Rapid, qualitative prediction of antimicrobial resistance by alchemical free energy methods

Philip W Fowler
Experimental Medicine
University of Oxford

http://fowlerlab.org
@philipwfowler

Oliver
Adams

Joshua
Carter

Dominykas
Lukauskis

Alice
Brankin

Philip
Fowler

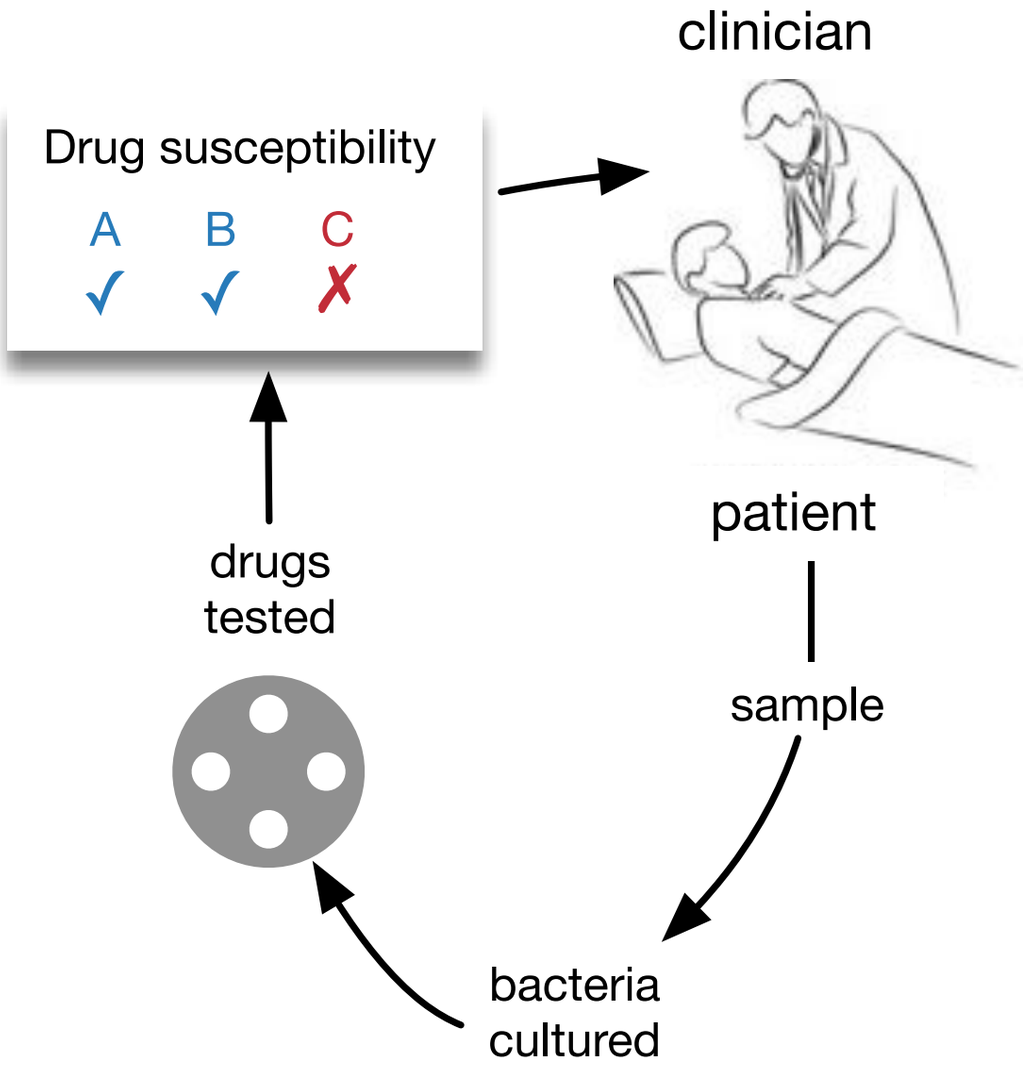
NUFFIELD DEPARTMENT of MEDICINE

UNIVERSITY OF OXFORD

I find it incredible that doctors must still prescribe antibiotics based only on their immediate assessment of a patient's symptoms, just like they used to when antibiotics first entered common use in the 1950s. When a test is used to confirm the diagnosis it is often based on a slow technology that hasn't changed significantly since the 1860s.

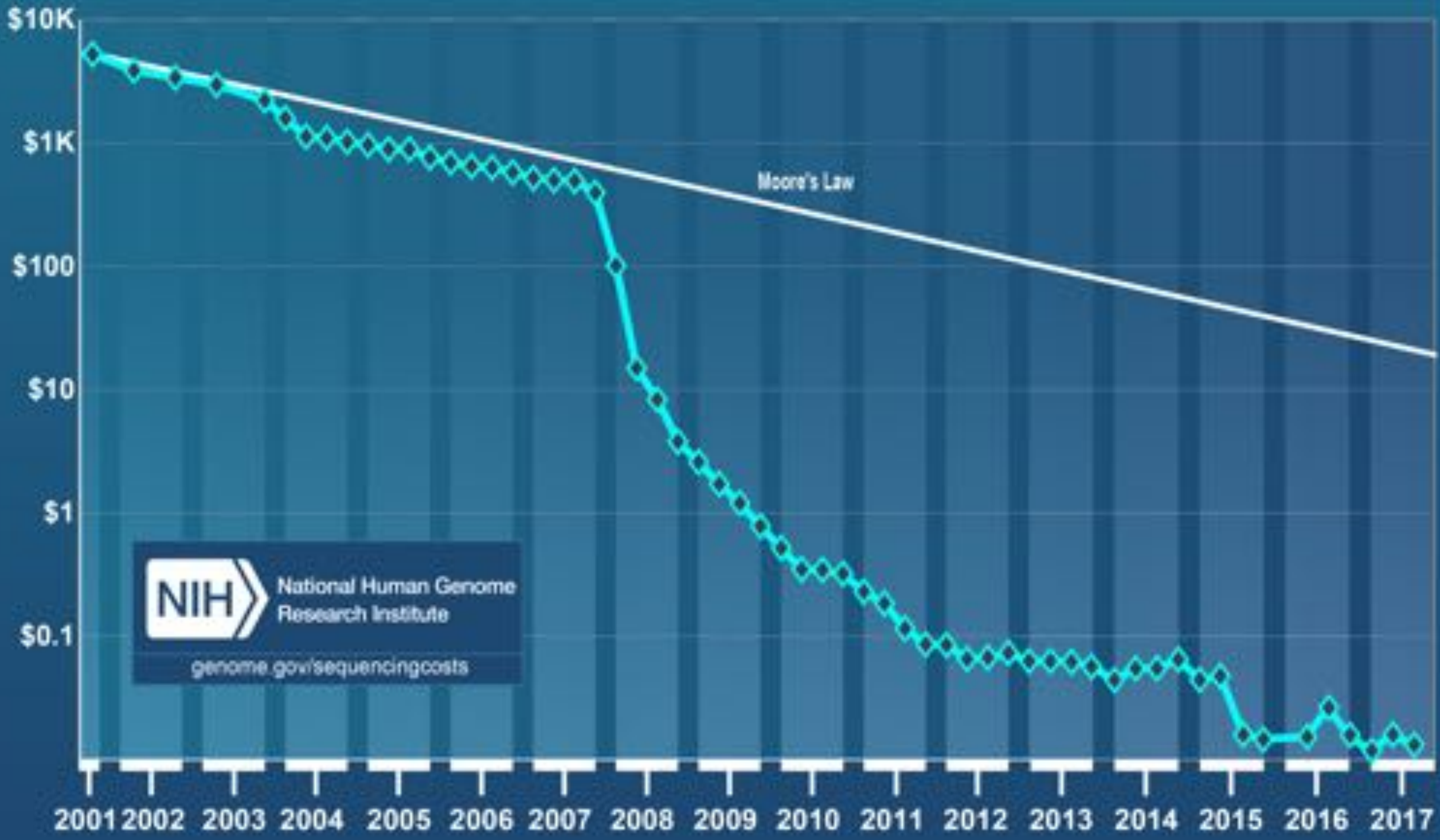
O'Neill report (UK, 2016)

Culture-based clinical microbiology



most pathogens: 1-2 days
M. tuberculosis: 6-8 weeks

Cost per Raw Megabase of DNA Sequence



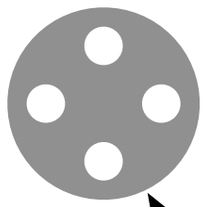
Culture-based clinical microbiology

clinician

Drug susceptibility

A	B	C
✓	✓	✗

drugs tested



patient

sample

bacteria cultured

genome sequenced

most pathogens: 1-2 days
M. tuberculosis: 6-8 weeks

Genetics-based clinical microbiology

Drug susceptibility

A	B	C
✓	✓	✗

previously seen mutation
TCGGG**A**CGGT

Look up mutations in phenotype database

Modernising Medical Microbiology

Whole Genome Sequencing replaced routine culture-based microbiology for TB in England in March 2017

BBC Sign in News Sport Weather iPlayer

NEWS

Home UK World Business Election 2017 Tech Science Health Ed

Health

British scientists in world-first TB breakthrough

24 March 2017 | Health | Share

British scientists claim major advance in TB treatment

British scientists say they have made a world-first breakthrough in the diagnosis of tuberculosis.

Researchers in Oxford and Birmingham say they can isolate different strains of the disease using a process called genome sequencing.

It means patients who may have waited months to get the right drugs can now be diagnosed in just a few days - so they have a greater chance of recovery.

Health Secretary Jeremy Hunt said the breakthrough "will save lives".

Public Health England
Tuberculosis Reference Laboratories

Public Health England

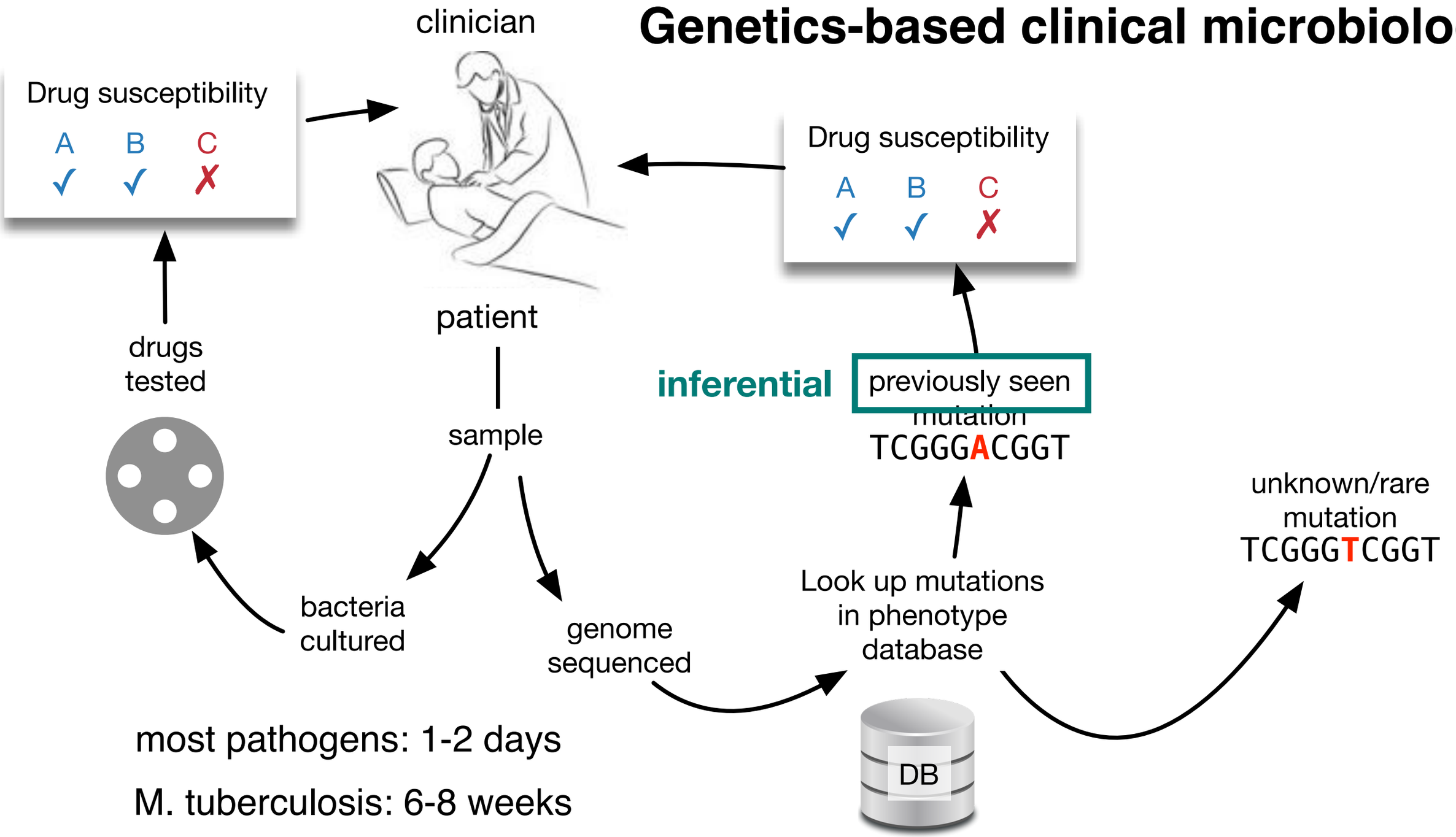
UNIVERSITY OF OXFORD

NHS

Walker, T. M., Cruz, A. L. G., Peto, T. E., Smith, E. G., Esmail, H., & Crook, D. W. (2017). Tuberculosis is changing. *Lancet Infec Disease*, 17(4), 359–361.

Culture-based clinical microbiology

Genetics-based clinical microbiology



Culture-based clinical microbiology

Genetics-based clinical microbiology

clinician

Drug susceptibility

A	B	C
✓	✓	✗

Predict the effect of previously unseen/rare mutations on the action of specific antibiotics

drugs tested

patient

sample

inferential

previously seen mutation
TCGGGACGGT

unknown/rare mutation
TCGGGTCGGT

bacteria cultured

genome sequenced

Look up mutations in phenotype database

DB

most pathogens: 1-2 days

M. tuberculosis: 6-8 weeks

Predict the effect of previously unseen/rare mutations on the action of specific antibiotics

..by considering the effect of the mutation on the chemistry and structure of the protein

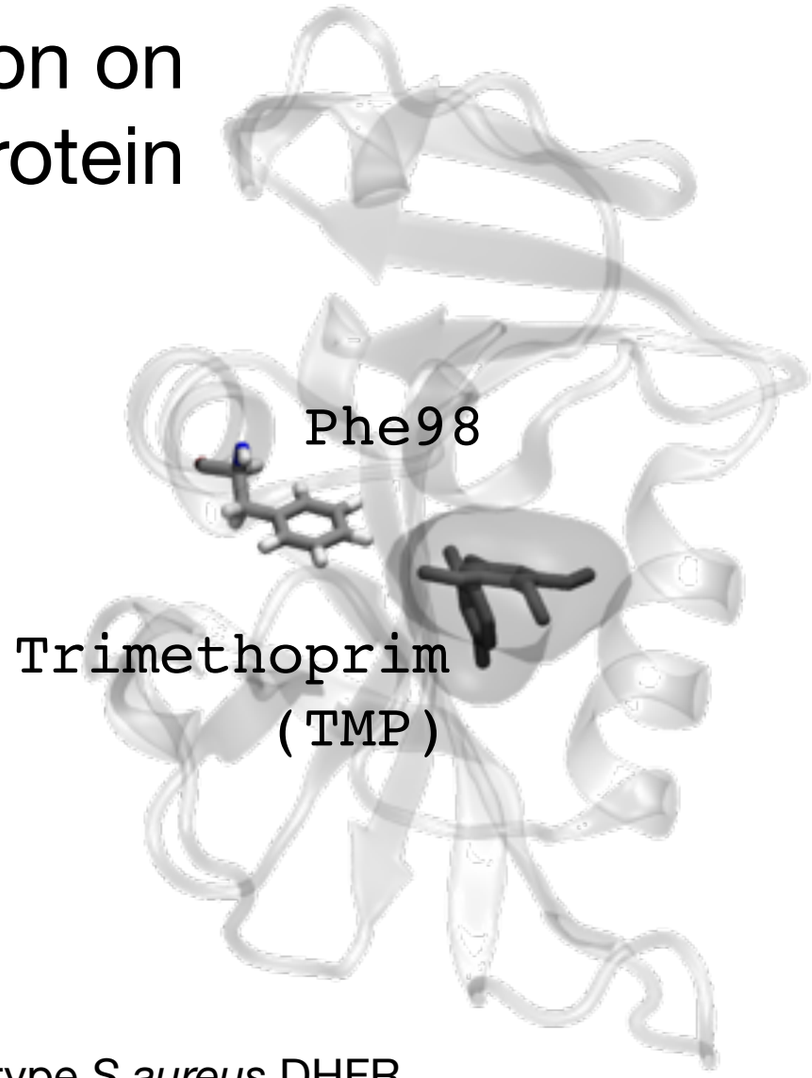
gene (*dfrB*) → protein (DHFR) →

..CUAUUUGAG.. → ..LFE..

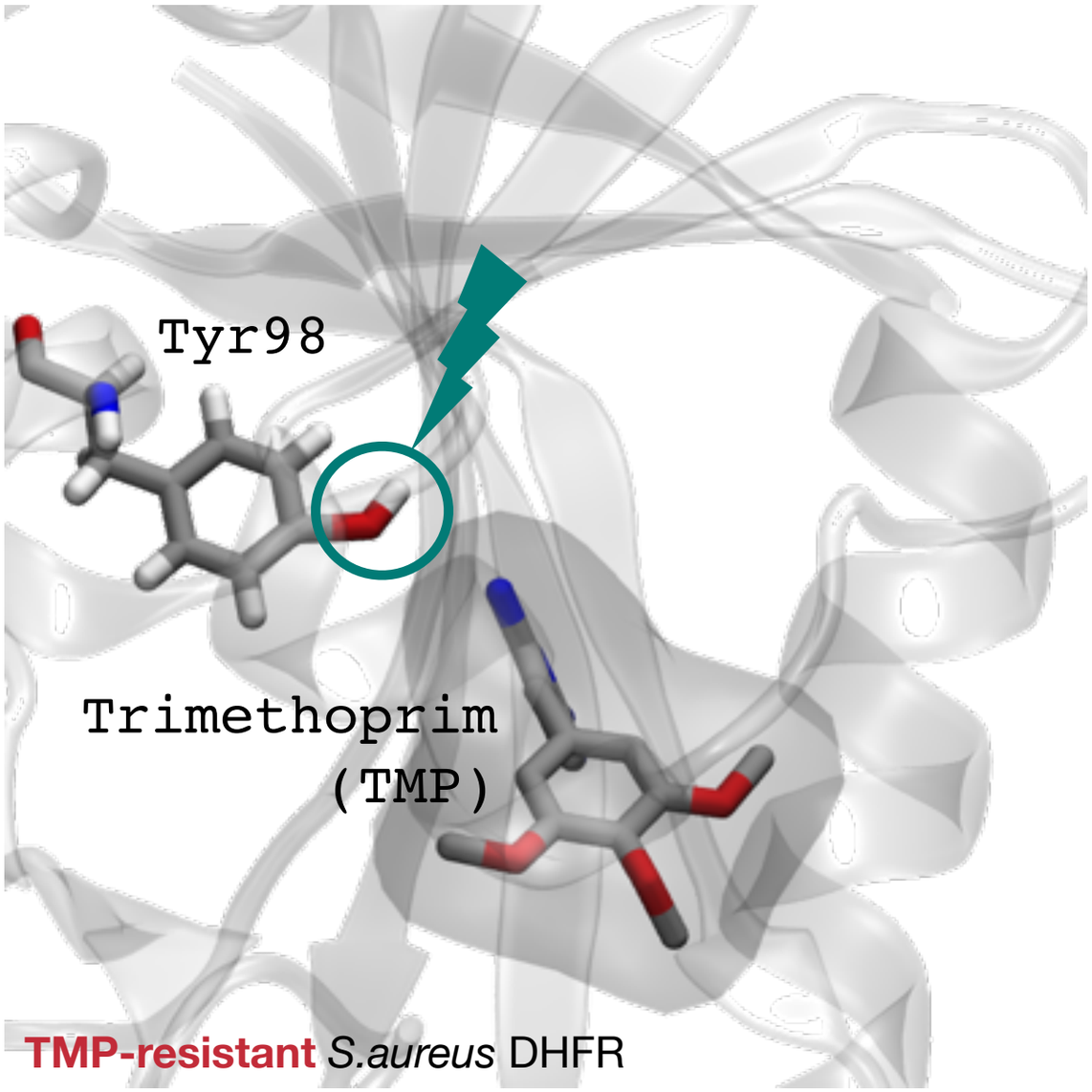
Leu Phe Glu

nucleotides

amino acids



Wildtype *S.aureus* DHFR



TMP-resistant *S.aureus* DHFR

Hypothesis: The **mutation** causes the **antibiotic** to **bind less well** to the protein (whilst not significantly affecting the binding of the natural substrate)

Calculate how the binding free energy of the antibiotic changes (ΔΔG) upon introducing the mutation

Computational Chemistry methods for calculating binding free energies

$$\Delta G = \Delta H - T\Delta S$$

Accuracy

computational docking

1. Empirical scoring functions

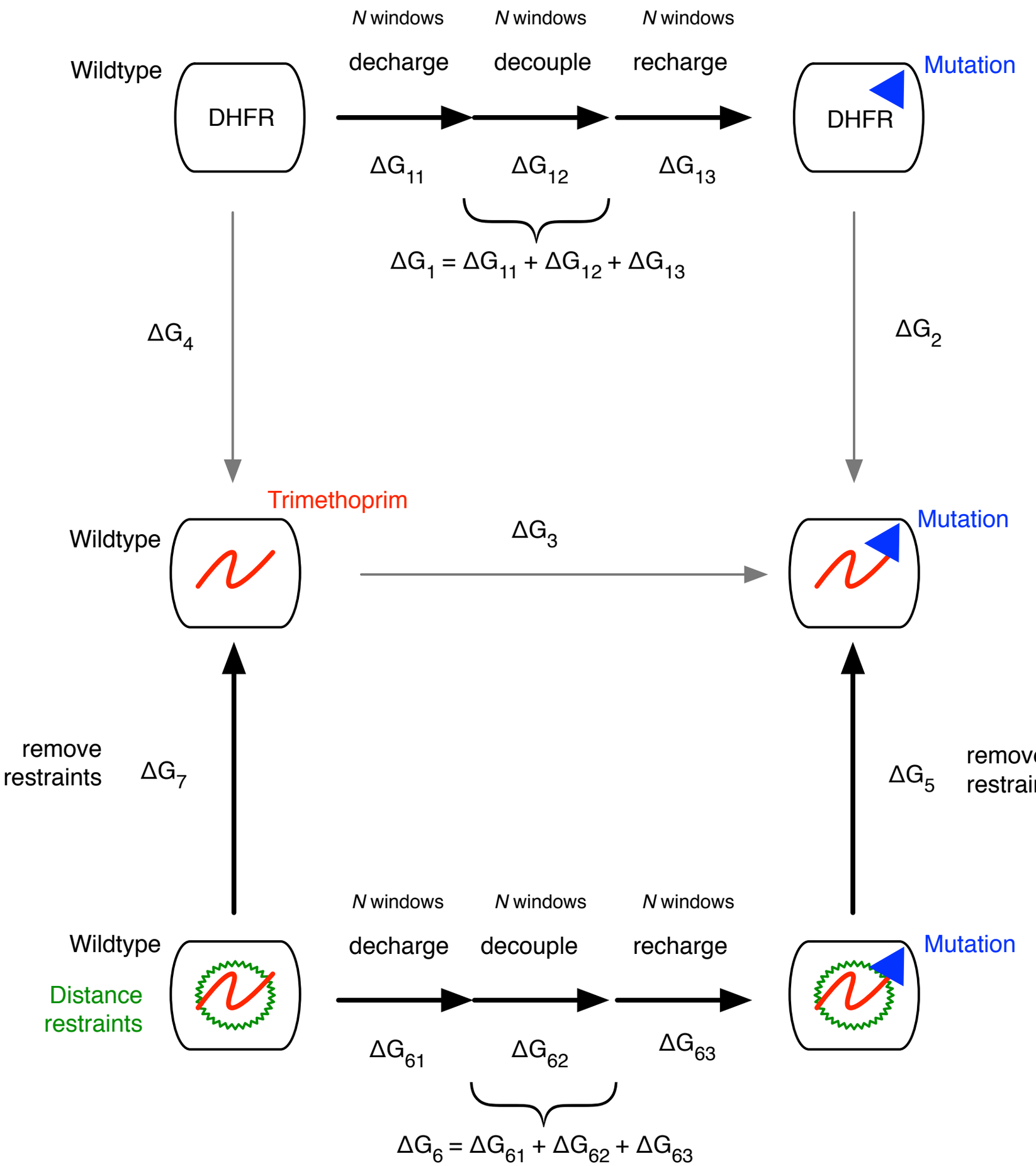
2. End-point methods

3. *Alchemical* methods

rely on **molecular dynamics** simulations

10^{-4} 10^{-2} 10^{0} 10^{2} 10^{4}

Computer time per prediction (CPUhours)



