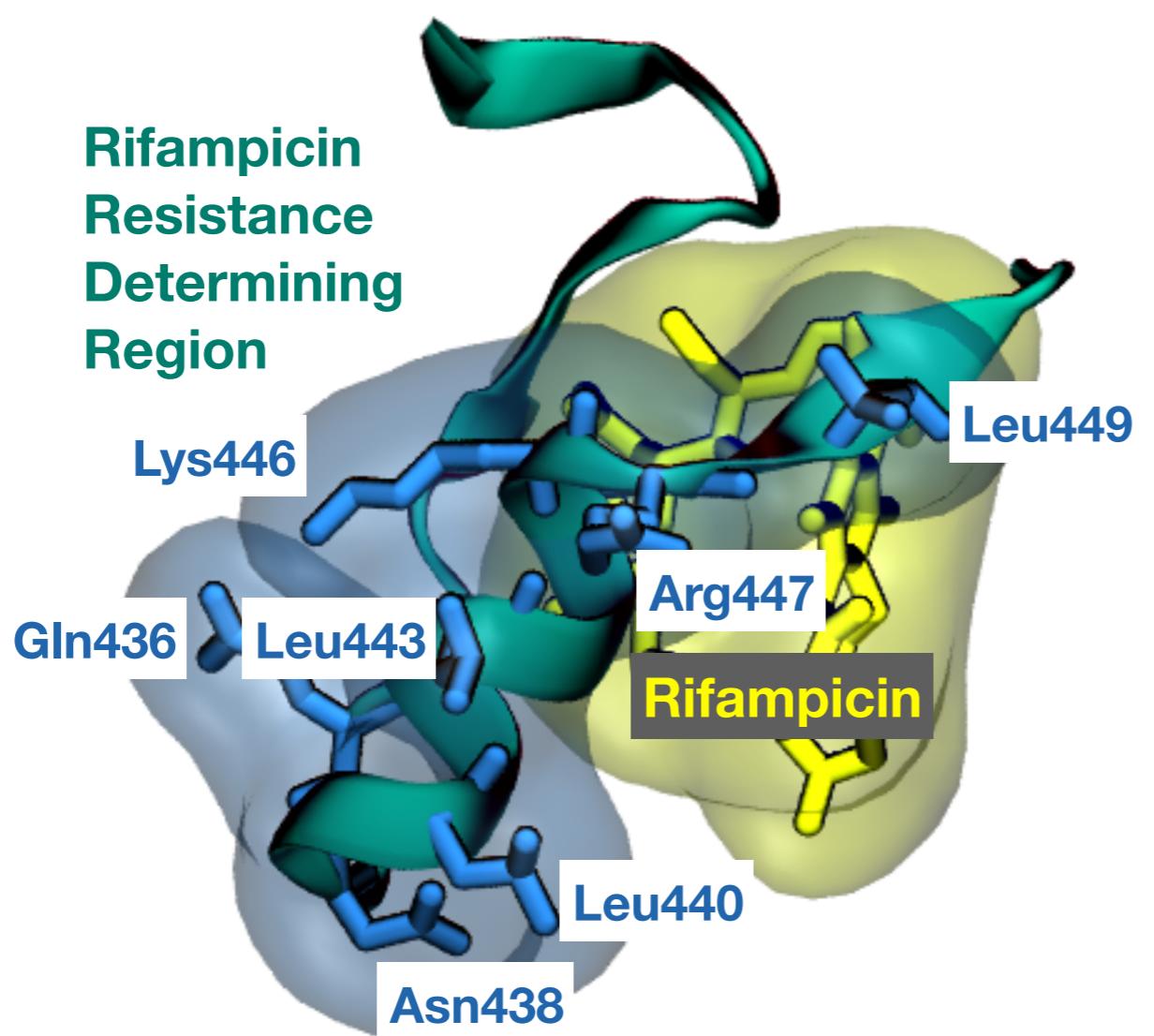


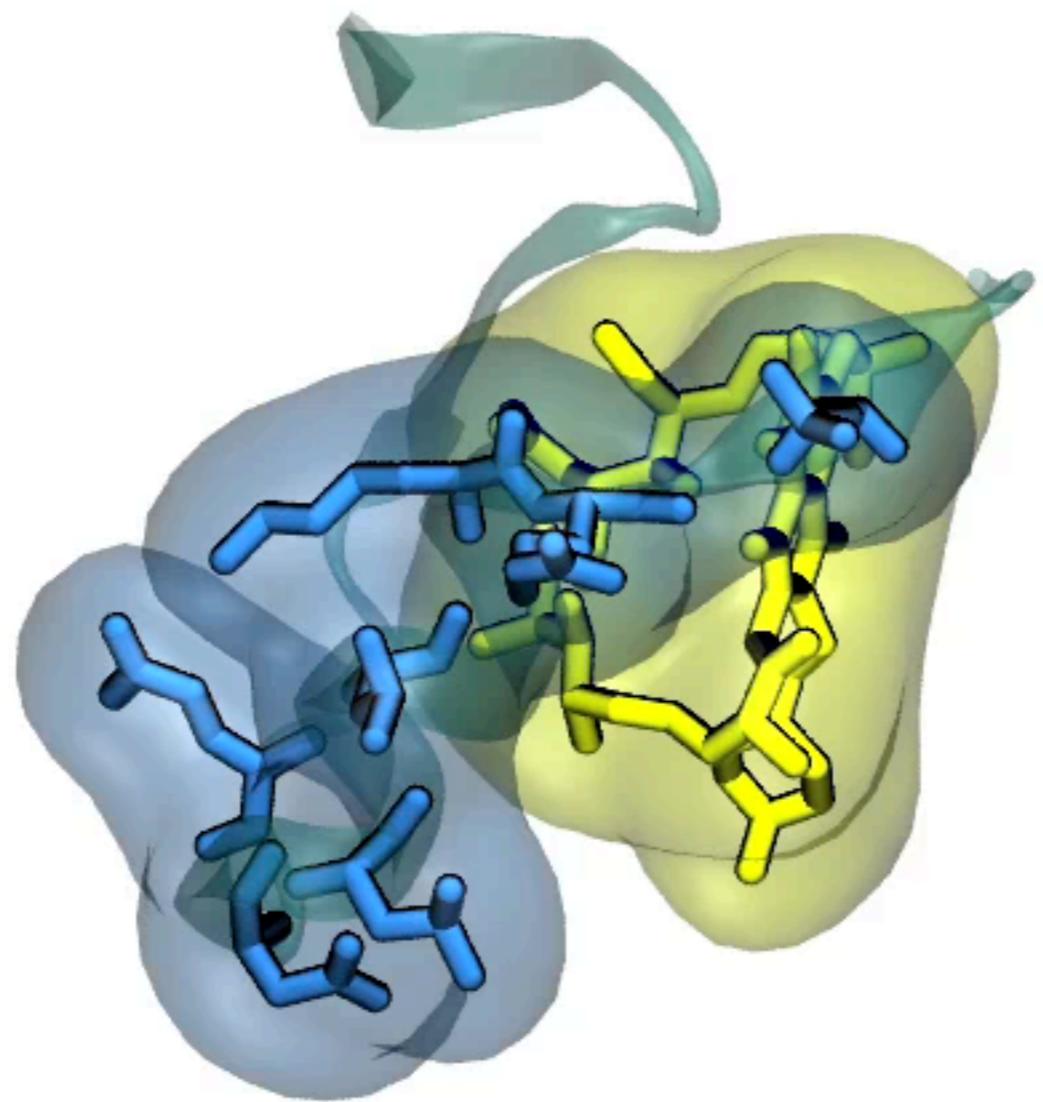
**Rifampicin
Resistance
Determining
Region**



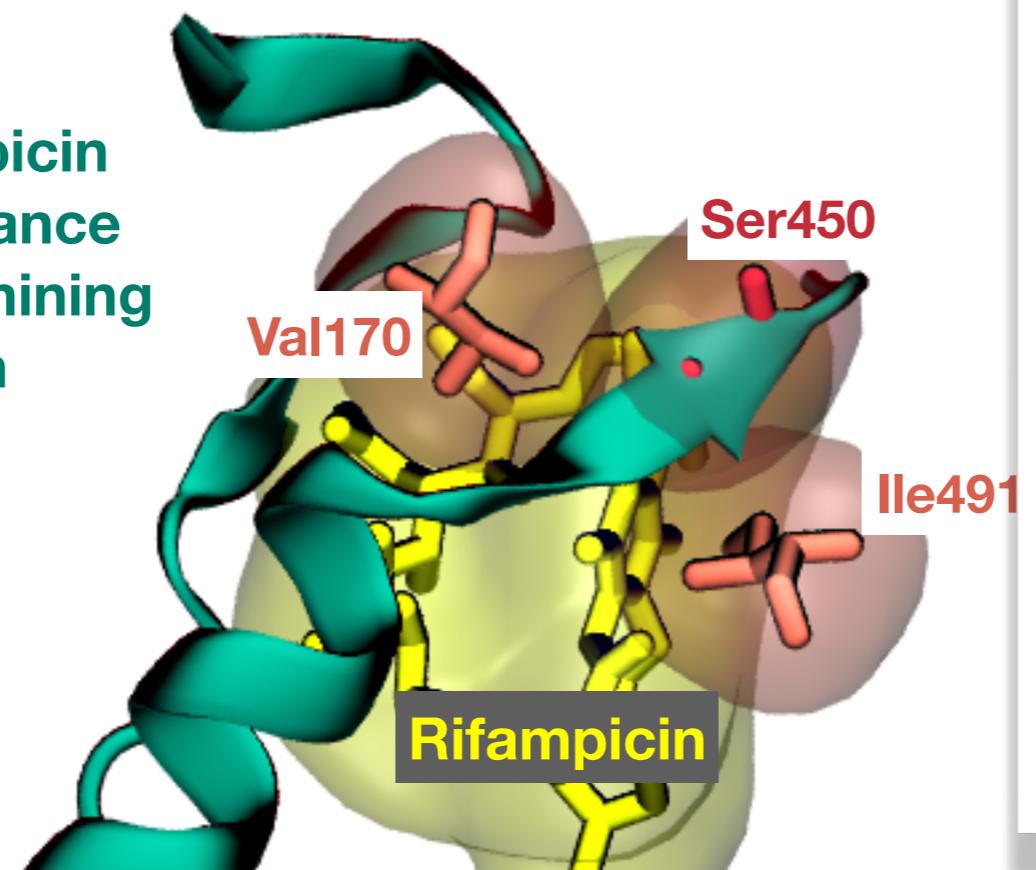








Rifampicin Resistance Determining Region



Detection of Drug-Resistant Tuberculosis by Xpert MTB/RIF in Swaziland

TO THE EDITOR: Tuberculosis is a major global health problem that has worsened with the increasing emergence of *Mycobacterium tuberculosis* (MTB) complex strains that are resistant to rifampin (RIF) and isoniazid. As recommended by the World Health Organization (WHO), the timely detection of drug resistance with the use of rapid molecular diagnostic tests, such as the Xpert MTB/RIF assay (Cepheid), is essential for appropriate treatment of patients with tuberculosis and for limiting the further spread of multidrug-resistant disease.^{1,2}

We used 34-loci mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR) analysis and spoligotyping to perform classic genotypic analysis of MTB complex strains from the most recent (2009) national survey of tuberculosis-drug resistance in Swaziland, a country with a high prevalence of tuberculosis (945 cases per 100,000 persons, or approximately 1%).³ We found that 38 of 125 multidrug-resistant strains (30%) that were isolated during the survey carried the *rpoB* I491F mutation, which confers resistance to rifampin (Table 1; and the Supplementary Appendix, available with the full text of this letter at NEJM.org). This mutation, which was previously reported with low frequency in clinical isolates from Hong Kong and Australia,⁴ is not detected by the Xpert MTB/RIF assay.

Xpert MTB/RIF, a cartridge-based point-of-care assay, is designed to identify rifampin resistance mutations in an 81-bp region of *rpoB* (codons 426 to 452). Its inability to detect the *rpoB* I491F outbreak strain raises new challenges, since Xpert MTB/RIF is used throughout most of Swaziland as the first-line diagnostic test for tuberculosis and for multidrug-resistant tuberculosis, as recommended by the WHO.⁵ Thus, the circulation of strains with the *rpoB* I491F

mutation substantially reduces the sensitivity of Xpert MTB/RIF-based diagnosis in Swaziland and presumably results in underdiagnosis and potentially inadequate treatment. This is problematic in a country where an estimated 26% of adults are infected with the human immunodeficiency virus (HIV) and 80% of patients with tuberculosis are coinjected with HIV. In addition, coinjected patients are more likely than

Table 1. Mutations in *rpoB* in 125 Multidrug-Resistant Strains From the 2009 Survey Regarding Tuberculosis-Drug Resistance in Swaziland*

Mutation	Strains with Mutation no. (%)	Mutation in <i>rpoB</i> Hot-Spot Region†
D450F	1 (0.8)	Yes
D450F, N450D	5 (2.4)	D450F, yes; N450D, yes
D450W	1 (0.8)	Yes
G421R‡, H421F	1 (0.8)	G421R, yes; H421F, no
H445D	7 (5.8)	Yes
H451L	6 (4.8)	Yes
H451Y	6 (4.8)	Yes
H491F, R552C	1 (0.8)	H491F, no; R552C, no
H51F	16 (12.4)	No
Q452→K154del	1 (0.8)	Yes
S450L	38 (29.6)	Yes
S450W	1 (0.8)	Yes
Unmutated	1 (0.8)	No

* Numbers are listed according to numbering for the *Mycobacterium tuberculosis* H37Rv genome. Some strains carry two mutations.

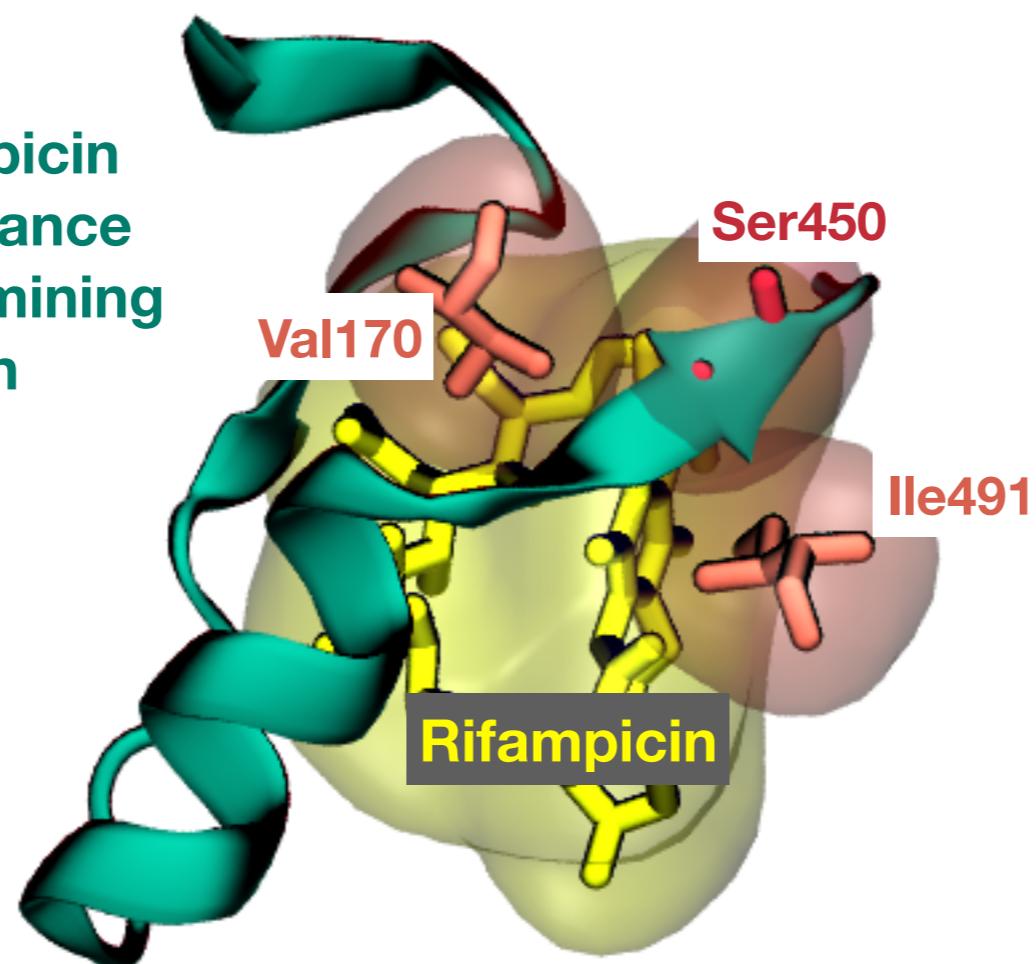
† The hot-spot region of *rpoB* ranges from codon 426 to codon 452.

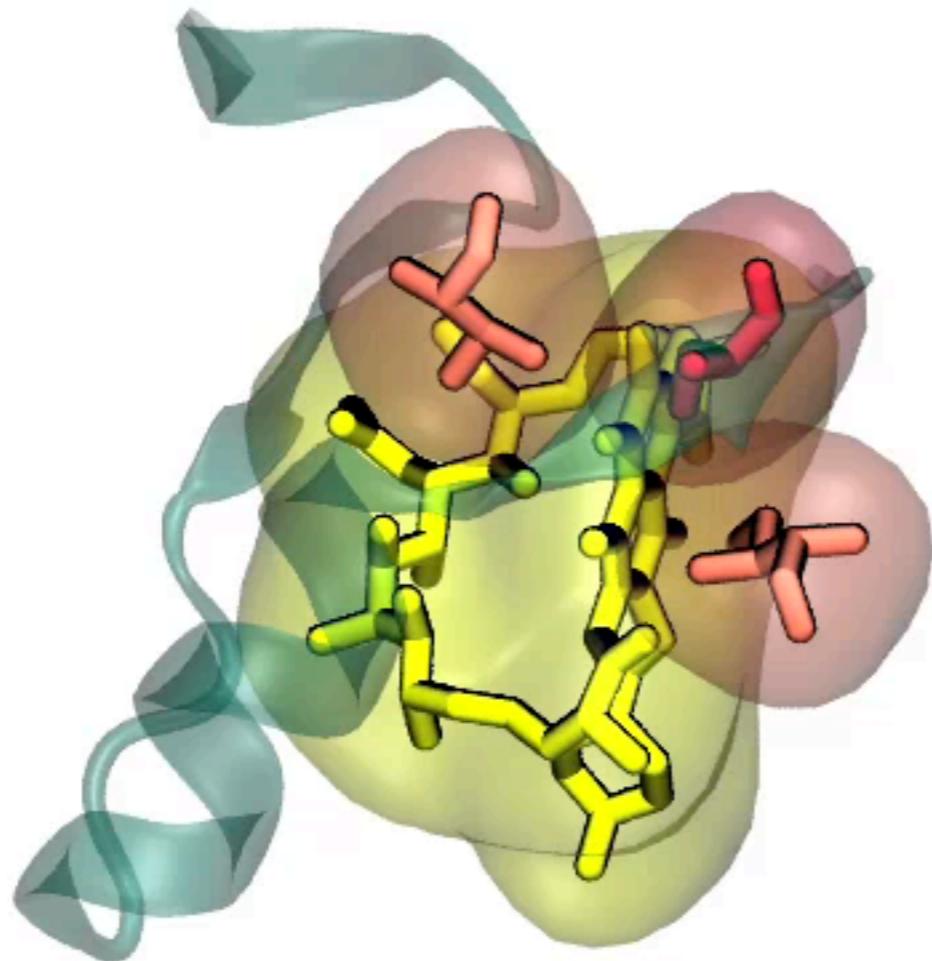
‡ This is a heterozygous single-nucleotide polymorphism.

THE NEW ENGLAND JOURNAL OF MEDICINE

“Thus, the circulation of strains with the *rpoB* I491F mutation substantially reduces the sensitivity of Xpert MTB/RIF-based diagnosis in Swaziland and presumably results in underdiagnosis and potentially inadequate treatment.”

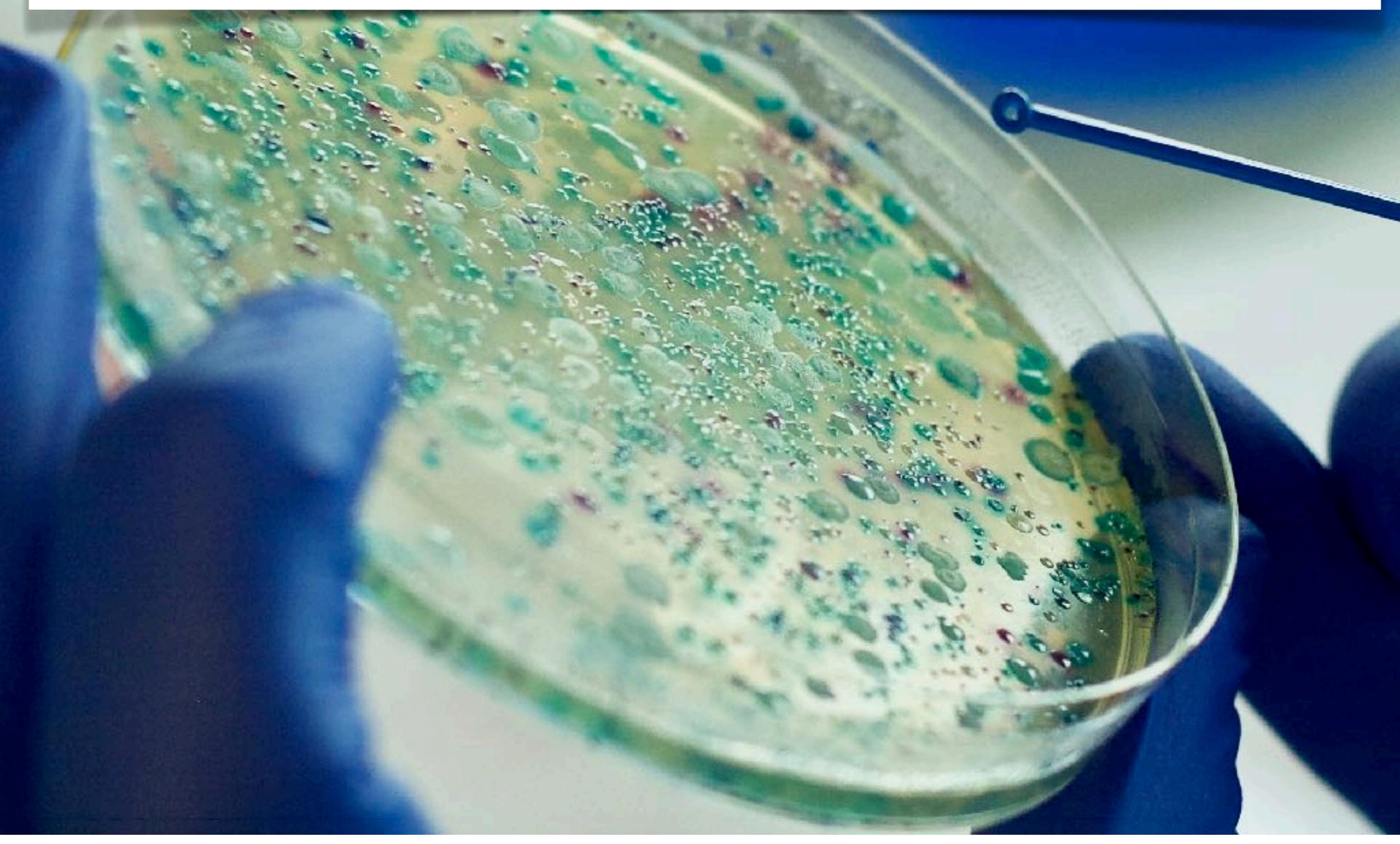
**Rifampicin
Resistance
Determining
Region**





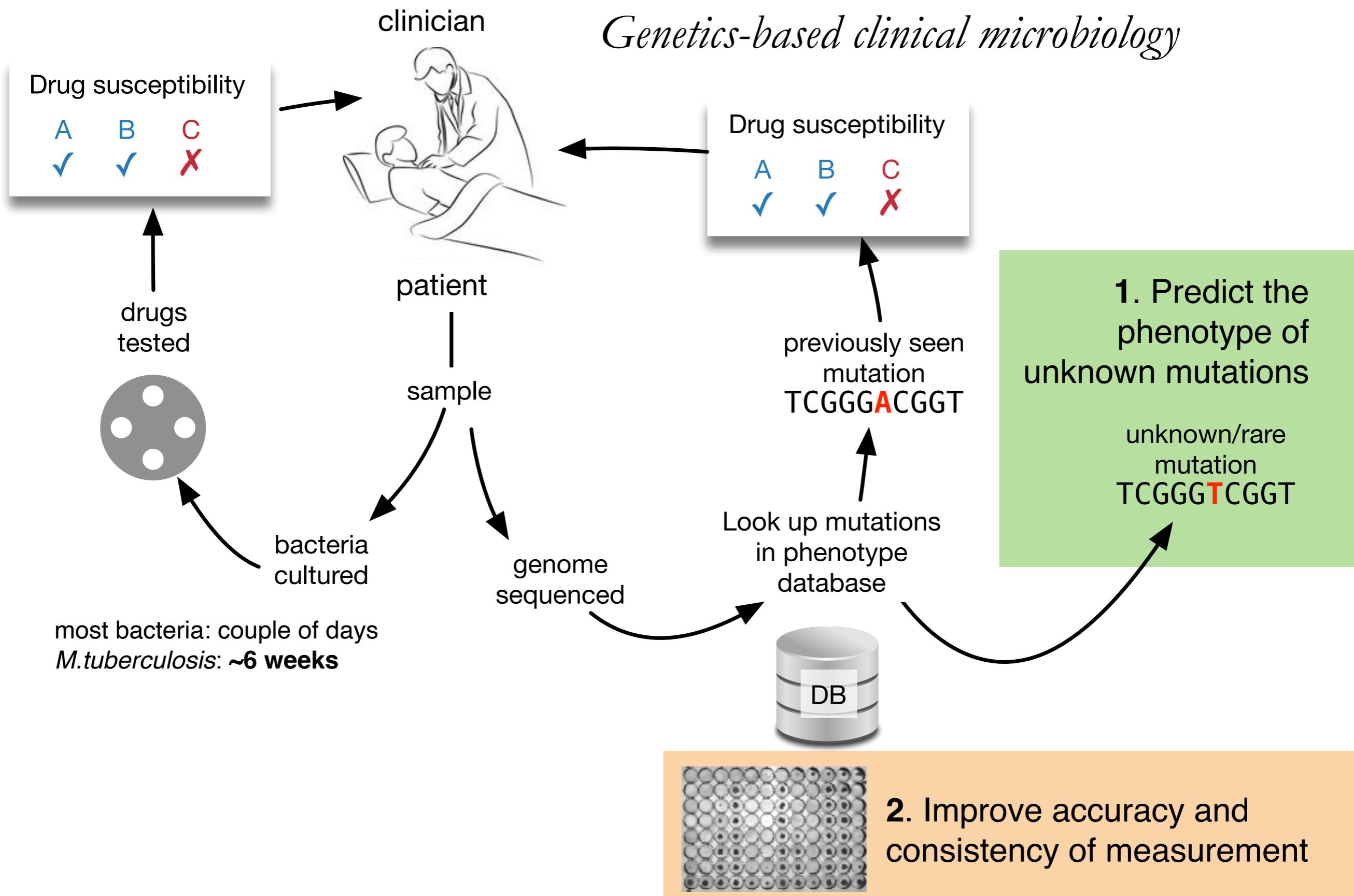
Rifampicin binds to the RNA polymerase, preventing the extension of RNA

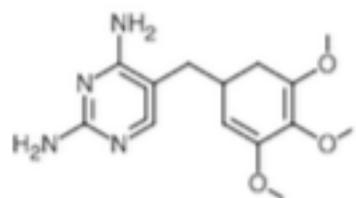
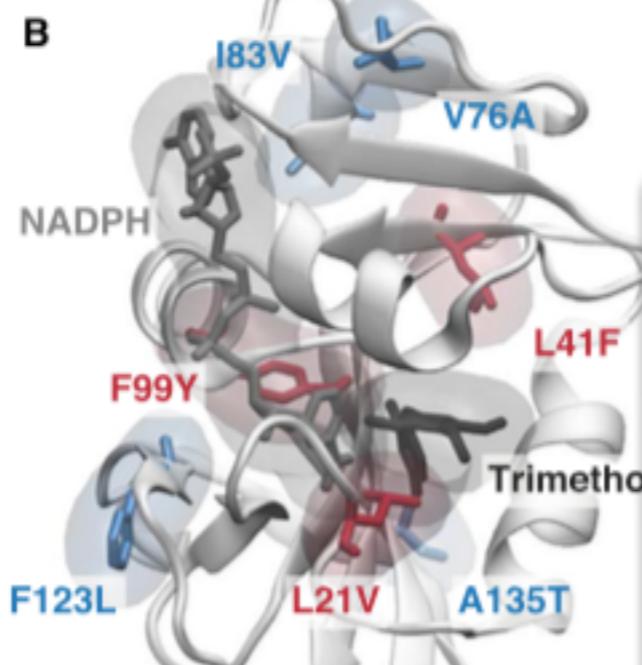
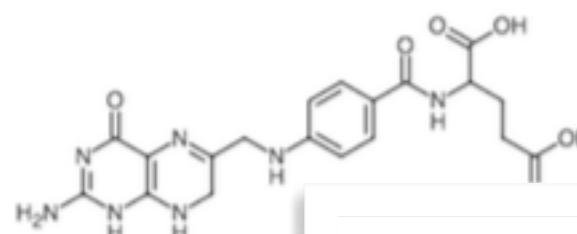
A large % of mutations that confer resistance occur in a 81bp region of *rpoB*