Python

pmx

datreant

alchemlyb

GROMACS 2019.x

Hamiltonian exchange

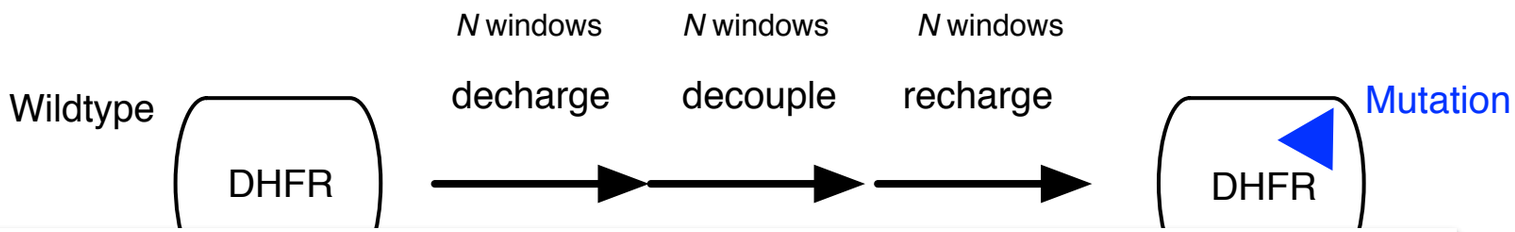
Soft-core vdW

AMBER

ff99SB*-ILDN

GAFF

UNIVERSITY OF OXFORD



Good python packages help enormously

Python

- pmx
- datreant
- alchemlyb

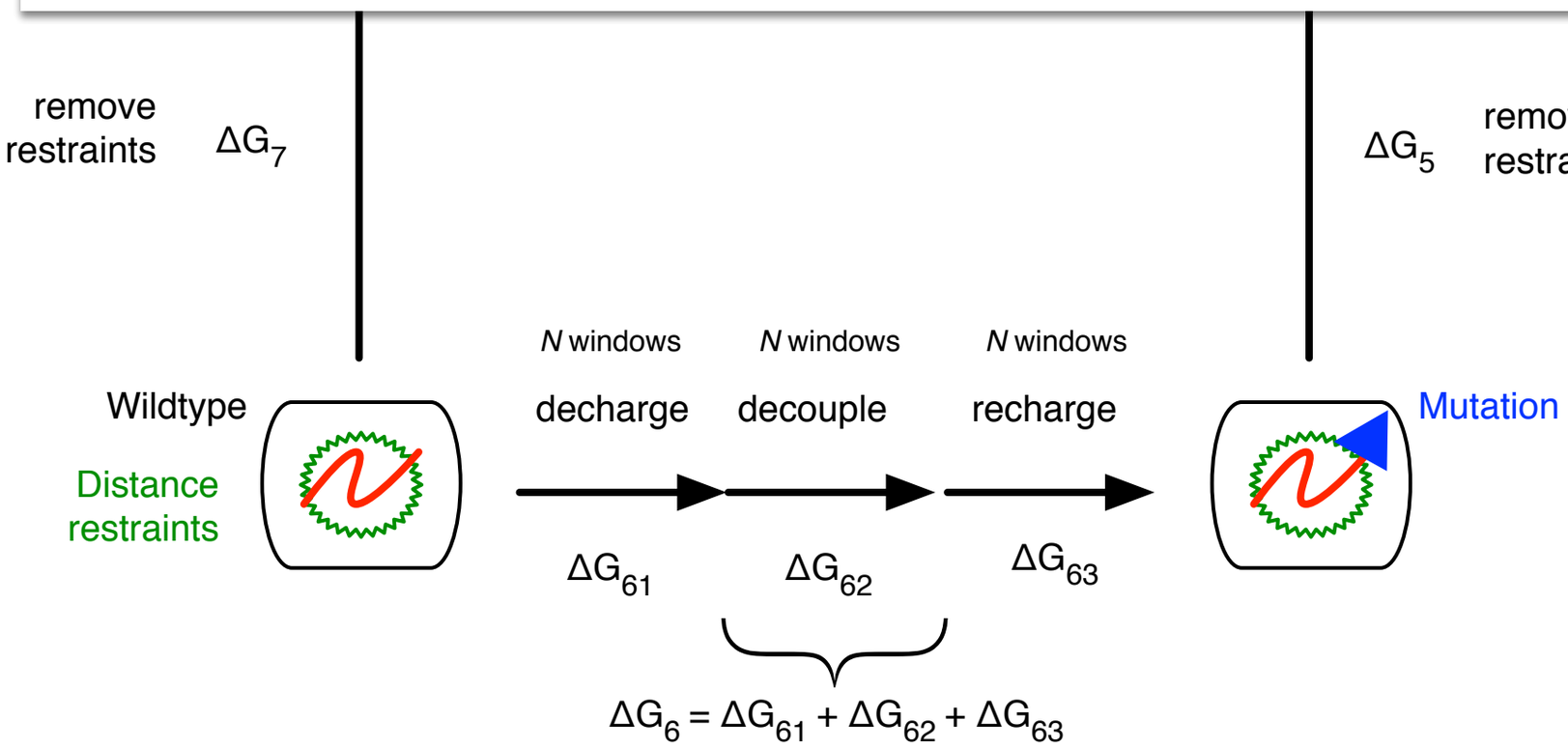
We are applying Free Energy methods, *not* developing them

GROMACS 2019.x

- Hamiltonian exchange
- Soft-core vdW

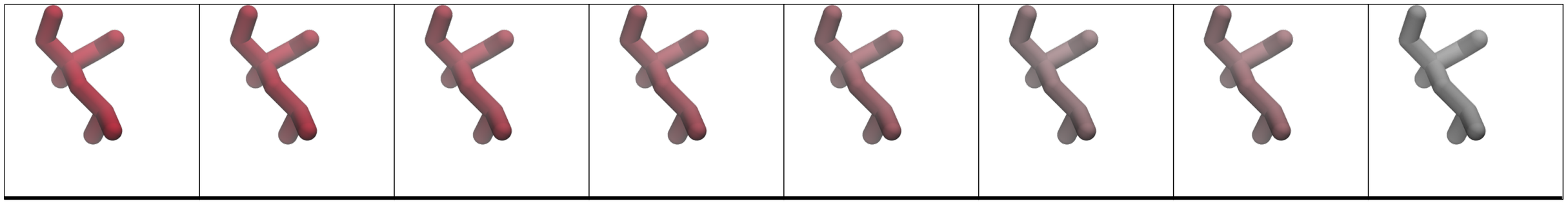
AMBER

- ff99SB*-ILDN
- GAFF

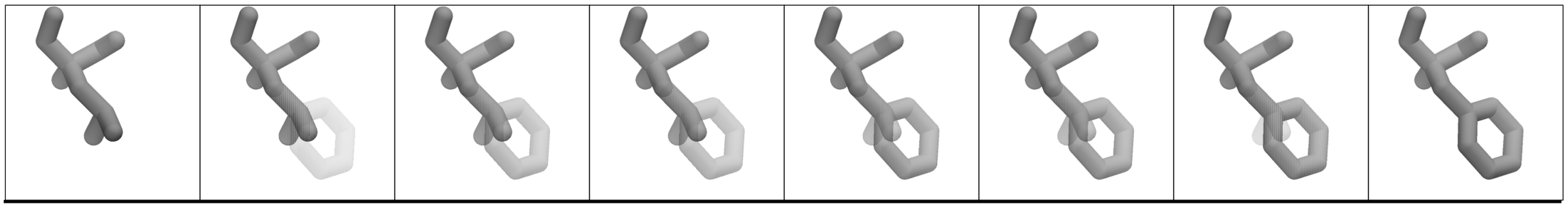


UNIVERSITY OF OXFORD

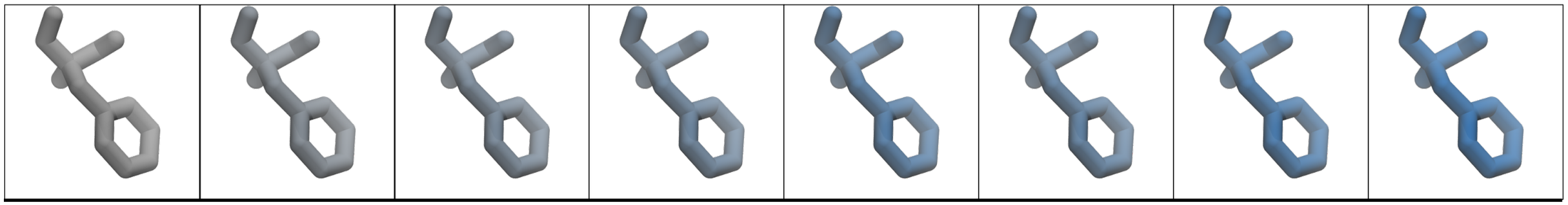
(a) remove **charges**



(b) Leucine → Phenylalanine



(c) add **charges**



Please cite this article in press as: Fowler et al., Robust Prediction of Resistance to Trimethoprim in Staphylococcus aureus, Cell Chemical Biology (2017), https://doi.org/10.1016/j.chembiol.2017.12.009

Cell Chemical Biology
Resource

Robust Prediction of Resistance to Trimethoprim in *Staphylococcus aureus*

Philip W. Fowler,[1,2,6,*] Kevin Cole,[3] N. Claire Gordon,[1] Angela M. Kearns,[3] Martin J. Llewelyn,[2,4] Tim E.A. Peto,[1,4] Derrick W. Crook,[1,4] and A. Sarah Walker[1,4]
[1]Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, Headley Way, Oxford OX3 9DU, UK
[2]Department of Infectious Diseases and Microbiology, Royal Sussex County Hospital, Brighton, Brighton and Sussex Medical School, Brighton BN1 9PS, UK
[3]Antimicrobial Resistance and Healthcare Associated Infections Reference Unit, Public Health England, Colindale NW9 5EQ, UK
[4]Senior author
[5]Twitter: @philipwfowler
[6]Lead Contact
*Correspondence: philip.fowler@ndm.ox.ac.uk
https://doi.org/10.1016/j.chembiol.2017.12.009

SUMMARY

The rise of antibiotic resistance threatens modern medicine; to combat it new diagnostic methods are required. Sequencing the whole genome of a pathogen offers the potential to accurately determine which antibiotics will be effective to treat a patient. A key limitation of this approach is that it cannot classify rare or previously unseen mutations. Here we demonstrate that alchemical free energy methods, a well-established class of methods from computational chemistry, can successfully predict whether mutations in *Staphylococcus aureus* dihydrofolate reductase confer resistance to trimethoprim. We also show that the method is quantitatively accurate by calculating how much the most common resistance-conferring mutation, F99Y, reduces the binding free energy of trimethoprim and comparing predicted and experimentally measured minimum inhibitory concentrations for seven different mutations. Finally, by considering up to 32 free energy calculations for each mutation, we estimate its specificity and sensitivity.

INTRODUCTION

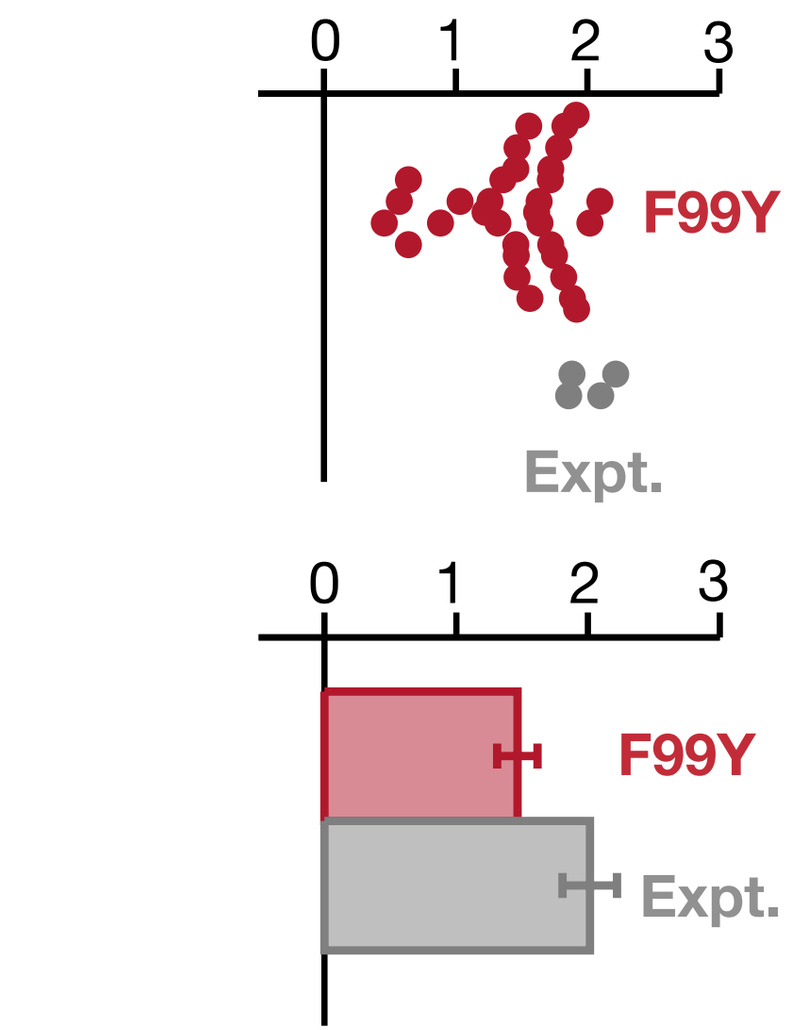
Resistance of bacteria to the antibiotics used to treat them is a substantial and growing global threat to human health (Davies, 2013; World Economic Forum, 2013). Measures to counter the emergence of antibiotic resistance are restricted by the limitations of conventional diagnostic microbiology. This predominantly still relies on culture-based, phenotypic identification of bacteria, followed by growth in the presence of different antibiotic concentrations to detect resistance. The process is labor intensive, takes days or even weeks depending on the growth rate of the organism in question, is expensive, and open to subjective interpretation. Genetic approaches, particularly those based on sequencing the entire genome of a pathogen (Didelot et al., 2012; Köser et al., 2014), have the potential to be faster and cheaper. Inferring the phenotype of an infecting pathogen from whole-genome sequence data by considering known resistance genes or mutations has already been shown to be reasonably accurate for a range of pathogens (Gordon et al., 2014; Walker et al., 2015; Pankhurst et al., 2016; Bradley et al., 2015), and has recently been implemented in the UK for the routine diagnosis of *Mycobacterium tuberculosis* infections (Walker et al., 2017). New mutations, however, continually arise, and a genetics-based clinical microbiology service therefore also needs to be able to *predict* the effect of novel mutations. In this paper we demonstrate that molecular-based computational chemistry methods can predict whether individual protein mutations confer resistance to an antibiotic.

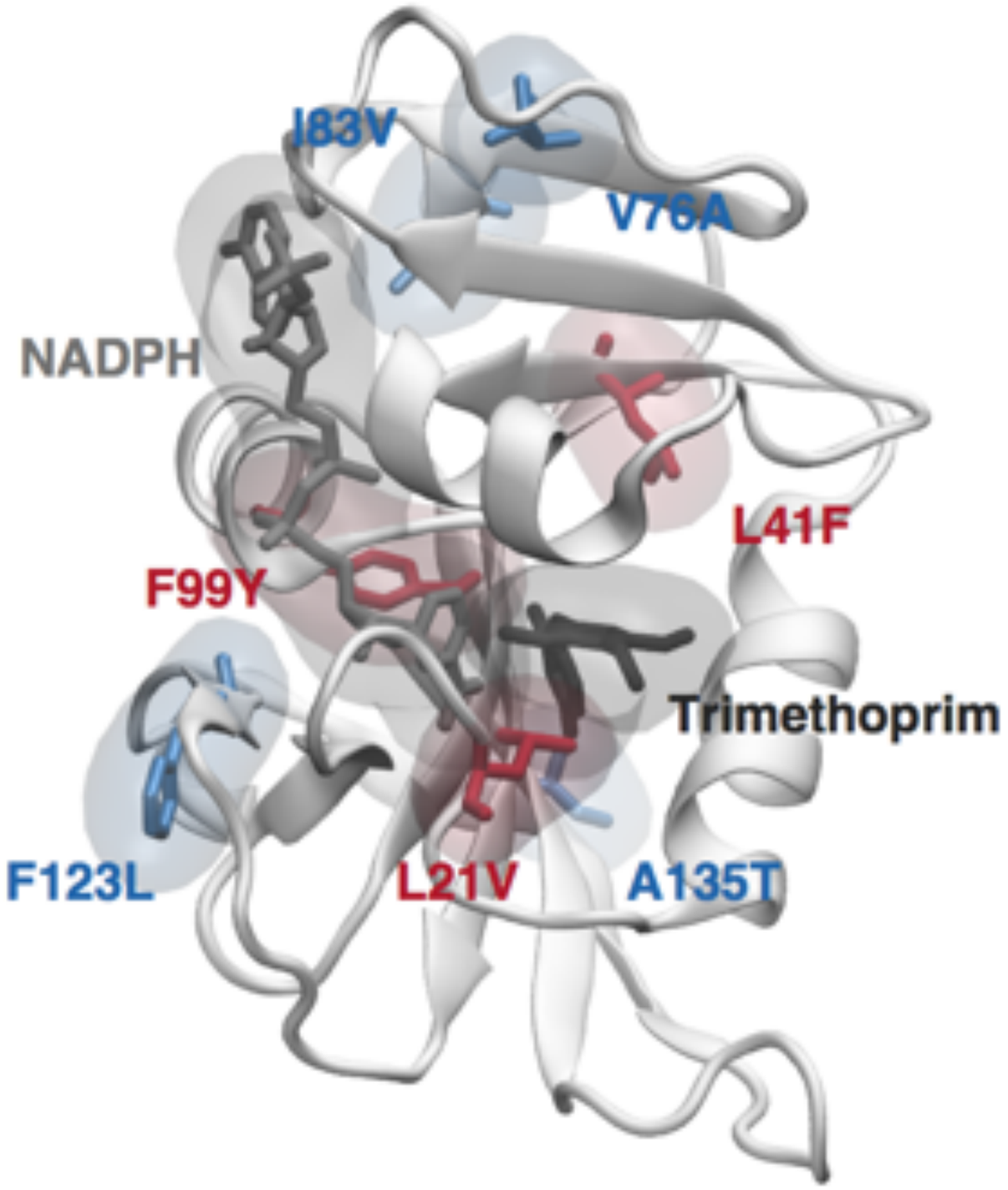
As proof of principle we have investigated the effect of mutations to *Staphylococcus aureus* dihydrofolate reductase (DHFR) on the binding of the antibiotic trimethoprim (TMP) (Figure 1A). *S. aureus* is a clinically important Gram-positive pathogen and has been the focus of much research due to the development of methicillin- and vancomycin-resistant strains, known as MRSA and VRSA, respectively. TMP, usually administered as co-trimoxazole (trimethoprim-sulfamethoxazole), has a long history of treating *S. aureus* infections (Tong et al., 2015), including common skin and soft tissue infections caused by MRSA strains (Nurjadi et al., 2014). TMP competes with the natural substrate, dihydrofolic acid (DHA) (Figure 1A), for binding to DHFR, thereby preventing DHFR catalyzing the conversion of DHA to tetrahydrofolic acid. Since tetrahydrofolate is essential for the biosynthesis of thymidylate, purine nucleotides, and some amino acids, arresting the production of DHA inhibits bacterial growth. Resistance to TMP in *S. aureus* can either arise from mutations in the chromosomal gene *dfrB*, or from the introduction of other naturally resistant genes (*dfrA*, *dfrG*, and *dfrK*) via plasmids (Lowy, 2003; Nurjadi et al., 2014). Here we focus on seven mutations in the *dfrB* 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

Susceptible **Resistant**

mutation makes it easier for antibiotic to bind

mutation makes it harder for antibiotic to bind





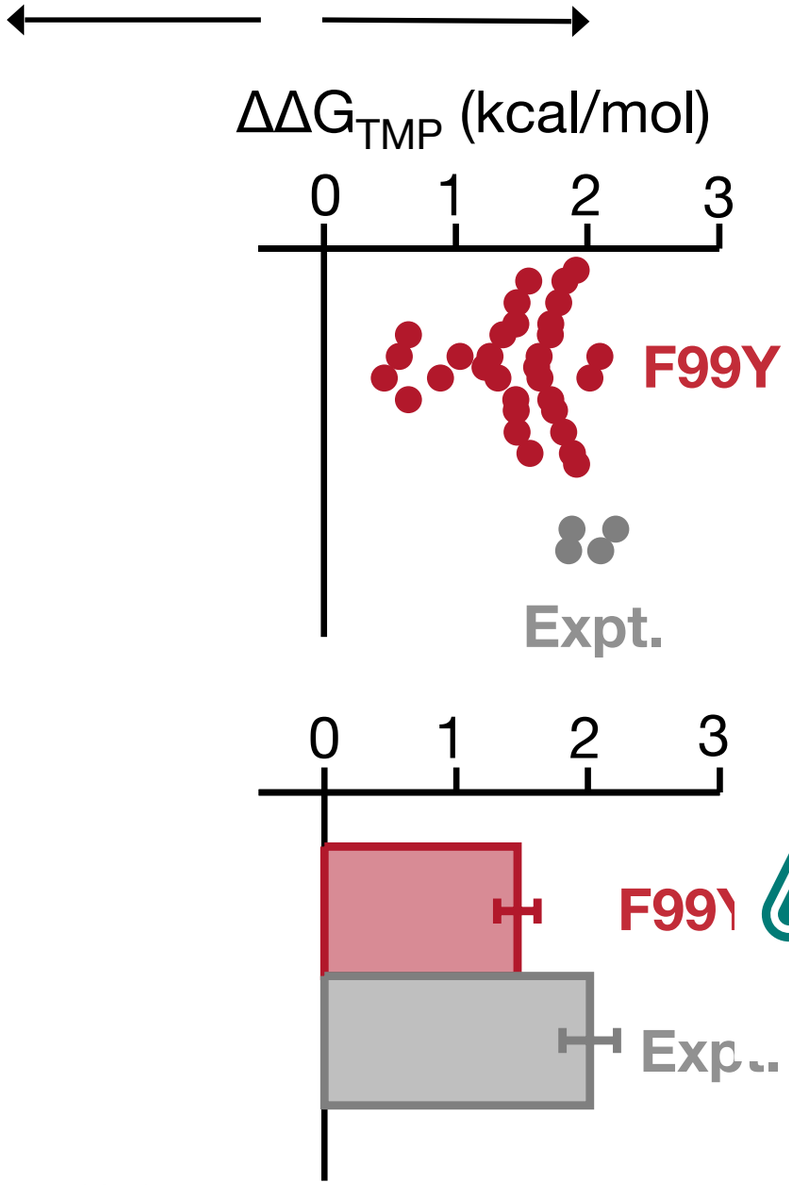
error bars 95% CI (t-statistic)

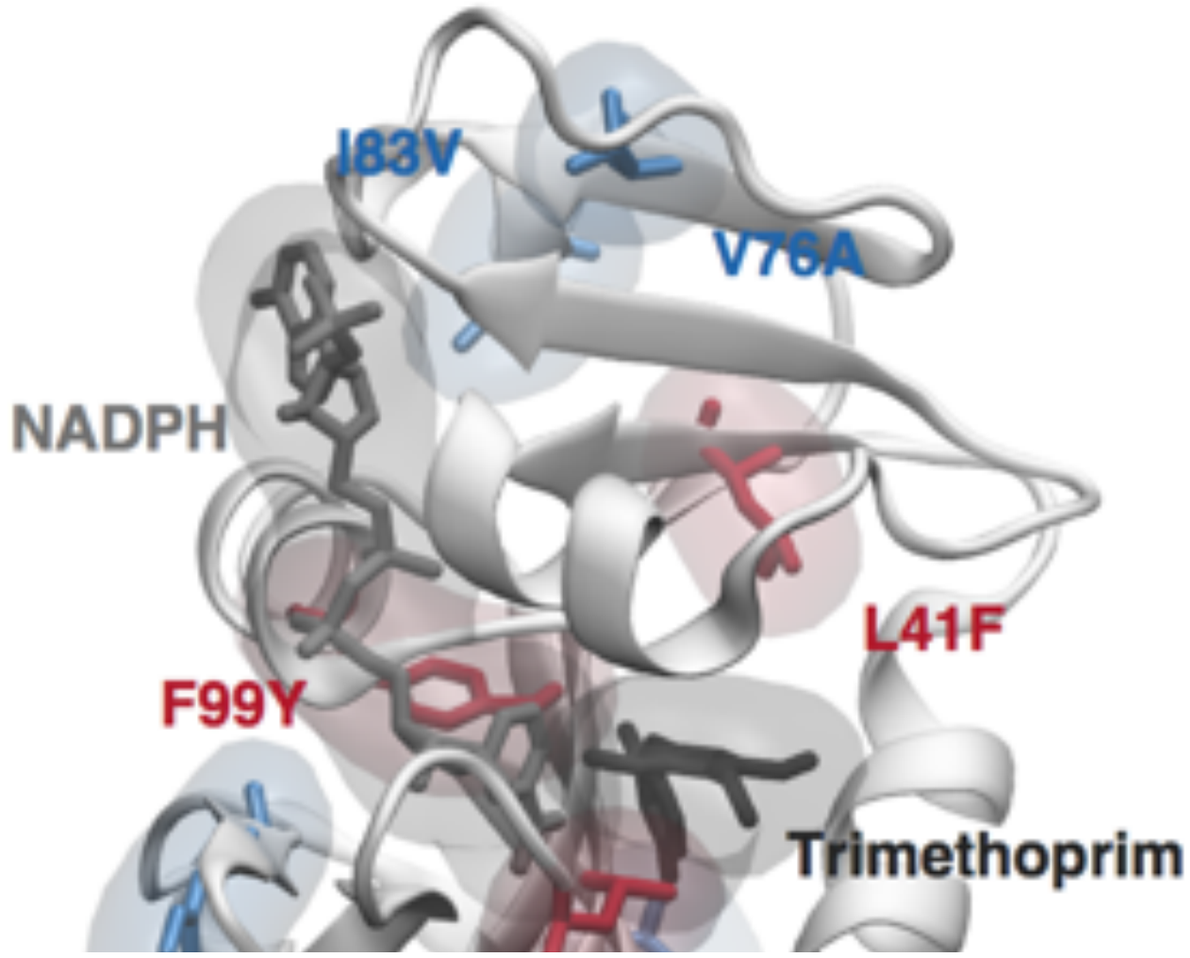
Experimental value from four independently published isothermal titration calorimetry studies