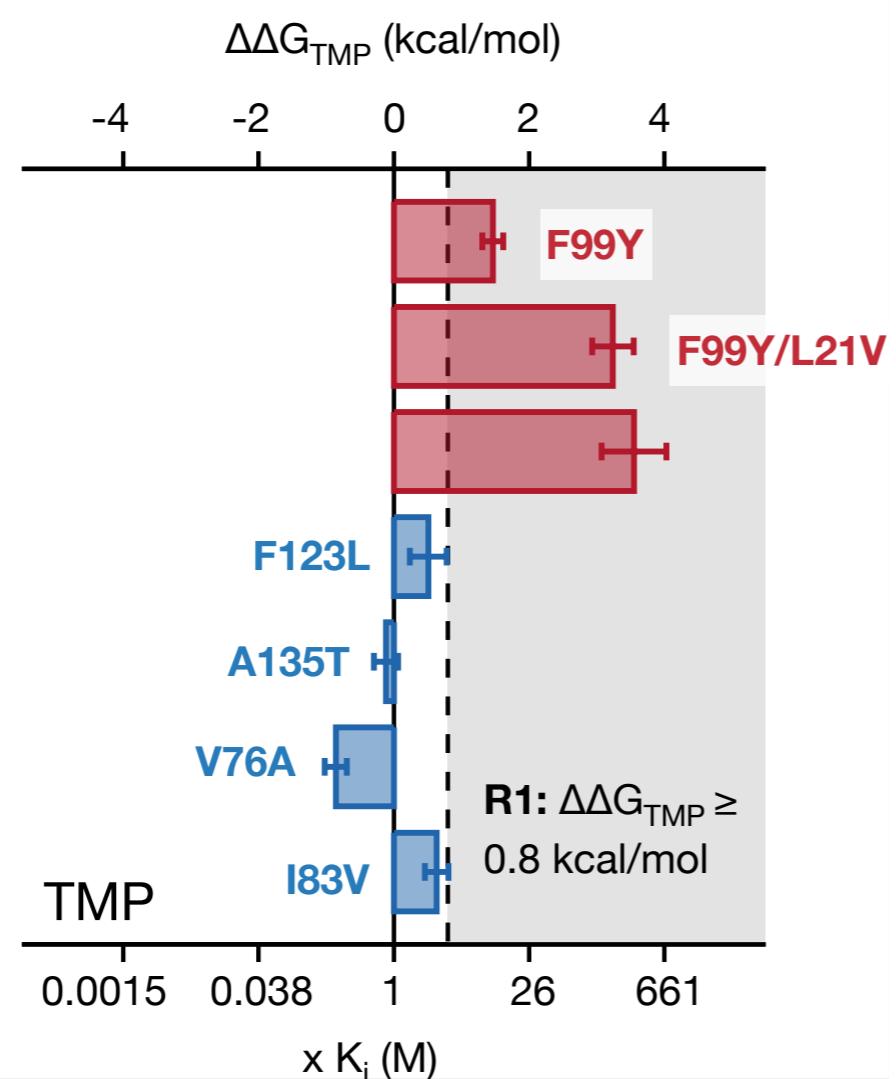
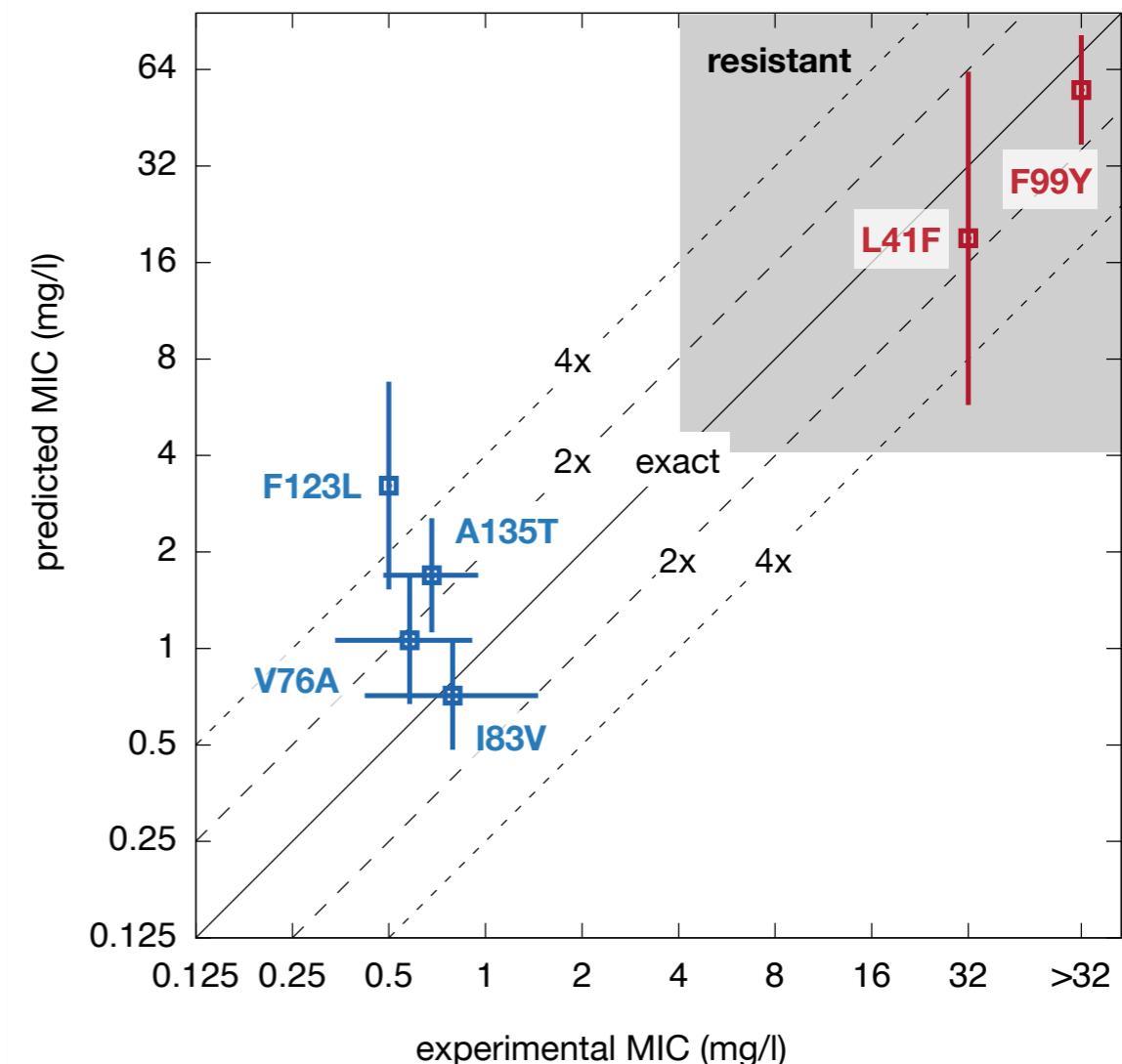
Traditional clinical microbiology

My research

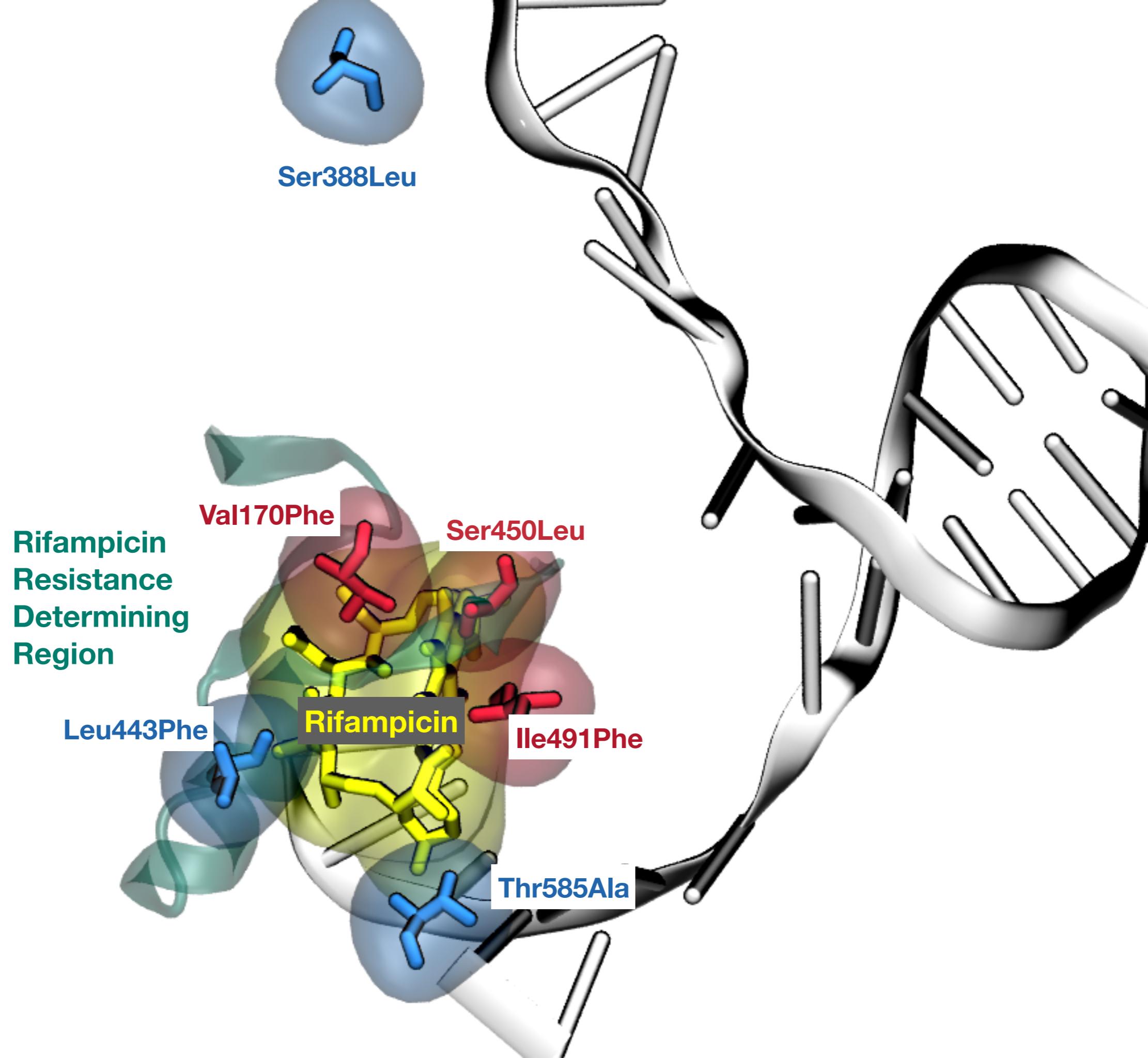


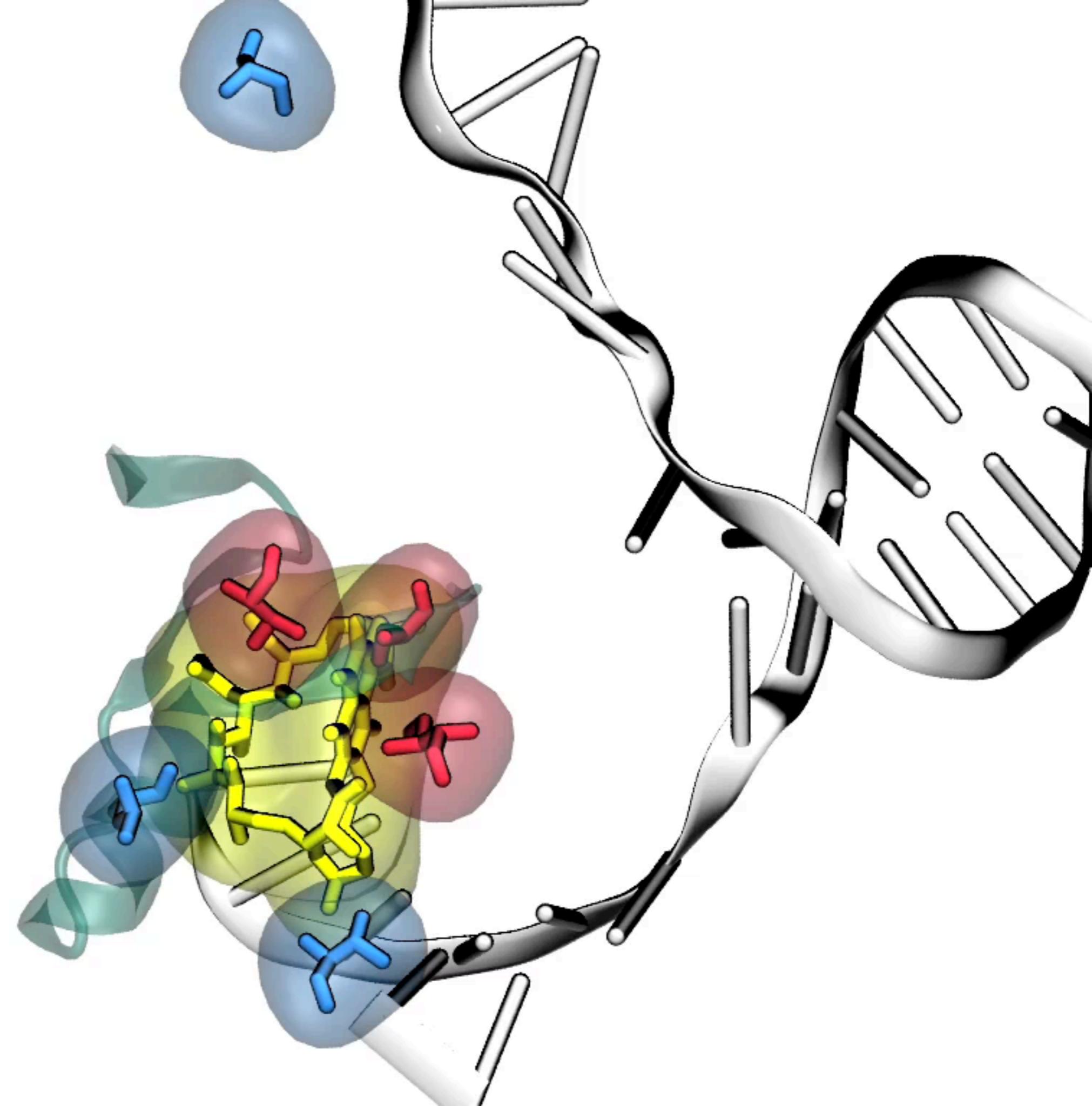
A trimethoprim (TMP)**dihydrofolic acid (DHA)** $\Delta\Delta G_{\text{TMP}}$ (kcal/mol)

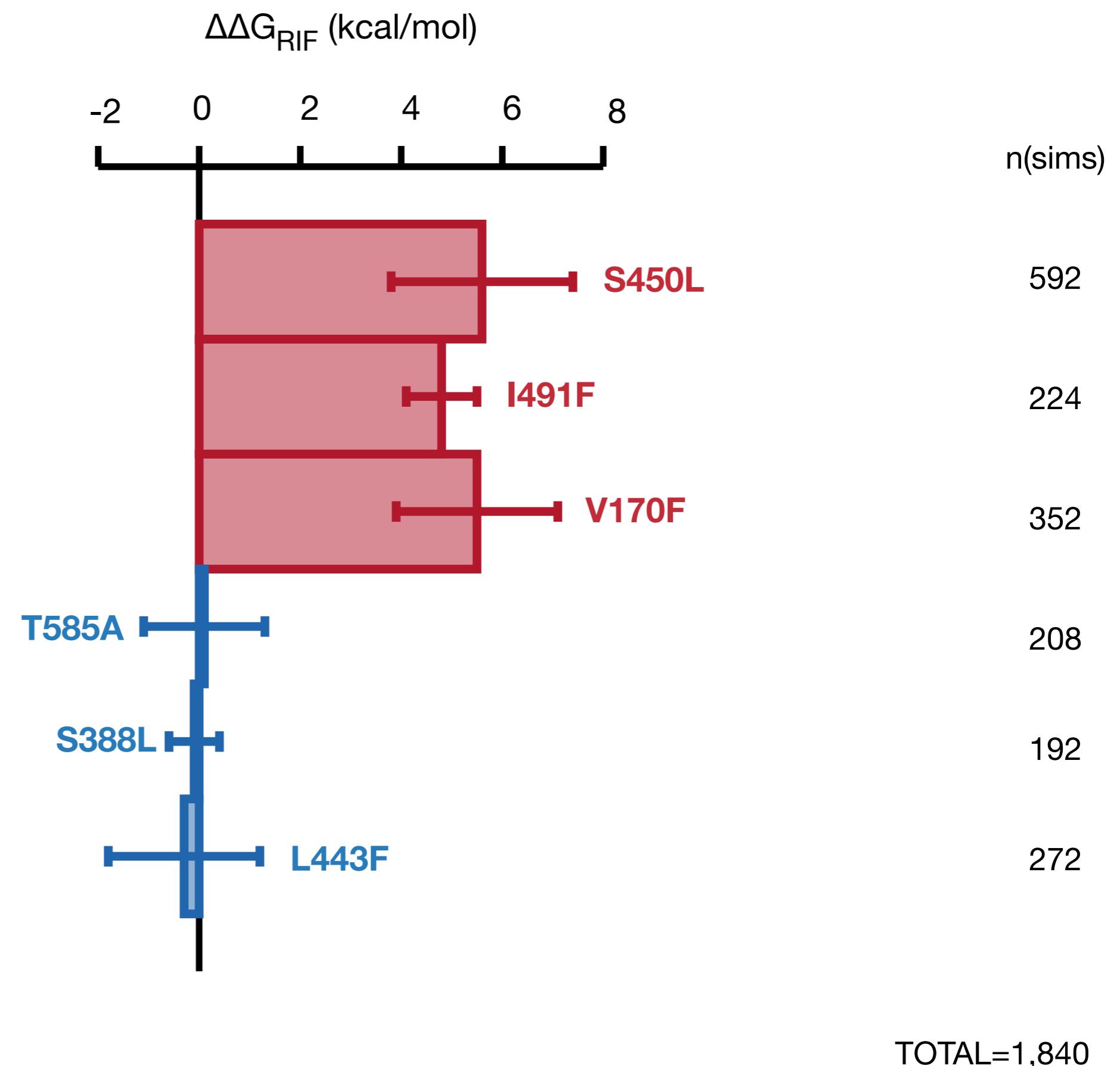
-4 -2 0 2 4

**e to**

from mutations in the chromosomal gene *dhfr*, or from the introduction of other naturally resistant genes (*dfrA*, *dfrG*, and *dfrK*) via plasmids (Iwaya, 2003; Nurjati et al., 2014). Here we focus on seven mutations in the *dhfr* chromosomal gene. We have chosen this gene for five reasons: (1) a series of resistance-conferring and no-effect mutations have been identified via whole-genome sequencing of isolates from patient infections (Gordon et al., 2014), as well as by more traditional methods, (2) the most common resistance-conferring mutation is a very small chemical change (Phe → Tyr) and this is therefore a challenging test for any predictive approach, (3) DHFR is a small, soluble protein that has been well studied, (4) several experimental







Rifampicin binds to the RNA polymerase, preventing the extension of RNA

A large % of mutations that confer resistance occur in a 81bp region of *rpoB*

Proximal mutations *reduce* how well rifampicin binds to the RNAP





- Tier-1 High Performance Supercomputer
- 4,920 compute nodes each with 2x12 core CPUs = 118,080 cores, 314TB memory
- These calculations used 90,000 CPU hours courtesy of CCPBioSim
- Notional cost £3,360



- Tier-0 High Performance Supercomputer
- 22,636 compute nodes each with 16 cores = 362,240 cores with 1.38 PB memory
- Also 4,228 nodes with 1x GPU and 8 cores.

**Rifampicin
Resistance
Determining
Region**

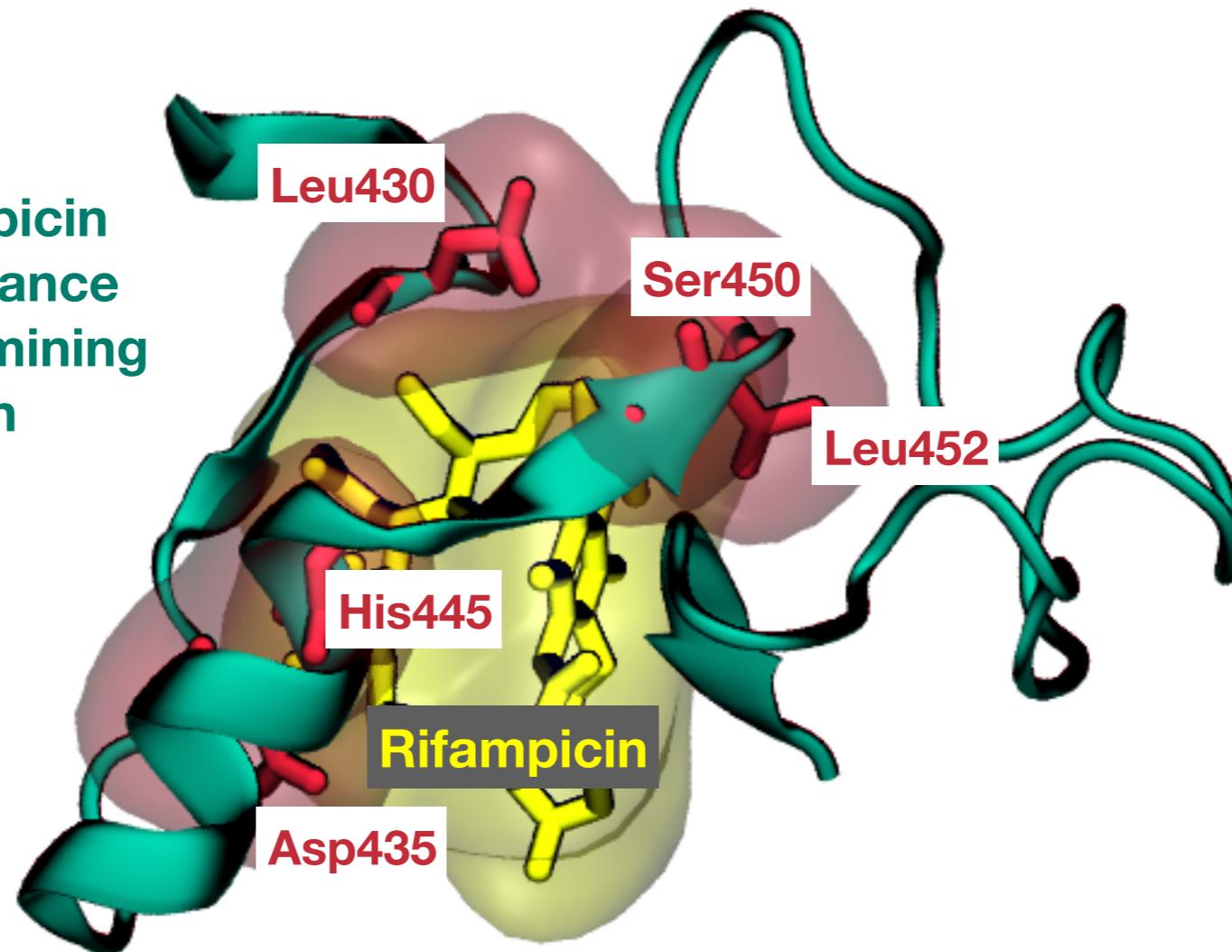


TABLE 1. Relative fitness of induced H37Rv rifampin-resistant mutant alleles

SSCP pattern	No. of replicates	Mean no. of generations of H37Rv		Location of mutation	Relative fitness	
		Sensitive	Resistant			
His445	A ₁	4	8.65	6.95	His ₅₂₆ →Tyr	0.80
	A ₂	4	5.90	2.80	His ₅₂₆ →Tyr	0.78
	A ₃	3	6.73	4.62	His ₅₂₆ →Asp	0.42
Ser450	B ₁	5	7.88	8.30	Ser ₅₃₁ →Leu	1.05
	B ₂	3	7.73	7.43	Ser ₅₃₁ →Leu	0.93
	B ₃	3	6.83	6.29	Ser ₅₃₁ →Leu	0.89
	B ₄	3	9.60	4.83	Ser ₅₃₁ →Leu	0.50
His445	C ₁	2	5.68	10.04	His ₅₂₆ →Arg	0.56
	C ₂	5	6.40	1.35	His ₅₂₆ →Arg	0.21

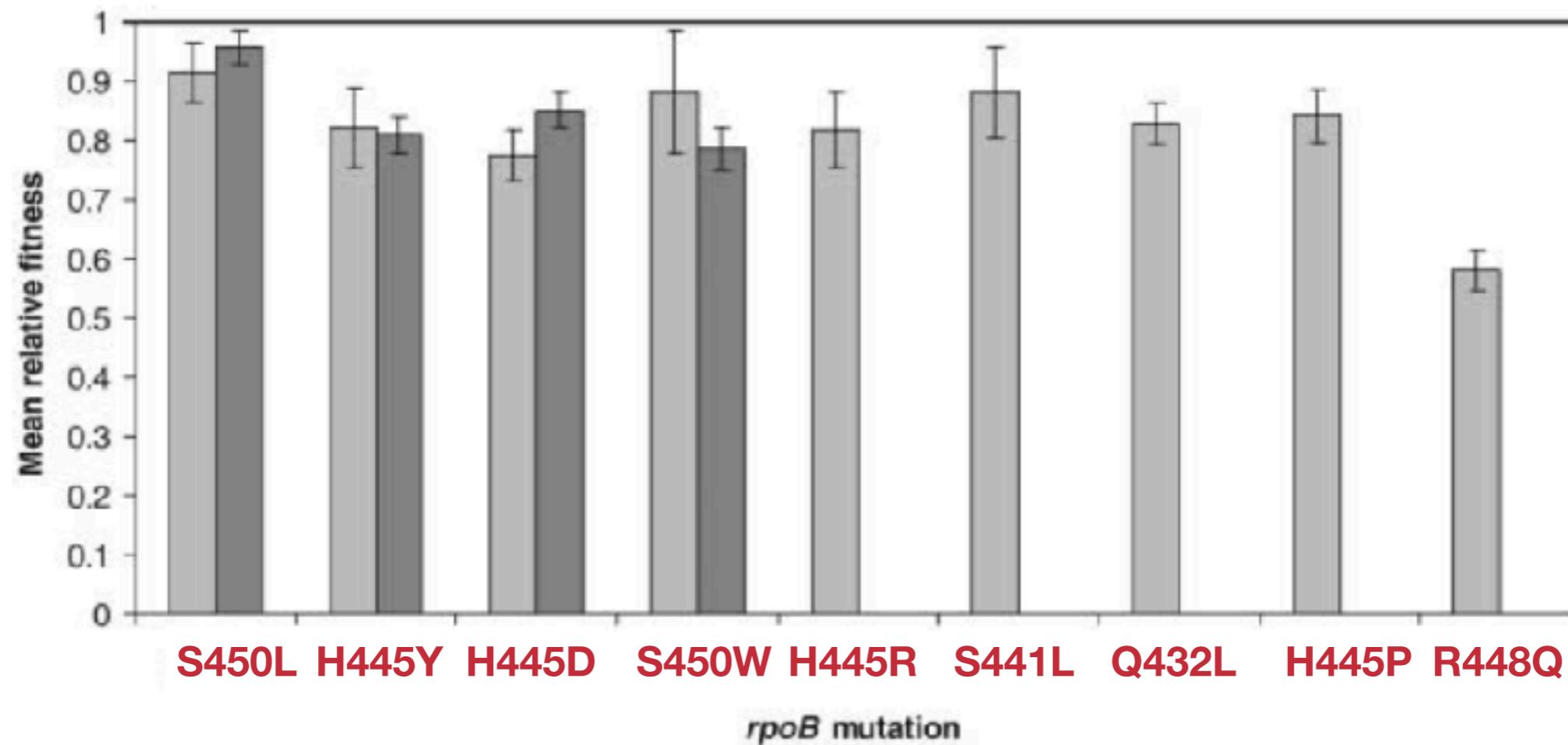
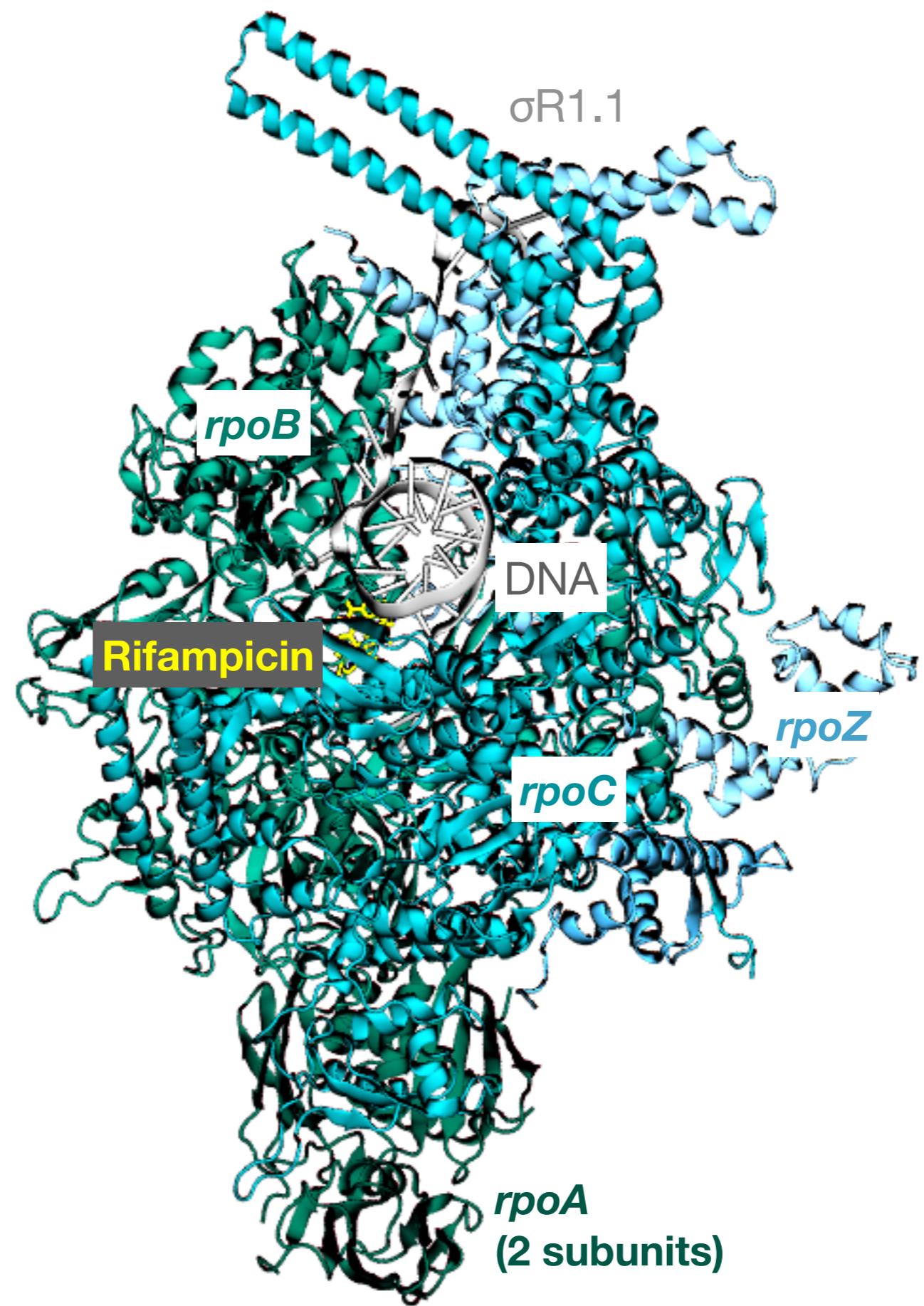


Fig. 1. Relative competitive fitness of laboratory-derived rifampin-resistant mutants of *M. tuberculosis*. All mutants had a statistically significant fitness cost (error bars indicate 95% confidence intervals). This cost was less in *rpoB* S531L mutants than in other *rpoB* mutants, irrespective of the strain background. Light gray bars, CDC1551 mutants; dark gray bars, T85 mutants. Y, Tyr; W, Trp; P, Pro.



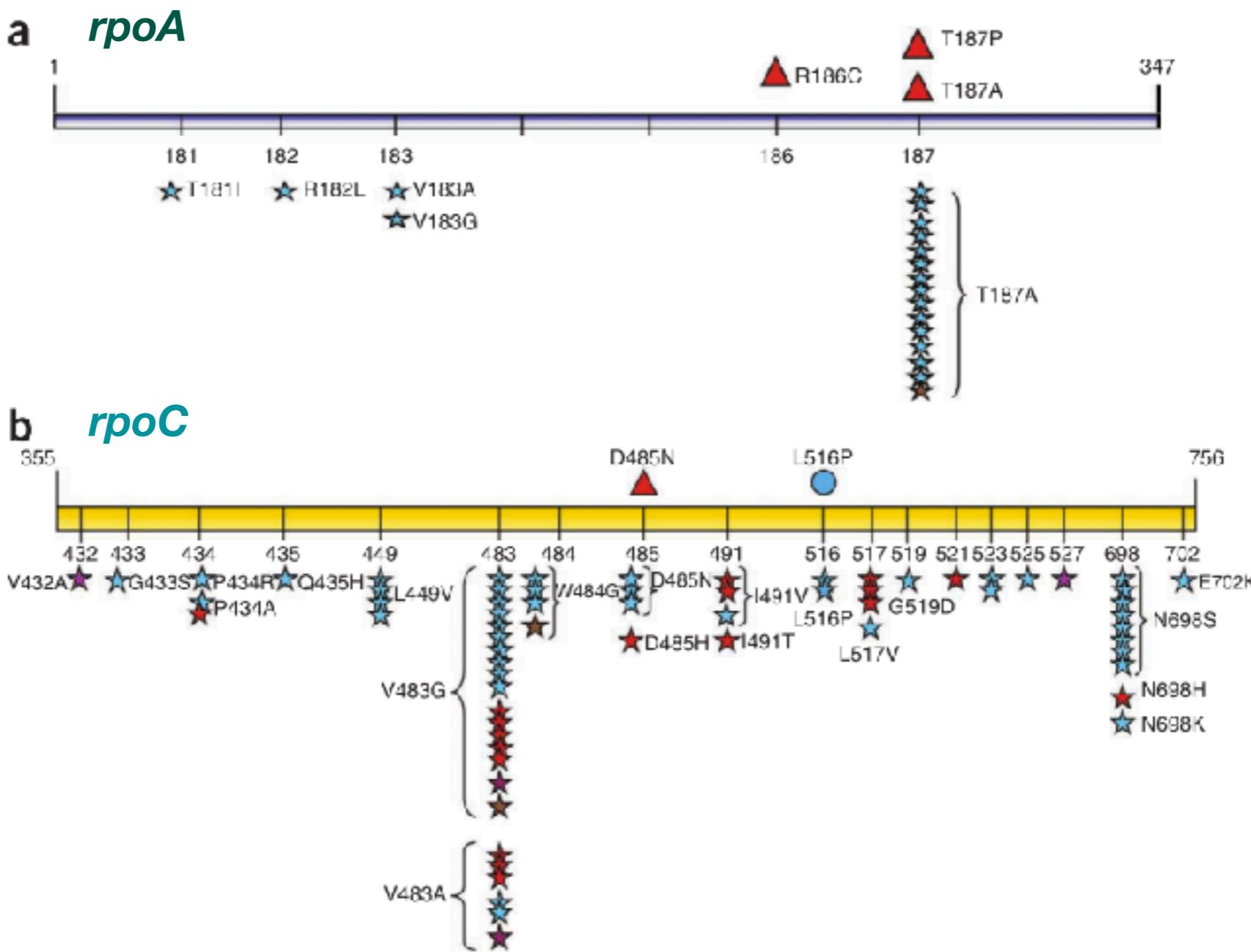


Figure 1 Putative compensatory mutations in *rpoA* and *rpoC* of *M. tuberculosis*. **(a,b)** Mutations identified after genome sequencing of experimentally evolved strains (circle) or paired clinical isolates (triangles) are indicated above the gene diagrams of *rpoA* **(a)** and *rpoC* **(b)**. Mutations identified by screening a global and a high-burden collection of MDR strains are indicated by stars below the gene diagrams. Colors indicate the respective strain lineage (blue, lineage 2; red, lineage 4; brown, lineage 5; pink, lineage 1). Some of these mutations occurred in multiple lineages or affect the same codon position.