Susceptible **Resistant**

mutation makes it easier for antibiotic to bind

mutation makes it harder for antibiotic to bind





n>1 makes calculating errors simple

error bars 95% CI (t-statistic)

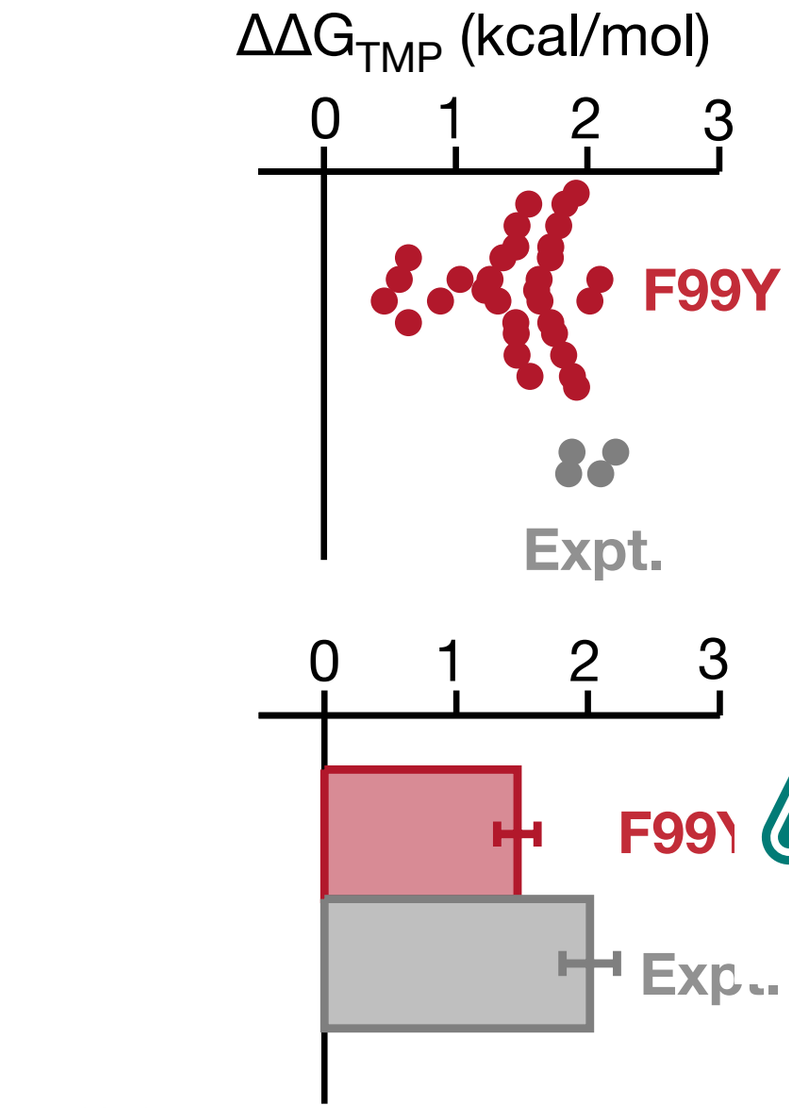
Experimental value from four independently published isothermal titration calorimetry studies

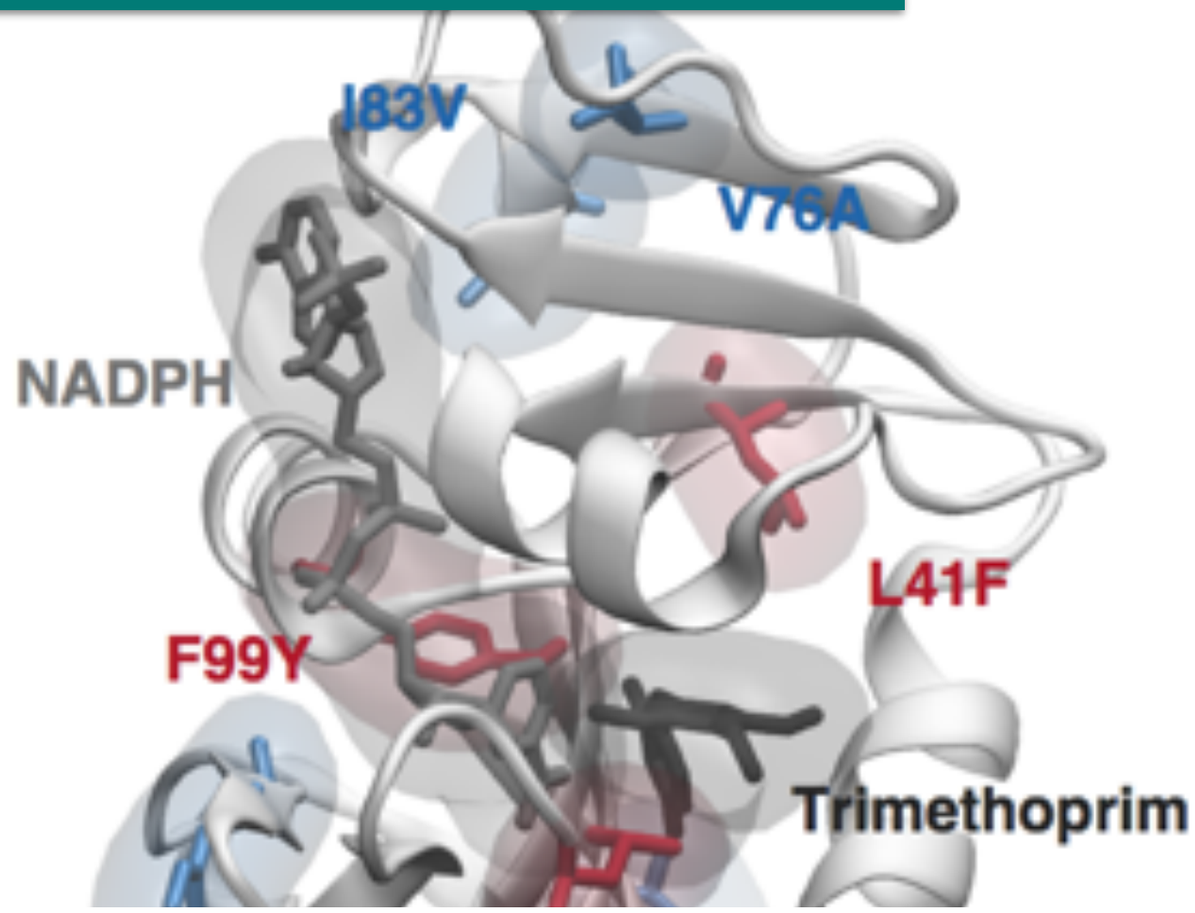
Quantitatively predict effect of F99Y mutation

Susceptible **Resistant**

mutation makes it easier for antibiotic to bind

mutation makes it harder for antibiotic to bind

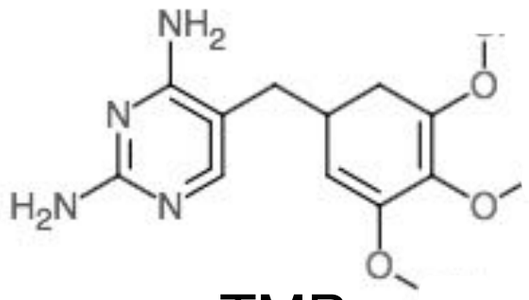




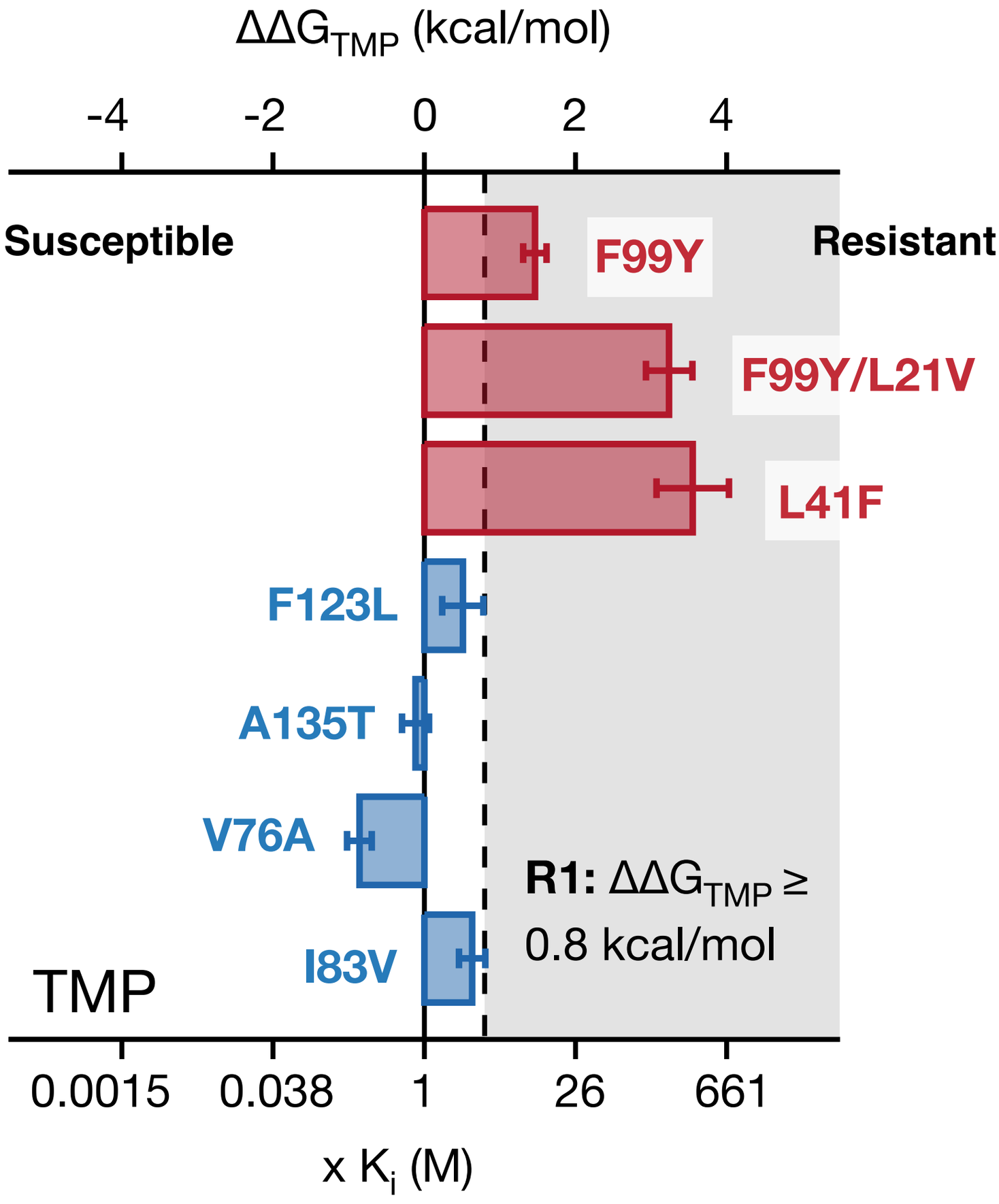
n>1 makes calculating errors simple

error bars 95% CI (t-statistic)

Experimental value from four independently published isothermal titration calorimetry studies

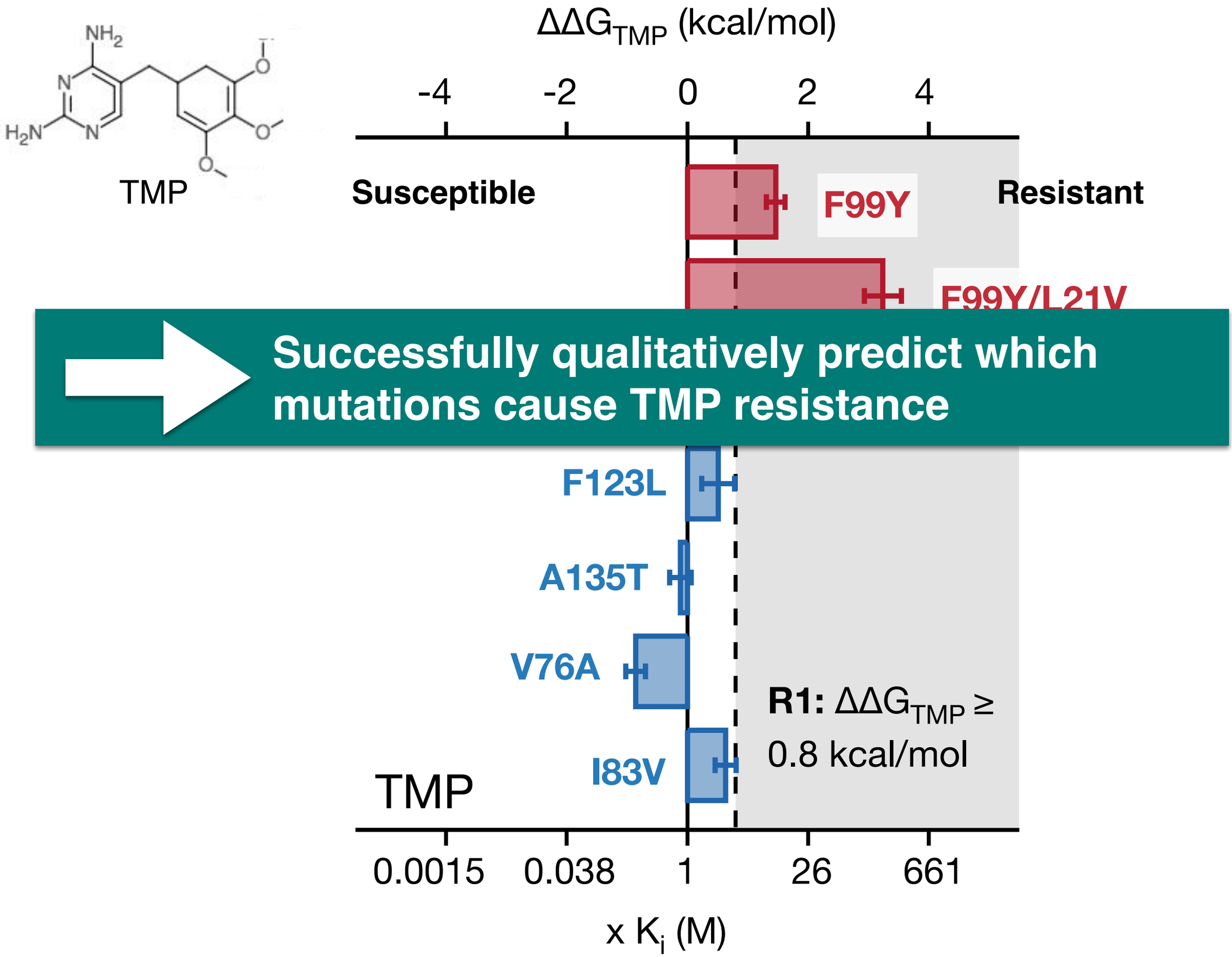


TMP



Each prediction: 32 repeats, 8x ΔG, 8-16x λ, 0.25 ns duration 3.2h, 4,000 CPUh

error bars 95% CI (t-statistic)



Successfully qualitatively predict which mutations cause TMP resistance

Each prediction: 32 repeats, 8x ΔG, 8-16x λ, 0.25 ns duration 3.2h, 4,000 CPUh

error bars 95% CI (t-statistic)

We can do amazingly well considering the λ simulations are only 250ps long

We can do amazingly well considering the λ simulations are only 250ps long

Clinically we care about predicting Resistant/Susceptible/Unknown not the accuracy and precision of ΔΔG

Just how fast and sloppy can we be and still make clinical predictions?

INTERNATIONAL STANDARD

ISO 20776-2

First edition
2007-07-01

Clinical laboratory testing and *in vitro* diagnostic test systems — Susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices —

Part 2:
Evaluation of performance of antimicrobial susceptibility test devices

Systèmes d'essais en laboratoire et de diagnostic in vitro — Sensibilité in vitro des agents infectieux et évaluation des performances des dispositifs pour antibiogrammes —

Partie 2: Évaluation des performances des dispositifs pour antibiogrammes

Reference number
ISO 20776-2:2007(E)

3.6.3
very major discrepancy
VMD
test result by the reference method interpreted as R and an AST device result of S

Another representation of the concept:

$$\frac{N_{\text{VMD}} \times 100}{N_{\text{RREF}}}$$

Major discrepancy = S sample predicted R

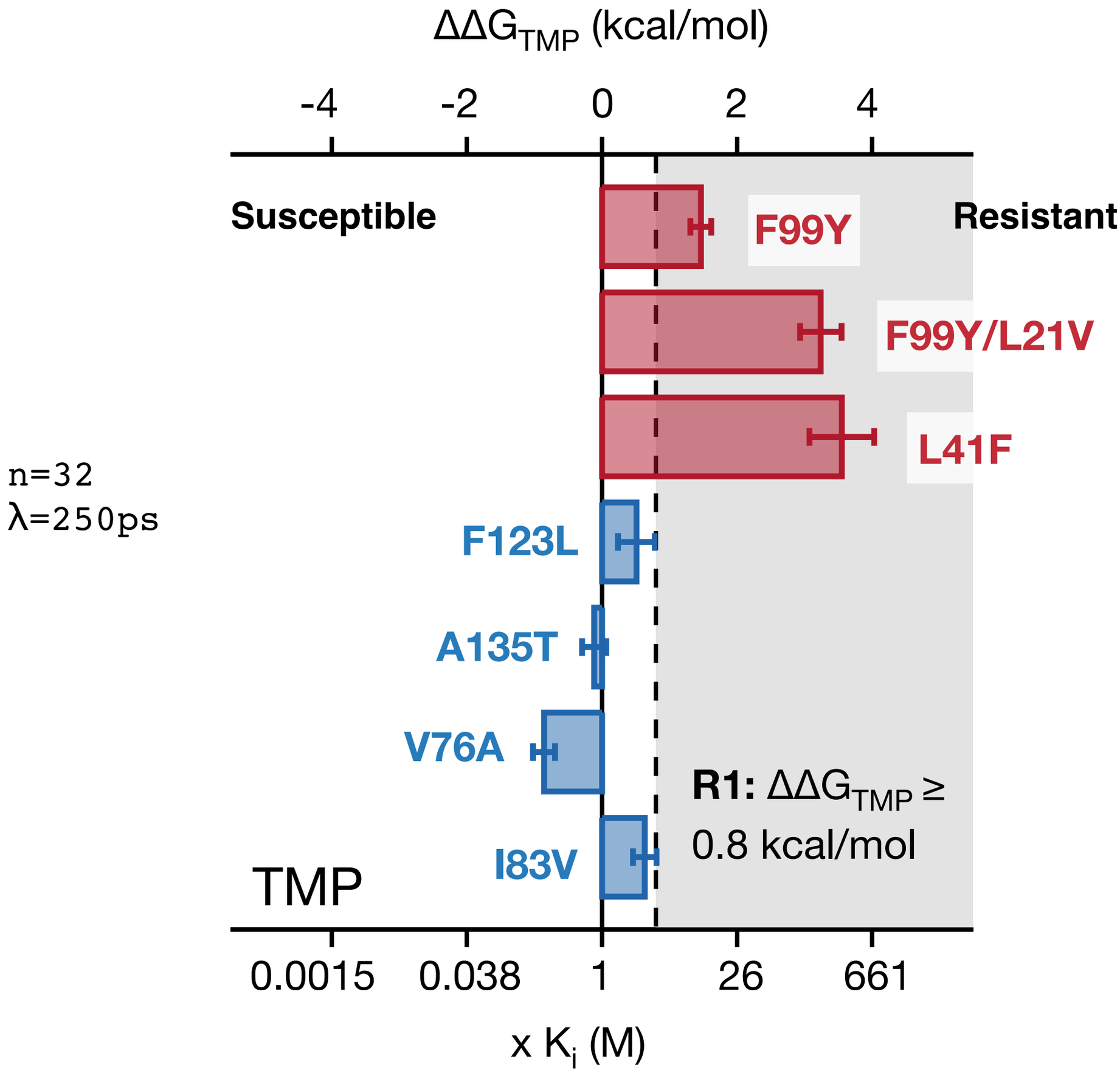
where

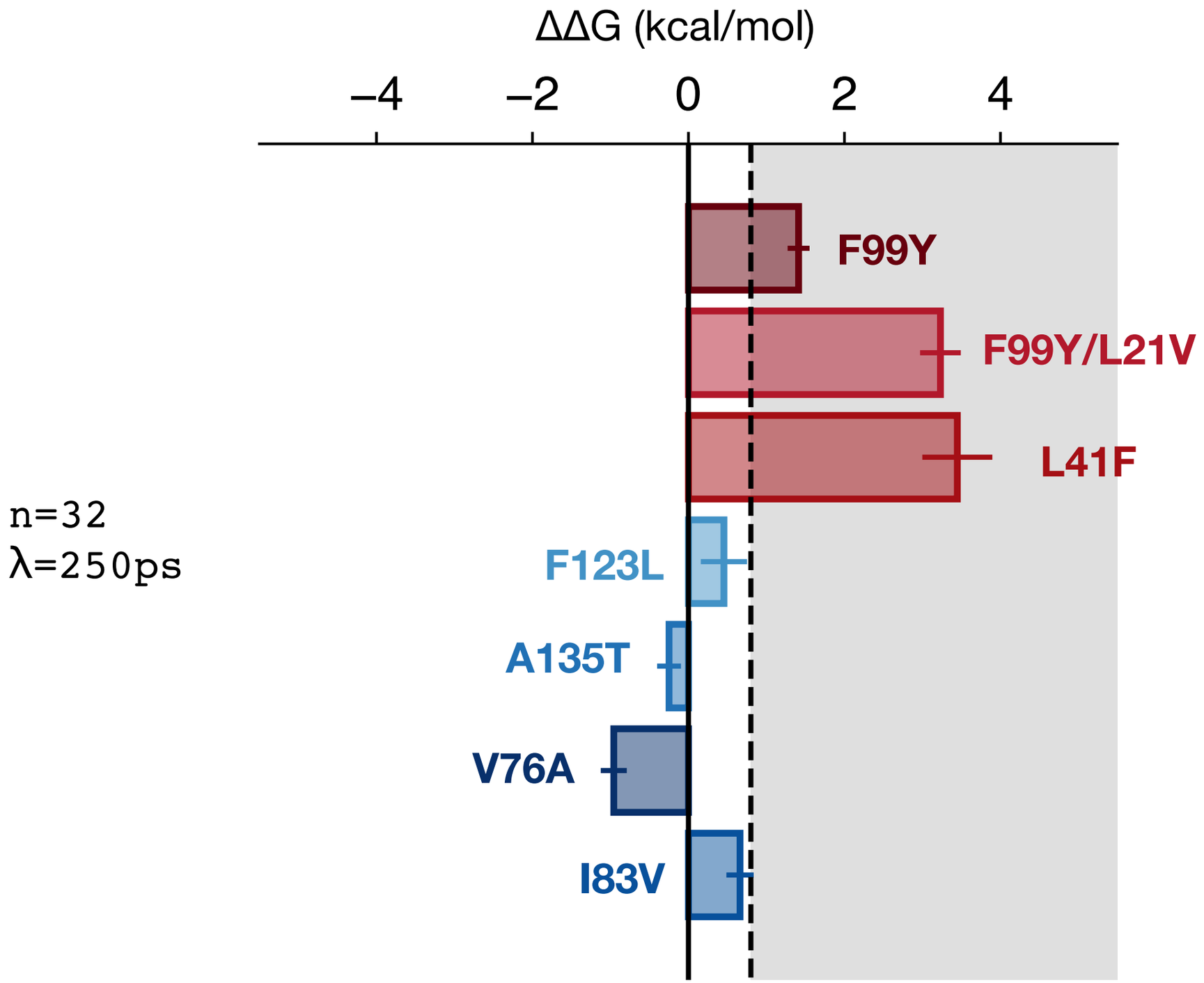
N_{VMD} is the number of tests that result in a VMD;

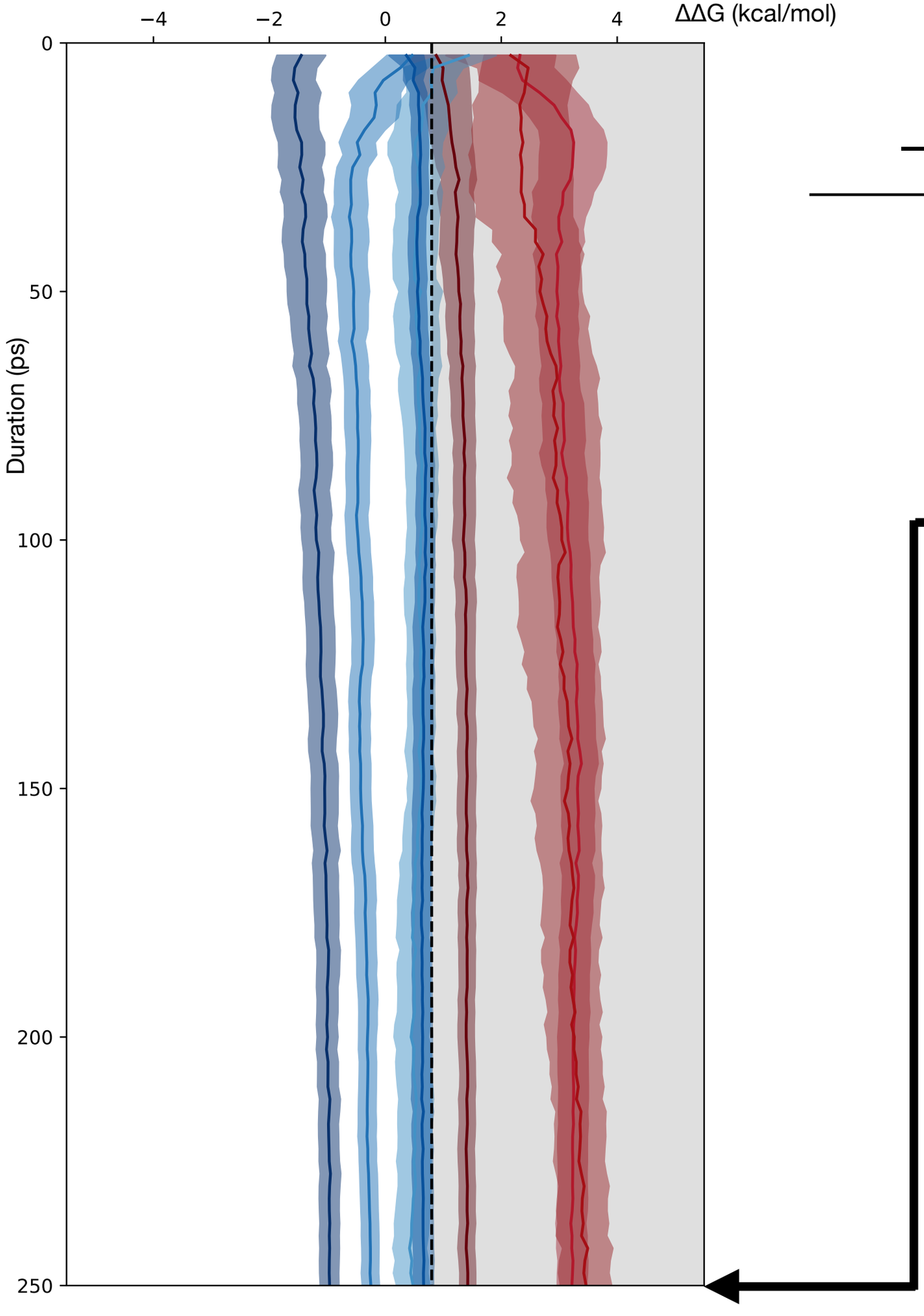
N_{RREF} is the number of resistant bacterial isolates as determined by the reference method (ISO 20776-1)

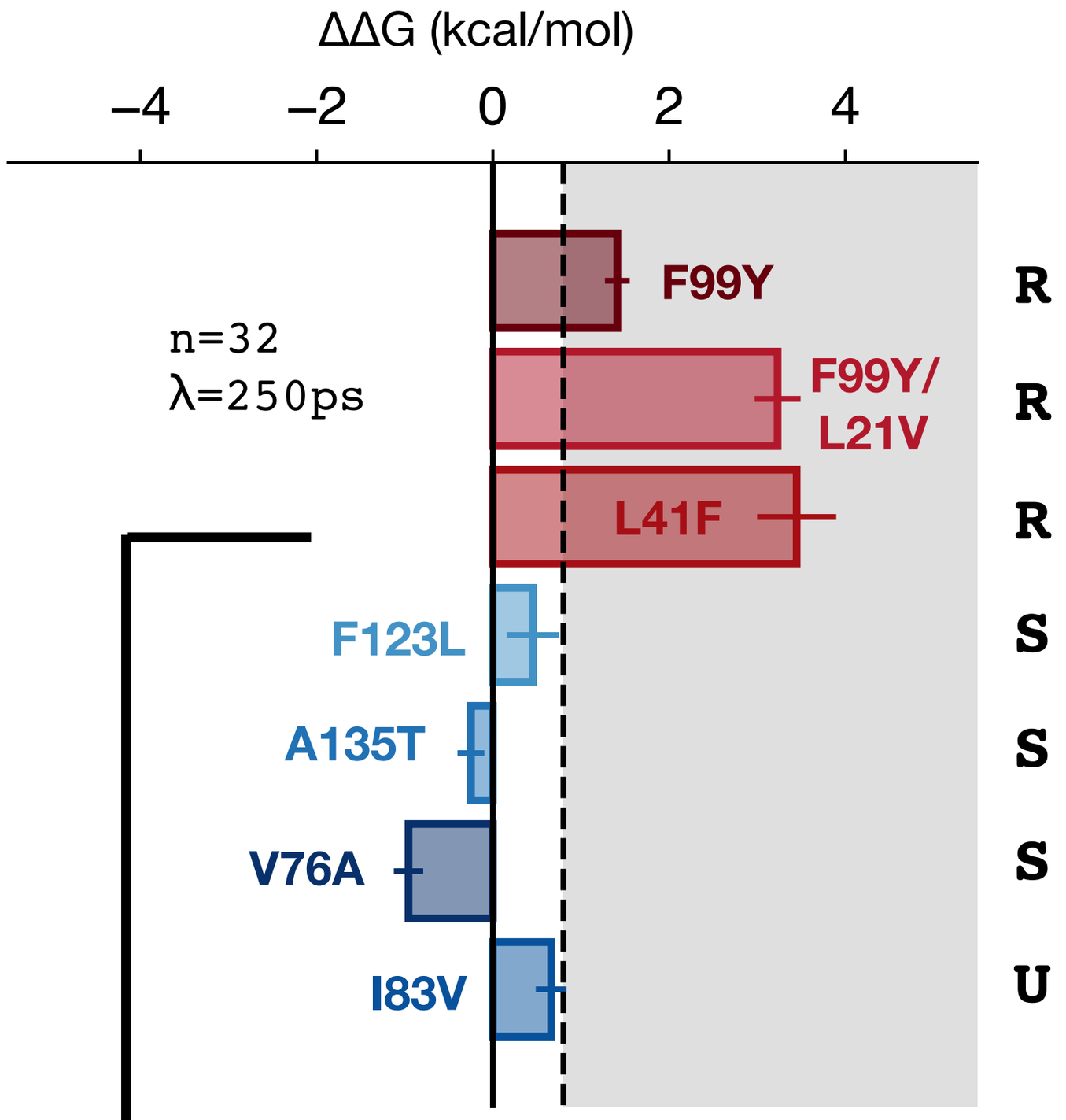
NOTE The overall VMD is expressed as a percentage.

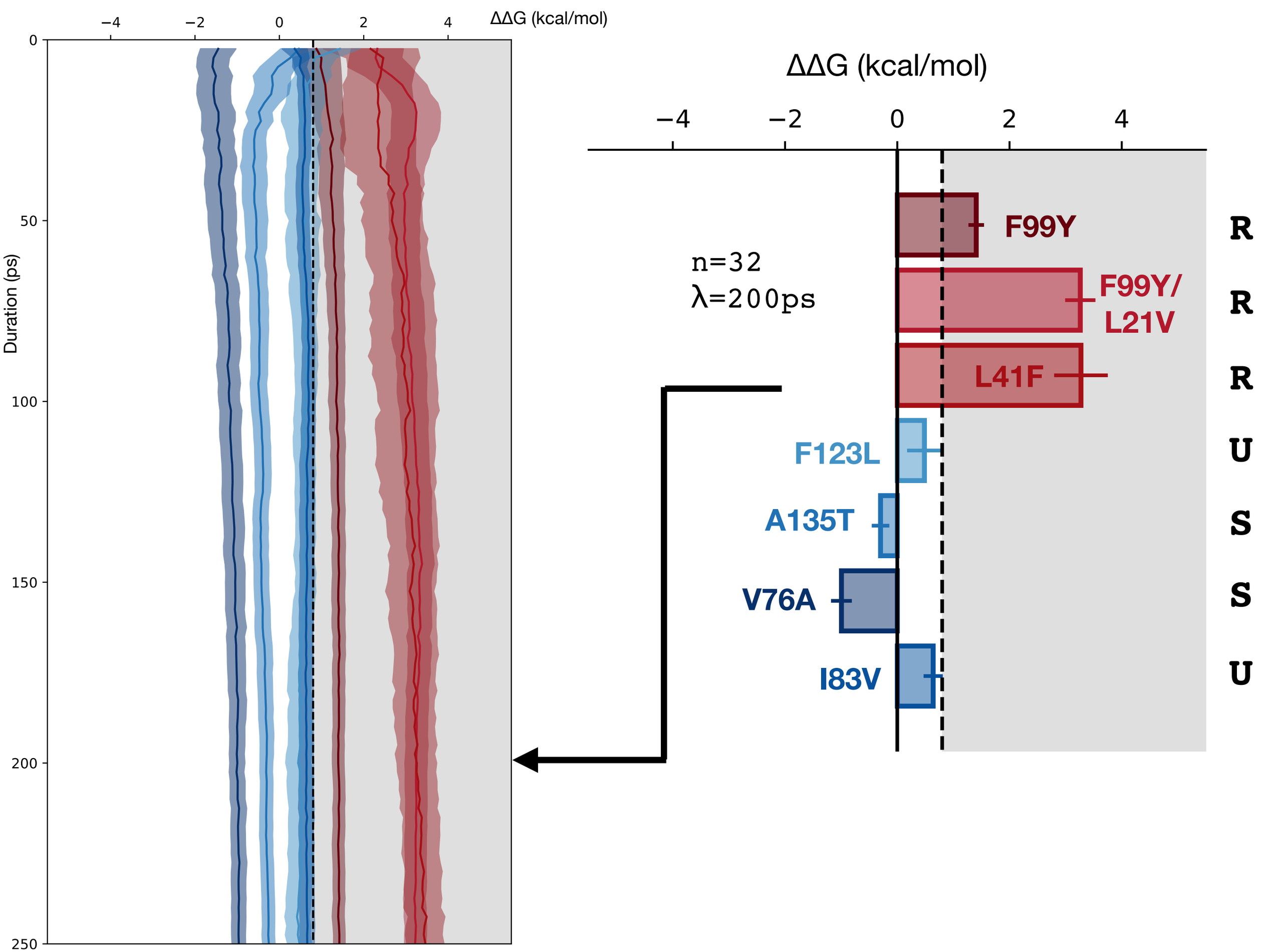
BP AST devices should have an overall CA $\geqslant$ 90 % when compared to the reference method result(s) and have overall VMD and MD $\leqslant$ 3 % each. Analysis should be made of MD and VMD to determine whether particular species are affected and require limitations for use of the device with that bacterial species and particular antimicrobial agents.

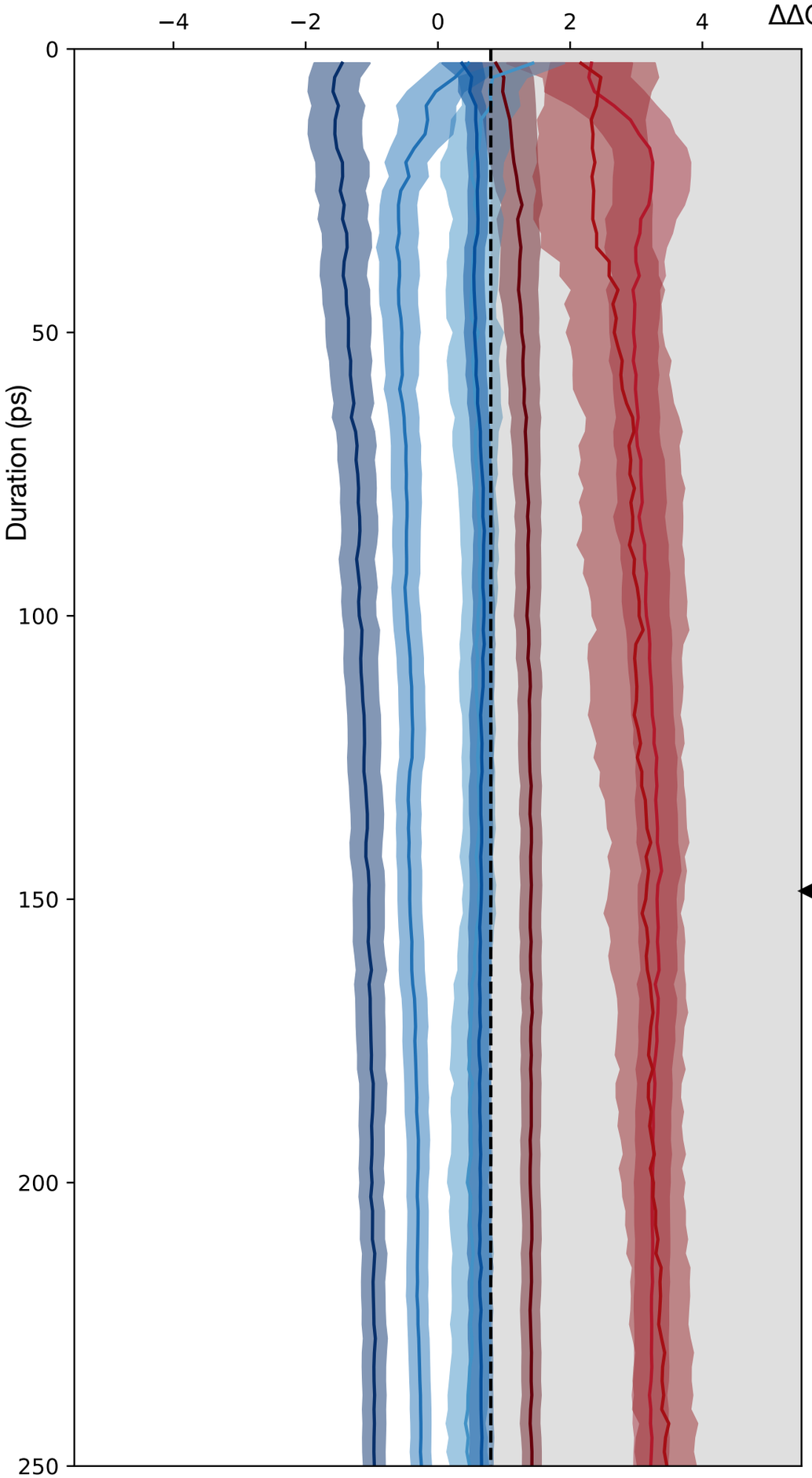

ΔΔG_TMP (kcal/mol)
-4
-2
0
2
4
Susceptible
Resistant
F99Y
F99Y/L21V
L41F
F123L
A135T
V76A
I83V
R1: ΔΔG_TMP ≥ 0.8 kcal/mol
TMP
0.0015
0.038
1
26
661
x K_i (M)
n=32
λ=250ps

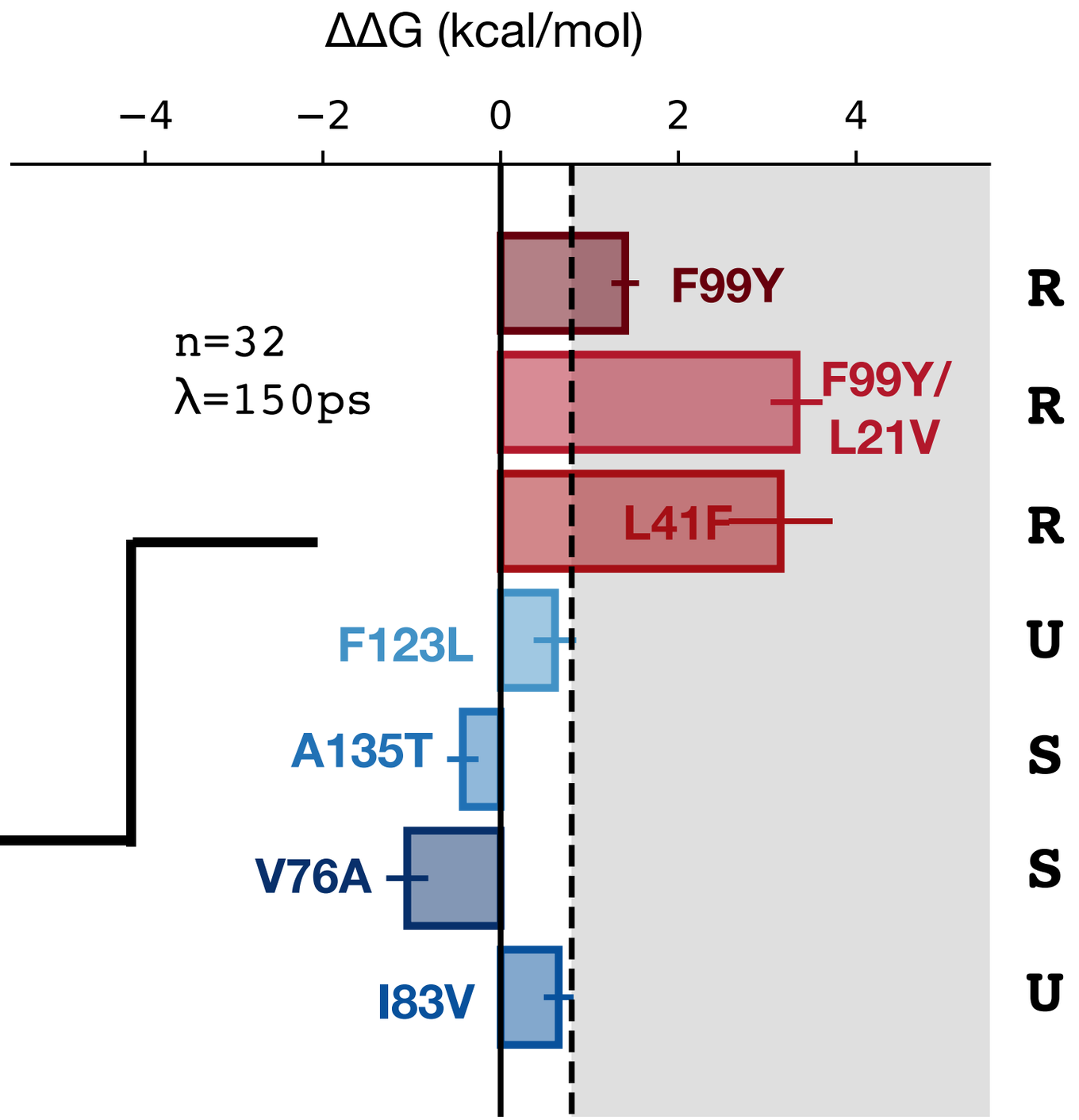

ΔΔG (kcal/mol)
–4
–2
0
2
4
F99Y
F99Y/L21V
L41F
F123L
A135T
V76A
I83V
n=32
λ=250ps

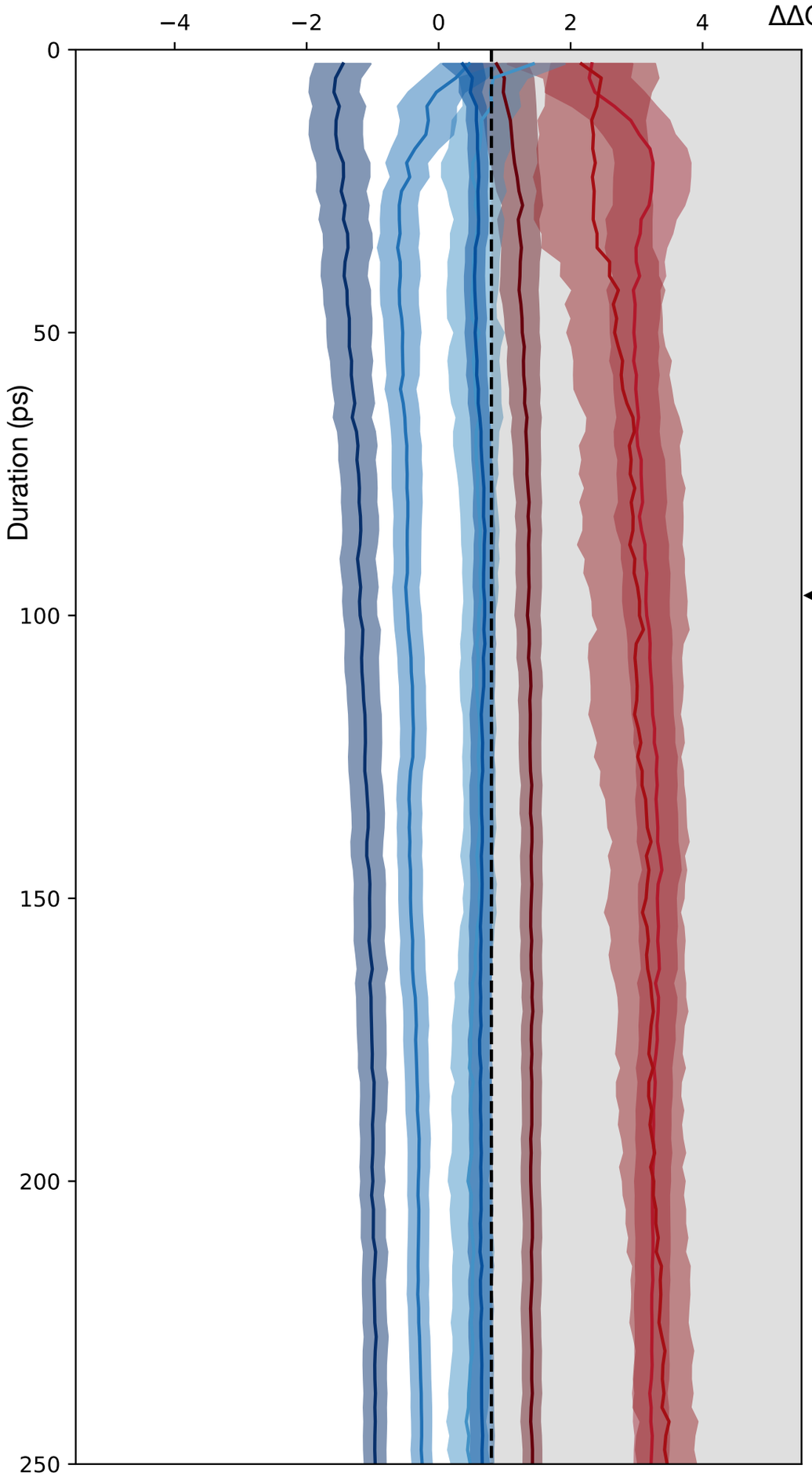

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

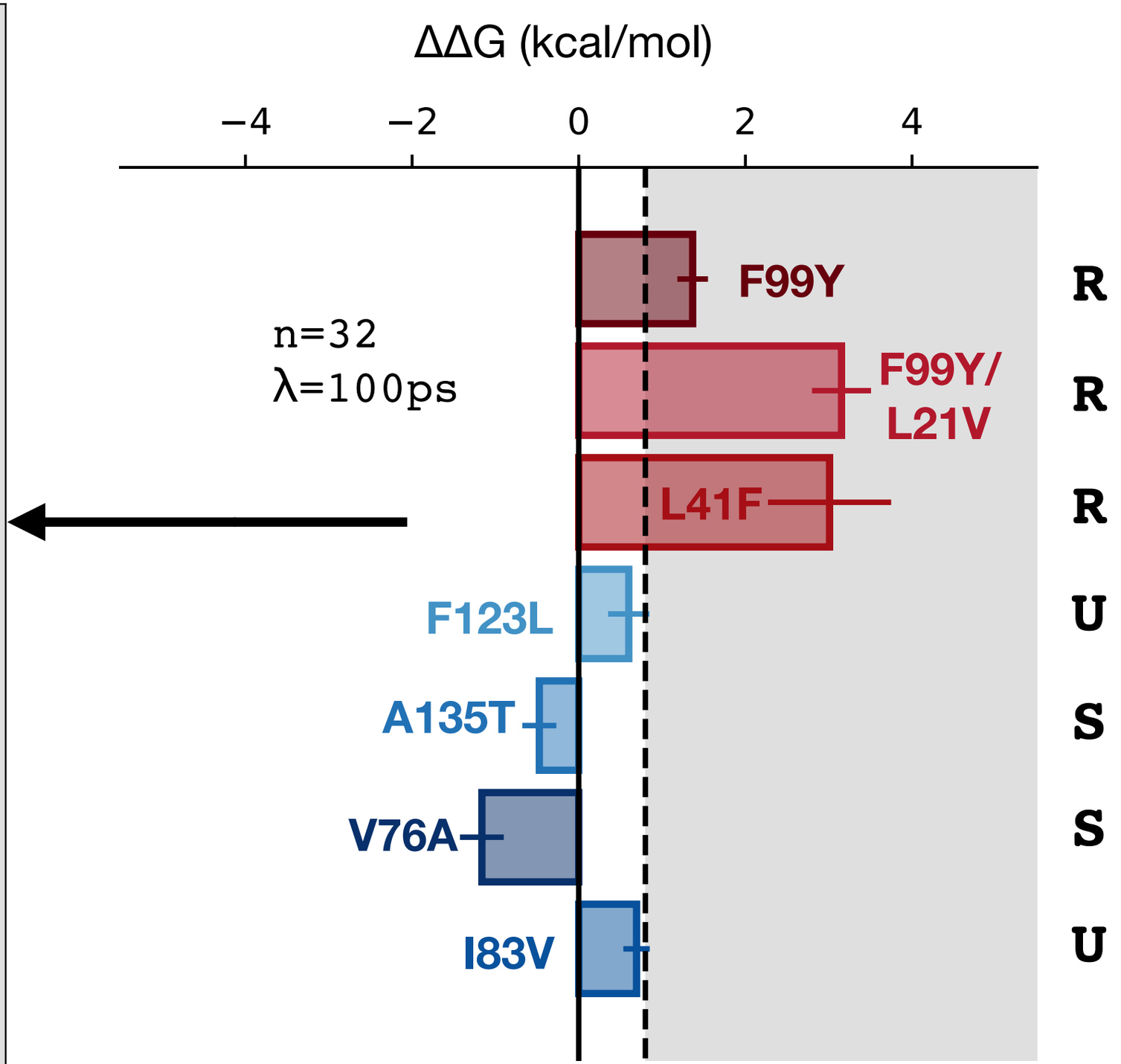

ΔΔG (kcal/mol)
−4
−2
0
2
4
n=32
λ=250ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
S
S
S
U