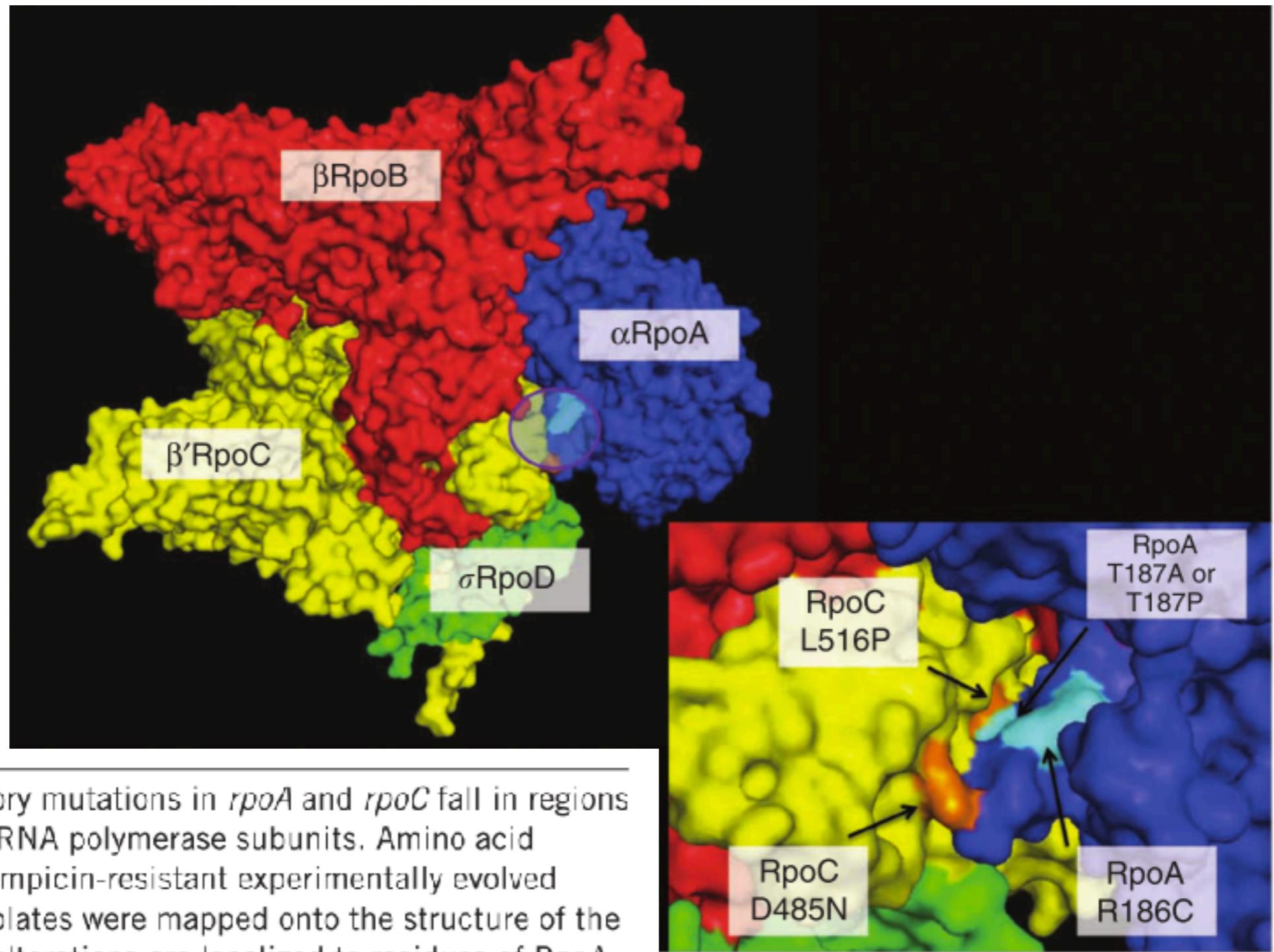
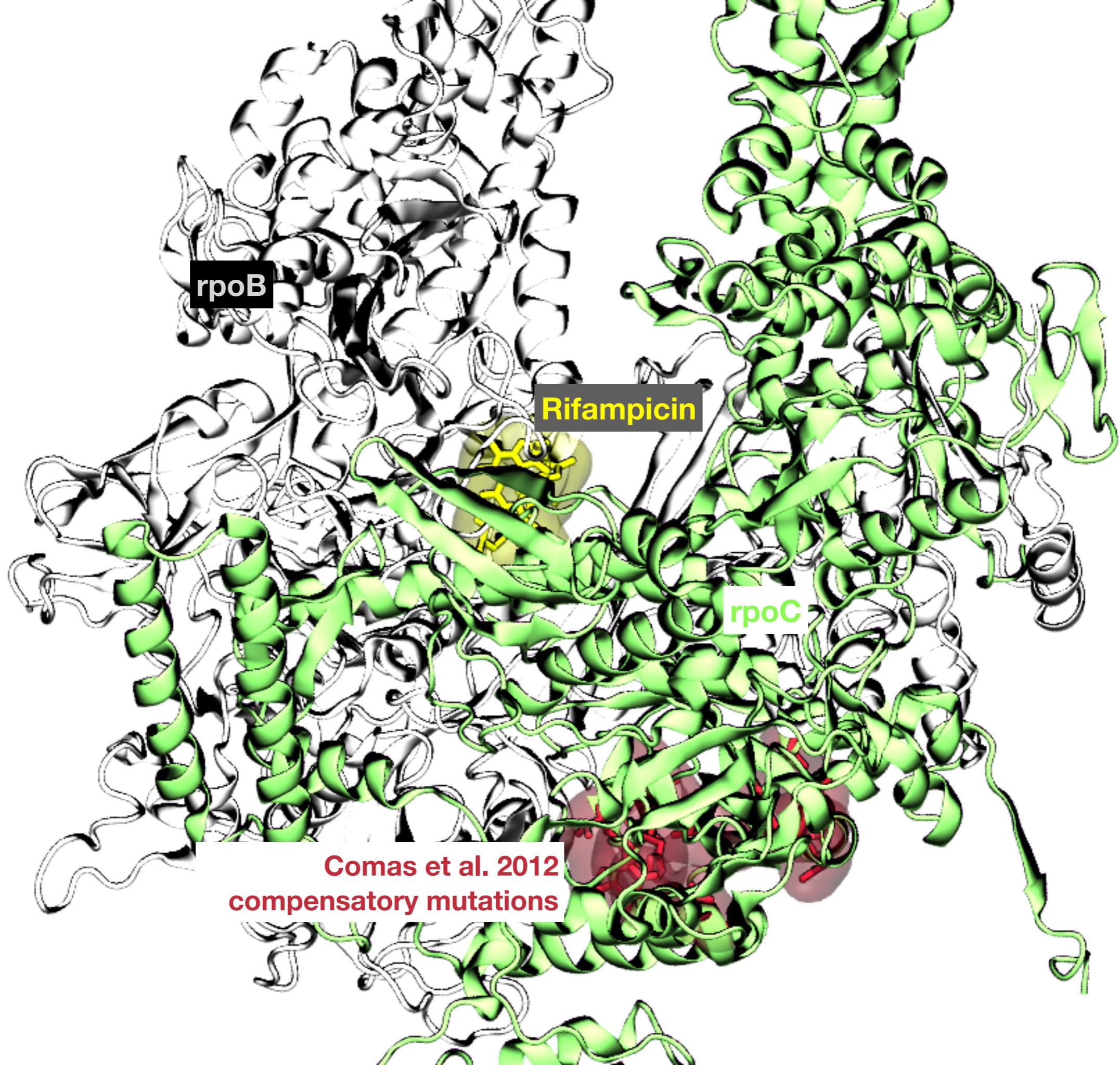
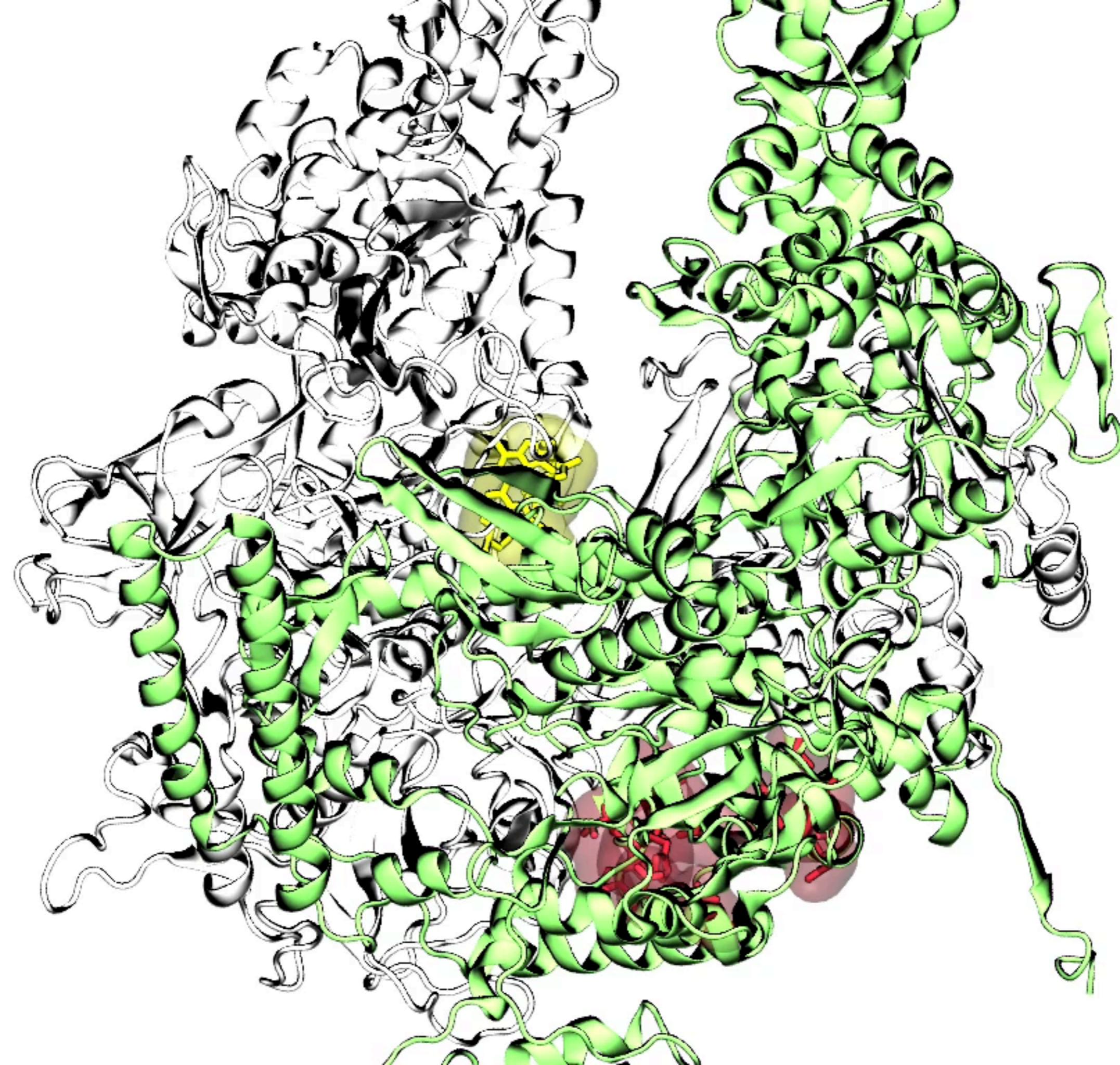
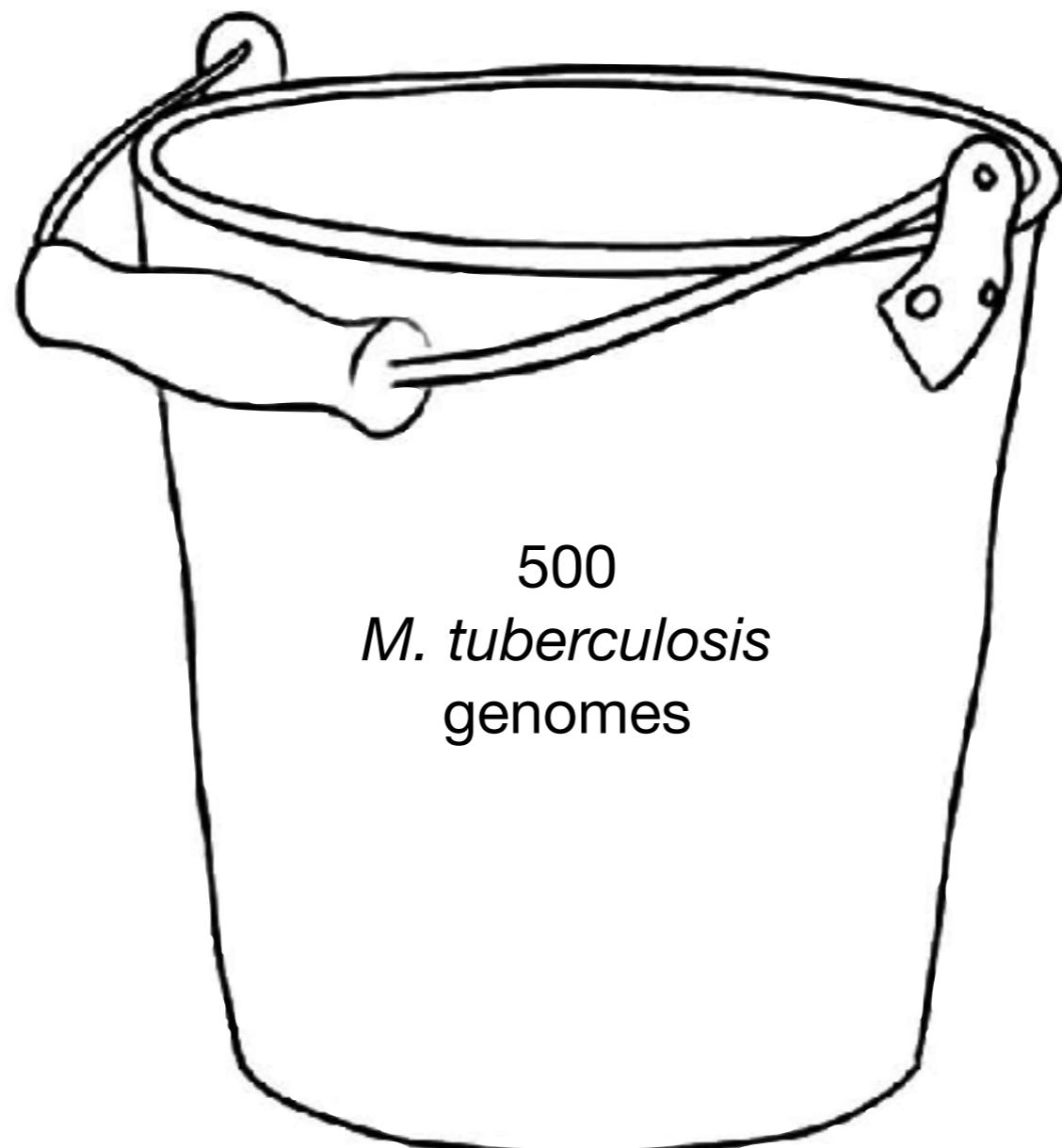


Figure 2 Putative compensatory mutations in *rpoA* and *rpoC* fall in regions encoding the interface of the RNA polymerase subunits. Amino acid substitutions identified in rifampicin-resistant experimentally evolved isolates and paired clinical isolates were mapped onto the structure of the *E. coli* RNA polymerase. The alterations are localized to residues of RpoA (light blue) and RpoC (orange) that are predicted to have roles in RNA polymerase subunit interaction. Residue numbers are indicated according to *M. tuberculosis* coordinates. RpoA (α subunit), blue; RpoB (β subunit), red; RpoC (β' subunit), yellow; RpoD (σ subunit), green.

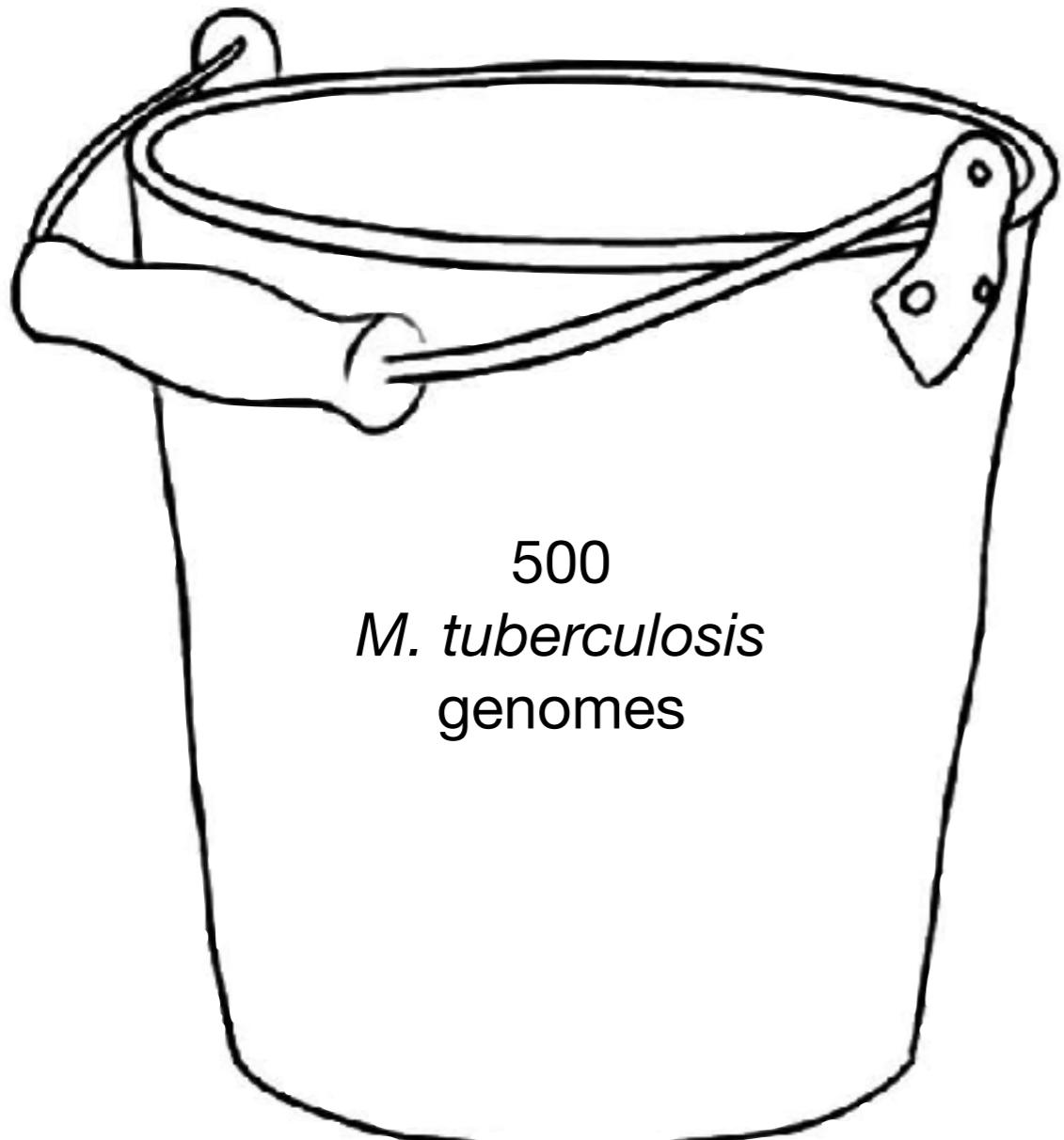




Comas et al. 2012



Comas et al. 2012



1. European Nucleotide Archive

28,000 genomes

2. CRyPTIC

4,500 genomes
(end May 2018)

1. European Nucleotide Archive



Comas et al. 2012



500
genomes

2. CRyPTIC



4,500 genomes
(end May 2018)

1. European Nucleotide Archive

EMBL-EBI 

Services Research Training About us

 ENA
European Nucleotide Archive

SRR2776873 [Search](#)
Examples: [BN000065](#), [histone](#)
[Advanced](#)
[Sequence](#)

Home [Search & Browse](#) [Submit & Update](#) [Software](#) [About ENA](#) [Support](#)

Search results for **SRR2776873** [Show more data from EMBL-EBI](#)

Read
Experiment (1)
Run (1)

Experiment (1 results found)
SRX1372111 Illumina HiSeq 2000 sequencing; GSM1917082: csoR_2; Mycobacterium tuberculosis; RNA-Seq
[View all 1 results](#)

Run (1 results found)
SRR2776873 Illumina HiSeq 2000 sequencing; GSM1917082: csoR_2; Mycobacterium tuberculosis; RNA-Seq
[View all 1 results](#)



New Results

Real-time search of all bacterial and viral genomic data

Phelim Bradley, Henk den Bakker, Eduardo Rocha, Gil McVean, Zamin Iqbal

doi: <https://doi.org/10.1101/234955>

This article is a preprint and has not been peer-reviewed [what does this mean?].

[Abstract](#)[Info/History](#)[Metrics](#) [Preview PDF](#)

Abstract

Genome sequencing of pathogens is now ubiquitous in microbiology, and the sequence archives are effectively no longer searchable for arbitrary sequences. Furthermore, the exponential increase of these archives is likely to be further spurred by automated diagnostics. To unlock their use for scientific research and real-time surveillance we have combined knowledge about bacterial genetic variation with ideas used in web-search, to build a DNA search engine for microbial data that can grow incrementally. We indexed the complete global corpus of bacterial and viral whole genome sequence data (447,833 genomes), using four orders of magnitude less storage than previous methods. The method allows future scaling to millions of genomes. This renders the global archive accessible to sequence search, which we demonstrate with three applications: ultra-fast search for resistance genes MCR1-3, analysis of host-range for 2827 plasmids, and quantification of the rise of antibiotic resistance prevalence in the sequence archives.

[? What is this page?](#)[Embed badge](#)[Share](#)

Real-time search of all bacterial and viral genomic data

Overview of attention for article published in bioRxiv, December 2017



About this Attention Score

In the top 5% of all research outputs scored by Altmetric

[MORE...](#)

Mentioned by

- 558 tweeters
- 1 Facebook page
- 1 Google+ user

Readers on

- 67 Mendeley

[What is this page?](#)

SUMMARY

[Twitter](#)[Facebook](#)[Google+](#)

Title Real-time search of all bacterial and viral genomic data

Published in bioRxiv, December 2017

DOI [10.1101/234955](https://doi.org/10.1101/234955)

Authors Phelim Bradley, Henk den Bakker, Eduardo Rocha, Gil McVean, Zamin Iqbal

[View on publisher site](#)

[Alert me about new mentions](#)

TWITTER DEMOGRAPHICS

MENDELEY READERS

ATTENTION SCORE IN CONTEXT

This research output has an Altmetric Attention Score of 333. This is our high-level measure of the quality and quantity of online attention that it has received. This Attention Score, as well as the ranking and number of research outputs shown below, was calculated when the research output was last mentioned on 28 May 2018.

ALL RESEARCH OUTPUTS

#24,671

of 11,279,987 outputs

OUTPUTS FROM biorXIV

#51

of 26,395 outputs

OUTPUTS OF SIMILAR AGE

#1,553

of 287,859 outputs

OUTPUTS OF SIMILAR AGE FROM biorXIV

#8

of 3,428 outputs

Altmetric has tracked 11,279,987 research outputs across all sources so far. Compared to these this one has done particularly well and is in the 99th percentile: It's in the top 5% of all research outputs ever tracked by Altmetric.

Searching a snapshot of publically available bacterial WGS datasets from the ENA/SRA (N=455,632) Dec 2016.

This is a proof-of-concept demonstration of the BIGSI search index for microbial genomes. We have indexed the complete bacterial and viral whole-genome sequence content of the European Nucleotide Archive as of December 2016. See [our paper](#).

Thanks to CLIMB for hosting.

You can use this to search for samples with a given gene, plasmid, or SNP. Queries must be at least 61bp in length. Species metadata provided by analysis by Bracken + Kraken.

More info at <https://bigsi.readme.io/> and <http://github.com/phelimb/bigsi>.

CGGTCAAGTCCTGTTCTTGTGGCGAGTGTGCCGTTTGACCGCGACCGCCAATCTTACCTTTTGATAAAA

Proportion of query k-mers threshold:

e.g. [MCR-1, OXA-1](#)

169 results

- ▶ 100% of query k-mers found in ERR1149371 (Escherichia coli : 91.1%; Shigella flexneri : 3.94%;)
- ▶ 100% of query k-mers found in ERR1163331 (Escherichia coli : 57.53%; Shigella boydii : 39.08%;)
- ▶ 100% of query k-mers found in ERR1163291 (Escherichia coli : 91.1%; Shigella flexneri : 3.94%;)
- ▶ 100% of query k-mers found in ERR1046133 (Salmonella enterica : 94.87%; Escherichia coli : 4.99%;)
- ▶ 100% of query k-mers found in ERR1609312 (Escherichia coli : 53.99%; Shigella dysenteriae : 38.14%;)
- ▶ 100% of query k-mers found in ERR1609249 (Escherichia coli : 89.11%; Shigella boydii : 6.74%;)
- ▶ 100% of query k-mers found in ERR1609215 (Escherichia coli : 97.47%; Shigella boydii : 0.83%;)

[Code](#)[Issues 0](#)[Pull requests 0](#)[Projects 0](#)[Wiki](#)[Insights](#)[Settings](#)

a Python class to interrogate BIGISI

[Edit](#)[Add topics](#)

47 commits

2 branches

0 releases

1 contributor

MIT

Branch: [master](#) ▾[New pull request](#)[Create new file](#)[Upload files](#)[Find file](#)[Clone or download](#) ▾ philipwfowler started adding by residue graphing

Latest commit 9206e94 on 5 Apr

 examples	whoops, no update for a while	2 months ago
 pygsi	started adding by residue graphing	2 months ago
 .DS_Store	whoops, no update for a while	2 months ago
 .gitignore	updated examples for new 63-kmer window approach	3 months ago
 CHANGES.txt	adding setup.py	5 months ago
 LICENSE	Initial commit	5 months ago
 README.md	Update README.md	3 months ago
 setup.py	improved setup.py	5 months ago

[README.md](#)

pygsi: a Python class to interrogate BIGISI

As discussed in the preprint below, whilst the Short Read Archive (SRA) holds nearly half a million bacterial and viral genomes, it is extremely difficult to, given a nucleotide sequence X, answer the question "how many times has X been deposited in the SRA?".

Bradley P, Den Bakker HC, Rocha EPC, McVean G, Iqbal Z. Real-time search of all bacterial and viral genomic data. 2017. [bioRxiv](#)

Consider a trivial case where we wish to find out all the single amino acid (i.e. single triplet) differences in a stretch of 12 amino acids i.e. (amino acid pos 1-12 incl, or bases 1-36). We therefore need to provide pygsi with amino acids -9 to 22 incl. (bases -29 to 66) and so the first fishing sequence is

base#	1	2	3
	123456789012345678901234567890123456		
-----000-----.			

where '.' is a nucleotide that is not included in the fishing sequence defined by '-' and 'o' and we can find all variation using the `permute_position()` method of the `pygsi` class. Then we move on to consider the second amino acid position

base#	1	2	3
	123456789012345678901234567890123456		
.....000.....			

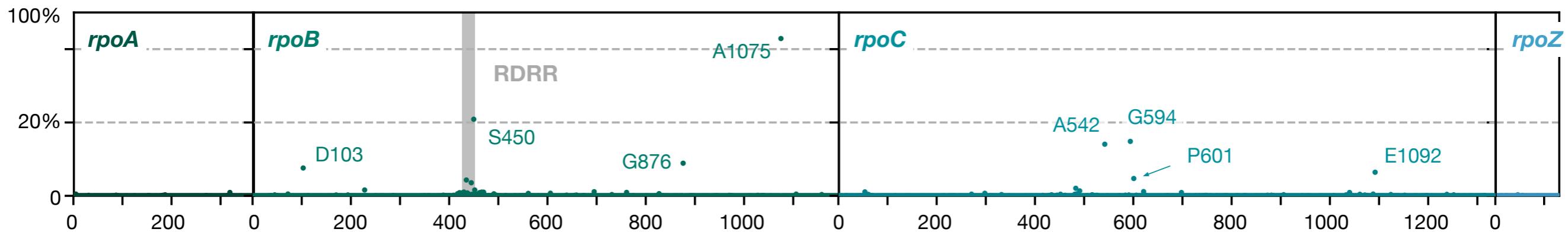
and then the third

base#	1	2	3
	123456789012345678901234567890123456		
.....000.....			

all the way up to the twelfth

base#	1	2	3
	123456789012345678901234567890123456		
.....			000-----

1. Proportion of sequences with any mutation



A novel multi SNP based method for identifying subspecies and associated lineages and sub-lineages of the *Mycobacterium tuberculosis* complex by whole genome sequencing

Authors

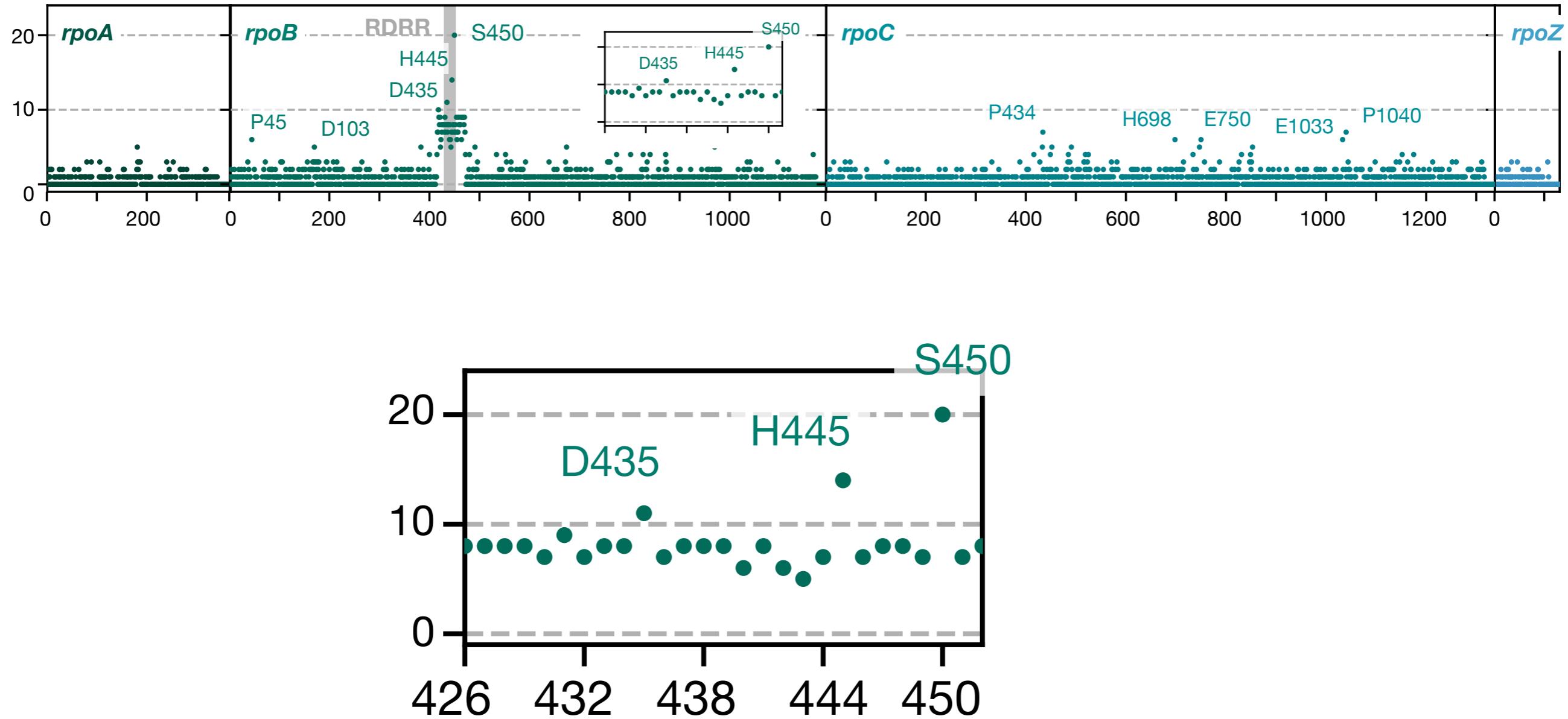
S Lipworth*, R Jajou*, H de Neeling, P Bradley, W van der Hoek, Z Iqbal, G Smith, T Peto, D Crook, T Walker*, D van Soolingen*

* denotes contributed equally

Abstract (limit 150 words)

The clinical phenotype of zoonotic tuberculosis, its contribution to the global burden of disease and prevalence is poorly understood and probably underestimated. This is partly because currently available laboratory and *in silico* tools have not been calibrated to accurately distinguish between all subspecies of the *Mycobacterium tuberculosis* complex (*Mtbc*). We here present the first such tool, SNPs to Identify TB ('SNP-IT'). Applying SNP-IT to a collection of clinical genomes from a UK reference laboratory, we demonstrate an unexpectedly high number of *M. avium* isolates. These are seen at a similar rate to *M. bovis* which attracts much health protection resource and yet *M. avium* cases have not been previously described in the UK. From an international perspective it is likely that *M. avium* is an underestimated zoonosis. As whole genome sequencing is increasingly integrated into the clinical setting, accurate subspecies identification with SNP-IT will allow the clinical phenotype, host range and transmission mechanisms of subspecies of the *Mtbc* to be studied in greater detail.

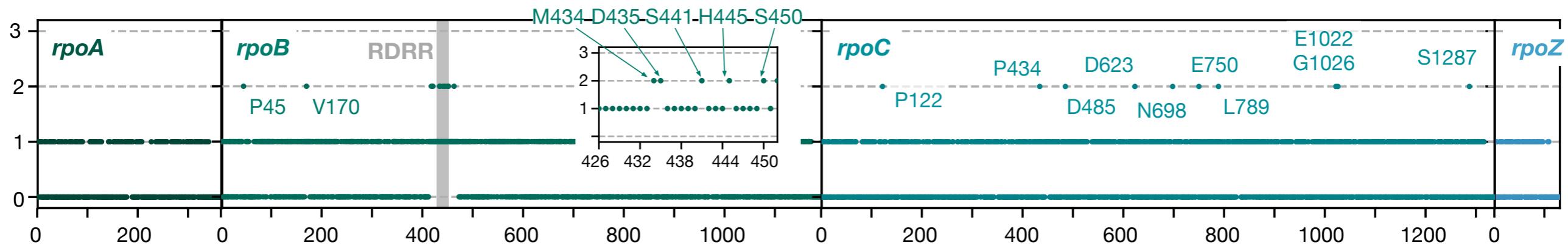
2. Number of distinct triplets observed



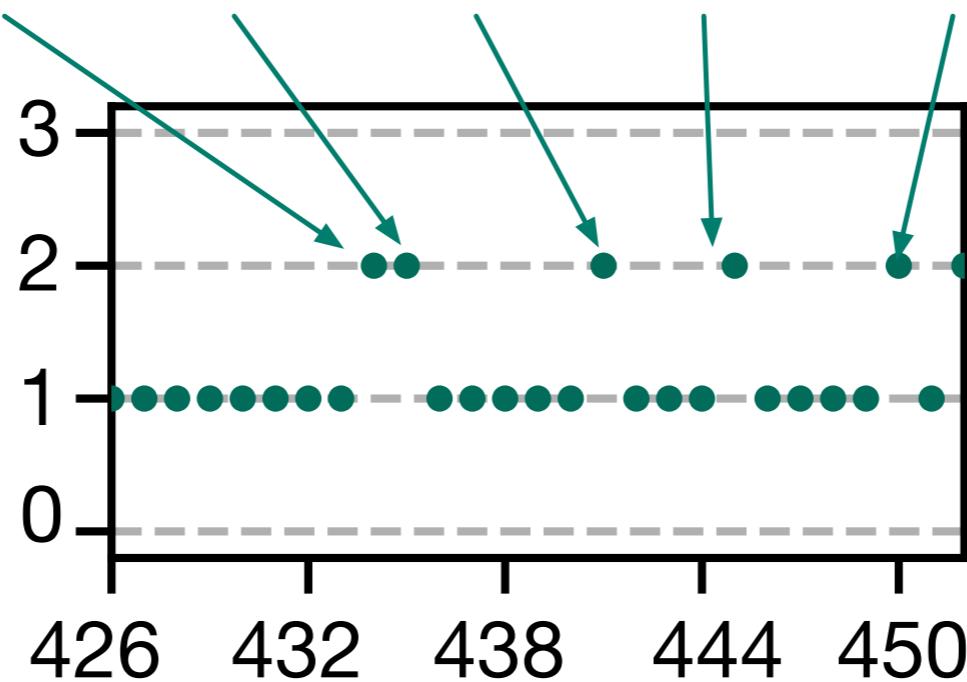
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	Thr	CGU	Arg
CUC	Leu						
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Only 5 Δ bp=1 non-synonymous mutations available..
 Ser (tcg) \rightarrow Leu, Trp, Pro, Thr, Ala

3. Maximum number of bases altered per triplet



M434 D435 S441 H445 S450

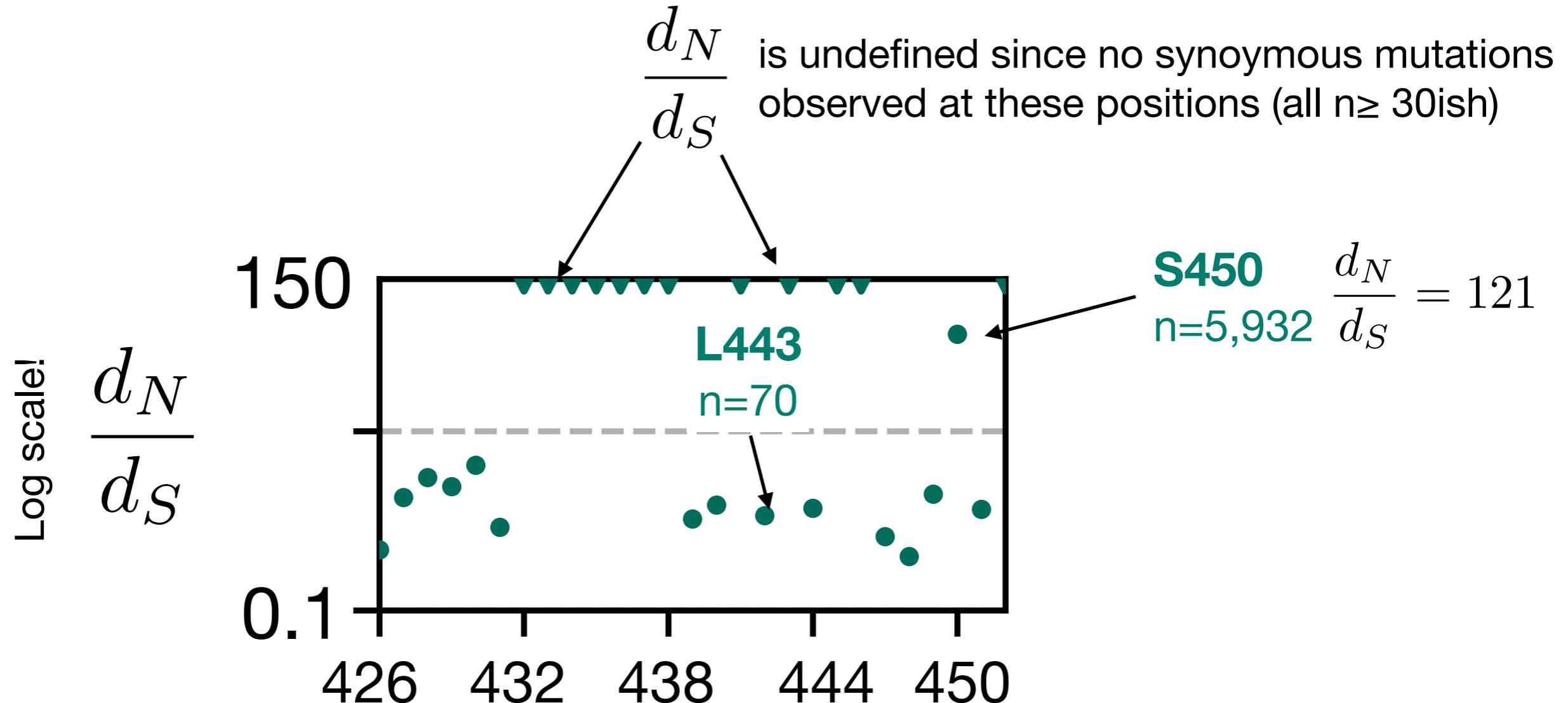


$$\frac{d_N}{d_S} = \frac{n(\text{non-synonymous mutations})^*}{n(\text{synonymous mutations})}$$

$\frac{d_N}{d_S} \sim 0$ Purifying selection

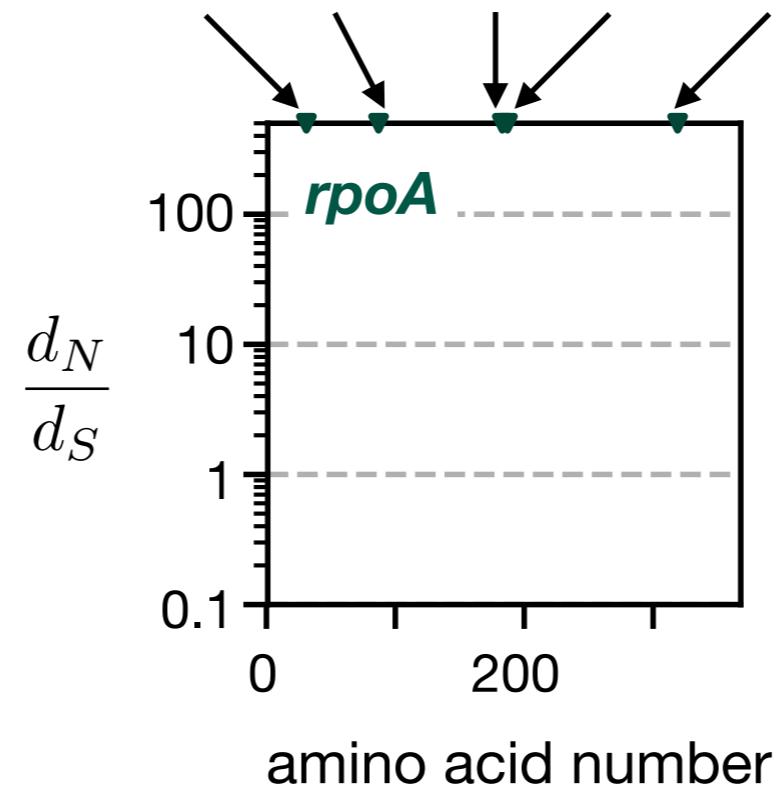
$\frac{d_N}{d_S} > 1$ Non-synonymous mutations conferring evolutionary advantage

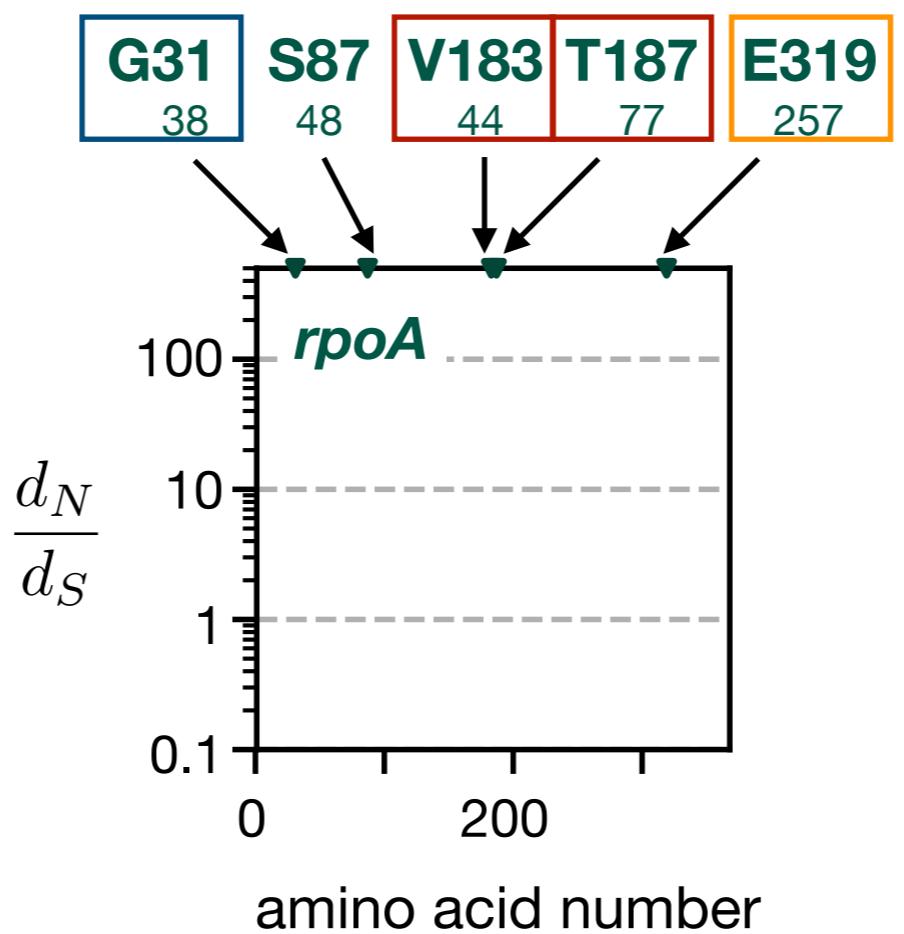
* adjusted for the ratio of non-syn to syn mutations available to that codon