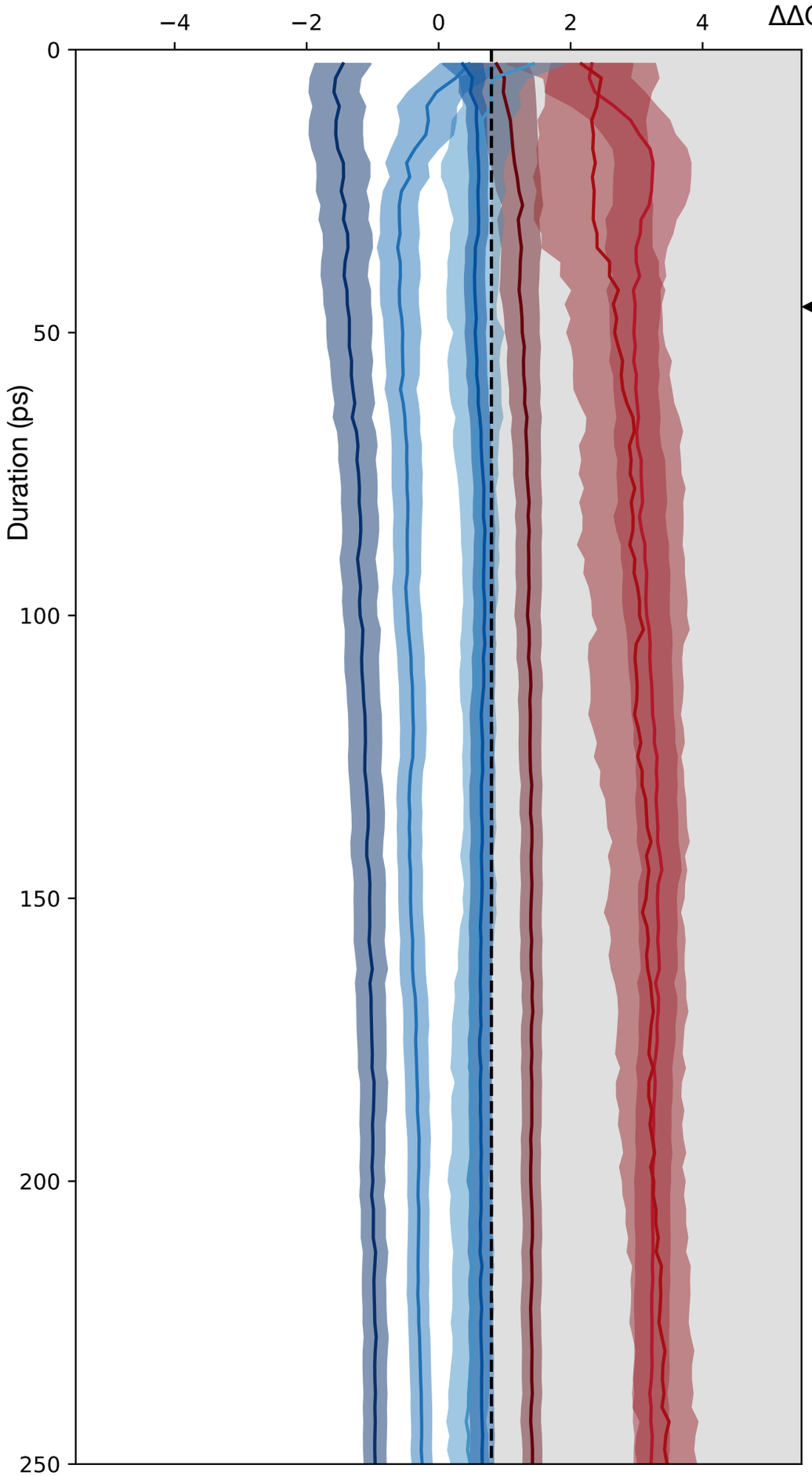

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250
ΔΔG (kcal/mol)
−4
−2
0
2
4
n=32
λ=200ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
U

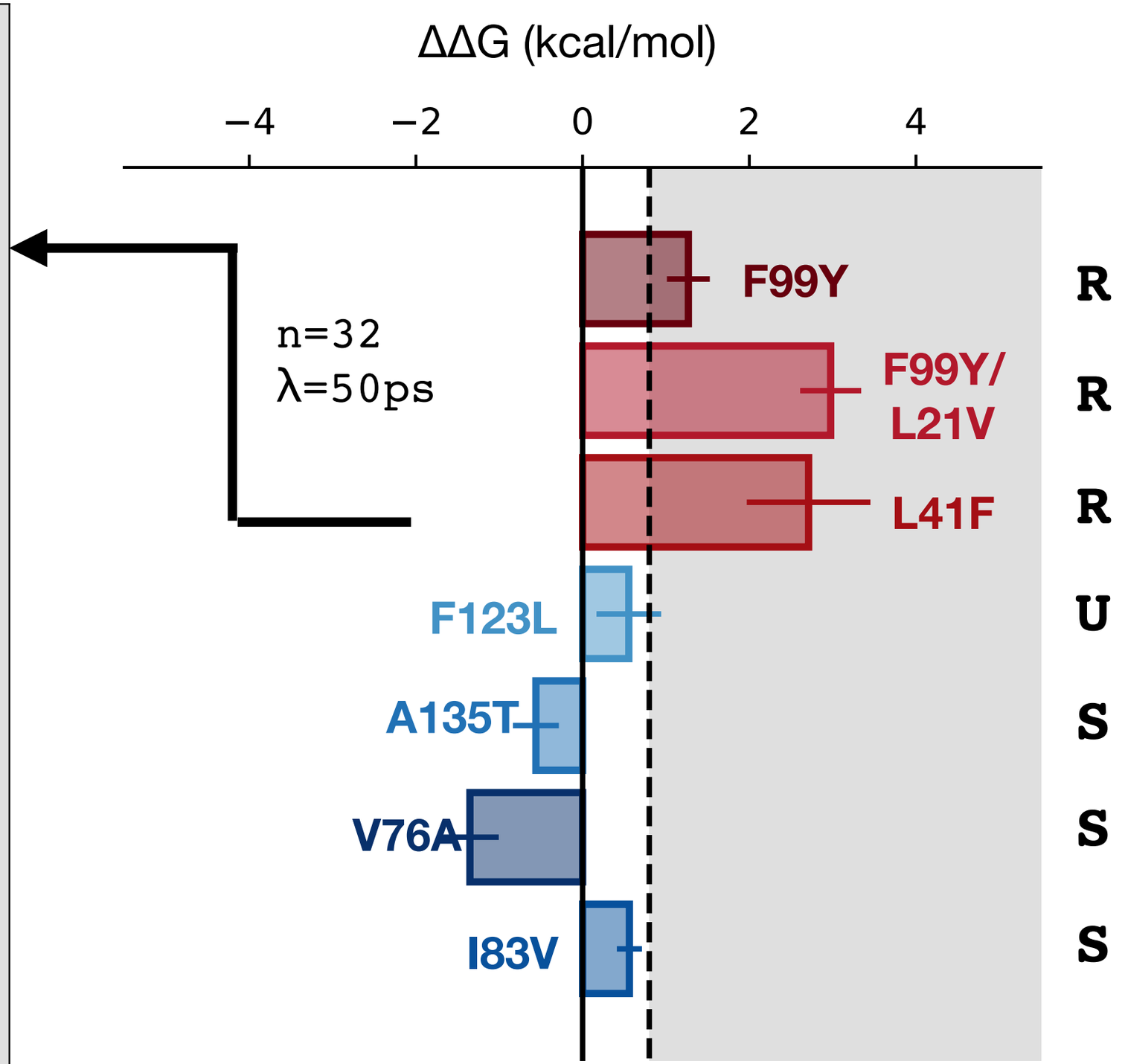

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

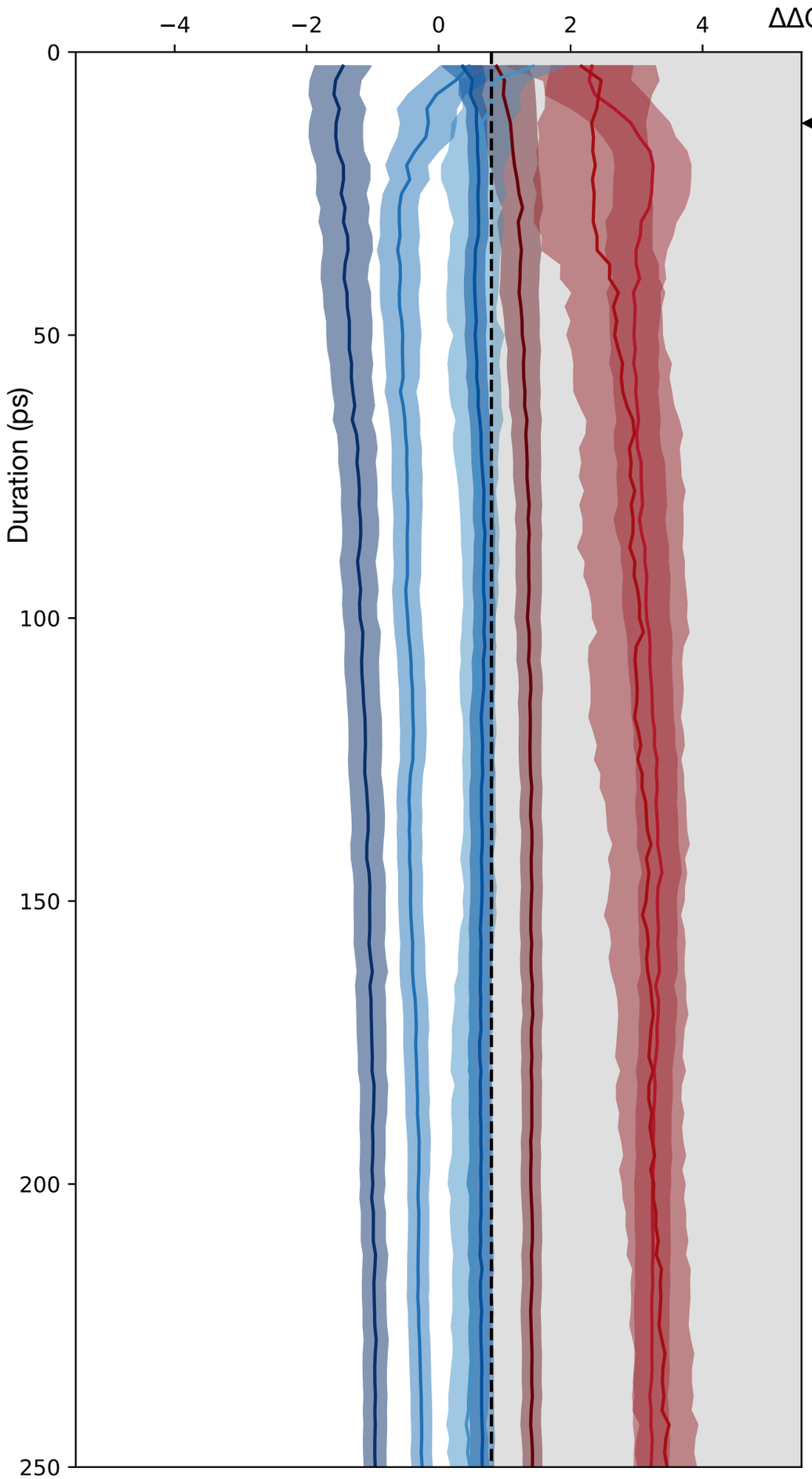

ΔΔG (kcal/mol)
−4
−2
0
2
4
n=32
λ=150ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
U

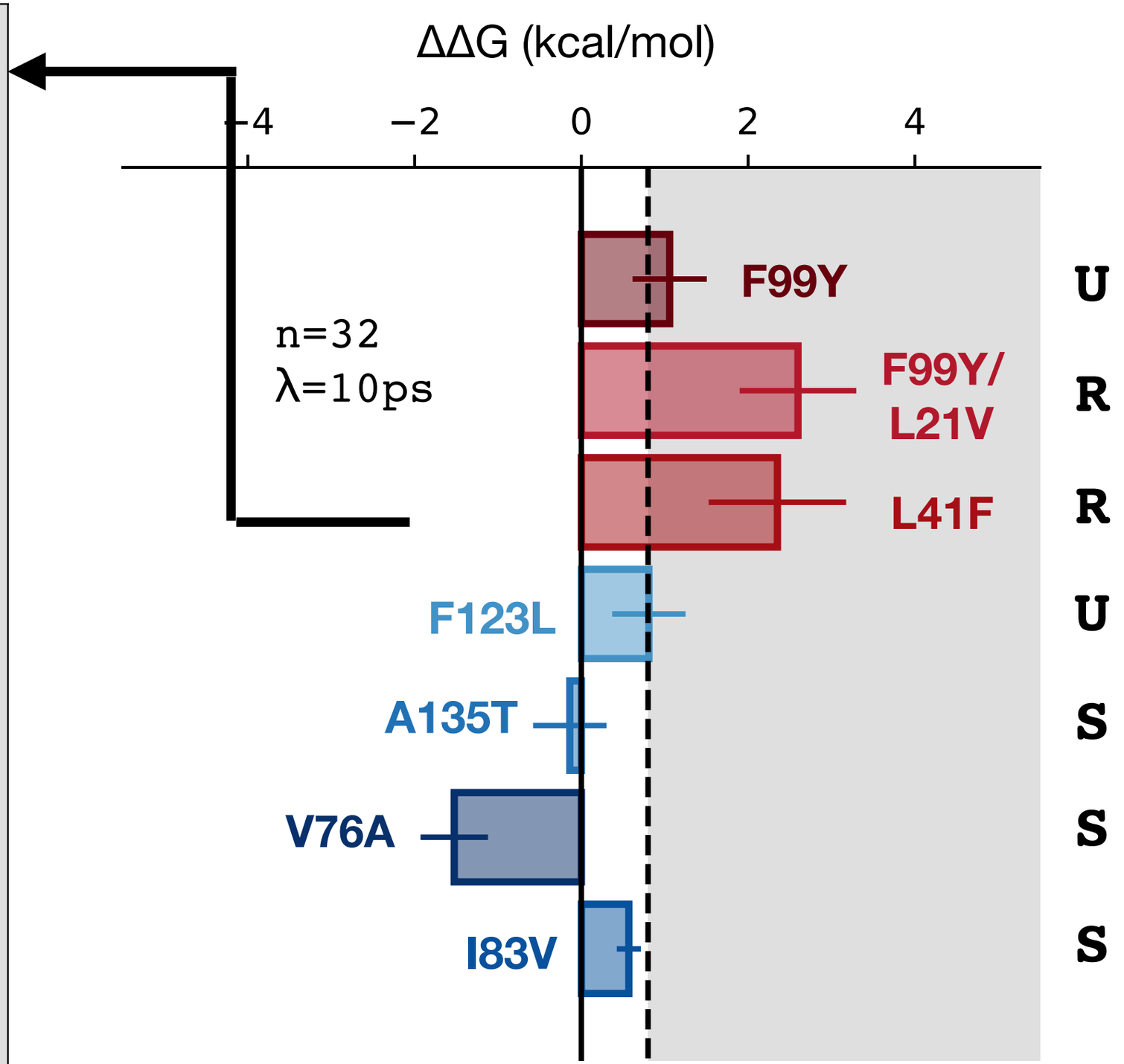

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

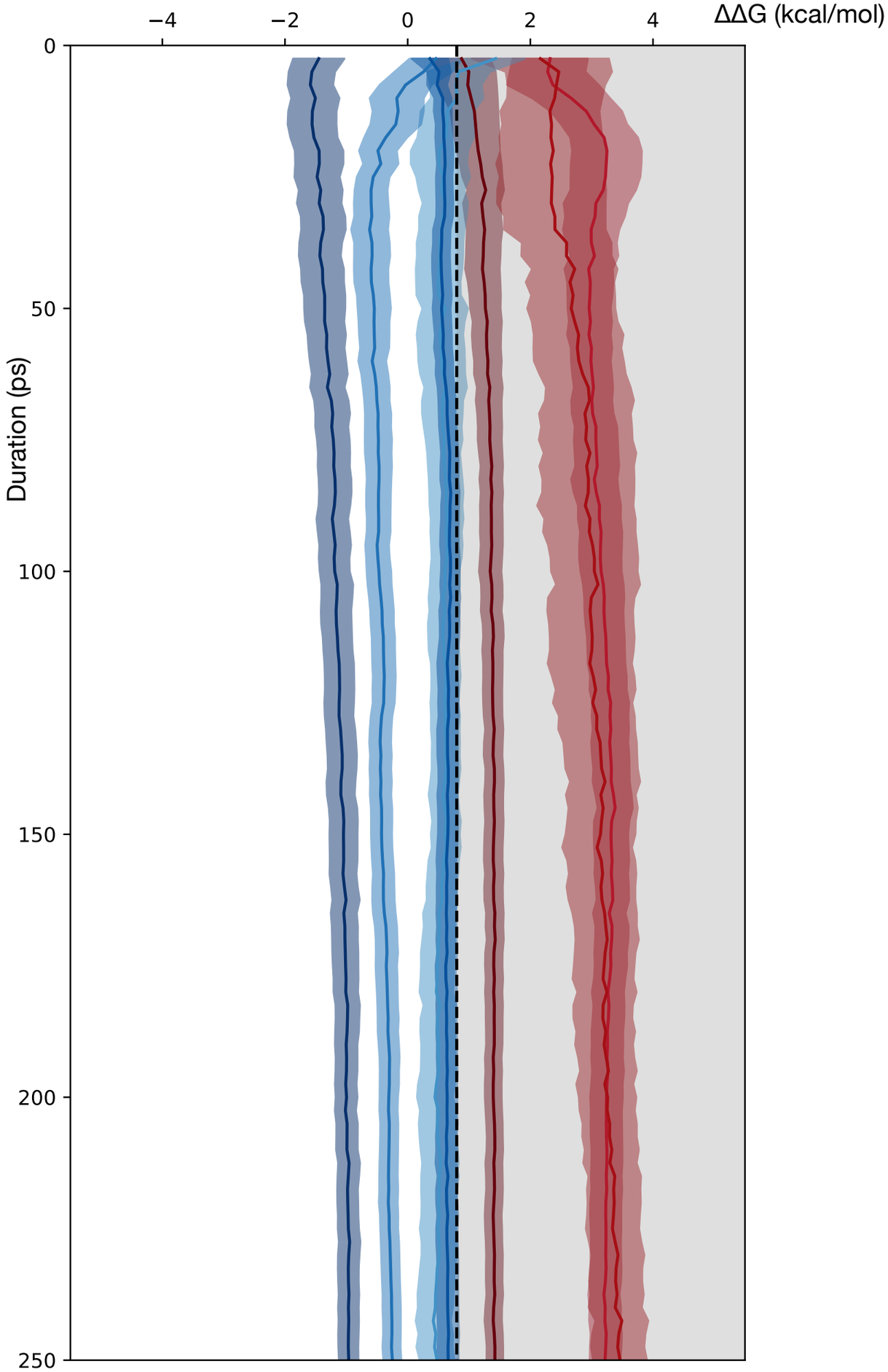

ΔΔG (kcal/mol)
−4
−2
0
2
4
n=32
λ=100ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
U