G31 S87 V183 T187 E319

38 48 44 77 257



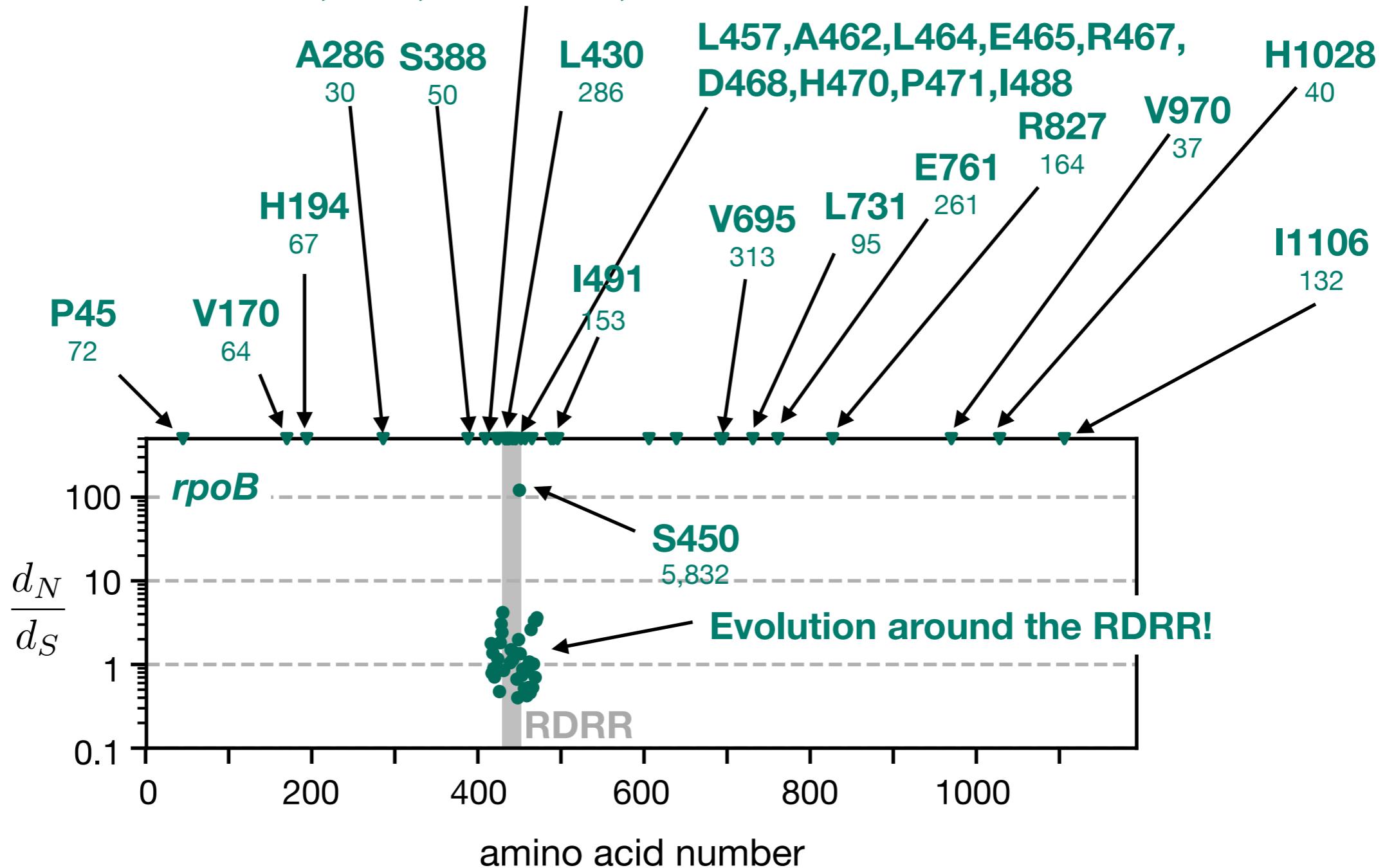


Comas et al. 2012

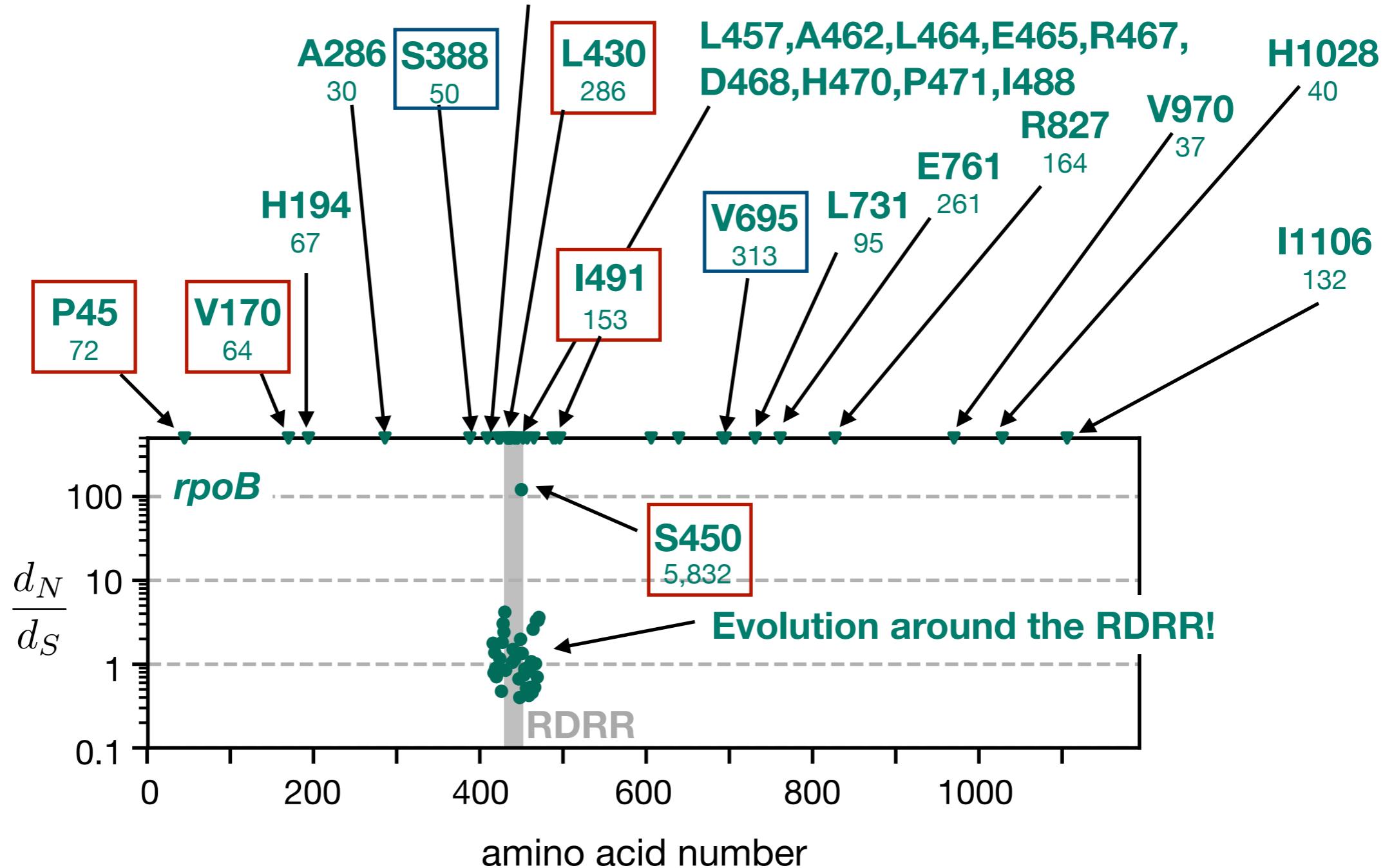
De Vos et al. 2013

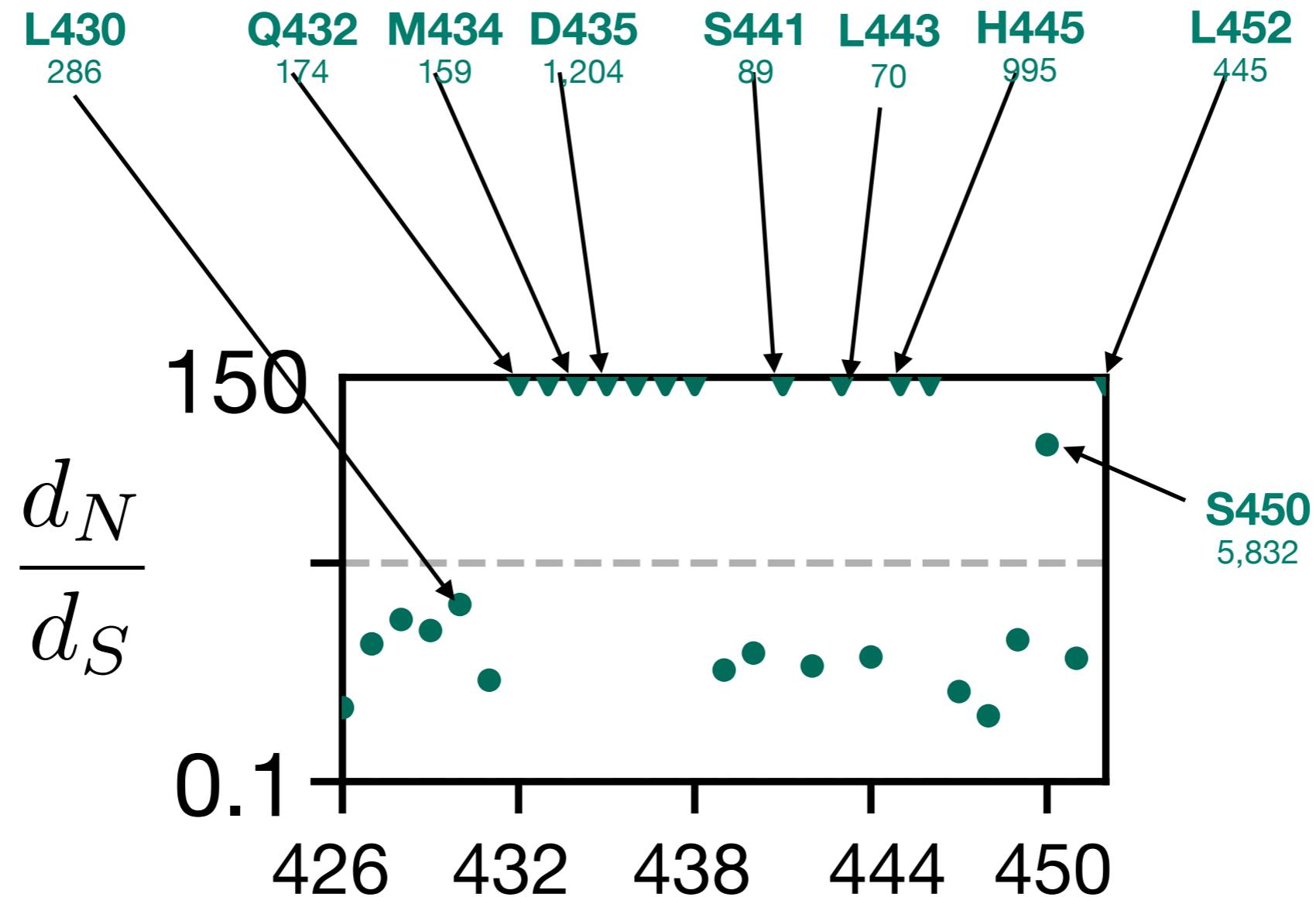
Li et al. 2016

P416,V418,K422-F425,T427-Q429

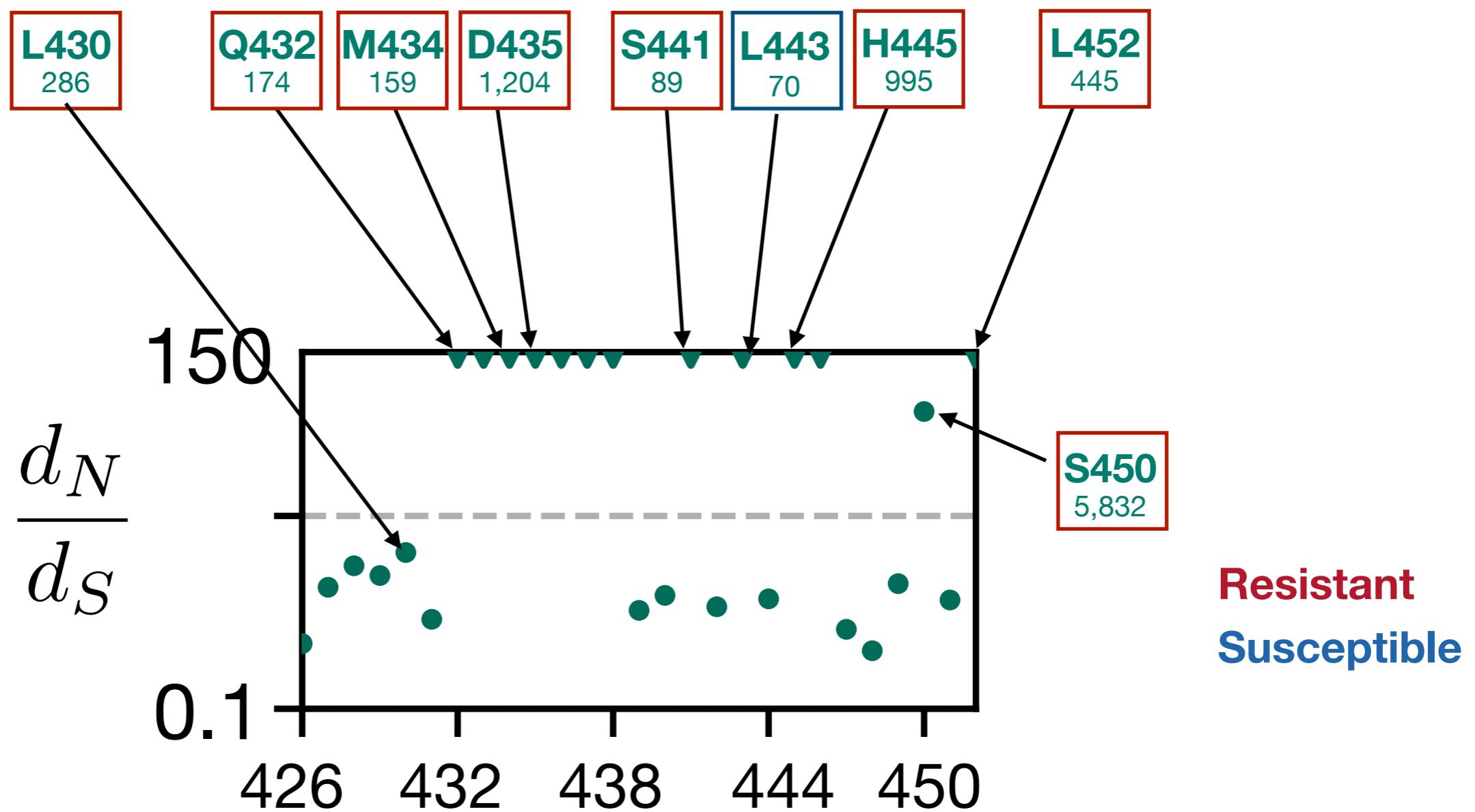


P416,V418,K422-F425,T427-Q429

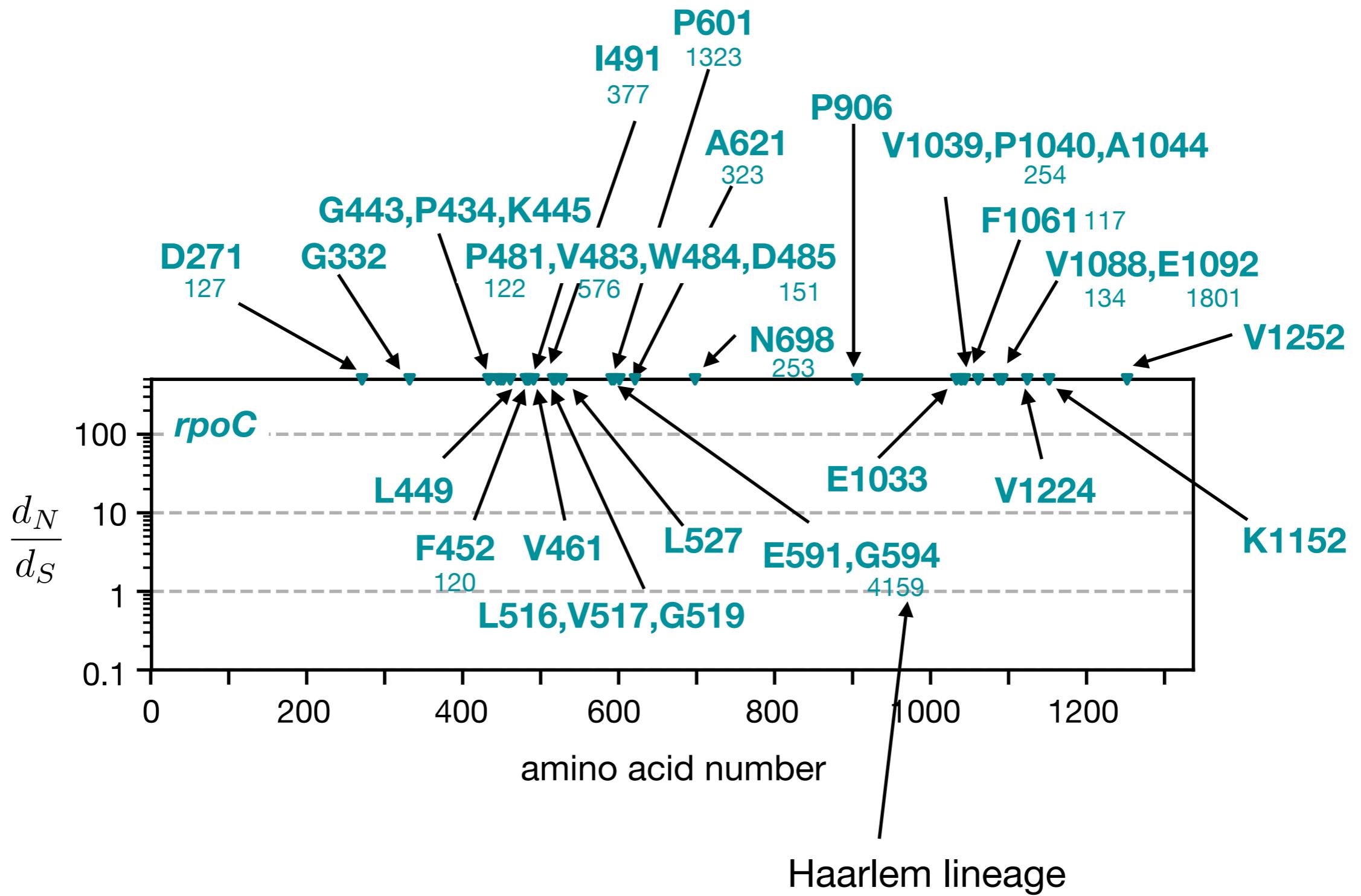




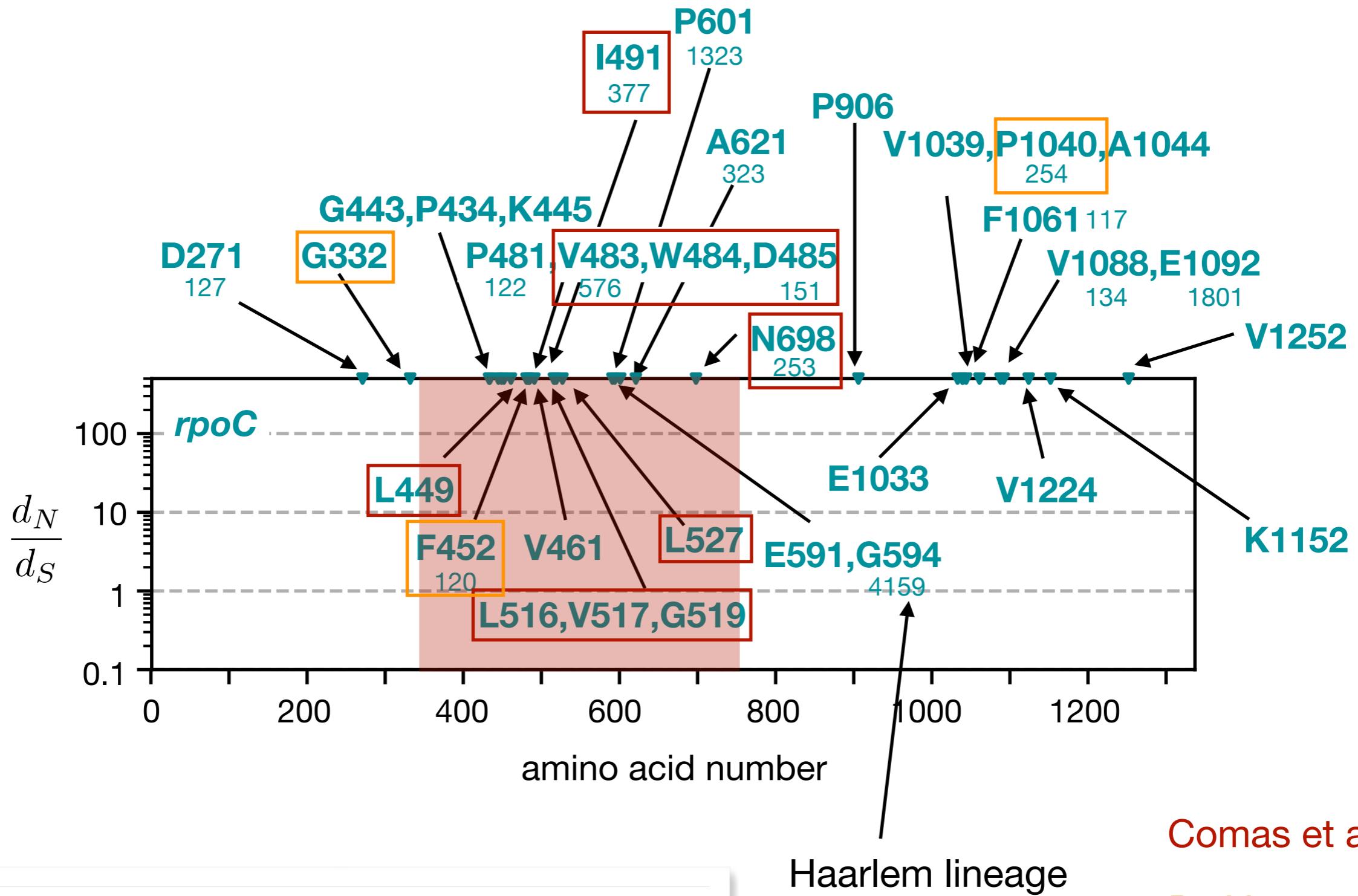
Points only labelled if in the current Walker/CRyPTIC genetic catalogue



Points only labelled if in the current Walker/CRyPTIC genetic catalogue



rpoC: 332,433,449,452,483,484,485,491,527,698,1040



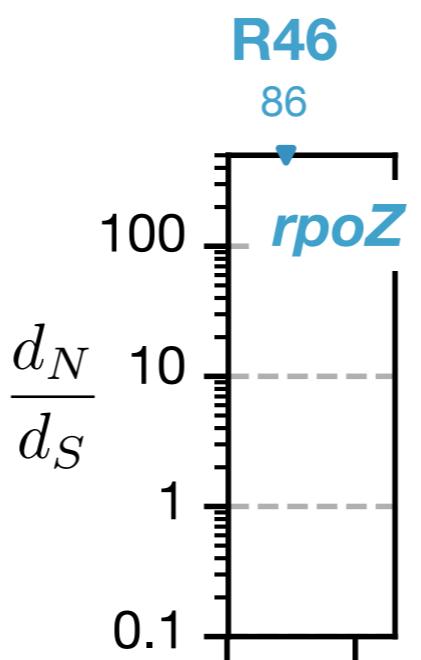
A novel multi SNP based method for identifying subspecies and associated lineages and sub-lineages of the *Mycobacterium tuberculosis* complex by whole genome sequencing

Authors

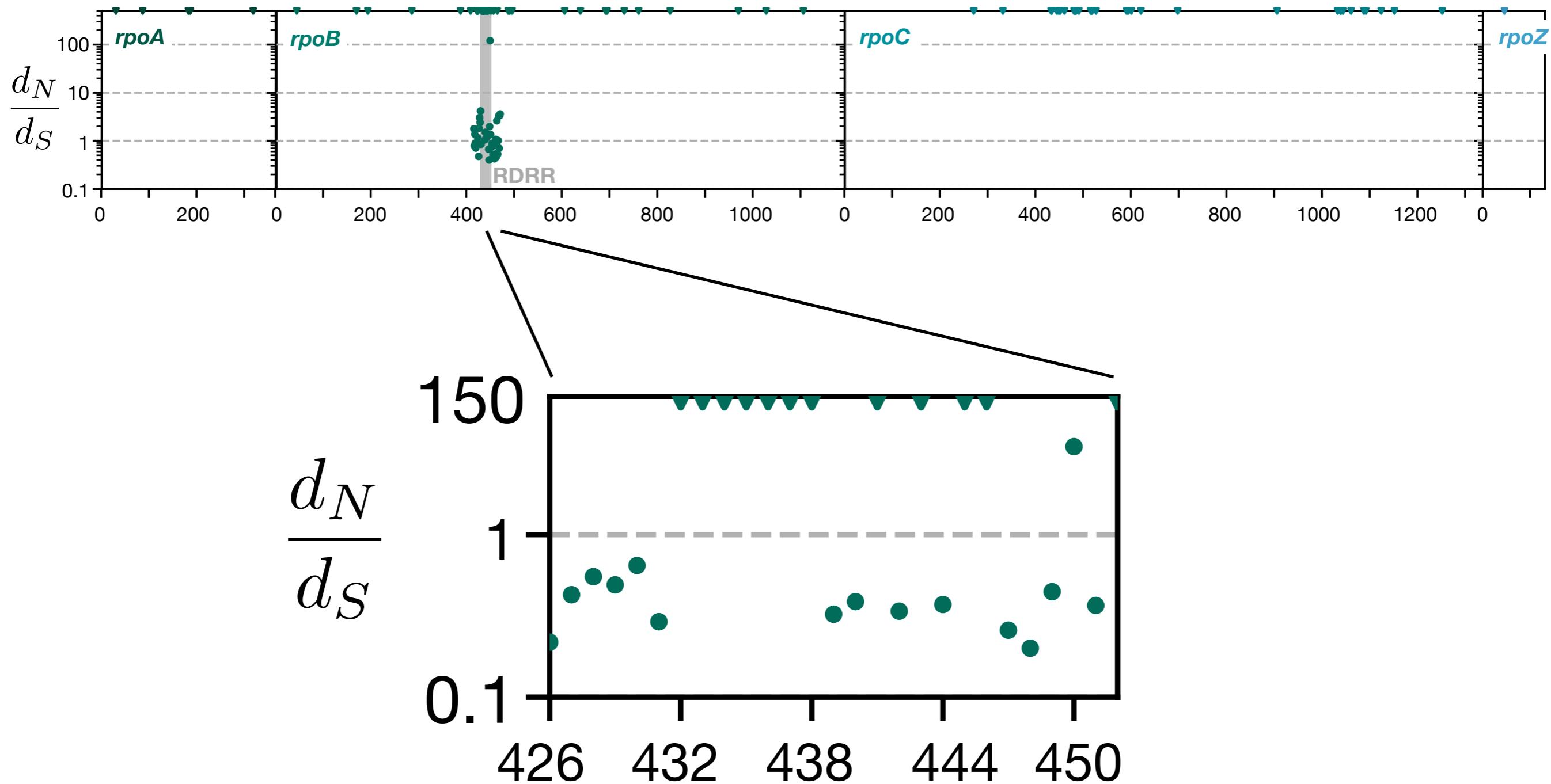
S Lipworth*, R Jajou*, H de Neeling, P Bradley, W van der Hoek, Z Iqbal, G Smith, T Peto, D Crook, T Walker*, D van Soolingen*

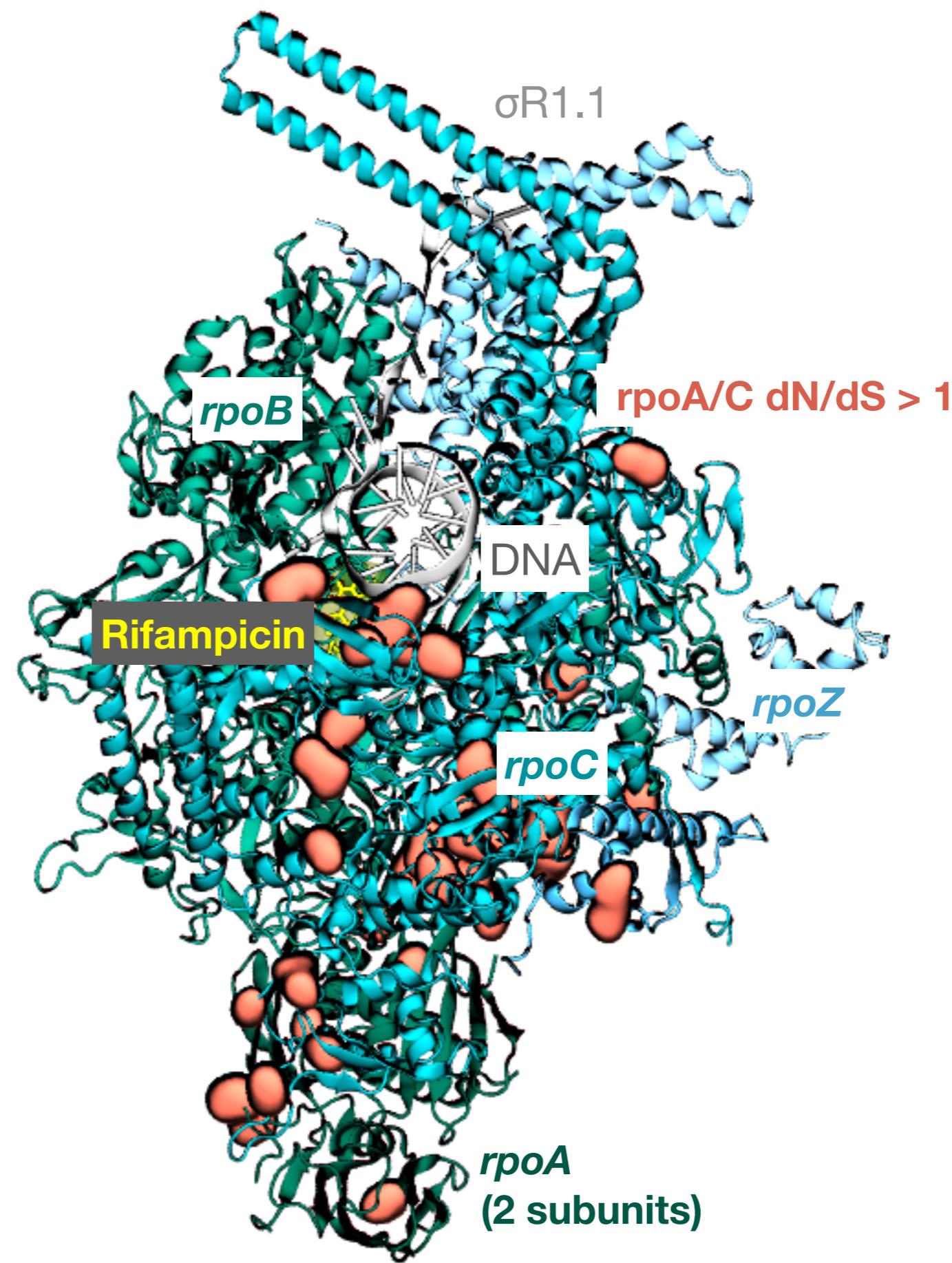
* denotes contributed equally

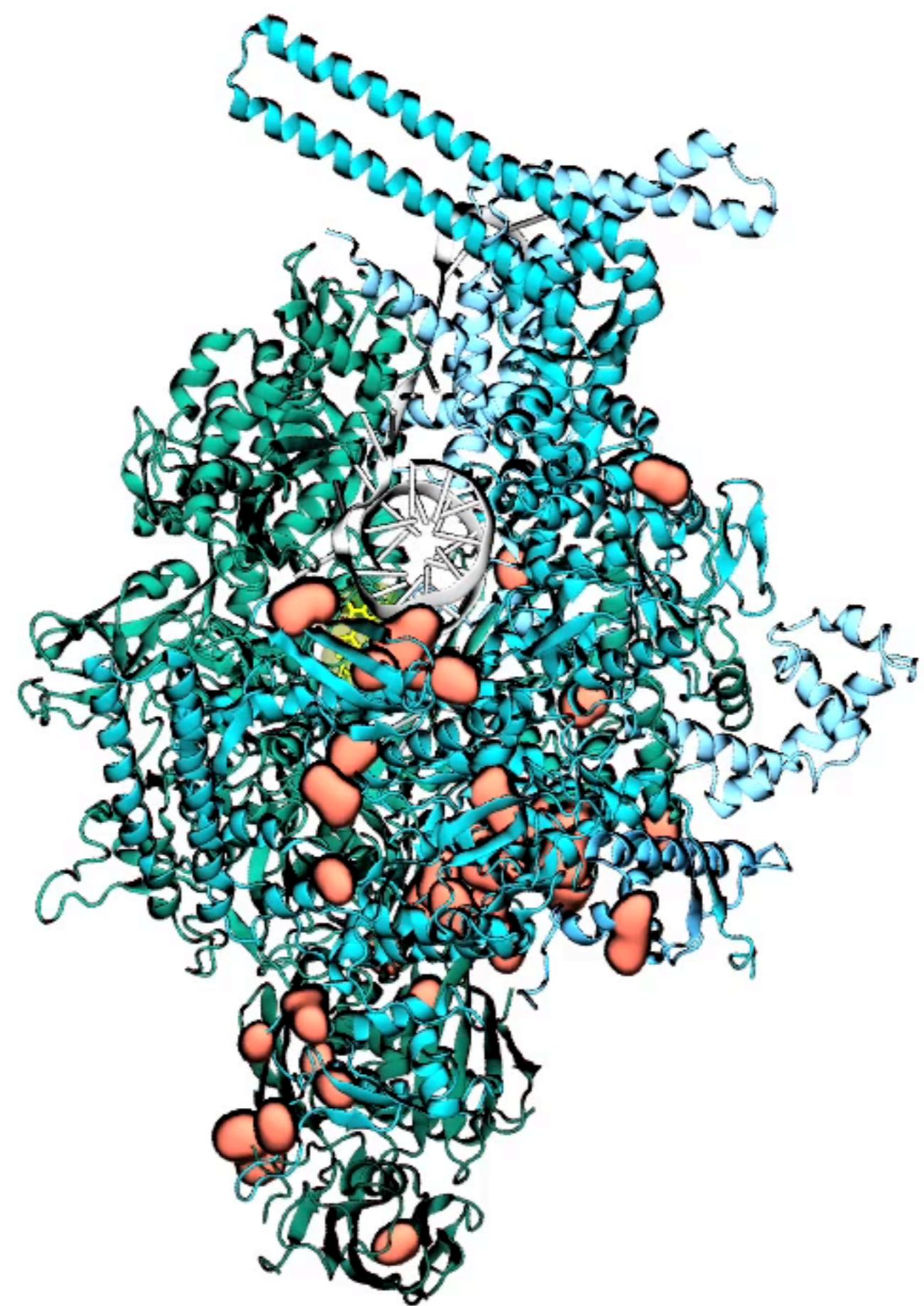
Abstract (limit 150 words)

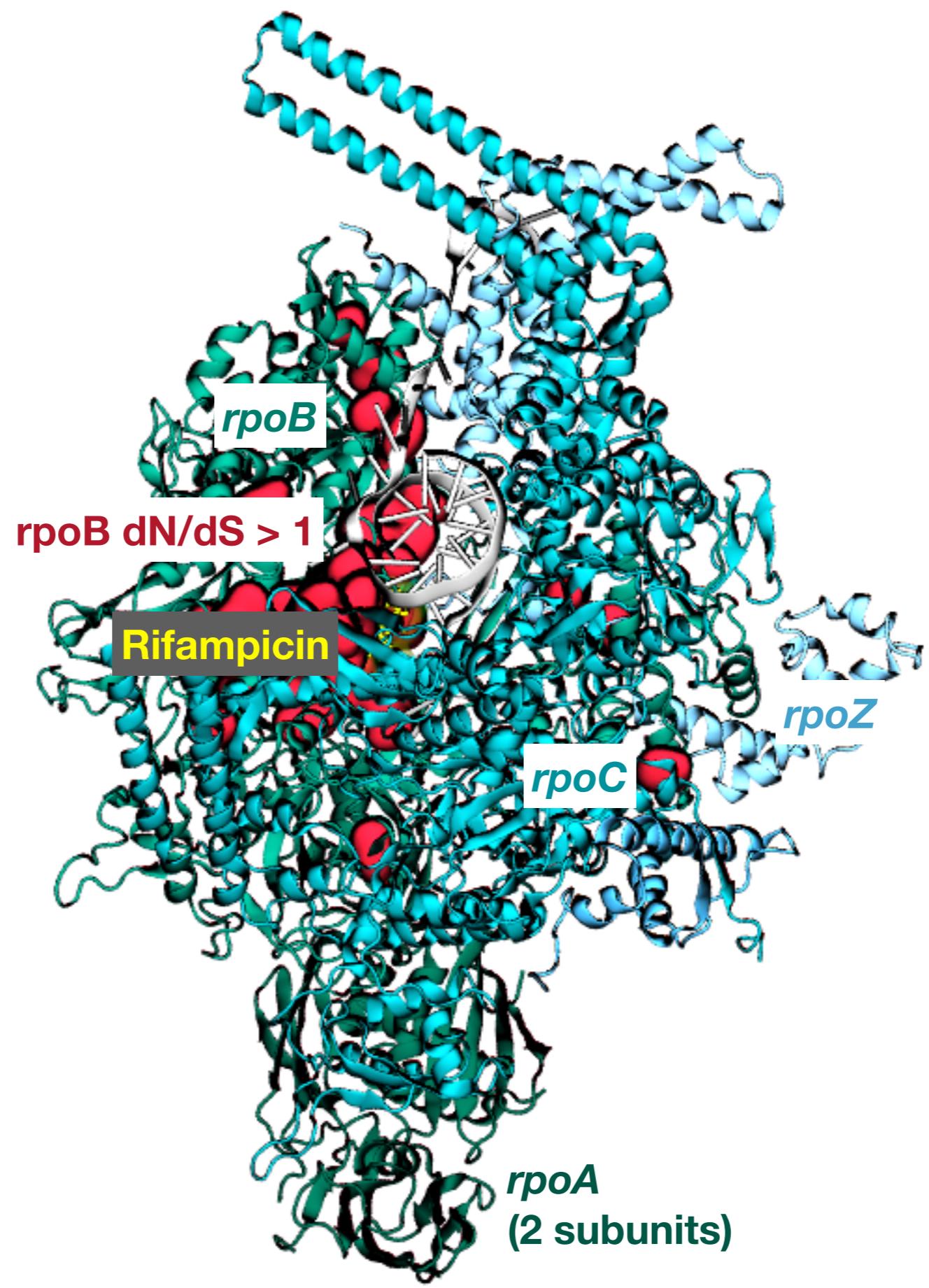


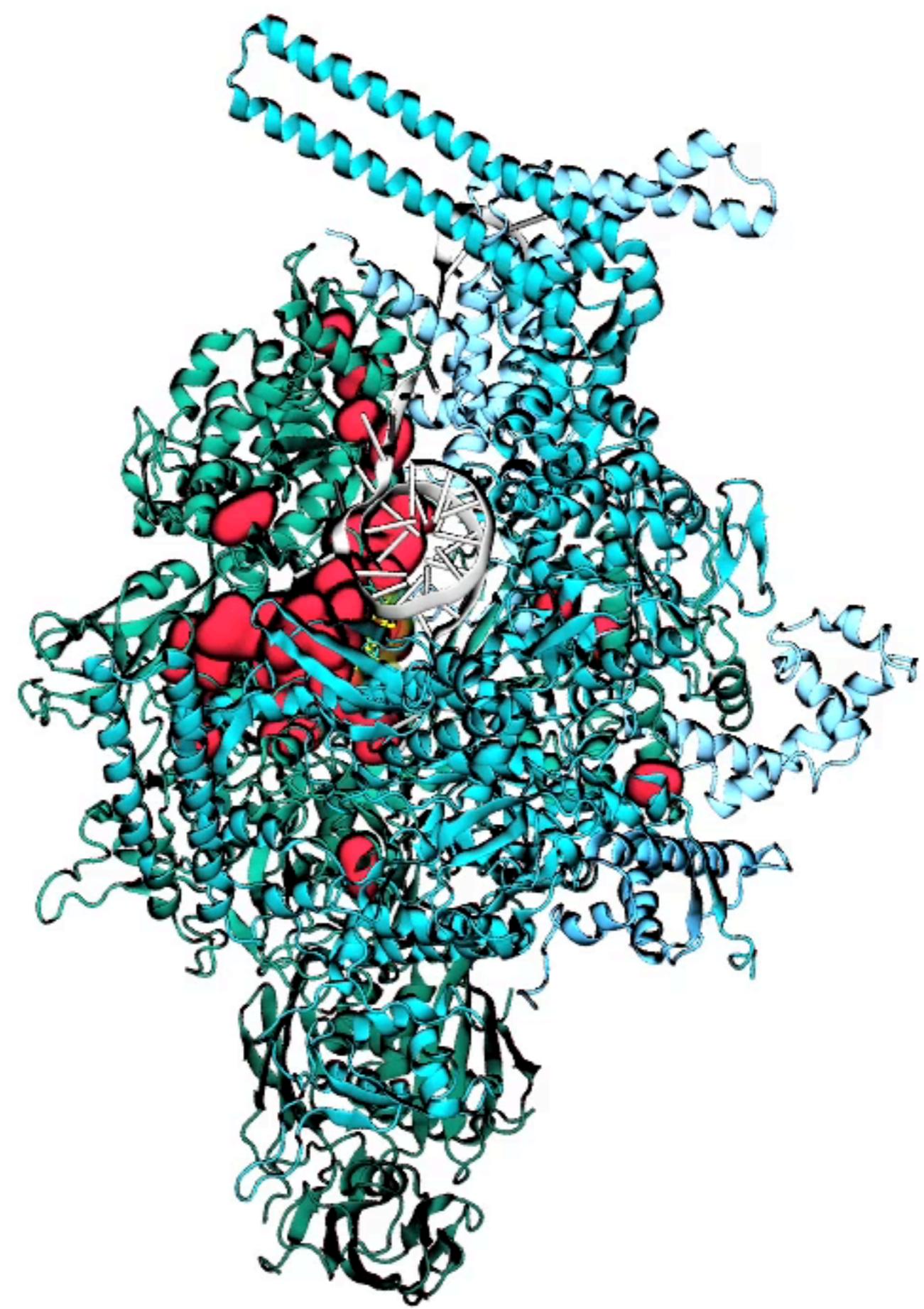
4. Ratio of non-synonymous to synonymous mutations (log scale)











Rifampicin binds to the RNA polymerase, preventing the extension of RNA

A large % of mutations that confer resistance occur in a 81bp region of *rpoB*

Proximal mutations *reduce* how well rifampicin binds to the RNAP

$\frac{d_N}{d_S}$ Identifies putative compensatory mutations and mutations that confer resistance



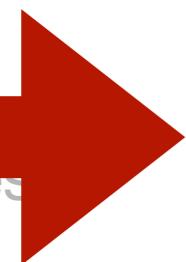
1. European Nucleotide Archive



Comas et al. 2012

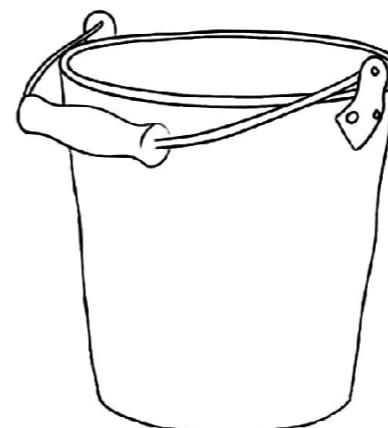


500
genomes



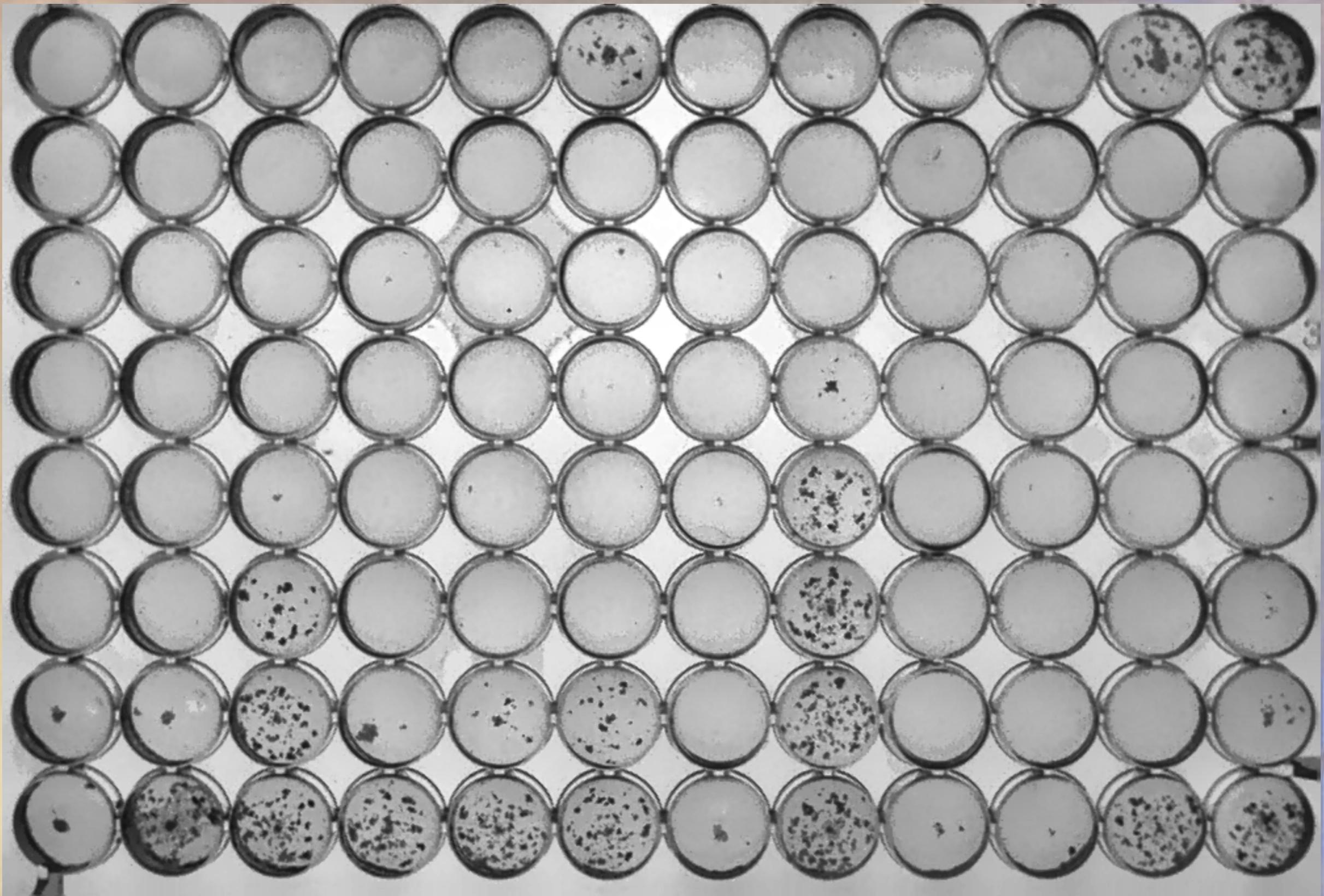
Can we measure the effect on the growth of *M. tuberculosis* of these compensatory mutations....?

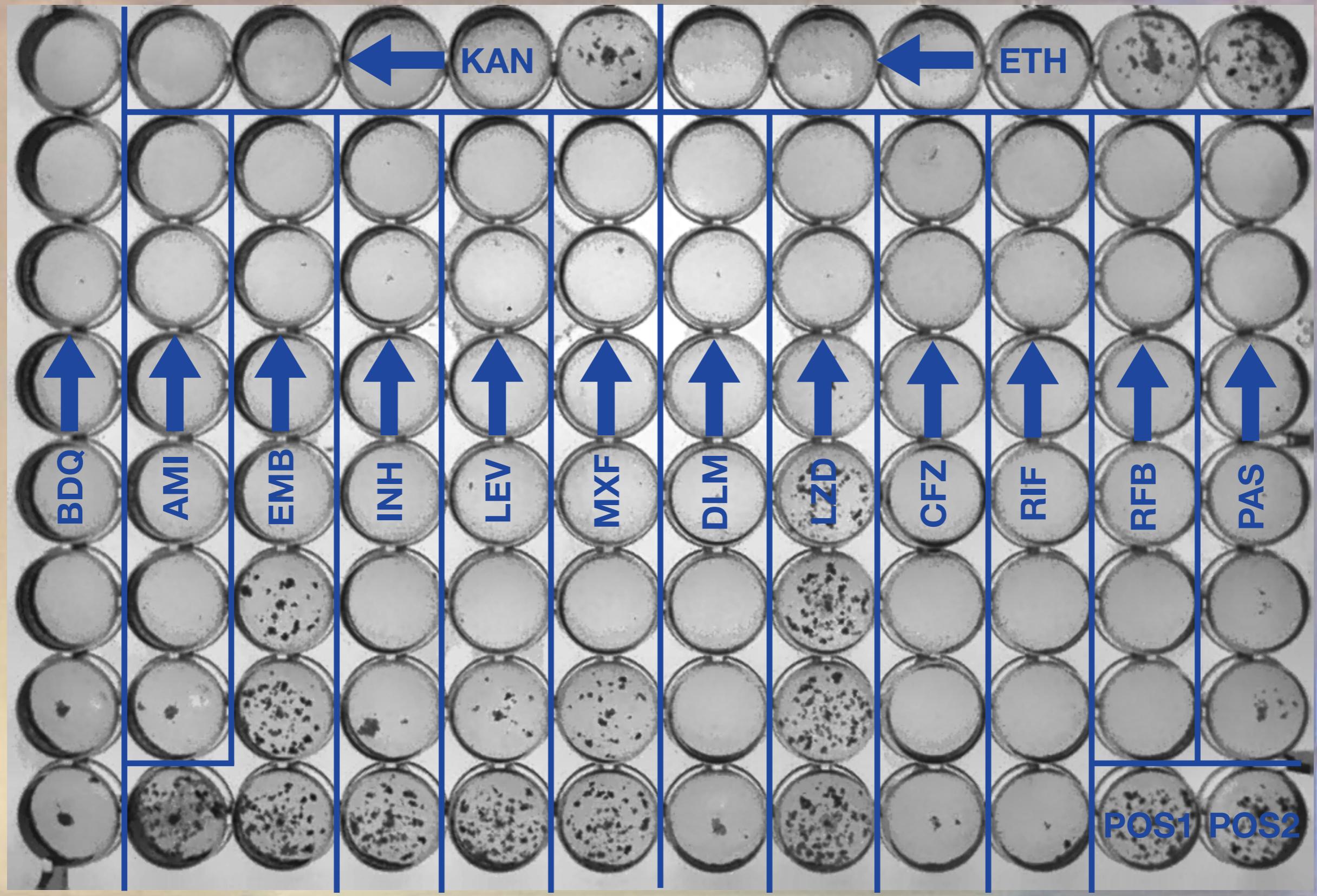
2. CRyPTIC

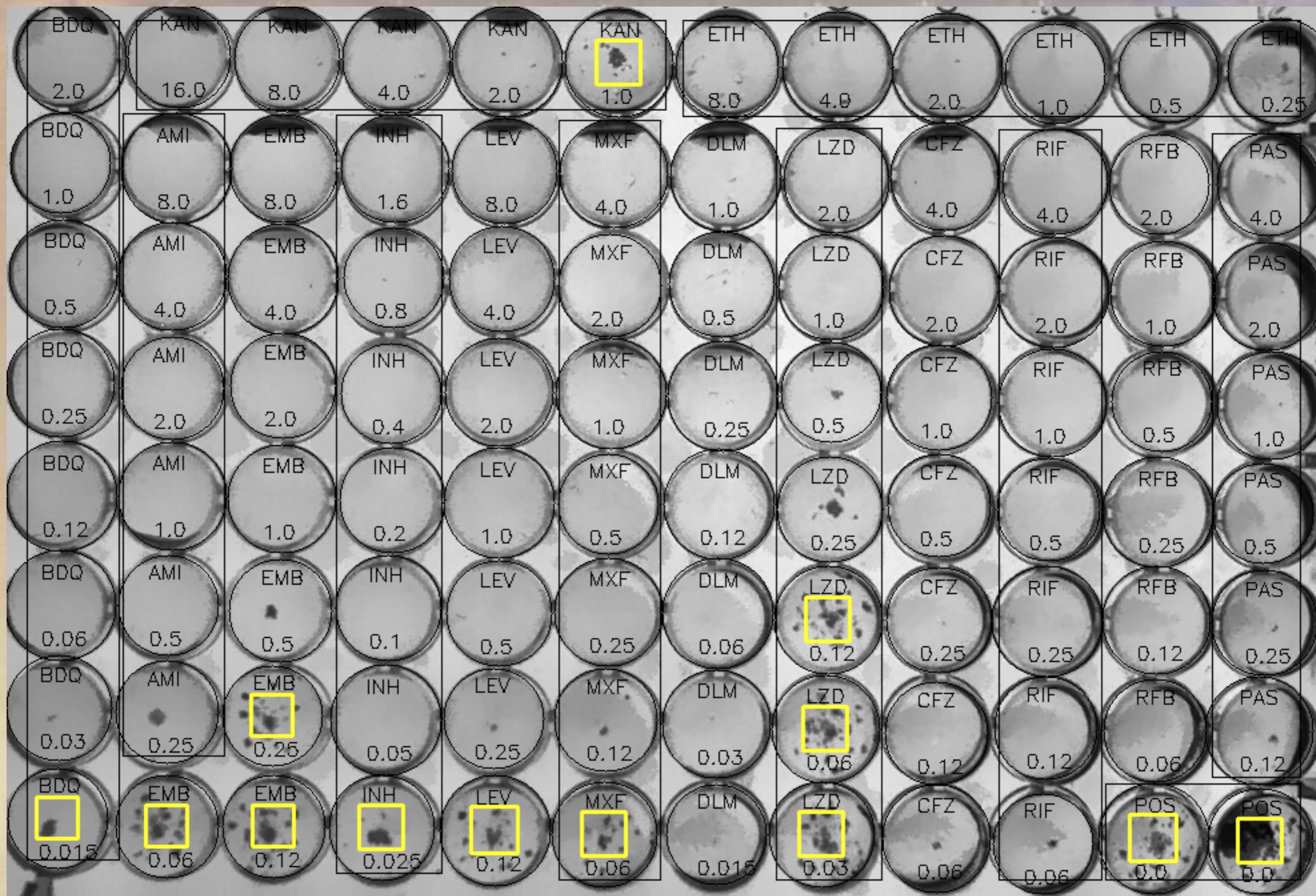


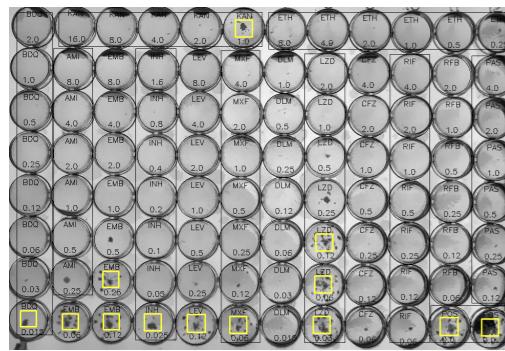
4,500 genomes
(end May 2018)



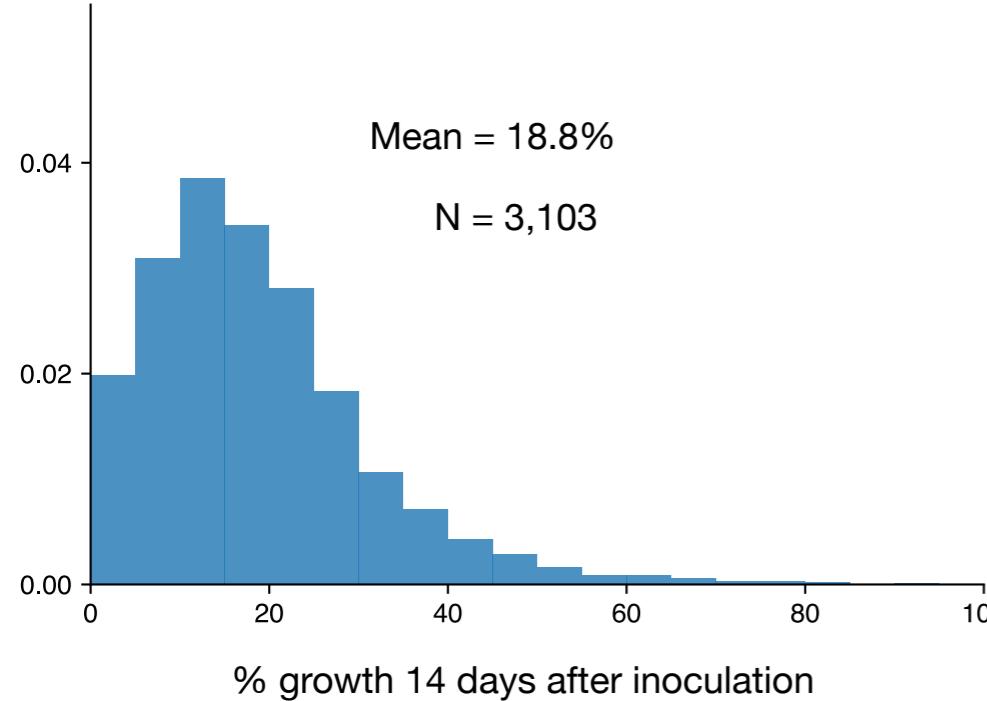




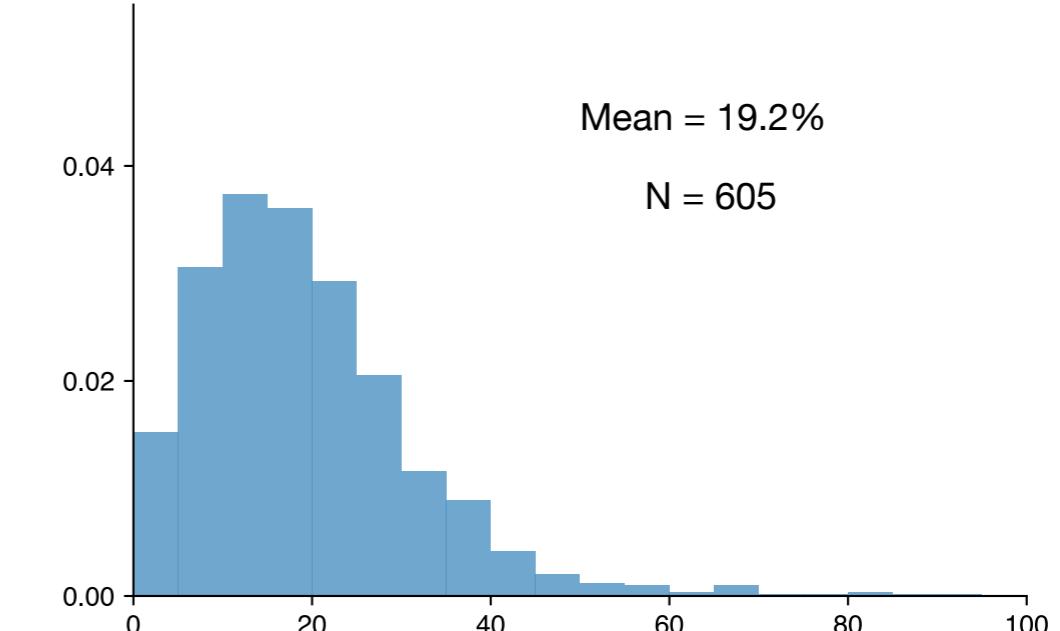




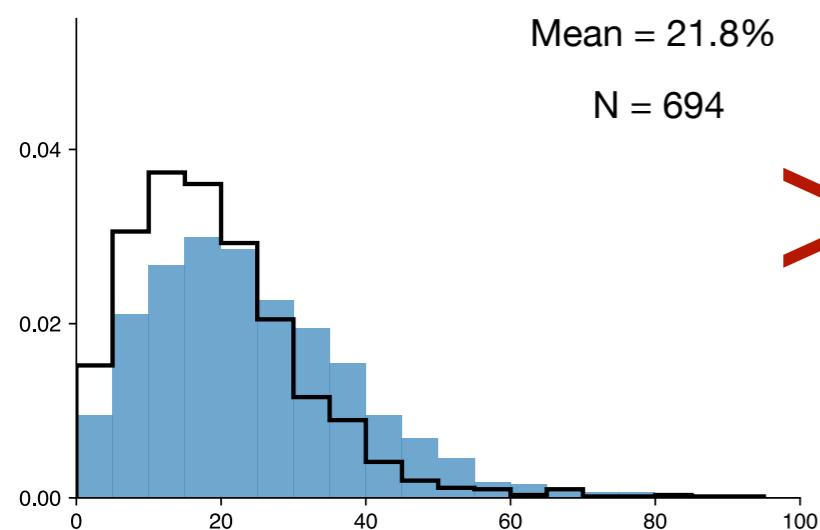
All good plates, whether genome exists or not



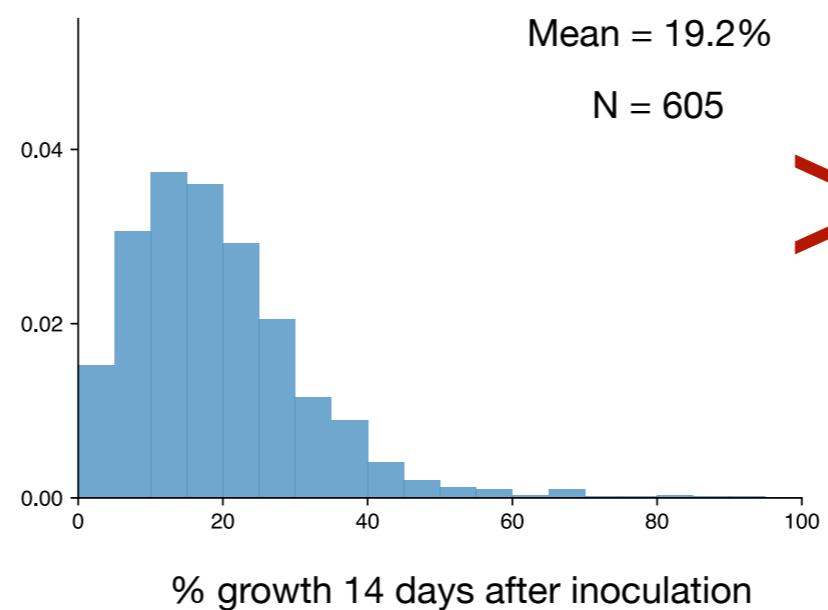
Plates predicted to be SSSS



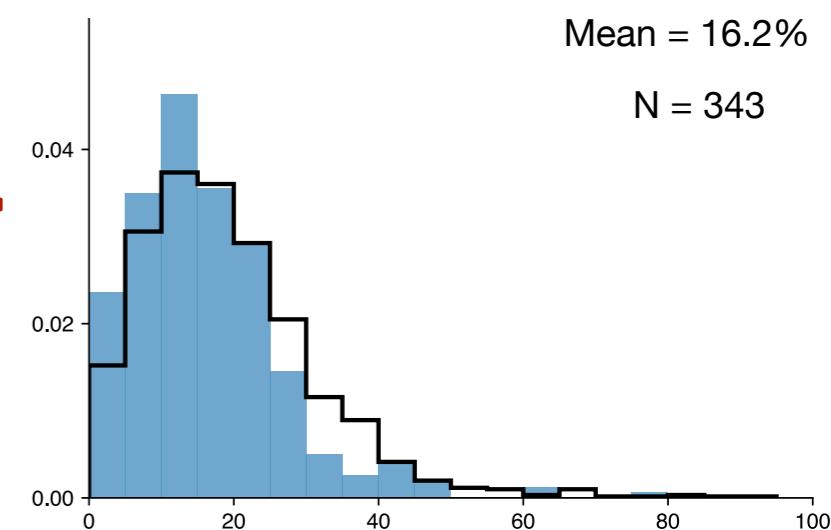
*rpoB_S450L &
rpoC* compensatory mutation



Plates predicted to be SSSS

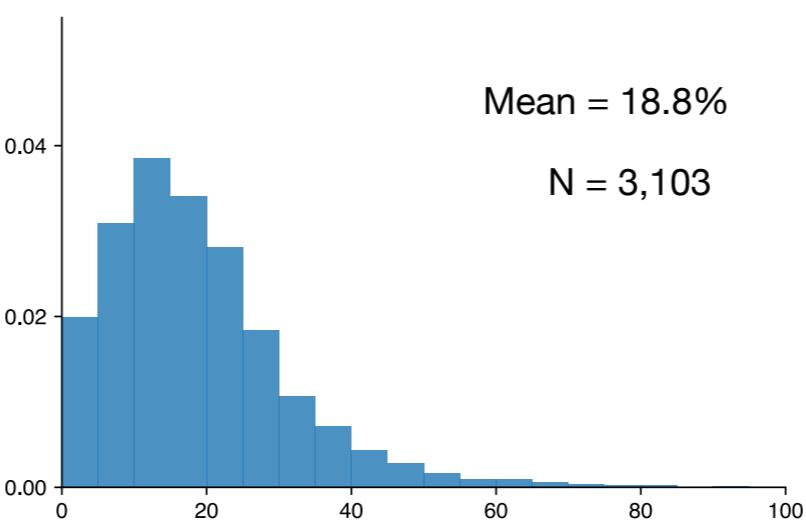


rpoB_S450L but
no compensatory mutation



¶

All good plates, whether genome exists or not



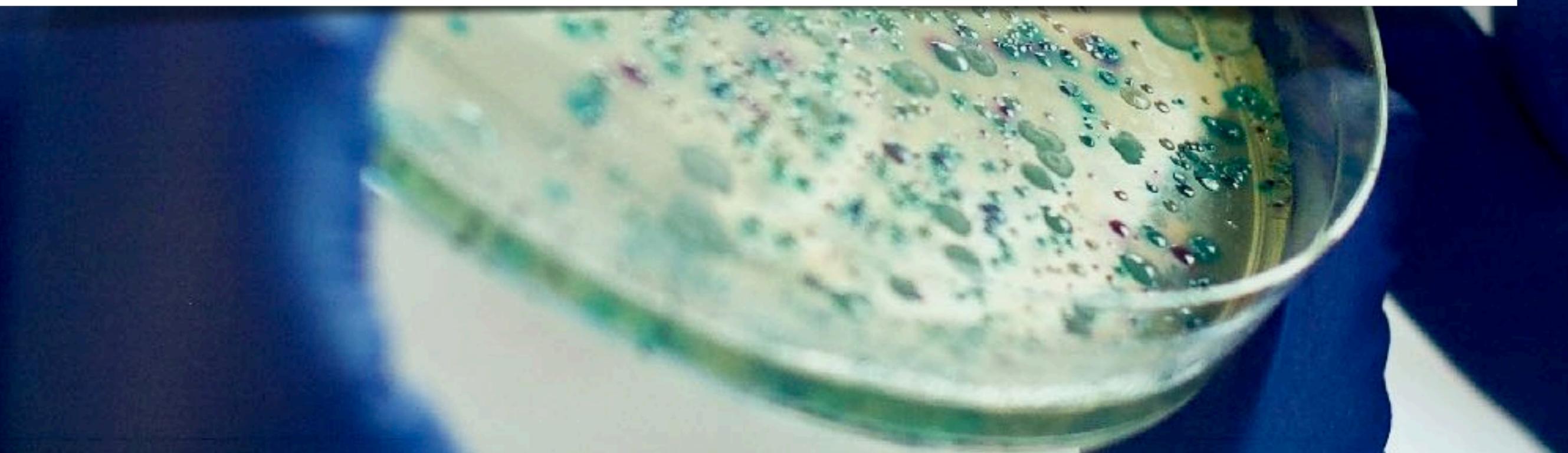
Rifampicin binds to the RNA polymerase, preventing the extension of RNA

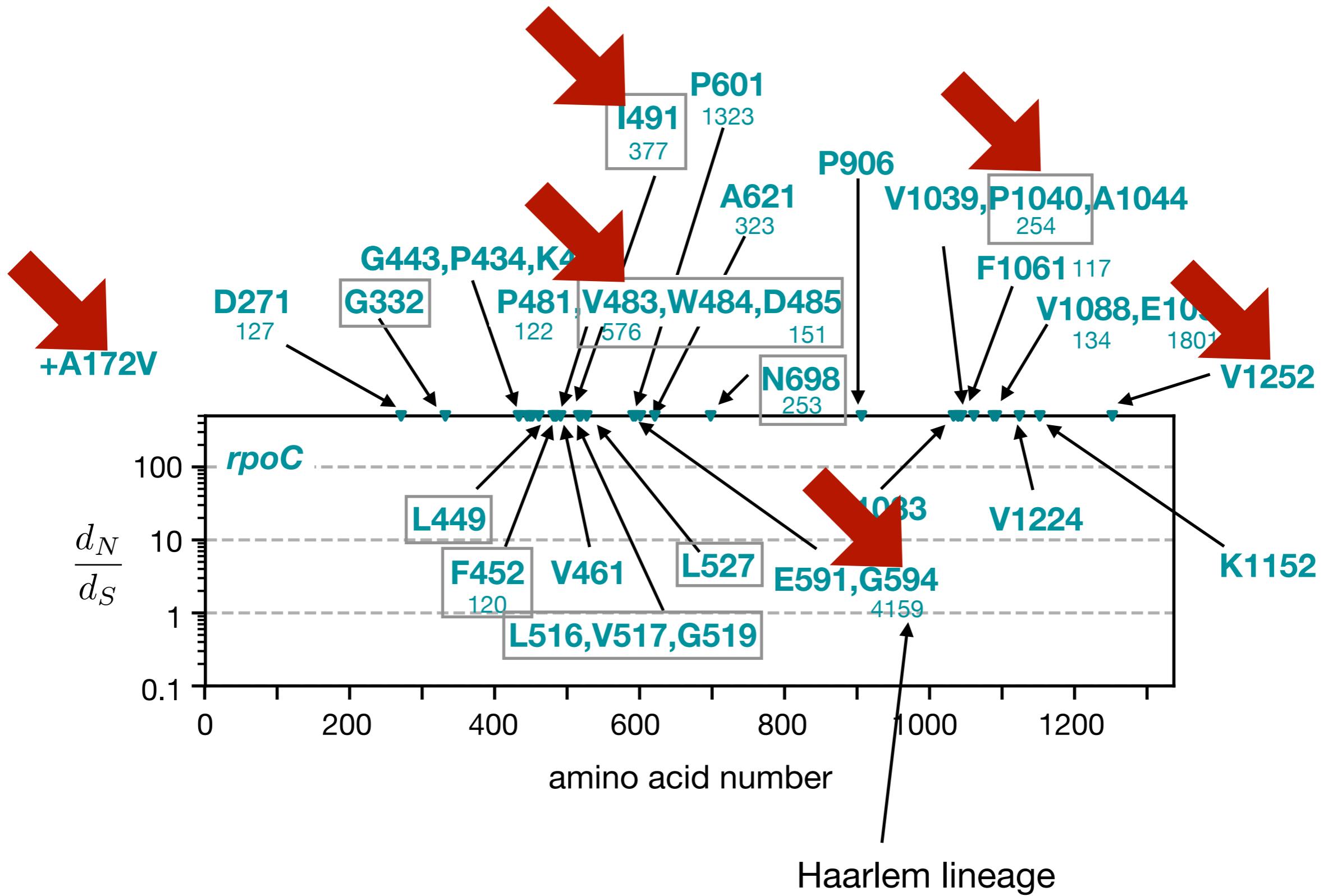
A large % of mutations that confer resistance occur in a 81bp region of *rpoB*

Proximal mutations *reduce* how well rifampicin binds to the RNAP

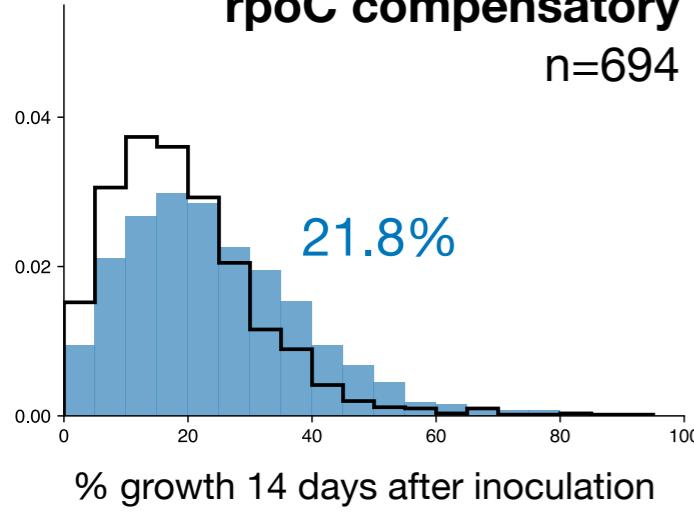
$\frac{d_N}{d_S}$ Identifies putative compensatory mutations and mutations that confer resistance

Strains with *rpoB_S450L* but no compensatory mutation grow more slowly than strains with a compensatory mutation

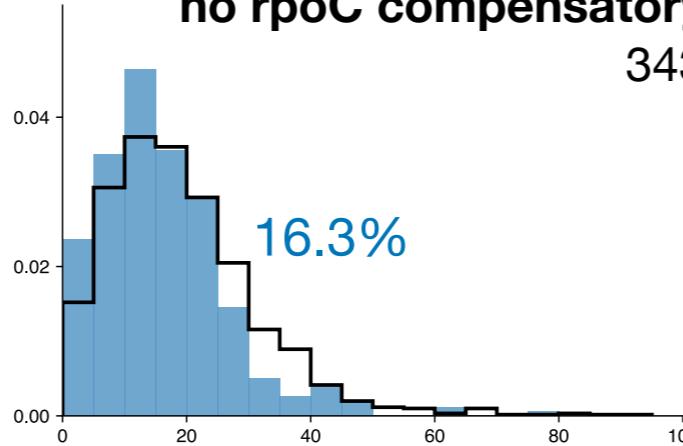




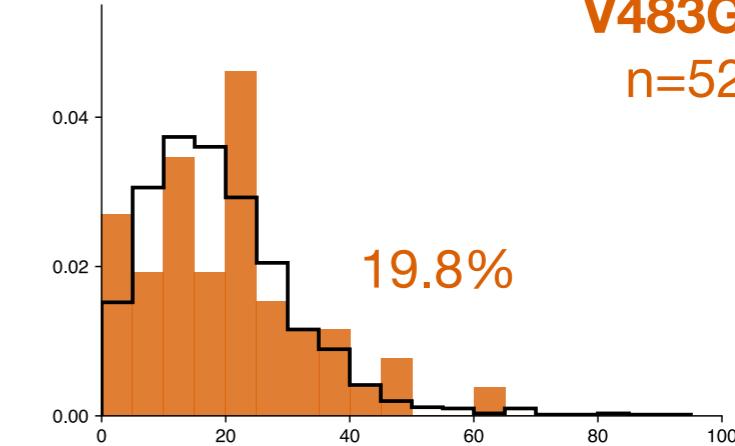
rpoB_S450L
rpoC compensatory
n=694



rpoB_S450L
no rpoC compensatory
343

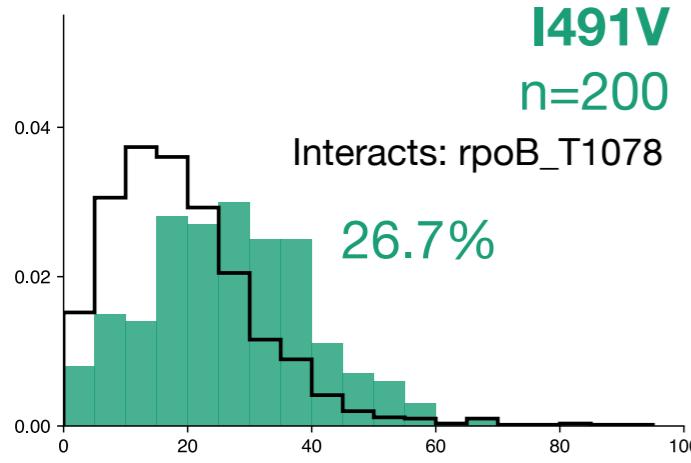


V483G
n=52



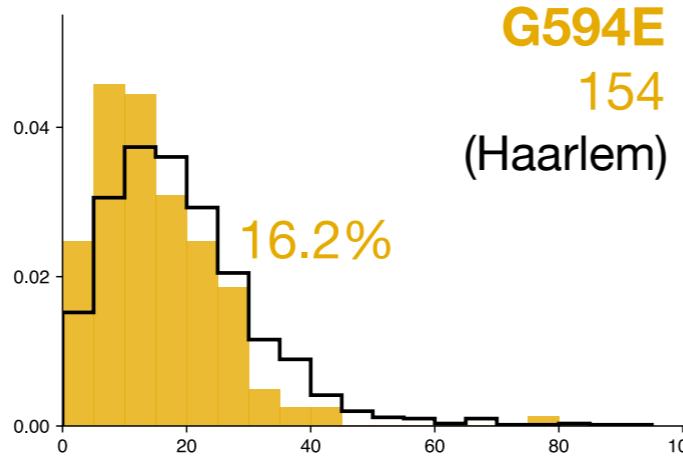
I491V
n=200

Interacts: rpoB_T1078
26.7%



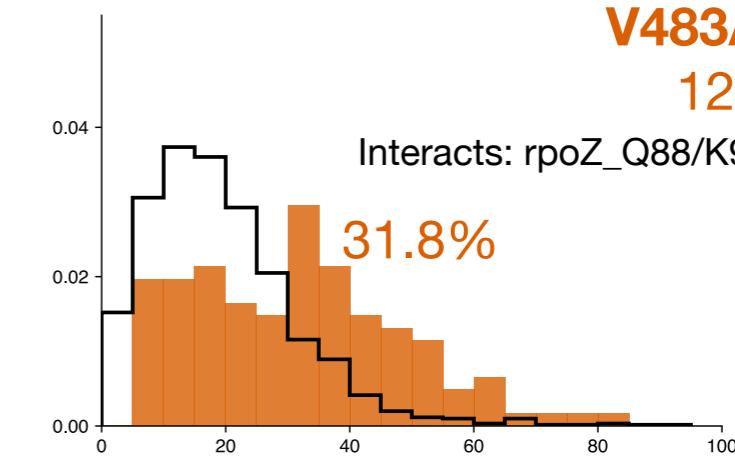
G594E
154
(Haarlem)