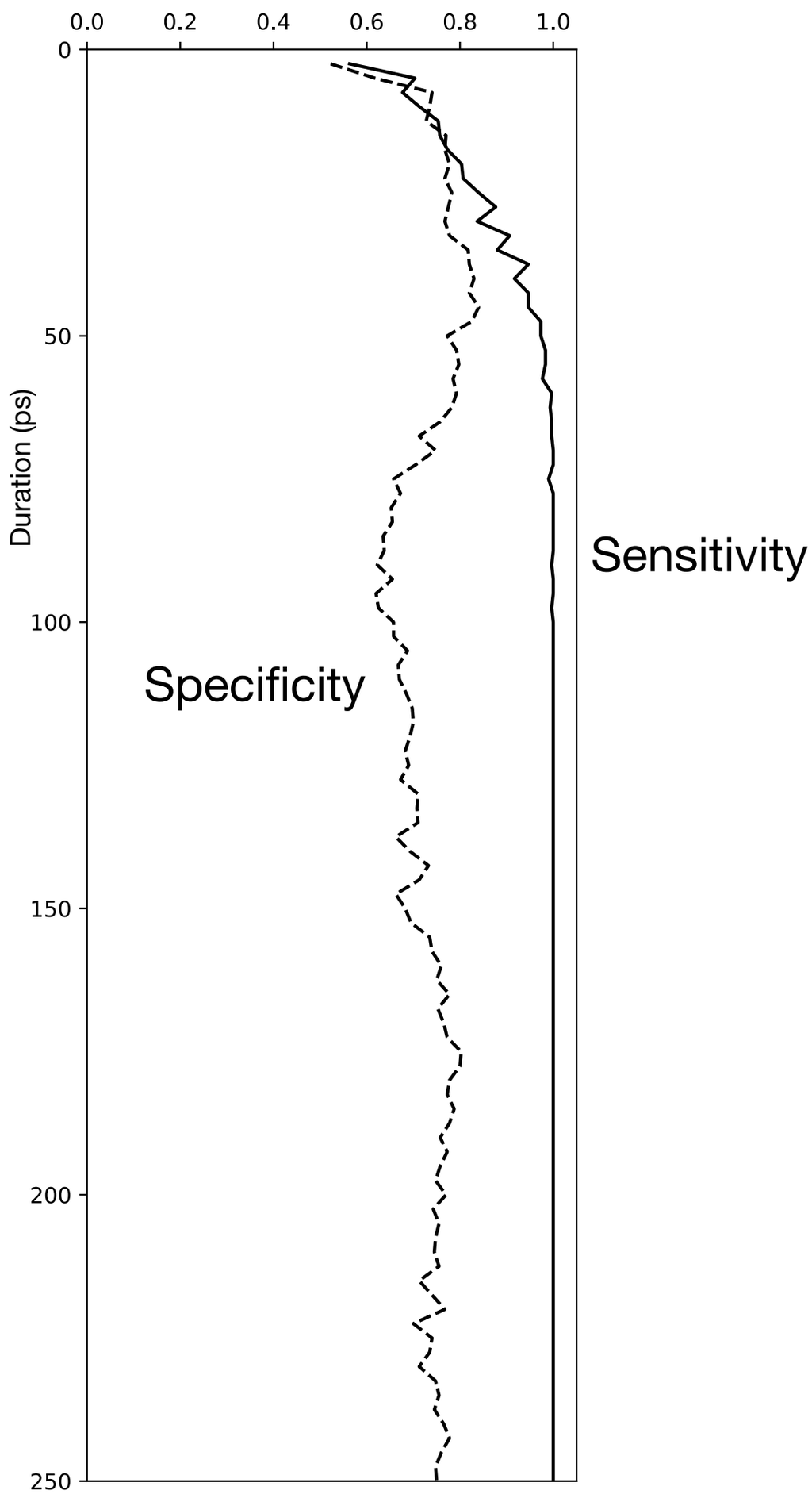

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

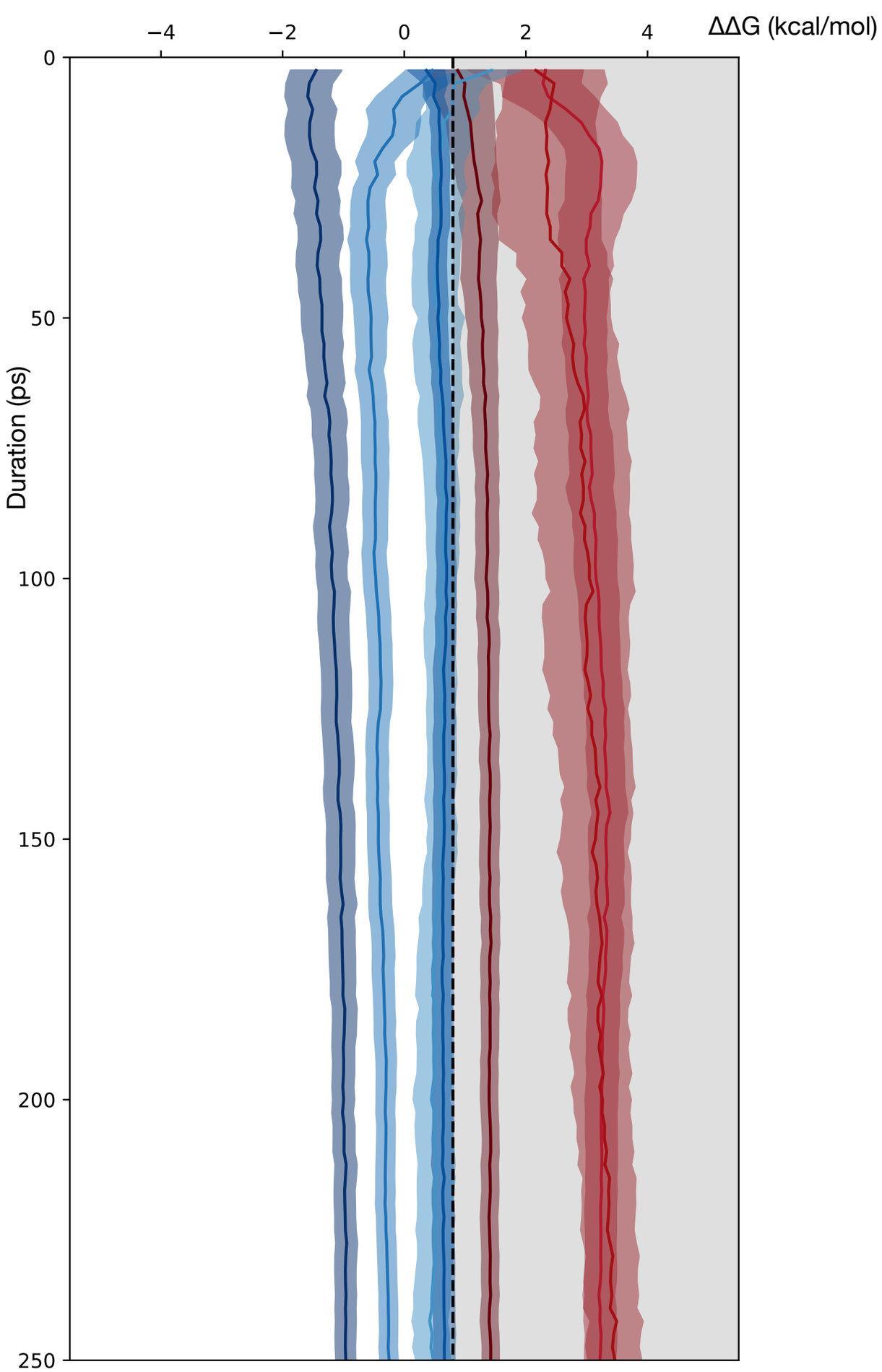

ΔΔG (kcal/mol)
−4
−2
0
2
4
n=32
λ=50ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
S

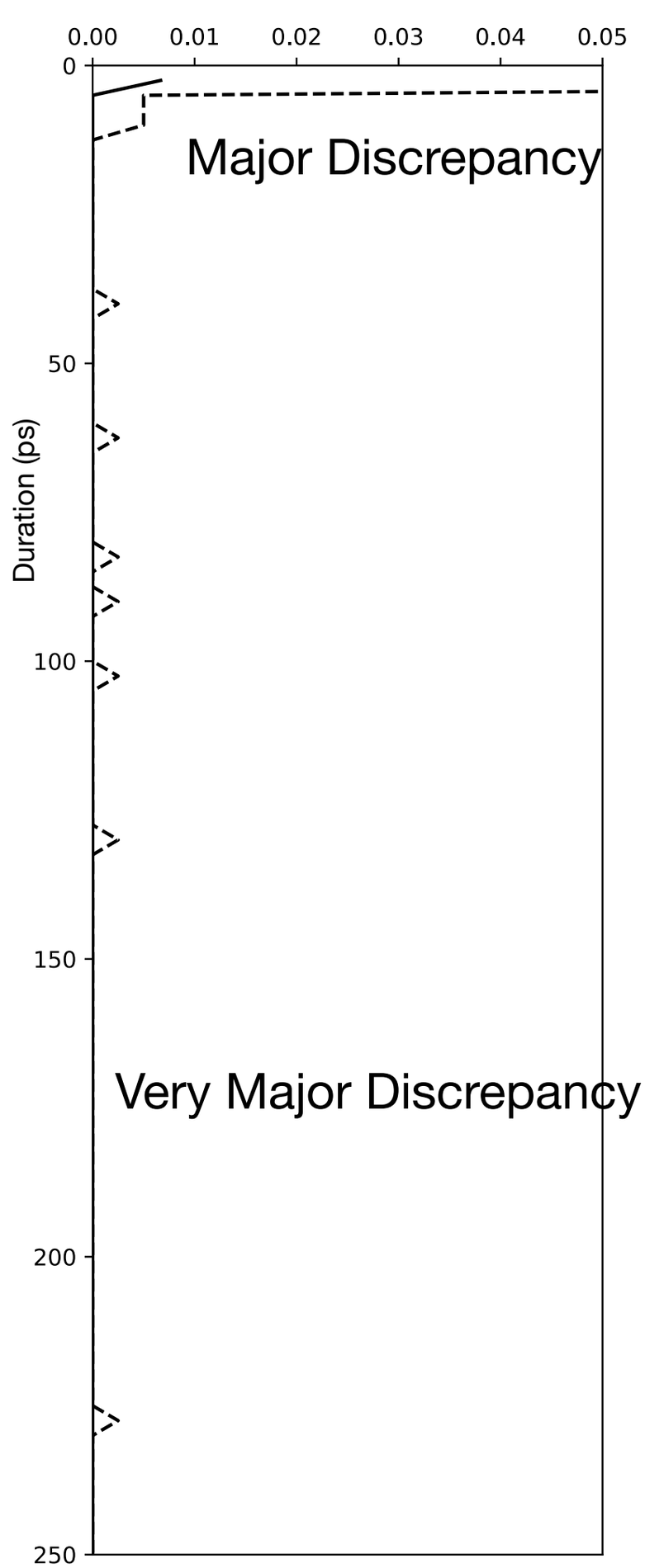

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

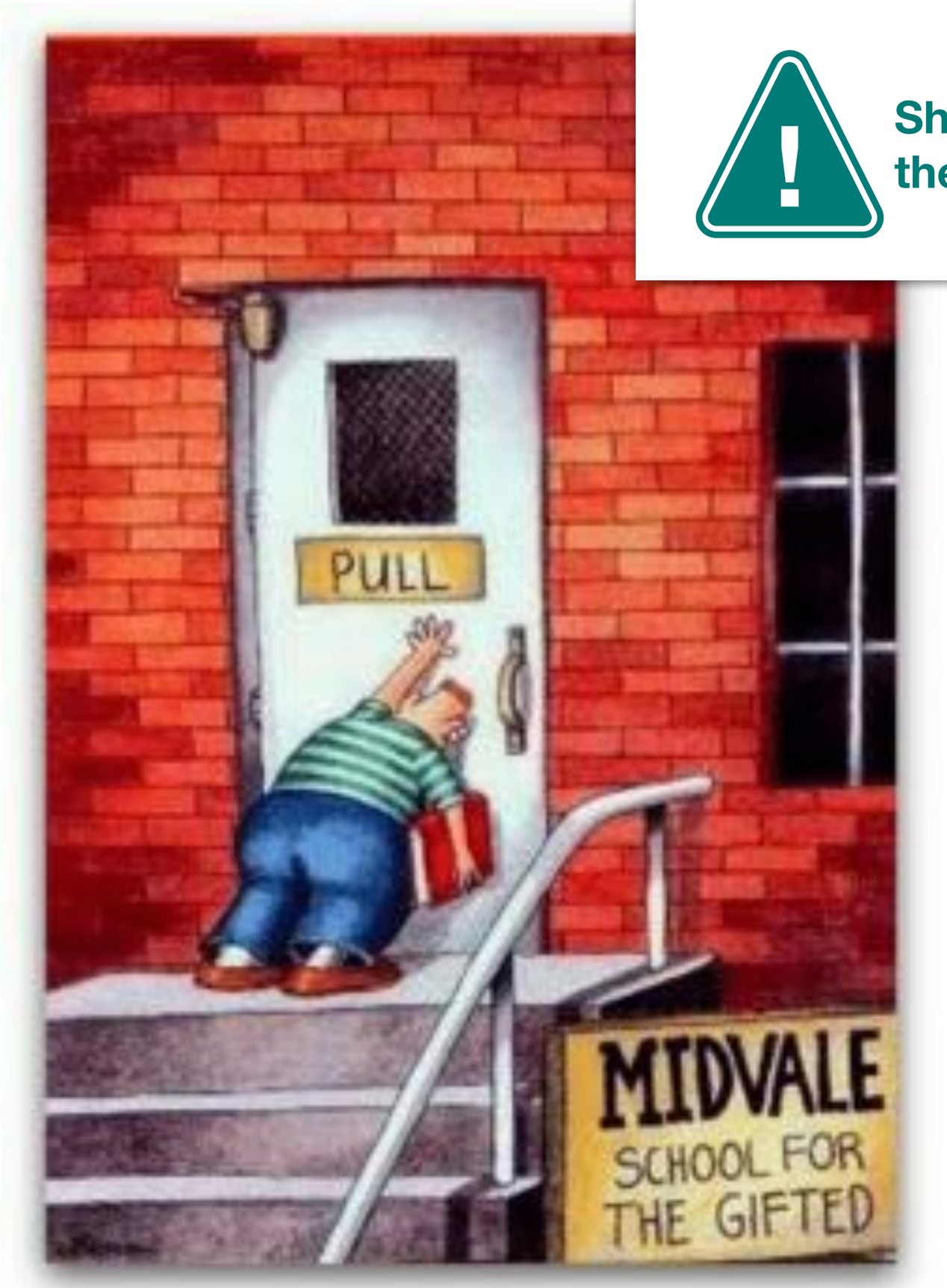

ΔΔG (kcal/mol)
−4
−2
0
2
4
n=32
λ=10ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
U
R
R
U
S
S
S


ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250


0.0
0.2
0.4
0.6
0.8
1.0
Duration (ps)
0
50
100
150
200
250
Specificity
Sensitivity


ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

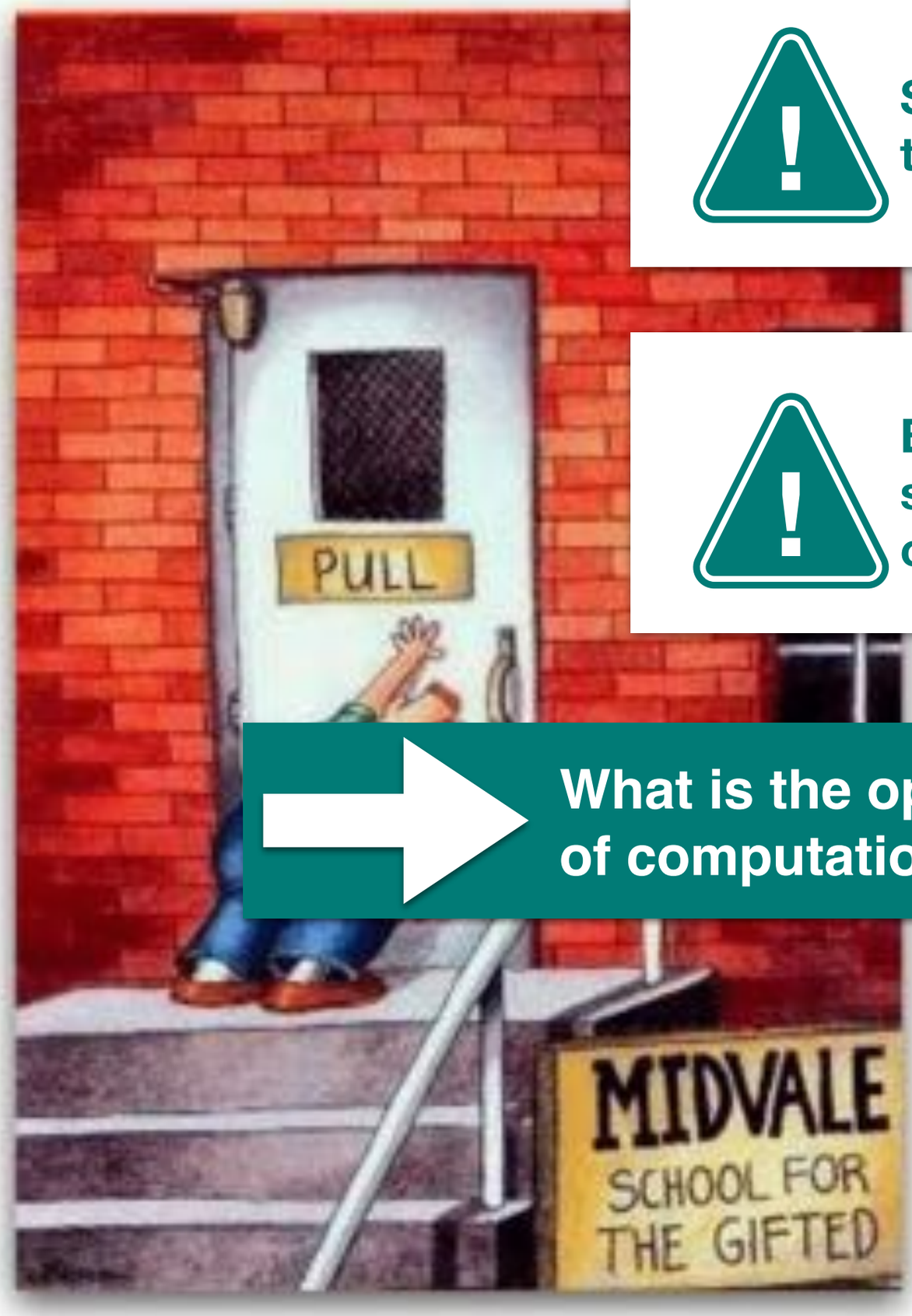

0.00
0.01
0.02
0.03
0.04
0.05
Duration (ps)
0
50
100
150
200
250
Major Discrepancy
Very Major Discrepancy



Shown that the shorter the λ simulations the larger the error.

Shown that the shorter the λ simulations the larger the error.

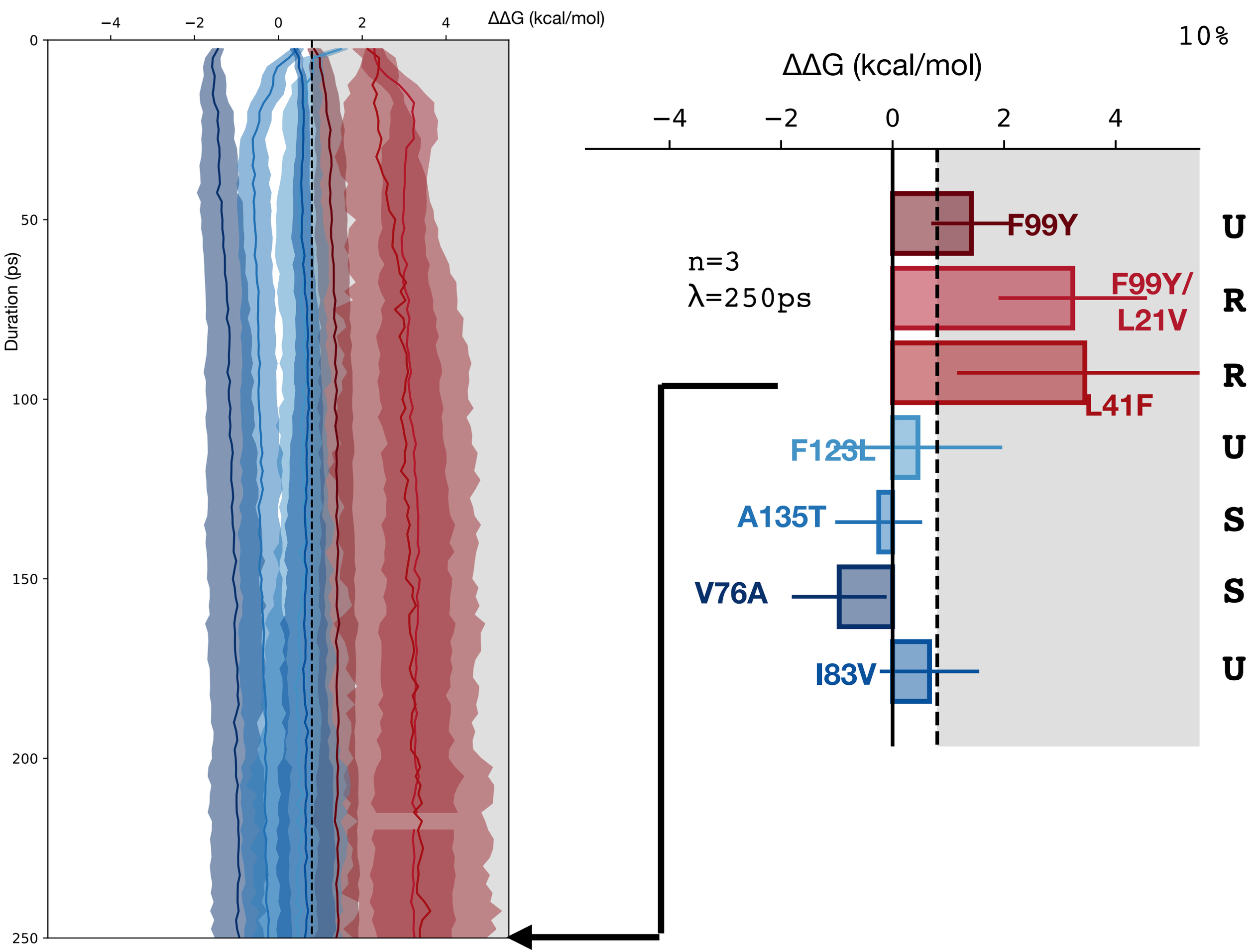
But it is surprising is how short a simulation will still give the correct clinical classification!

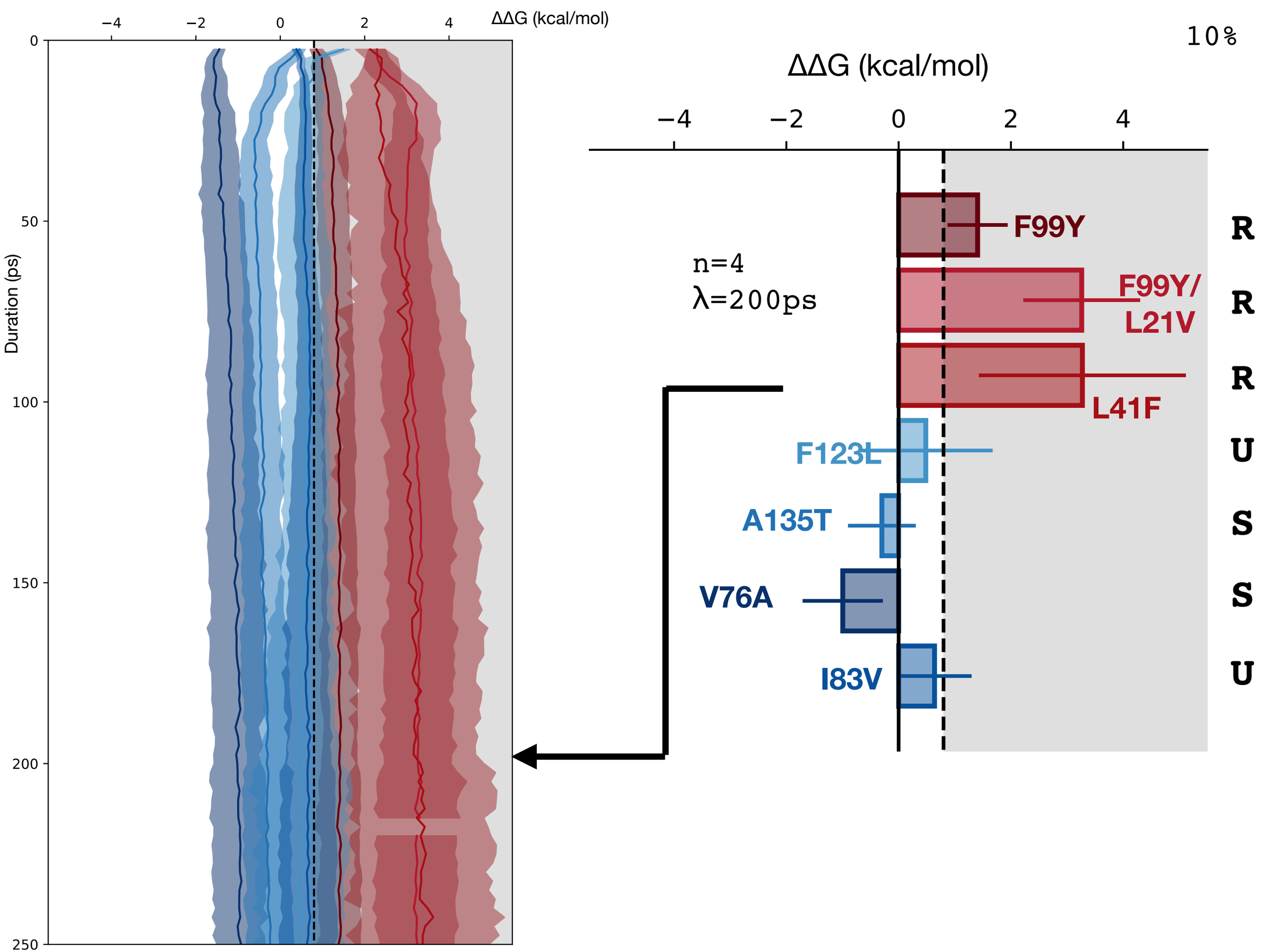


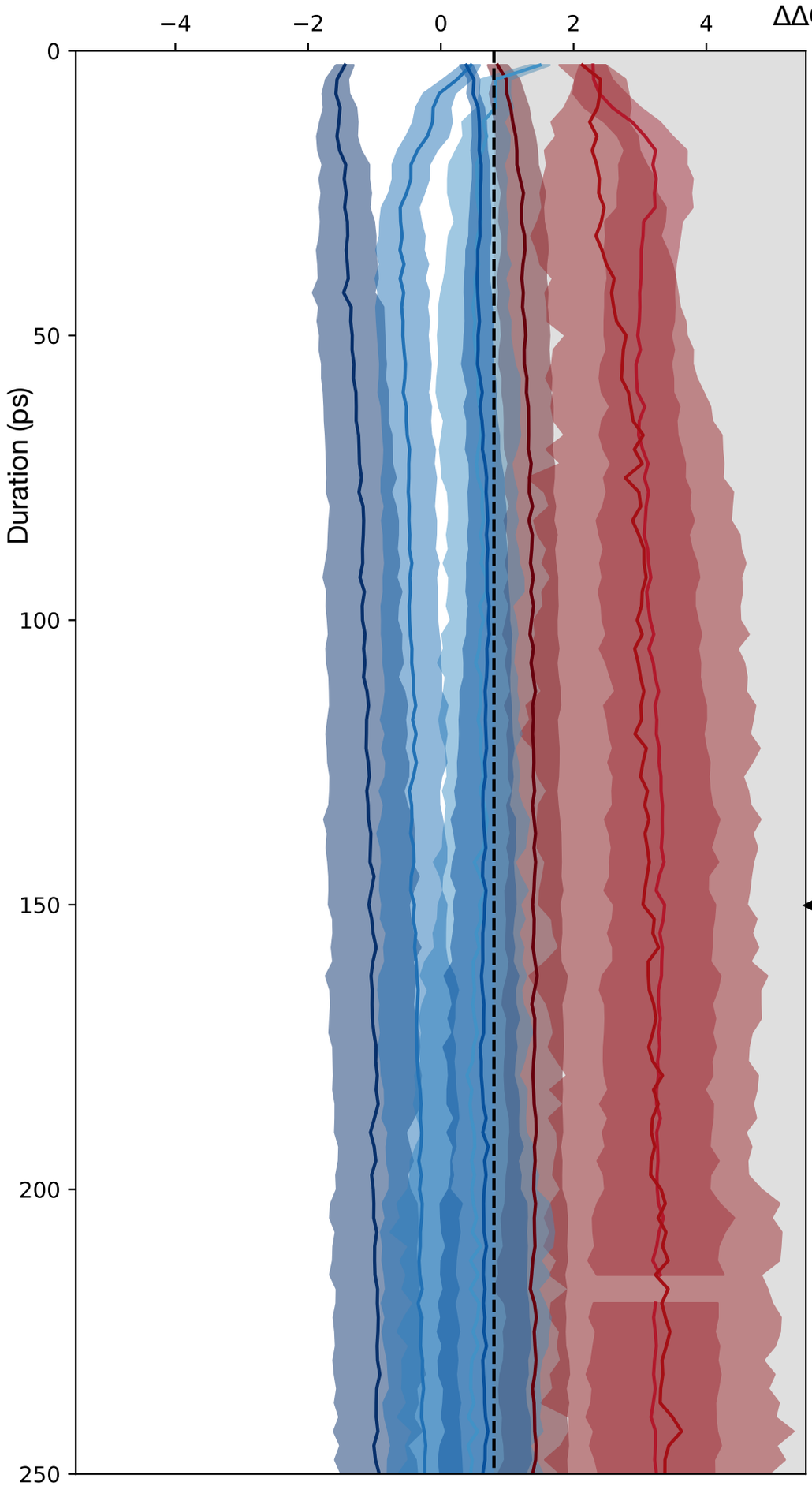
Shown that the shorter the λ simulations the larger the error.

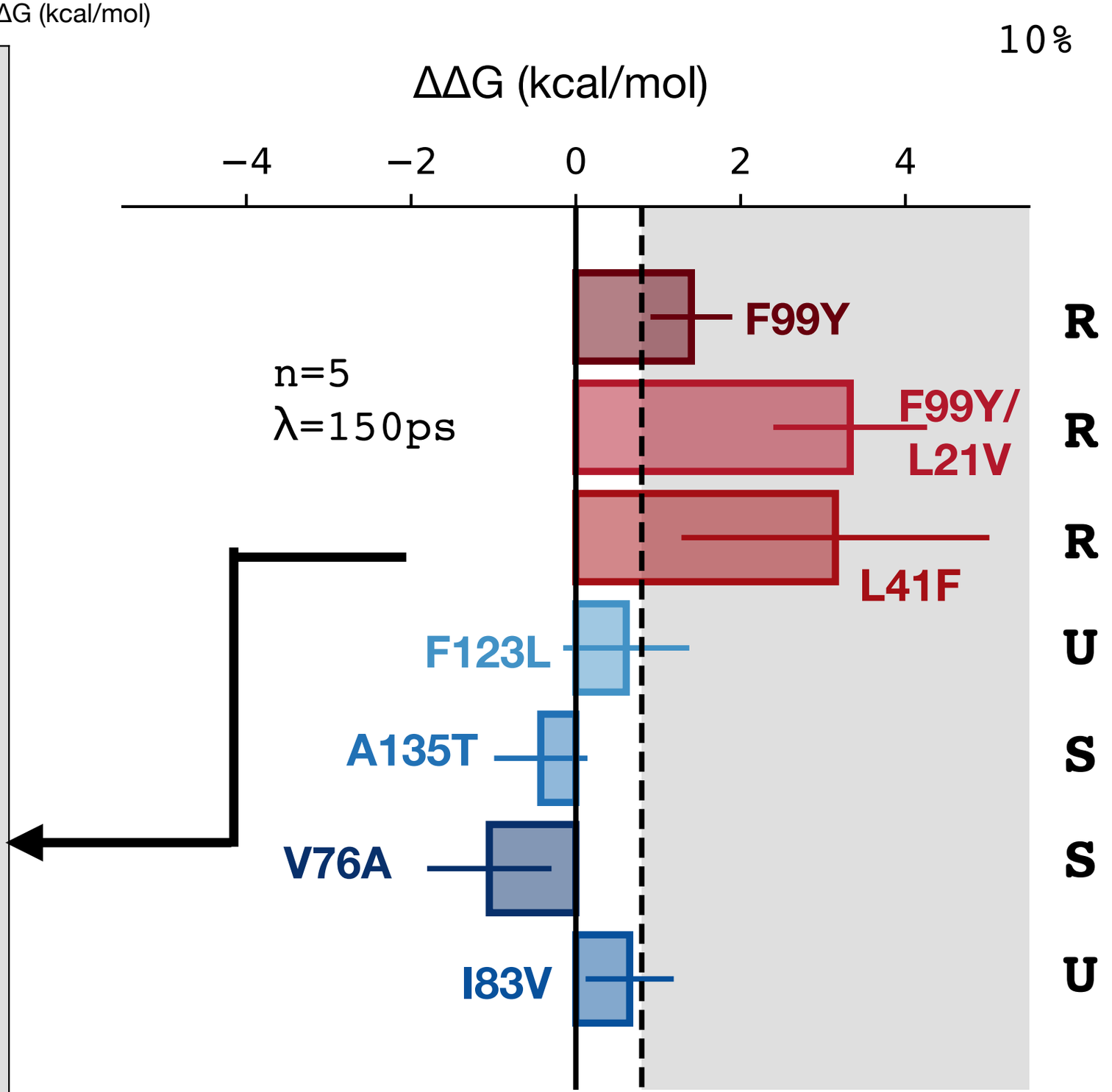
But it is surprising is how short a simulation will still give the correct clinical classification!

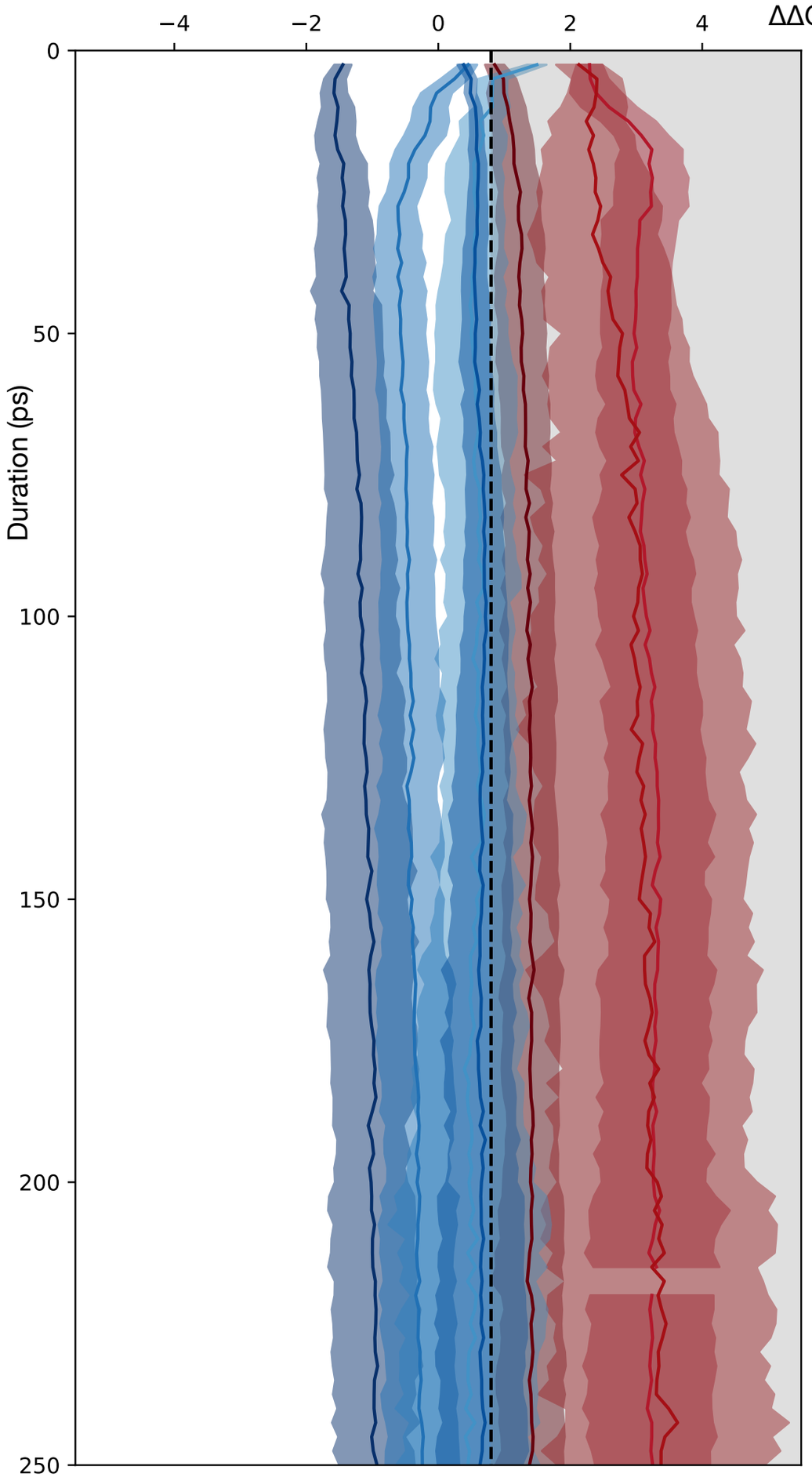
What is the optimal way to use a fixed amount of computational resource?

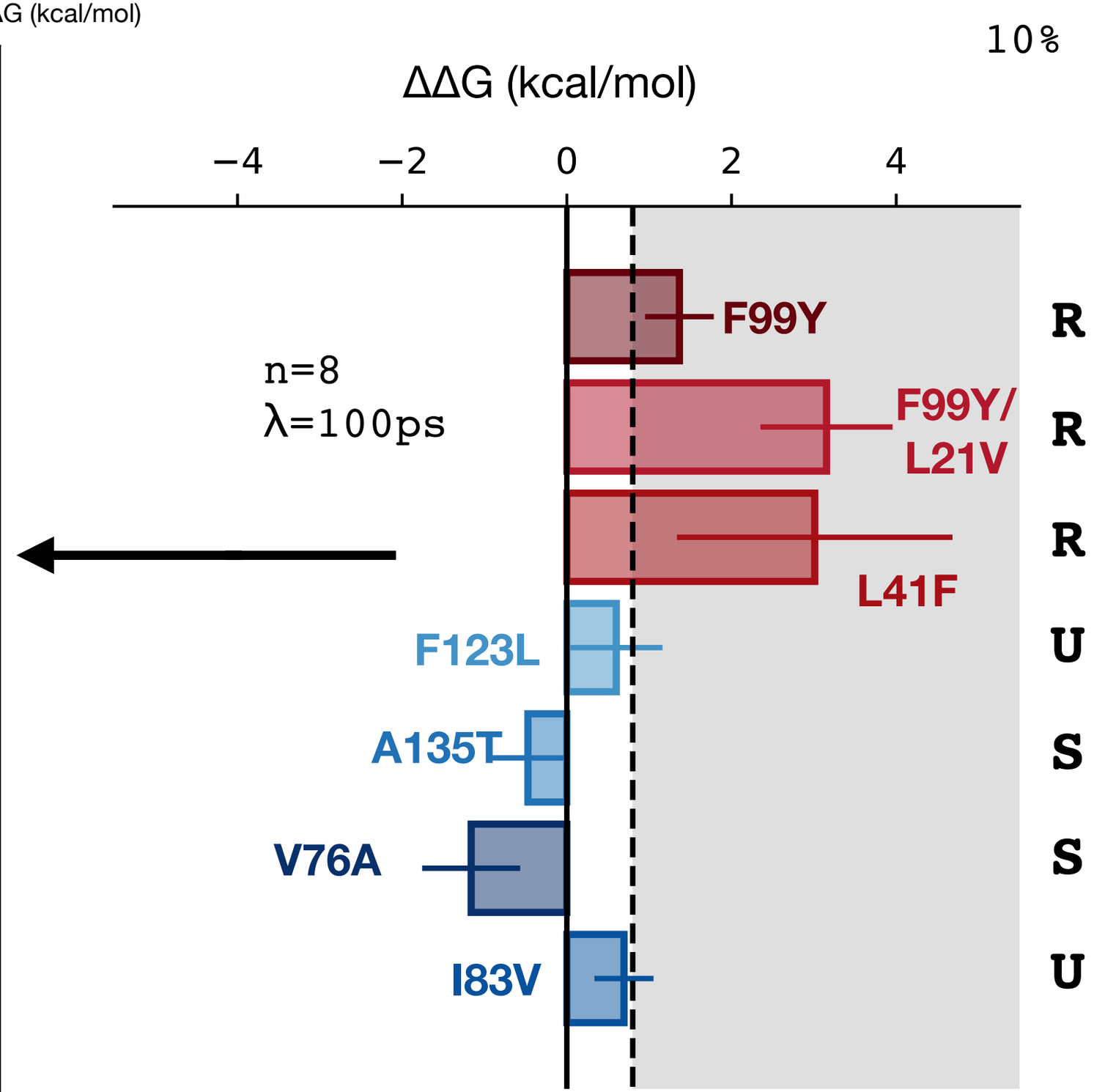

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250
10%
ΔΔG (kcal/mol)
−4
−2
0
2
4
n=3
λ=250ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
U
R
R
U
S
S
U

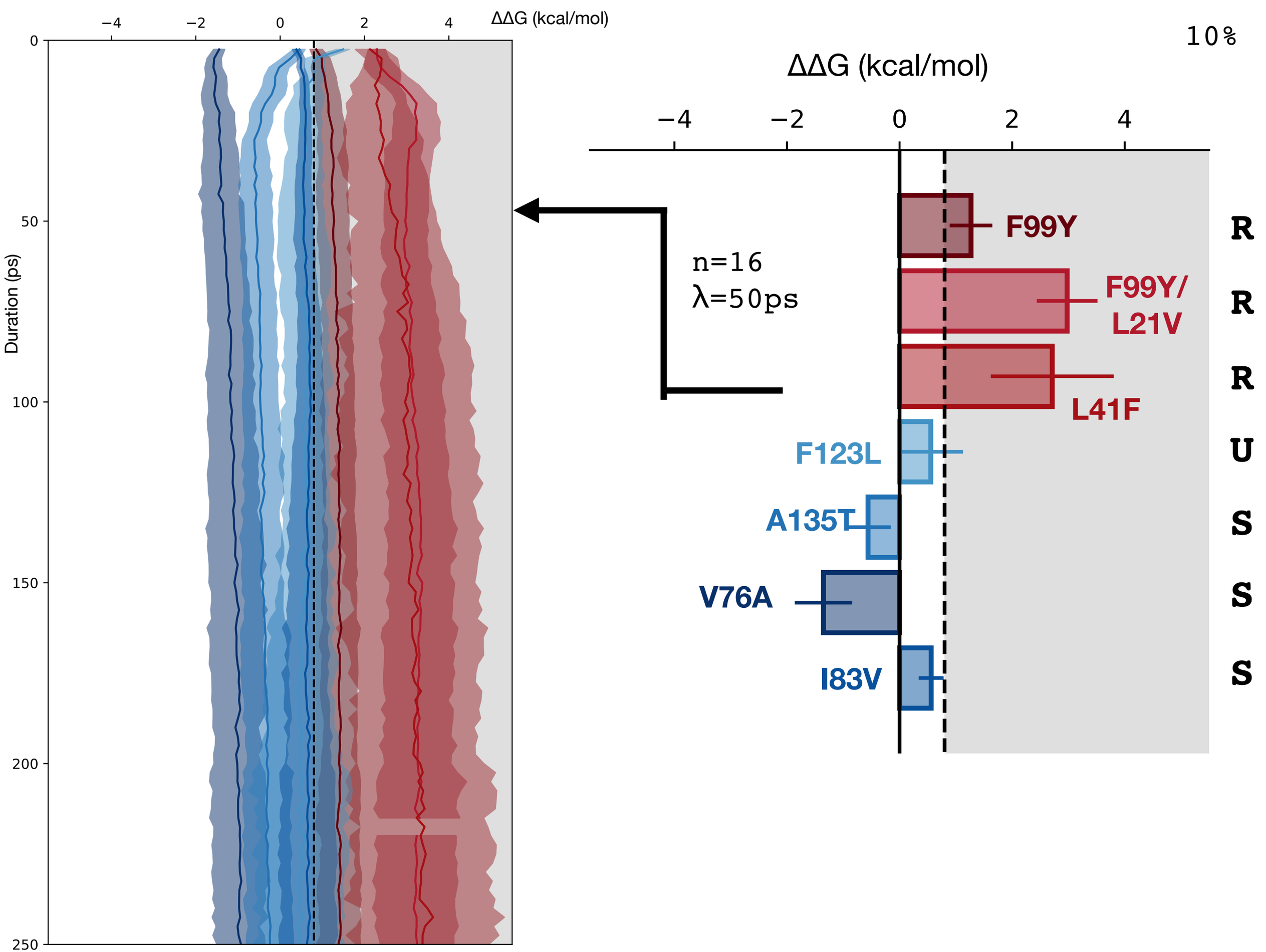

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250
10%
ΔΔG (kcal/mol)
−4
−2
0
2
4
n=4
λ=200ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
U

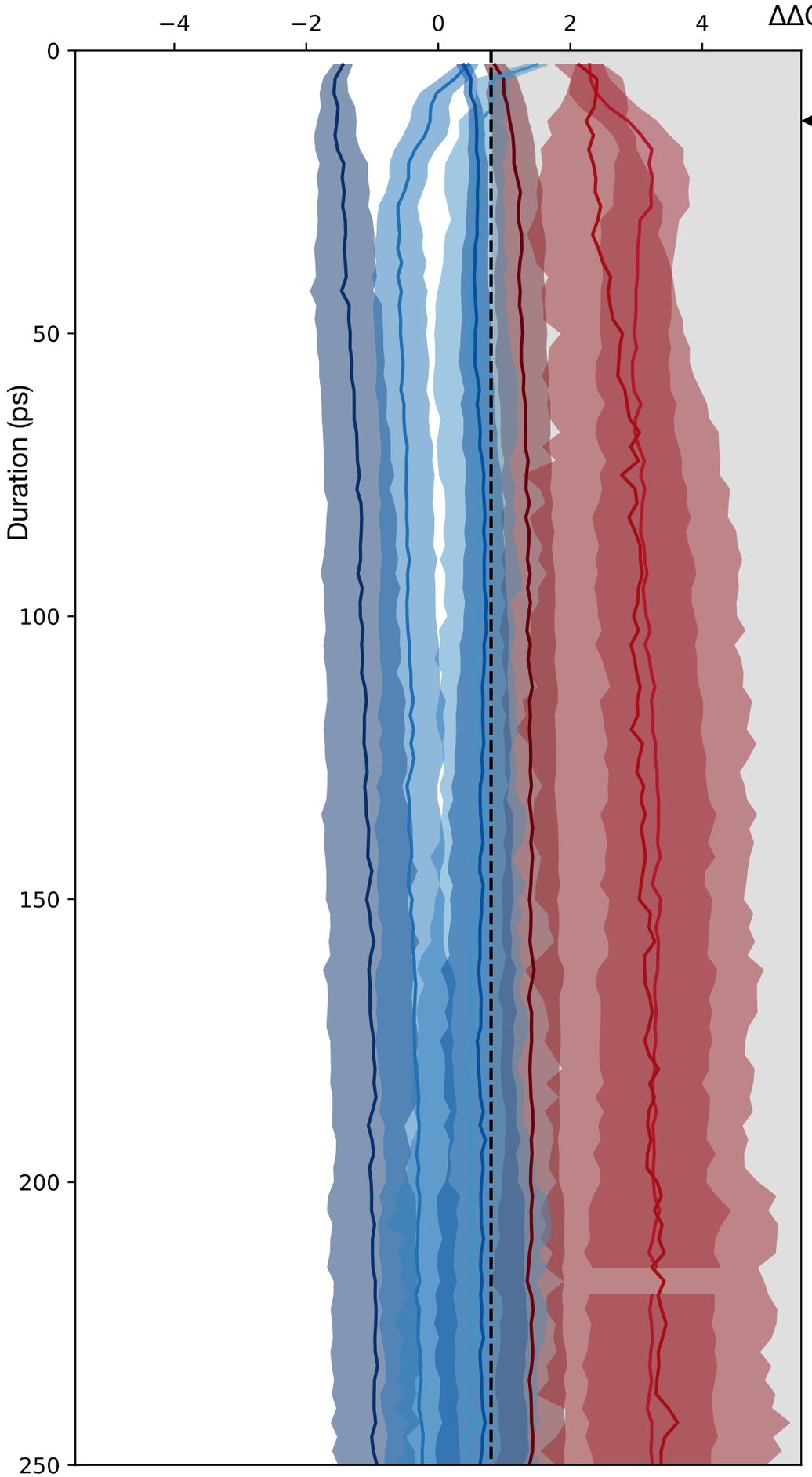

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

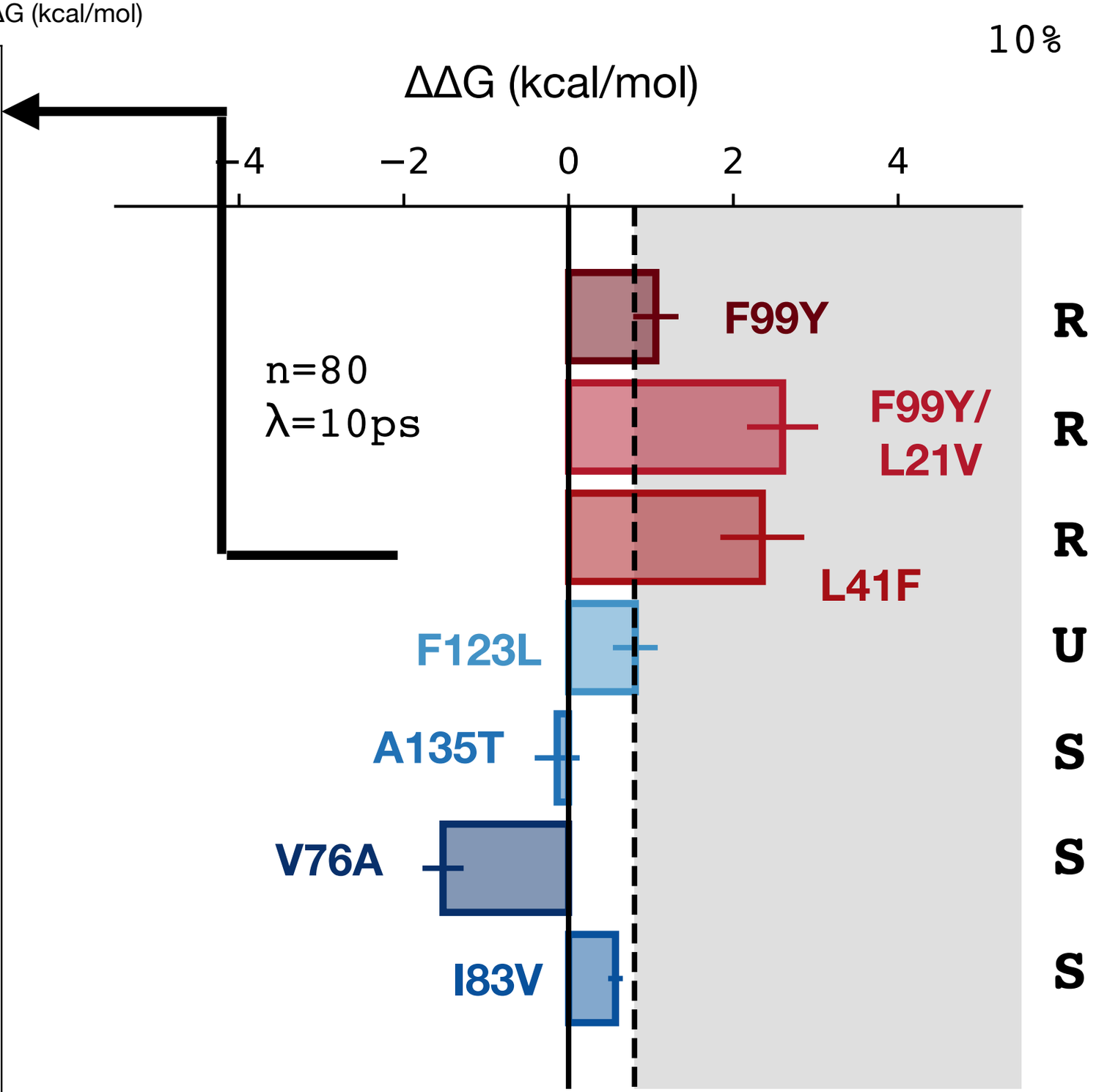

10%
ΔΔG (kcal/mol)
−4
−2
0
2
4
n=5
λ=150ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
U