16.2%



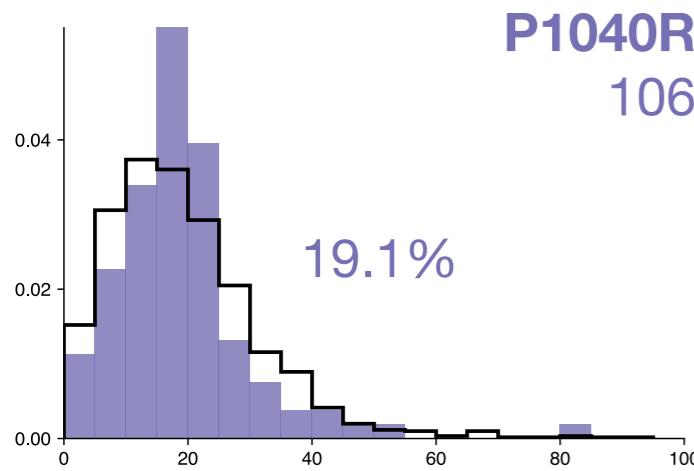
V483A
122

Interacts: rpoZ_Q88/K90
31.8%



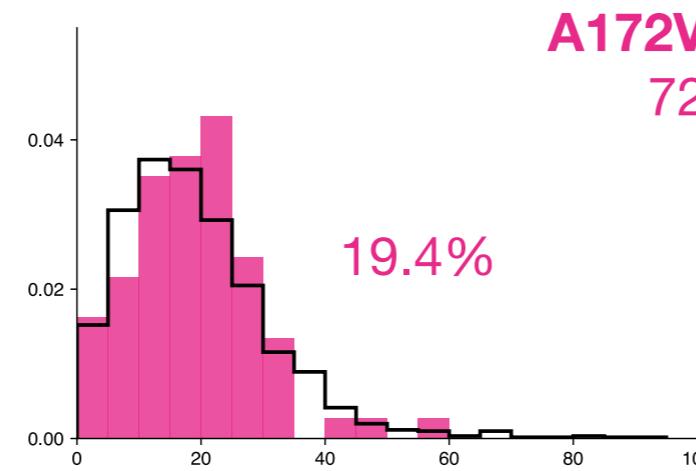
P1040R
106

19.1%



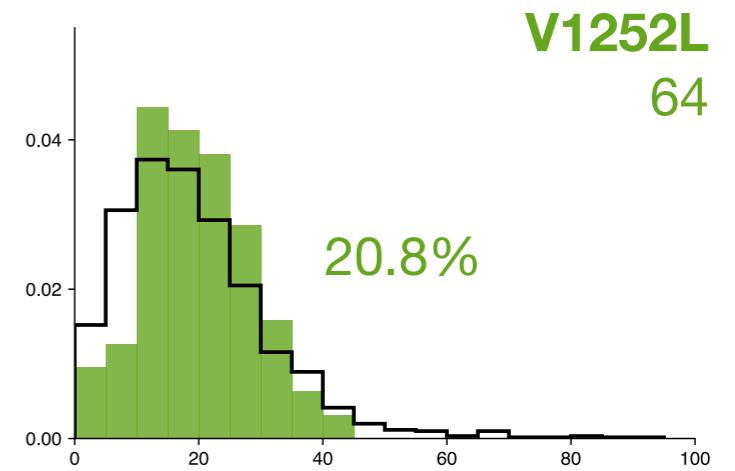
A172V
72

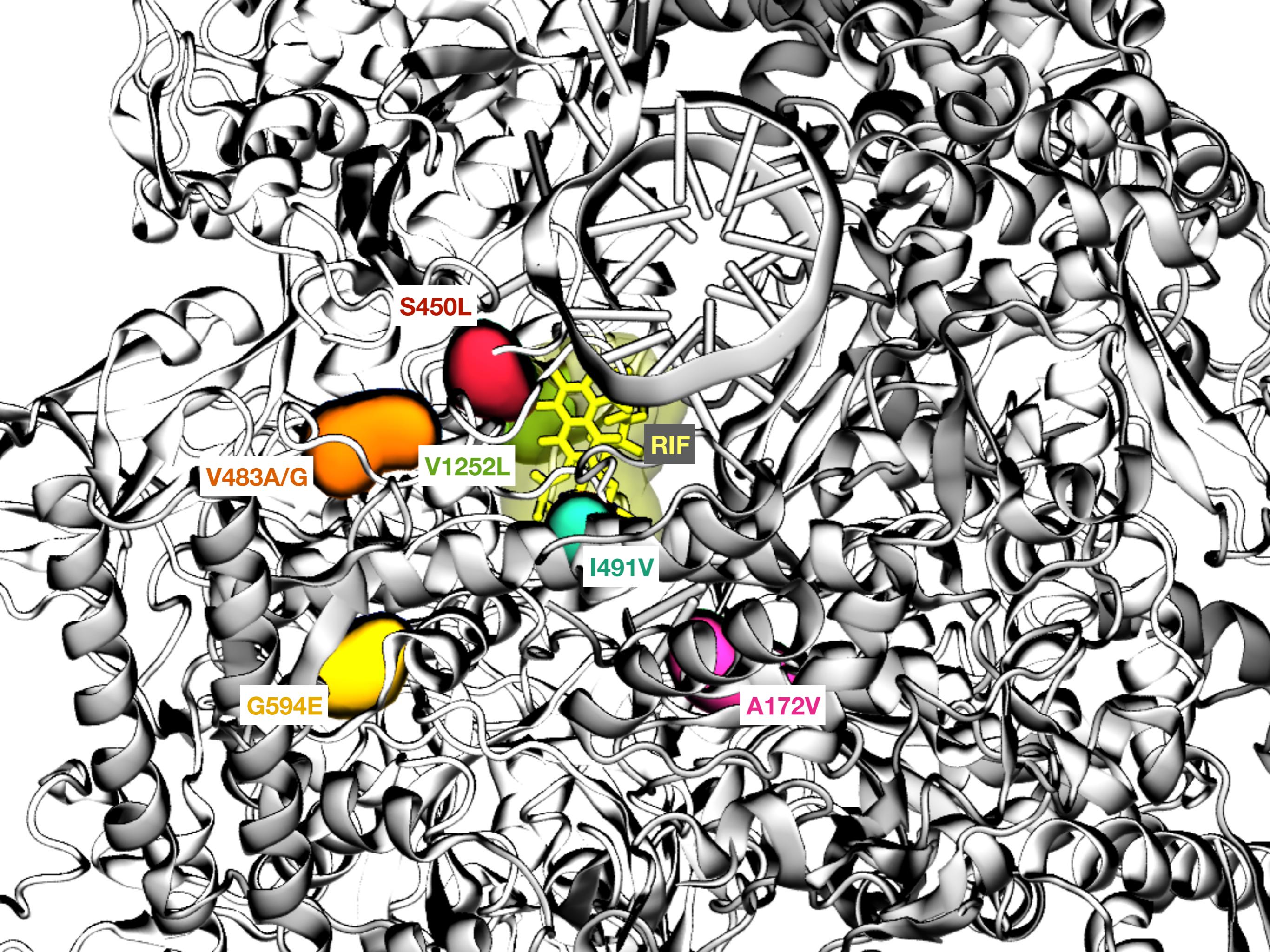
19.4%

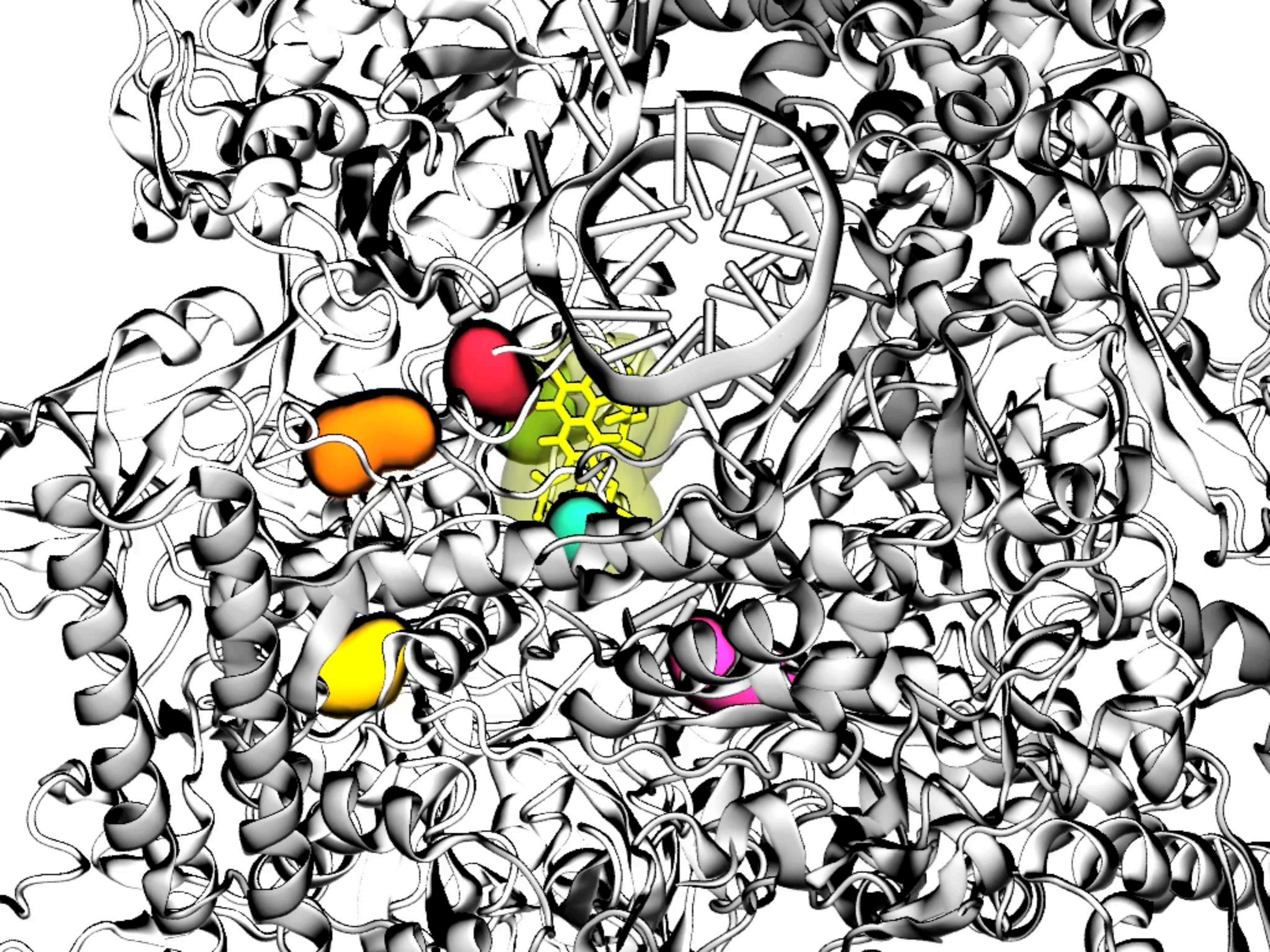


V1252L
64

20.8%







Rifampicin binds to the RNA polymerase, preventing the extension of RNA

A large % of mutations that confer resistance occur in a 81bp region of *rpoB*

Proximal mutations *reduce* how well rifampicin binds to the RNAP

$\frac{d_N}{d_S}$ Identifies putative compensatory mutations and mutations that confer resistance

Strains with *rpoB_S450L* but no compensatory mutation grow more slowly than strains with a compensatory mutation

Some compensatory mutations even appear to *enhance* growth

1. European Nucleotide Archive



Comas et al. 2012

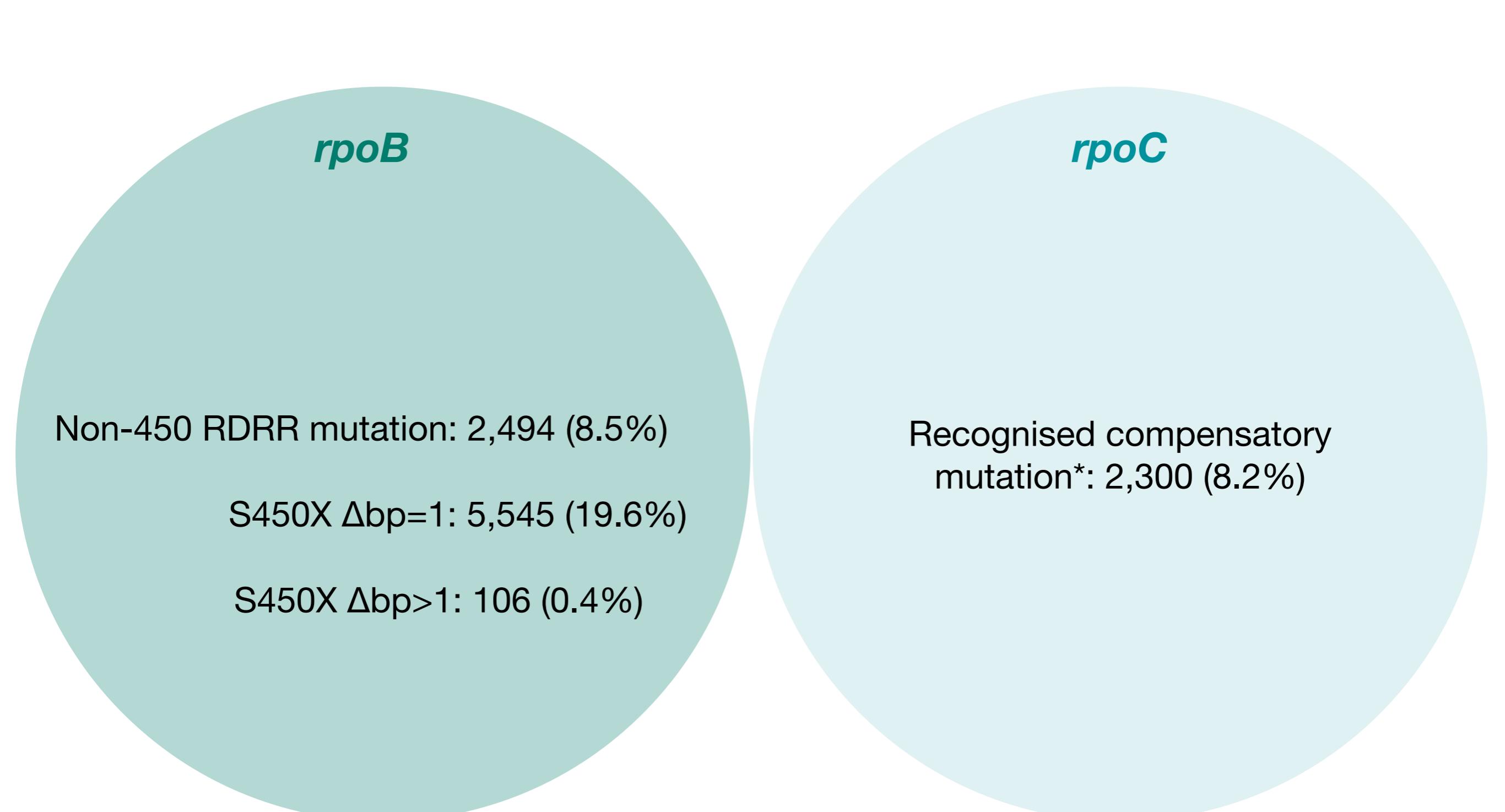


500
genomes

2. CRyPTIC



4,500 genomes
(end May 2018)



rpoB

Non-450 RDRR mutation: 2,494 (8.5%)

S450X Δ bp=1: 5,545 (19.6%)

S450X Δ bp>1: 106 (0.4%)

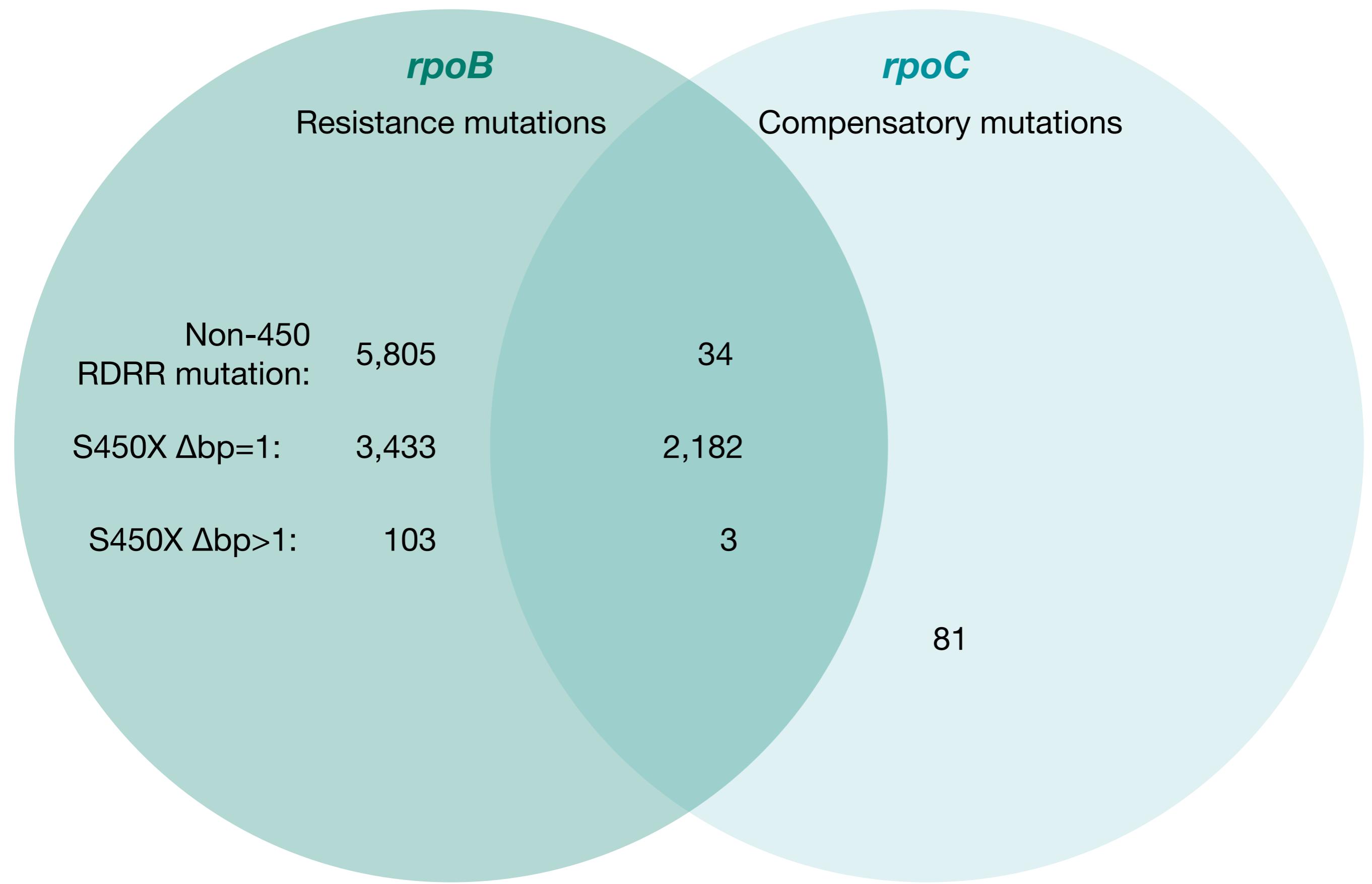
rpoC

Recognised compensatory
mutation*: 2,300 (8.2%)

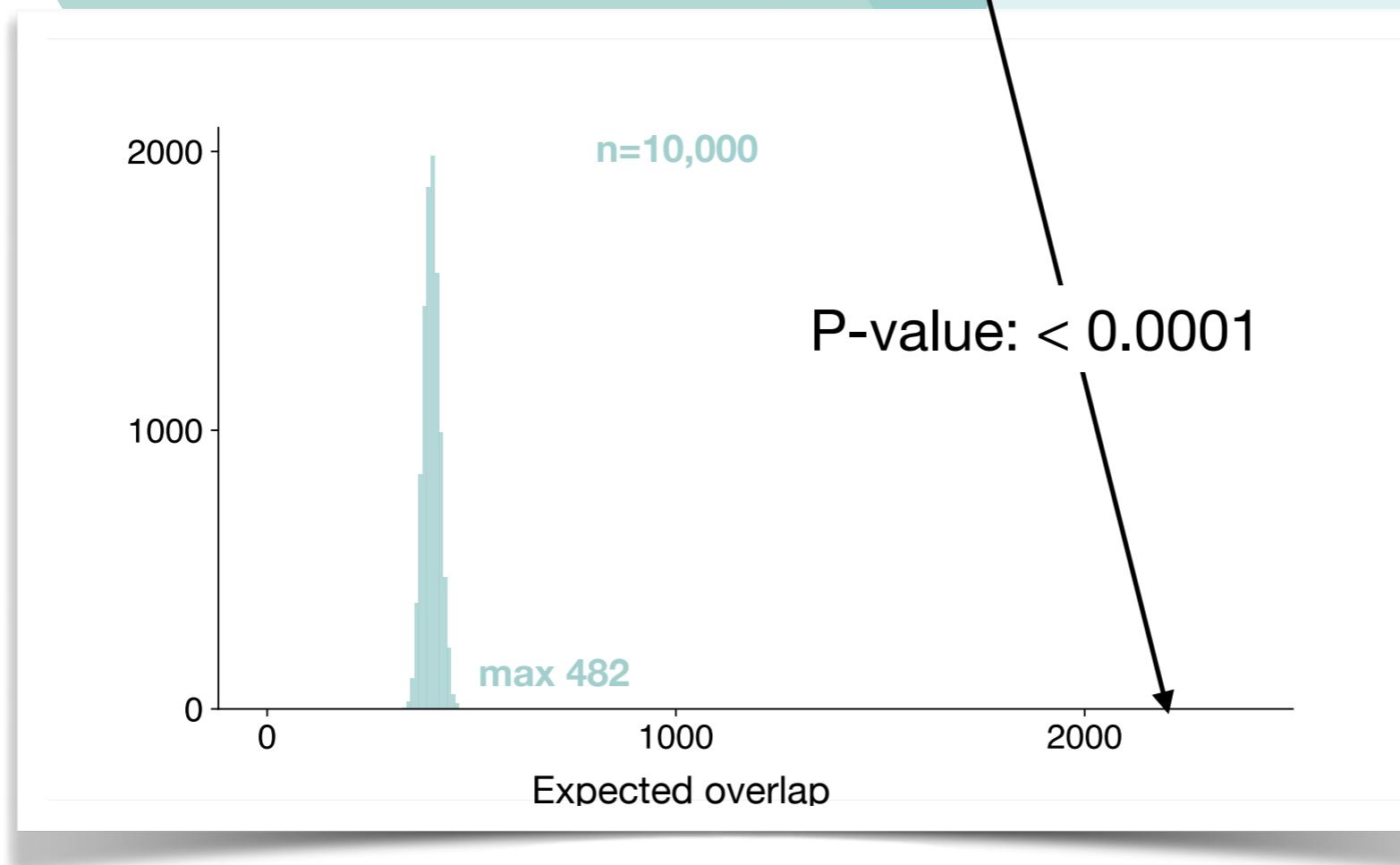
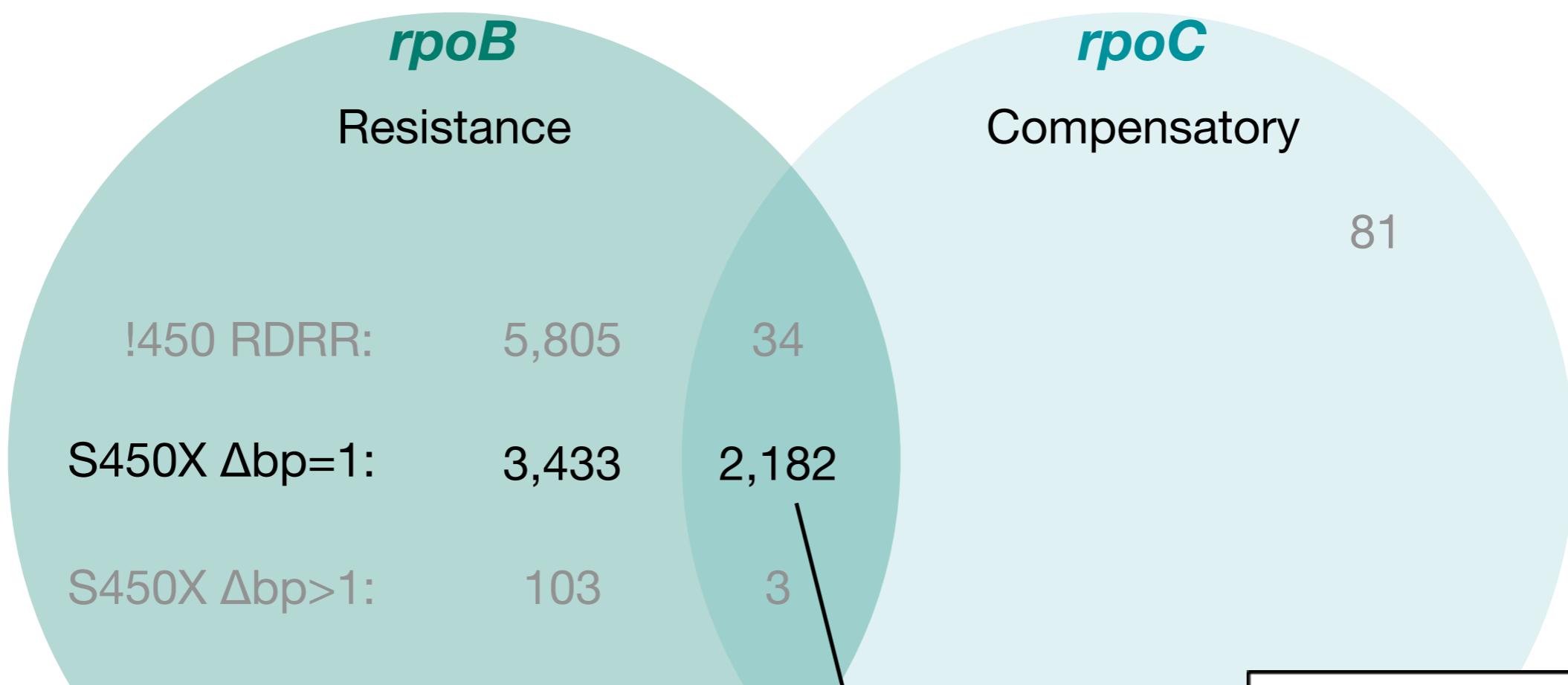
* seen in literature and dN/dS>1

rpoC: 332,433,434,445,449,452,483,484,485,491,516,517,519,527,698,1040

~28,000 genomes in the ENA

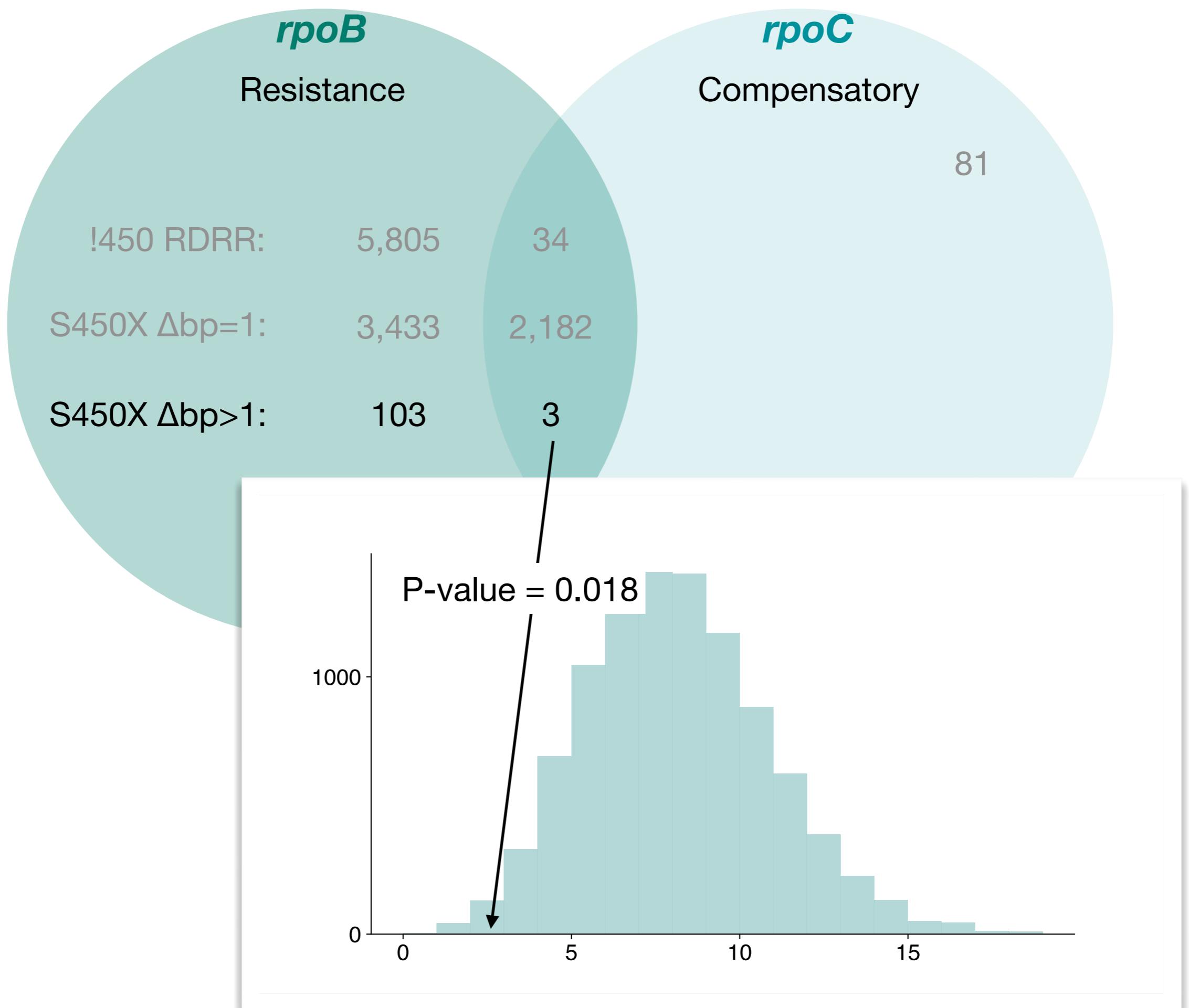


Hypothesis: Resistance conferring mutations and compensatory mutations are independent

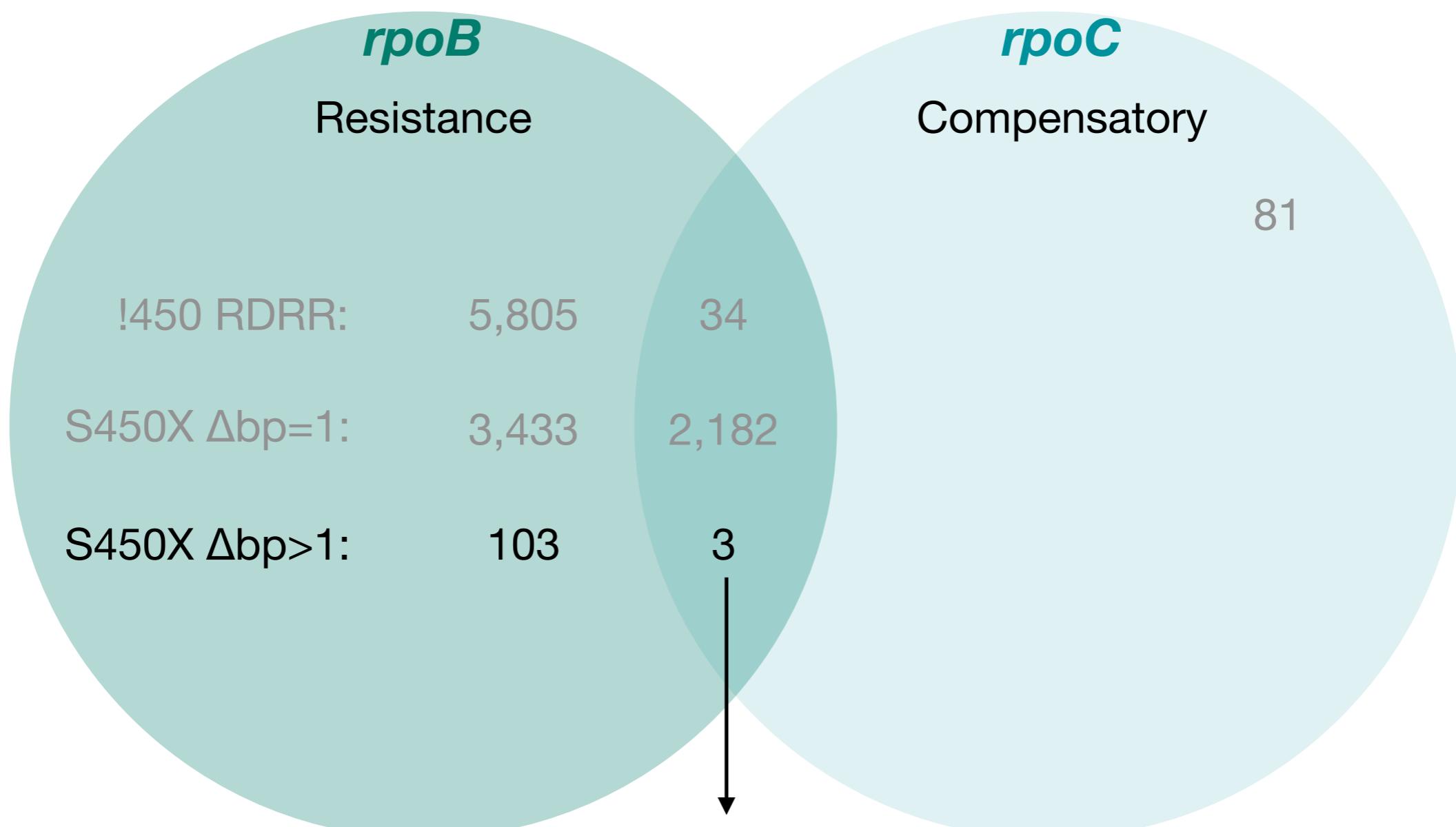


P-VALUE	INTERPRETATION
0.001	HIGHLY SIGNIFICANT
0.01	HIGHLY SIGNIFICANT
0.02	SIGNIFICANT
0.03	SIGNIFICANT
0.04	ON THE EDGE OF SIGNIFICANCE
0.049	ON THE EDGE OF SIGNIFICANCE
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	ON THE EDGE OF SIGNIFICANCE
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.08	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.09	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥0.1	THIS INTERESTING SUBGROUP ANALYSIS

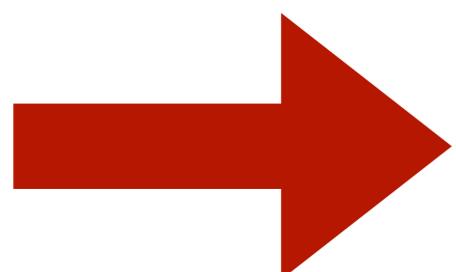
Hypothesis: Resistance conferring mutations and compensatory mutations are independent



~~Hypothesis: Resistance conferring mutations and compensatory mutations are independent~~



All 3 samples are heterozygous S450L Δbp=1/Δbp>1!