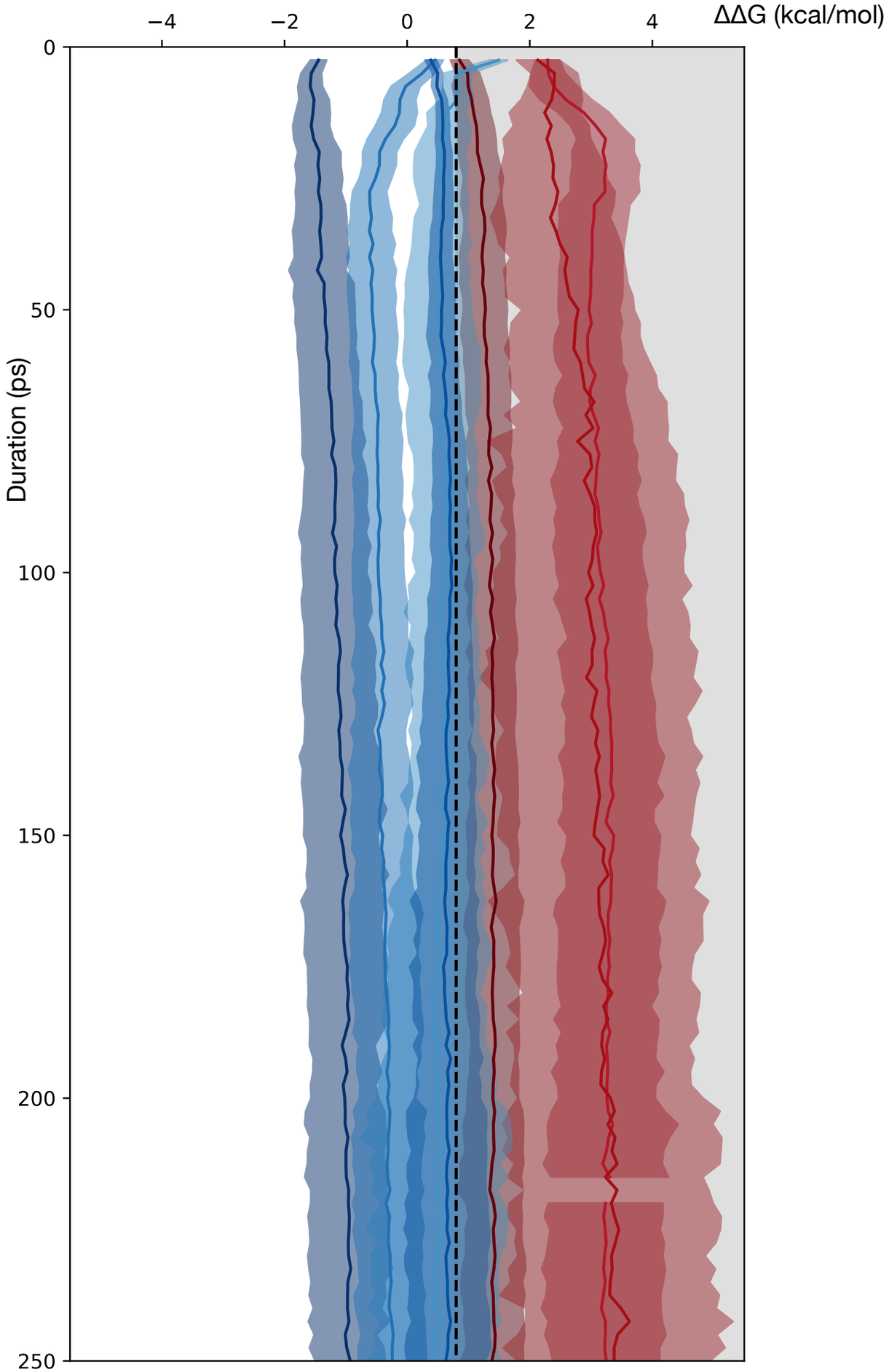

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

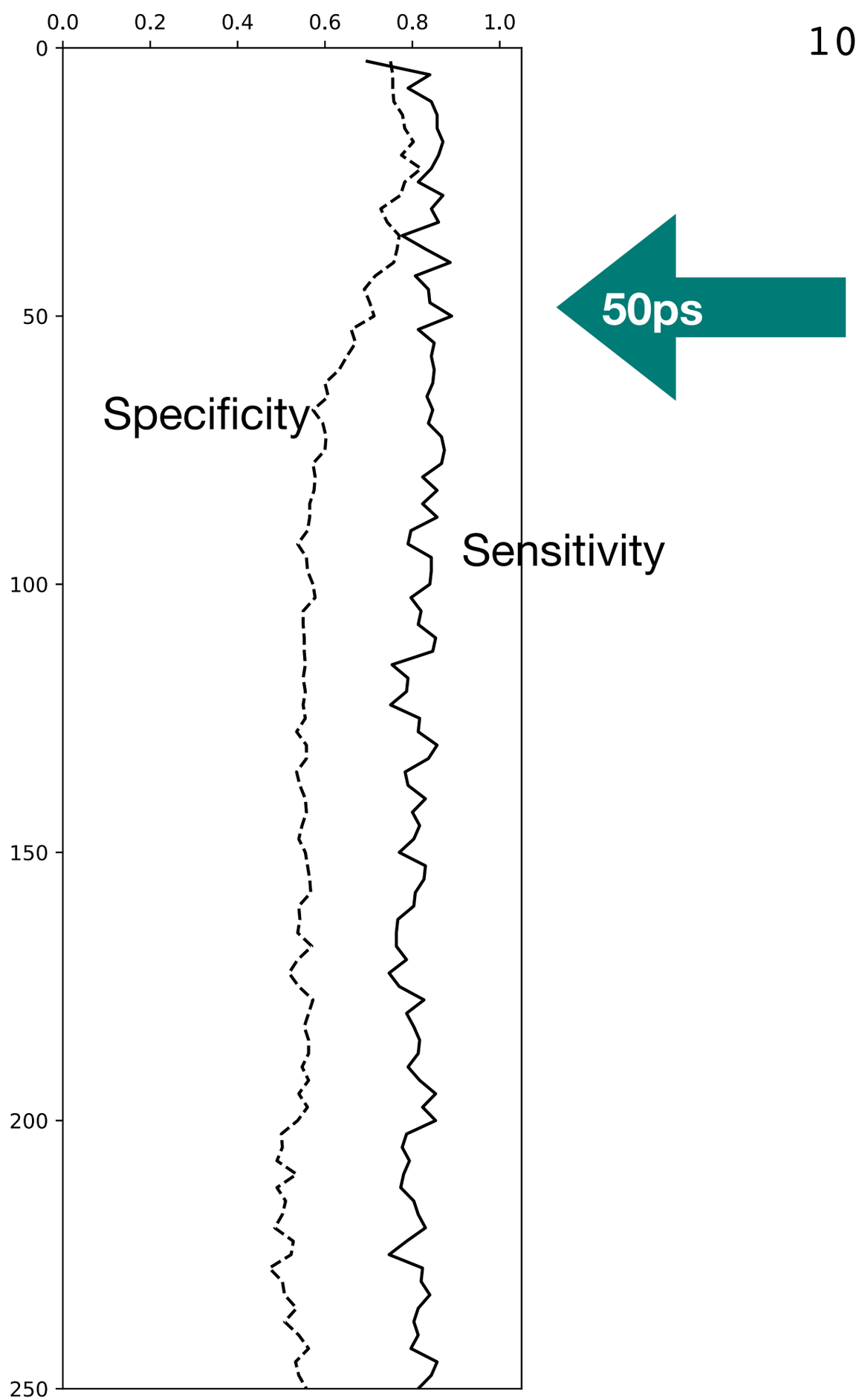

10%
ΔΔG (kcal/mol)
−4
−2
0
2
4
n=8
λ=100ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
U

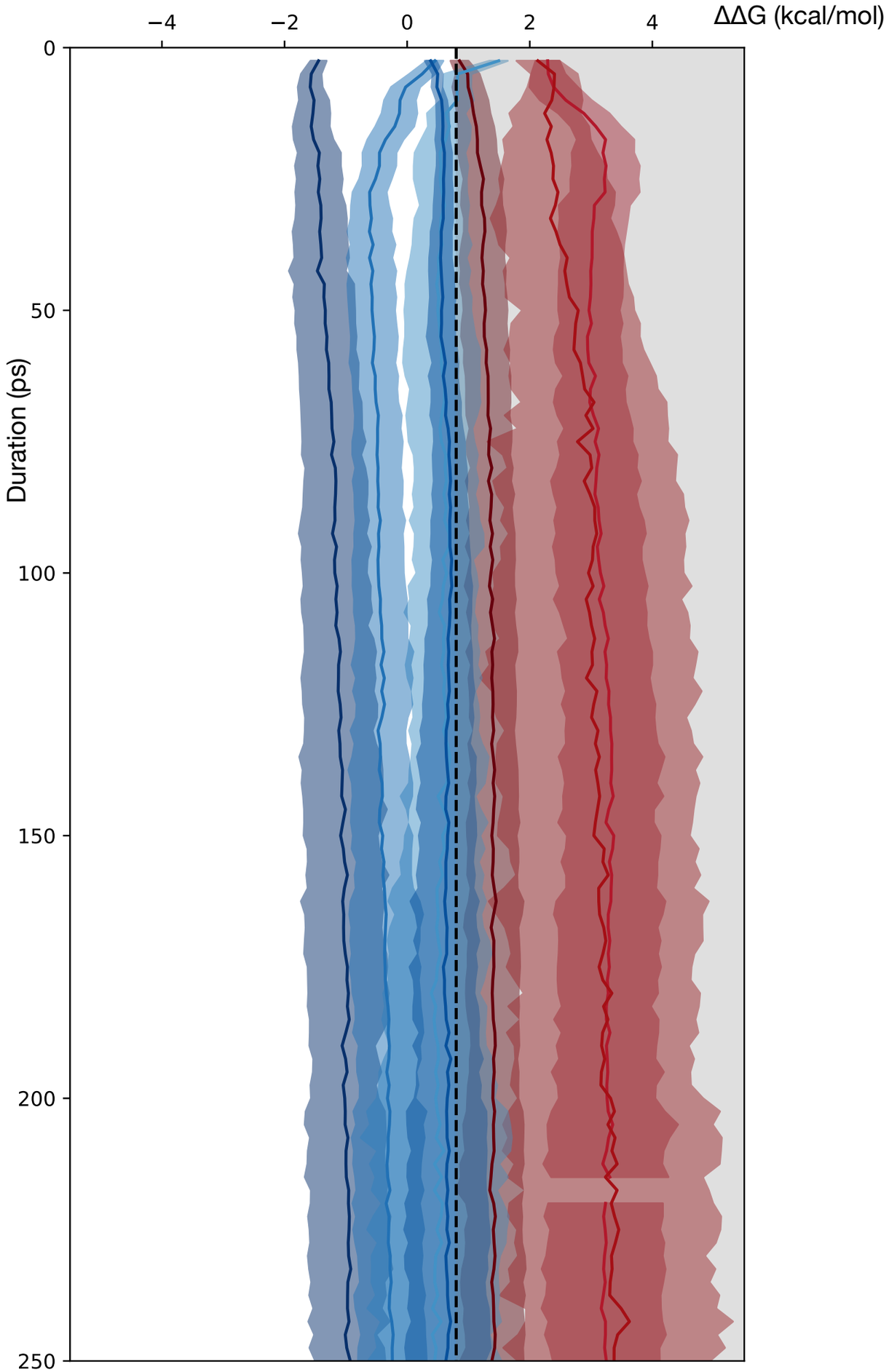

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250
n=16
λ=50ps
10%
ΔΔG (kcal/mol)
−4
−2
0
2
4
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
S

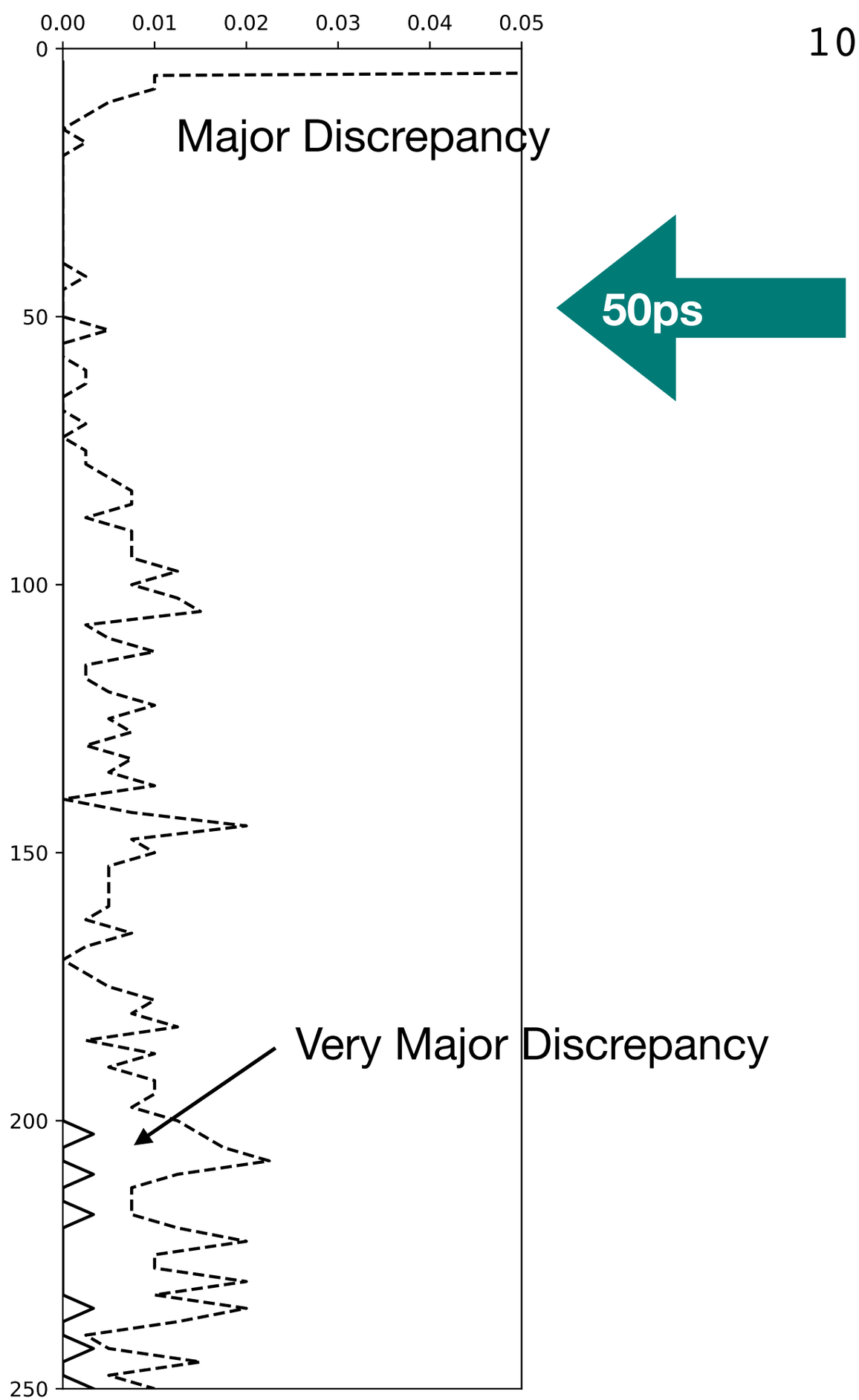

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250


10%
ΔΔG (kcal/mol)
−4
−2
0
2
4
n=80
λ=10ps
F99Y
F99Y/
L21V
L41F
F123L
A135T
V76A
I83V
R
R
R
U
S
S
S

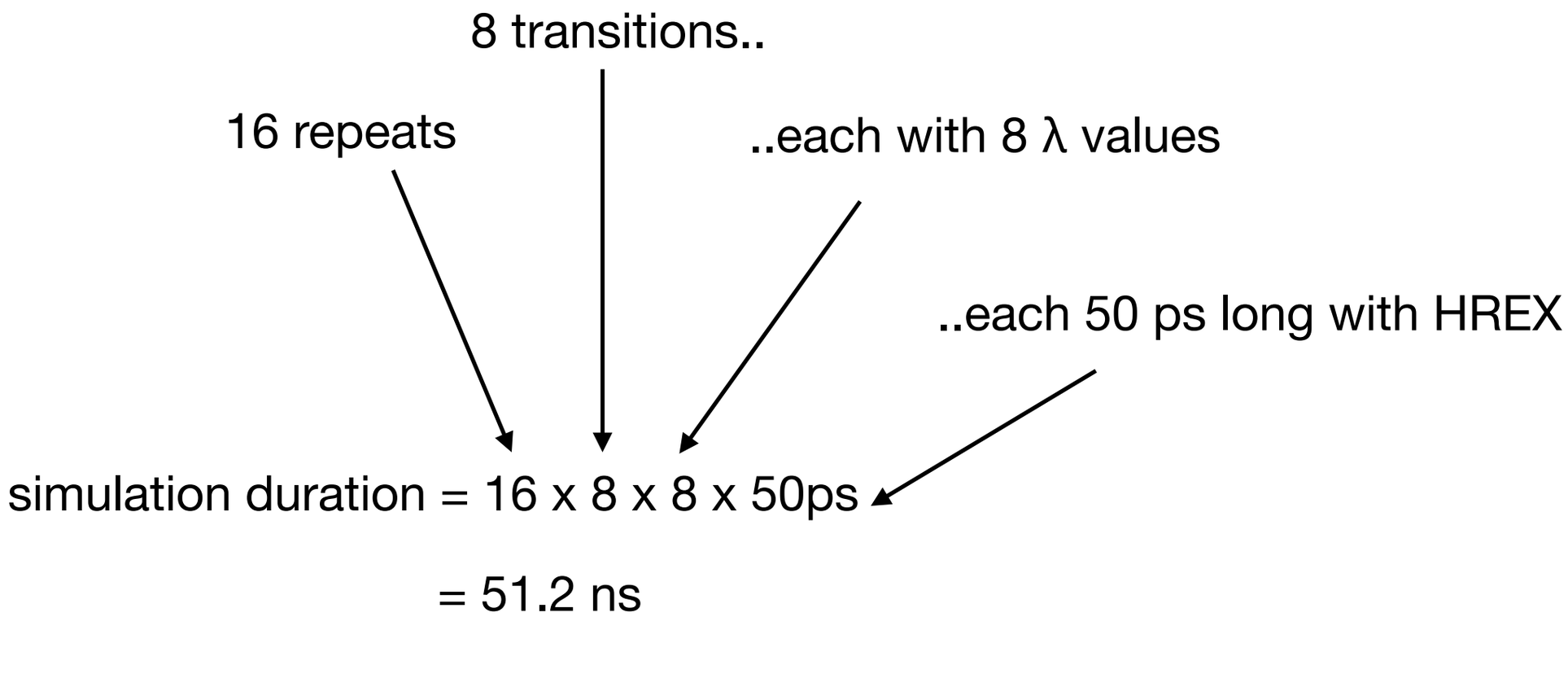

ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250

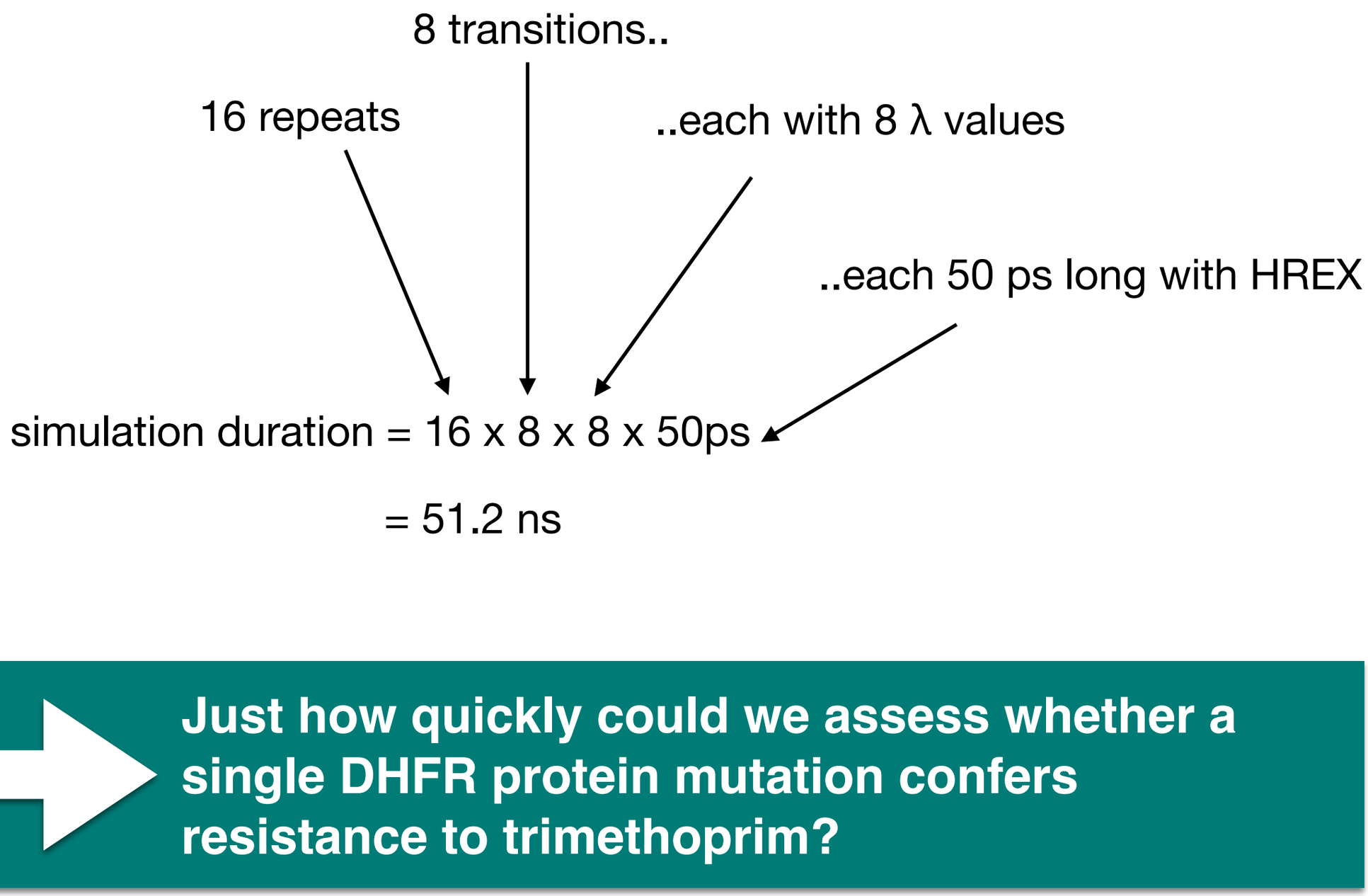

0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
Specificity
Sensitivity
50ps
10%


ΔΔG (kcal/mol)
−4
−2
0
2
4
Duration (ps)
0
50
100
150
200
250


0.00
0.01
0.02
0.03
0.04
0.05
0
50
100
150
200
250
Major Discrepancy
50ps
Very Major Discrepancy

10%


8 transitions..
16 repeats
..each with 8 λ values
..each 50 ps long with HREX
simulation duration = 16 x 8 x 8 x 50ps
= 51.2 ns



Just how quickly could we assess whether a single DHFR protein mutation confers resistance to trimethoprim?

GTX1080

9.6 ns /day

1 GPU with 4x OpenMP threads per λ simulation

(would take GROMACS longer to produce the TPR input files…)

7.5 minutes using 1,024 GPUs with 4,096 CPU cores

GTX1080

9.6 ns /day

1 GPU with 4x OpenMP threads per λ simulation

(would take GROMACS longer to produce the TPR input files…)

We could still do better since we've assumed average behaviour

Increase *n* until a definite classification (R/S) is made

Focus effort on free energy transitions which have more variability

Create a prior probability based on the antibiogram of the clinical sample

Use Machine Learning to "rule out" mutations that are highly likely to have no effect

NIHR | Oxford Biomedical Research Centre

Oxford University Hospitals NHS Foundation Trust

UNIVERSITY OF OXFORD

CRyPTIC

wellcome

BILL & MELINDA GATES foundation

Newton Fund

CompBioMed

amazon web services™

HECBioSim

archer

philip.fowler@ndm.ox.ac.uk
@philipwfowler
http://fowlerlab.org

Derrick Crook
Tim Peto
Sarah Walker
Tim Walker
Sarah Hoosdally
Ana Gibertoni-Cruz

Group


Oliver Adams


Dominykas Lukauskis


Alice Brankin


Joshua Carter

IGIB, New Delhi

Lipi Thukral

IPSB, Toulouse

Matthieu Chavent

Biochemistry

Simon Newstead

Oxford Brookes

Lucy Turner

BDI

Daniel Wilson
David Eyre

CRyPTIC

Daniela Cirillo

Grace Smith

Stefan Niemann

IBME

David Clifton
Tingting Zhu
Samaneh Kouchaki

EBI

Zam Iqbal
Martin Hunt

BASH THE BUG

All the Citizen Scientists

Helen Spiers

Grant Miller

Chris Lintott

BASH THE BUG

About Twitter

Help us fight antibiotic resistance!

TAKE ME TO THE ZOONIVERSE!

http://bashthebug.net
@bashthebug