If there is a mutation at S450, either see a *rpoC* compensatory mutation or a S450X Δbp>1 **but never both**

P-value = 0.0005

Rifampicin binds to the RNA polymerase, preventing the extension of RNA

A large % of mutations that confer resistance occur in a 81bp region of *rpoB*

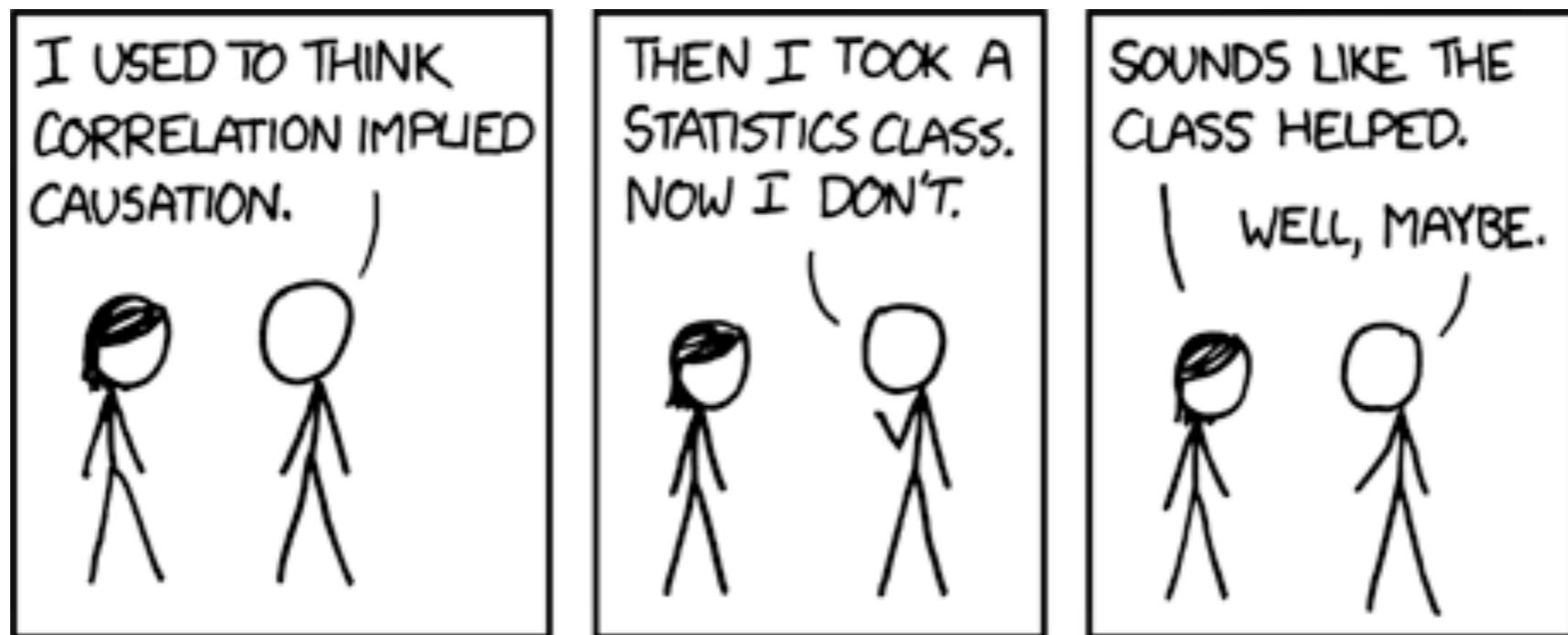
Proximal mutations *reduce* how well rifampicin binds to the RNAP

$\frac{d_N}{d_S}$ Identifies putative compensatory mutations and mutations that confer resistance

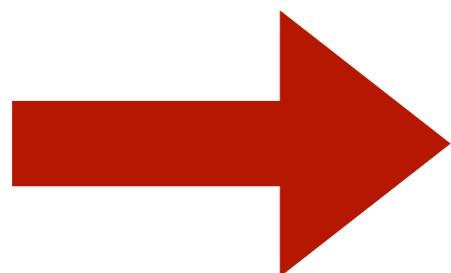
Strains with *rpoB_S450L* but no compensatory mutation grow more slowly than strains with a compensatory mutation

Some compensatory mutations even appear to *enhance* growth

Observe a small number of S450X Δbp>1 mutations; these *never* co-occur with compensatory mutations



<https://xkcd.com/552/>



So how did the S450X $\Delta\text{bp}>1$ get there? And why do we never see them in conjunction with an *rpoC* compensatory mutation?

Can we infer they are **revertants**?



Sarah Walker

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUА	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUА	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUА	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Rifampicin binds to the RNA polymerase, preventing the extension of RNA

A large % of mutations that confer resistance occur in a 81bp region of *rpoB*

Proximal mutations *reduce* how well rifampicin binds to the RNAP

$\frac{d_N}{d_S}$ Identifies putative compensatory mutations and mutations that confer resistance

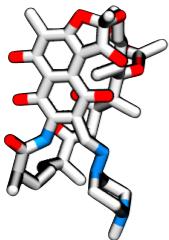
Strains with *rpoB_S450L* but no compensatory mutation grow more slowly than strains with a compensatory mutation

Some compensatory mutations even appear to *enhance* growth

Observe a small number of S450X Δbp>1 mutations; these *never* co-occur with compensatory mutations

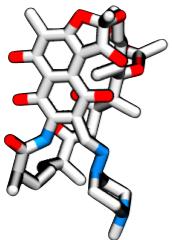
Infer they are reverting due to the fitness cost

1. Treat tuberculosis with rifampicin

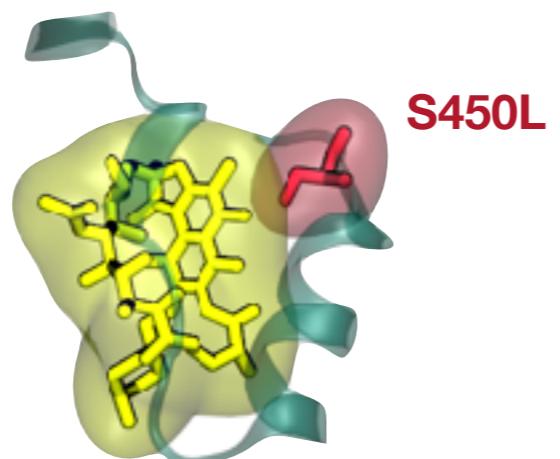


UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

1. Treat tuberculosis with rifampicin

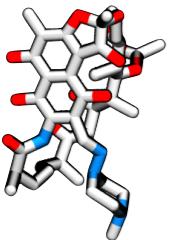


2. **S450L** mutation in rpoB arises
(reduces ΔG_{RIF} but at a fitness cost)

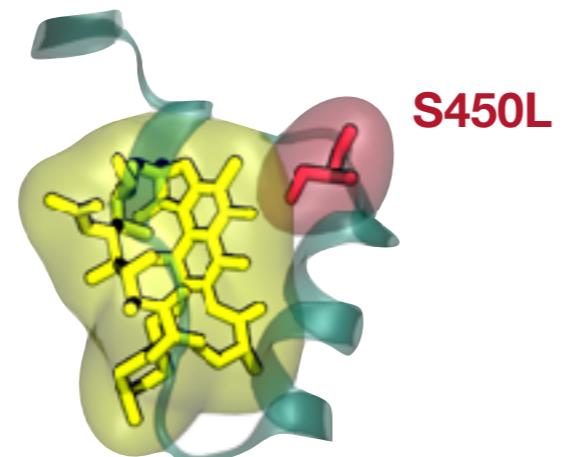


UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

1. Treat tuberculosis with rifampicin



2. S450L mutation in rpoB arises
(reduces ΔG_{RIF} but at a fitness cost)



82%

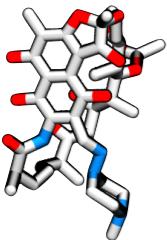
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

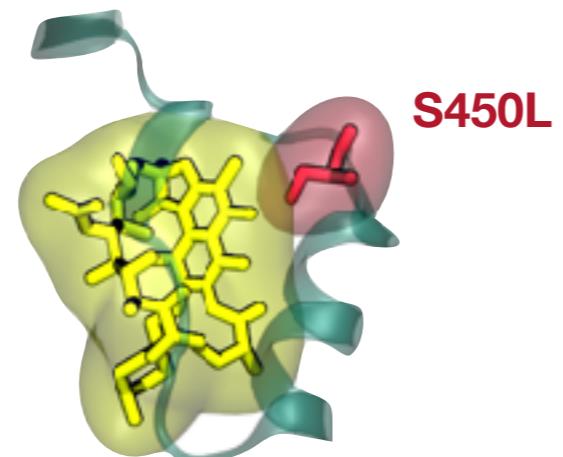
3a. Treatment stopped

Fitness cost leads to reversion;
New mutation may also be resistant
And has an unknown fitness cost

1. Treat tuberculosis with rifampicin



2. S450L mutation in *rpoB* arises
(reduces ΔG_{RIF} but at a fitness cost)



UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

82%

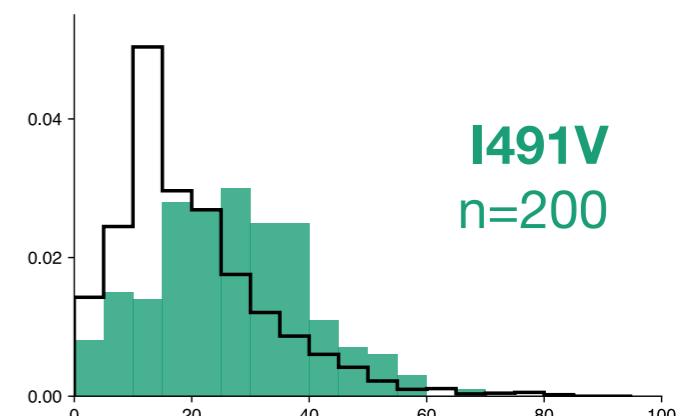
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

3a. Treatment stopped

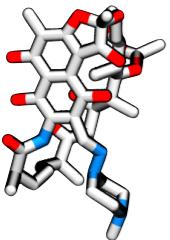
Fitness cost leads to reversion;
New mutation may also be resistant
And has an unknown fitness cost

3b. Treatment continues

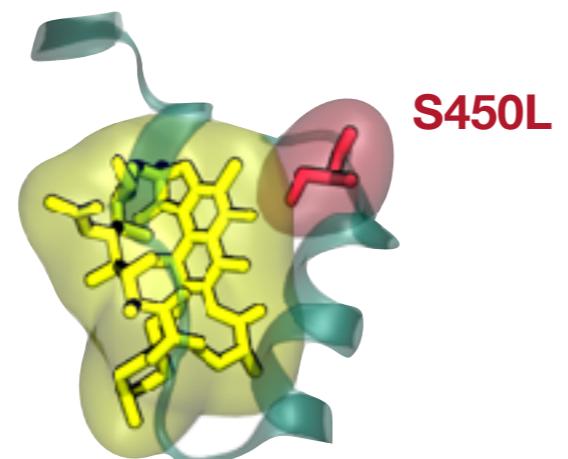
Compensatory mutation rises
abrogating fitness cost
(and may even be fitter)



1. Treat tuberculosis with rifampicin



2. S450L mutation in rpoB arises
(reduces ΔG_{RIF} but at a fitness cost)



82%

UUU Phe	UCU Ser	UAU Tyr	UGU Cys
UUC Phe	UCC Ser	UAC Tyr	UGC Cys
UUA Leu	UCA Ser	UAA Stop	UGA Stop
UUG Leu	UCG Ser	UAG Stop	UGG Trp
CUU Leu	CCU Pro	CAU His	CGU Arg
CUC Leu	CCC Pro	CAC His	CGC Arg
CUA Leu	CCA Pro	CAA Gln	CGA Arg
CUG Leu	CCG Pro	CAG Gln	CGG Arg
AUU Ile	ACU Thr	AAU Asn	AGU Ser
AUC Ile	ACC Thr	AAC Asn	AGC Ser
AUA Ile	ACA Thr	AAA Lys	AGA Arg
AUG Met	ACG Thr	AAG Lys	AGG Arg
GUU Val	GCU Ala	GAU Asp	GGU Gly
GUC Val	GCC Ala	GAC Asp	GGC Gly
GUA Val	GCA Ala	GAA Glu	GGA Gly
GUG Val	GCG Ala	GAG Glu	GGG Gly

UUU Phe	UCU Ser	UAU Tyr	UGU Cys
UUC Phe	UCC Ser	UAC Tyr	UGC Cys
UUA Leu	UCA Ser	UAA Stop	UGA Stop
UUG Leu	UCG Ser	UAG Stop	UGG Trp
CUU Leu	CCU Pro	CAU His	CGU Arg
CUC Leu	CCC Pro	CAC His	CGC Arg
CUA Leu	CCA Pro	CAA Gln	CGA Arg
CUG Leu	CCG Pro	CAG Gln	CGG Arg
AUU Ile	ACU Thr	AAU Asn	AGU Ser
AUC Ile	ACC Thr	AAC Asn	AGC Ser
AUA Ile	ACA Thr	AAA Lys	AGA Arg
AUG Met	ACG Thr	AAG Lys	AGG Arg
GUU Val	GCU Ala	GAU Asp	GGU Gly
GUC Val	GCC Ala	GAC Asp	GGC Gly
GUA Val	GCA Ala	GAA Glu	GGA Gly
GUG Val	GCG Ala	GAG Glu	GGG Gly

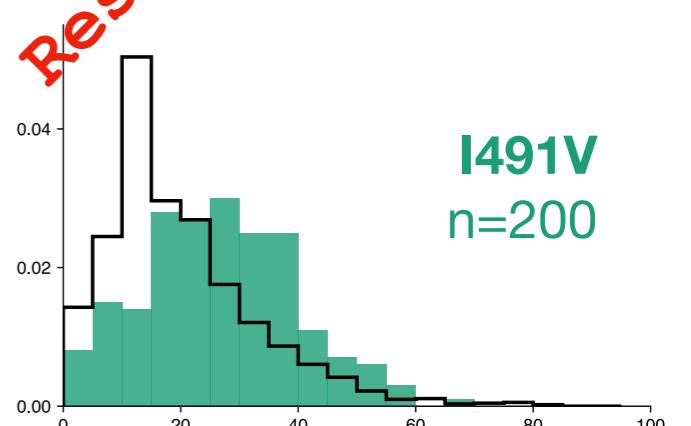
3a. Treatment stopped

Fitness cost leads to reversion;
New mutation may also be resistant
And has an unknown fitness cost

Resistance lost

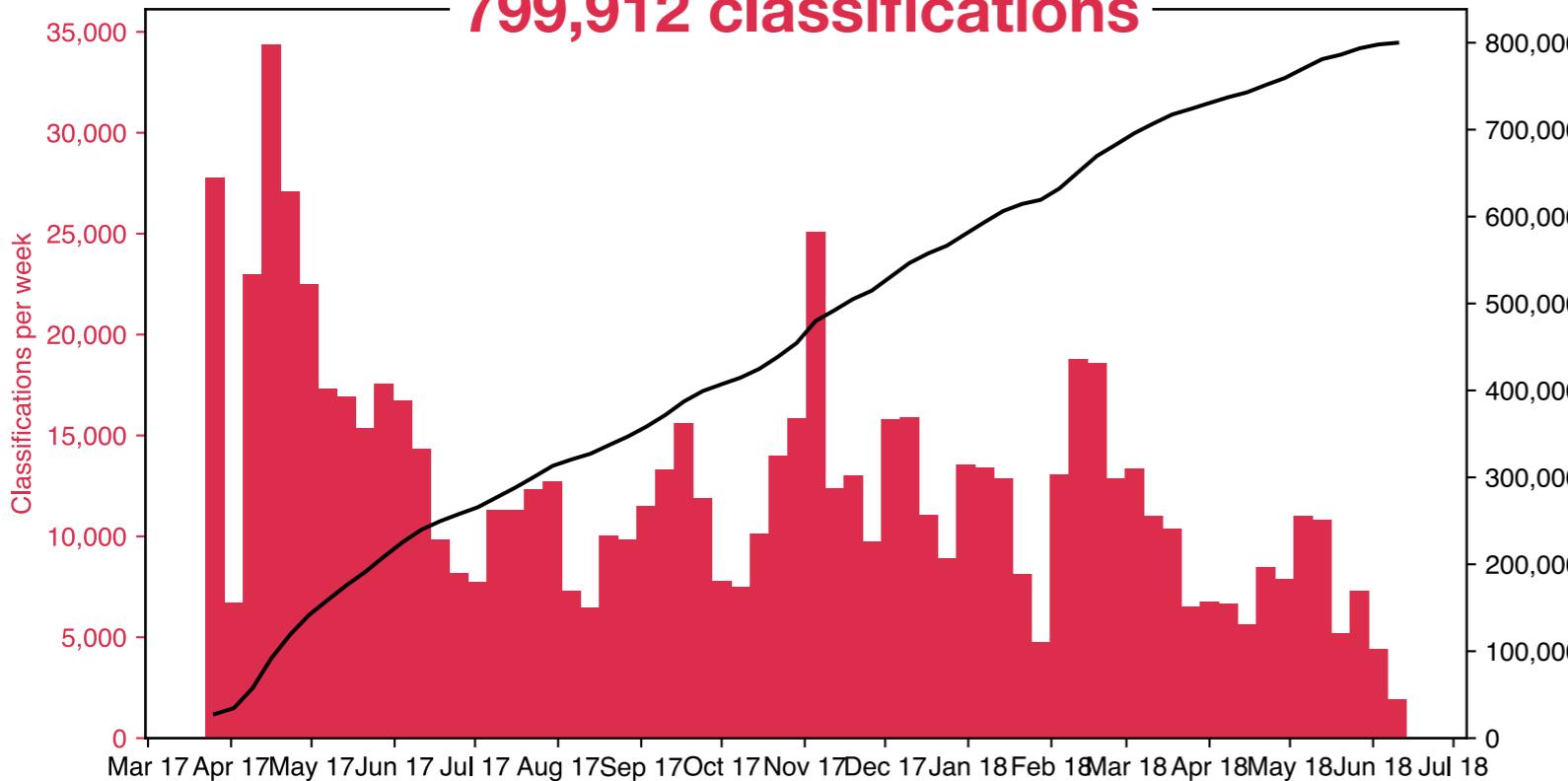
3b. Treatment continues

Compensatory mutation rises
abrogating fitness cost
(and may even be fitter)





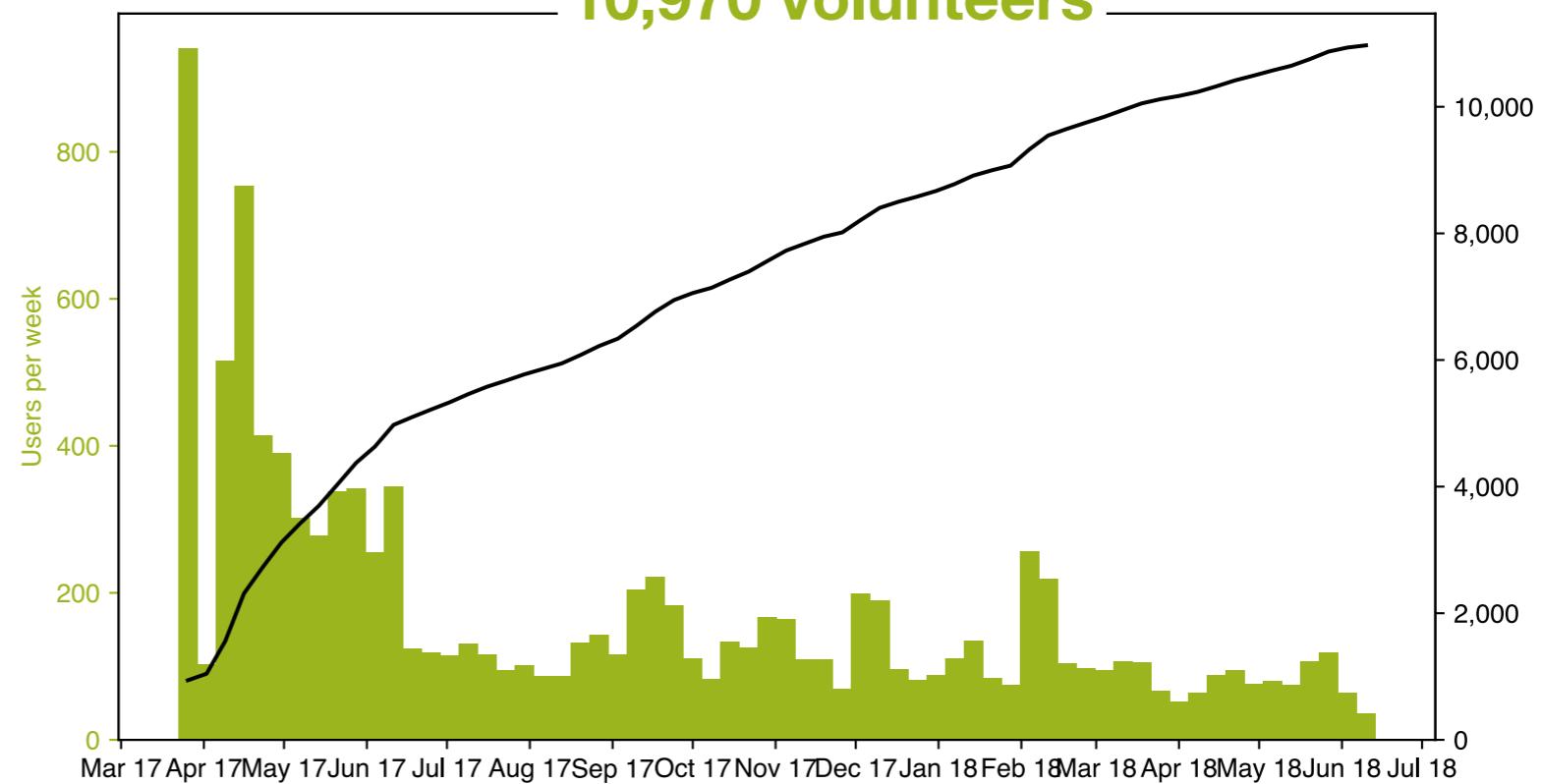
799,912 classifications

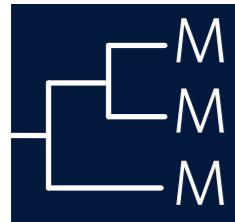


<http://bashthebug.net>
@bashthebug



10,970 volunteers





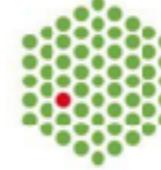
Derrick Crook
Tim Peto
Sarah Walker
Tim Walker
Sarah Hoosdally
Ana Gibertoni-Cruz
Clara Grazian
Josh Carter

<http://fowlerlab.org>

Oliver Adams
Alice Brankin
Dominykas Lukauskis

IGIB, New Delhi

Lipi Thukral



Zam Iqbal
Phelim Bradley
Martin Hunt

DEPARTMENT OF BIOCHEMISTRY
UNIVERSITY OF OXFORD



Simon Newstead



Daniela Cirillo
Grace Smith
Stefan Niemann

IPSB, Toulouse

Matthieu Chavent



Helen Spiers
Grant Miller
Chris Lintott

**BASH THE
BUG**

All the Citizen Scientists

Oxford Biomedical Research Centre

Oxford Radcliffe Hospitals NHS



NHS
National Institute for
Health Research



Funded by
NHS
National Institute for
Health Research



Science & Technology
Facilities Council

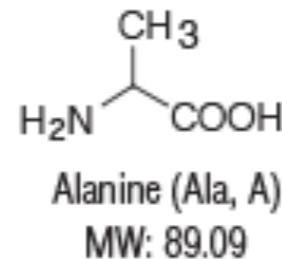
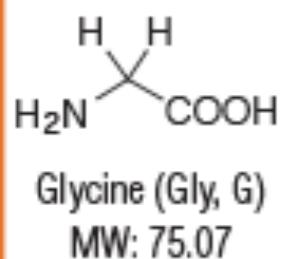
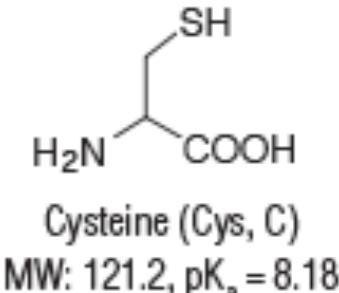
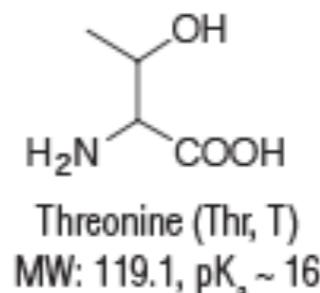
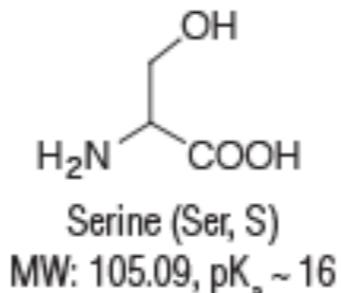
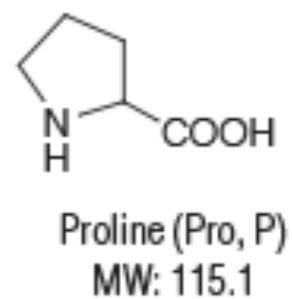
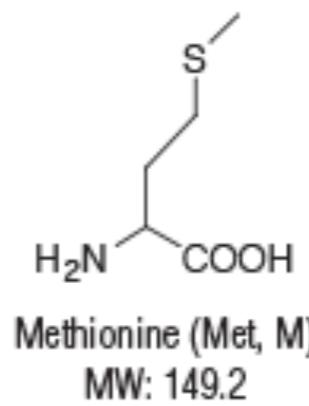
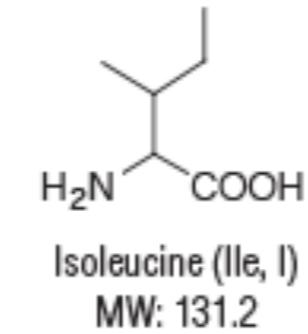
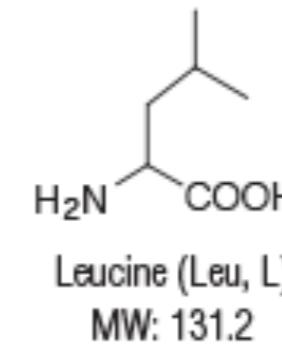
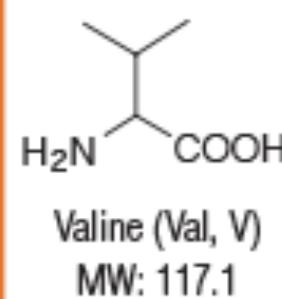
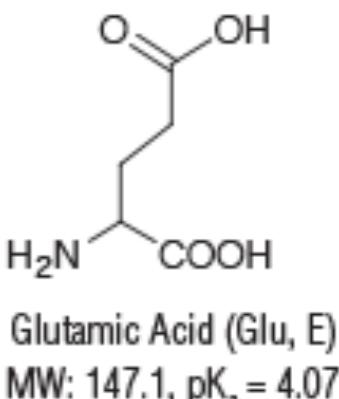
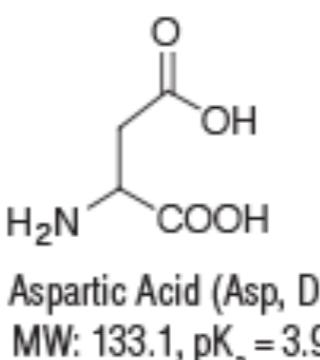
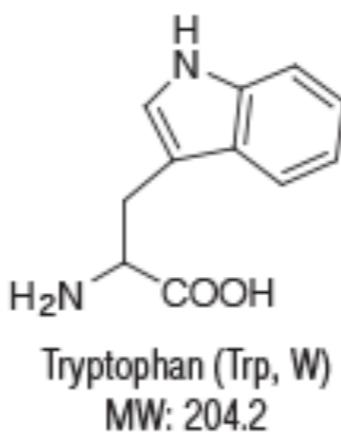
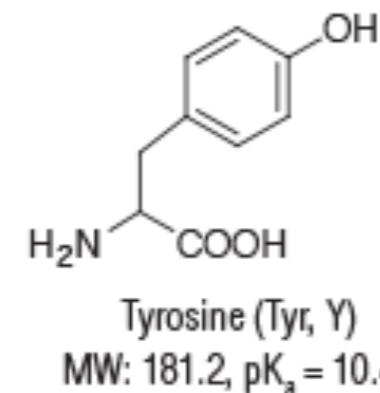
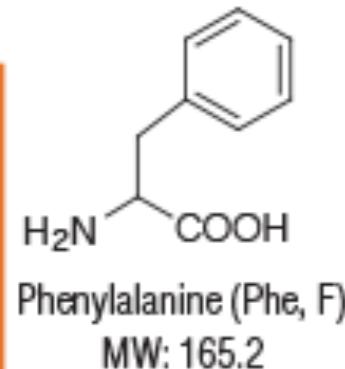
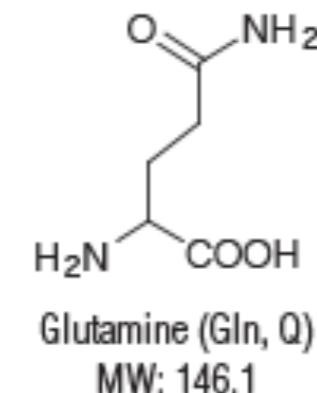
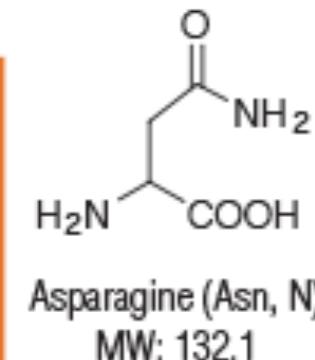
amazon
web services™

archer

CCP BioSim

CompBioMed

arc
advanced
research
computing

SMALL**NUCLEOPHILIC****HYDROPHOBIC****AROMATIC****AMIDE****BASIC**