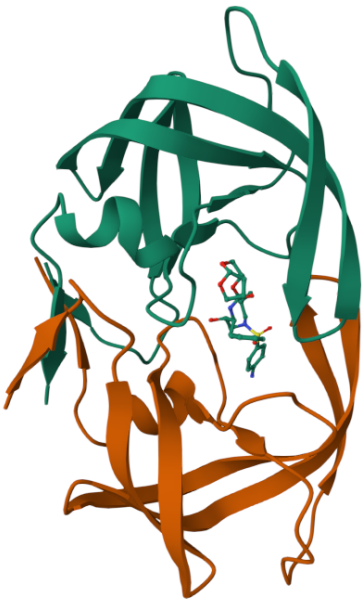
UNIVERSITY OF COPENHAGEN
DEPARTMENT OF MATHEMATICAL SCIENCES



Bioinformatics Project 2

Exploring the Impact of Multiple Mutations on Protein-Ligand Interactions on HIV Drug Resistance databaset

Author: Yinghan Jiang(wfz471)

Advisor: Amelie Stein

Date: Jun 2024

Abstract

Faculty: Faculty of Science

Author: Yinghan Jian

KUID : wfz471

Title: Exploring the Impact of Multiple Mutations on Protein-Ligand Interactions on HIV Drug Resistance dataset

Description: In this study, I explore a genotype-to-phenotype dataset for HIV drug resistance and conduct predictive analyses of multiple mutations using Rosetta. ΔΔG is calculated as the difference between the binding free energies of a mutant and the wild-type protein, it is an indicator of changes in binding affinity between proteins and various ligands due to mutations. By comparing the ΔΔG values for multiple mutations, this research aims to provide insights into how these mutations affect protein-ligand interactions. This approach is crucial for advancing drug design and therapeutic strategies, offering a deeper understanding of protein structure and function alterations.

Supervisor: Amelie Stein

Date: 17 06 2024

List of Abbreviations

HIV Human immunodeficiency virus

HIV-1 Human Immunodeficiency Virus Type 1

HIV-2 Human Immunodeficiency Virus Type 2

AIDS Acquired Immunodeficiency Syndrome

ART Antiretroviral Therapy

NRTI Nucleoside Reverse-transcriptase Inhibitor

NNRTI Non-nucleoside Reverse-transcriptase Inhibitor

PI Protease Inhibitor

INSTI Integrase Strand Transfer Inhibitor

PR Protease

DRV Darunavir

HIVDB HIV Drug Resistance Database

ΔΔG Delta Delta G

HPC High-Performance Computing

DeiC Danish e-Infrastructure Cooperation

List of Figures

List of Tables

Contents

1 Introduction

1.1 HIV-1 Drug Resistance

Acquired Immunodeficiency Syndrome (AIDS) is an immune deficiency condition caused by two types of the Human Immunodeficiency Virus (HIV), Human Immunodeficiency Virus Type 1 (HIV-1) and Human Immunodeficiency Virus Type 2 (HIV-2) [2]. Infections caused by HIV-1 are distributed worldwide, whereas HIV-2 affects populations in West Africa. This distribution pattern indicates a higher transmission rate of HIV-1 compared to HIV-2 [3], the drug resistance data used in this study are from HIV-1 virus isolates.

HIV infection can be treated with Antiretroviral therapy (ART), which includes four types of inhibitors: nucleoside reverse-transcriptase inhibitor (NRTI), non-nucleoside reverse-transcriptase inhibitor (NNRTI), protease inhibitor (PI), integrase strand transfer inhibitor (INSTI), and entry inhibitors.

Drug Classes	**Drugs**
PI	FPV, ATV, IDV, LPV, NFV, SQV, TPV, DRV
NRTI	3TC, ABC, AZT, D4T, DDI, TDF
NNRTI	EFV, NVP, ETR, RPV
INI	RAL, EVG, DTG, BIC, CAB

Table 1: List of HIV drug classes and corresponding drugs

Different types of inhibitors influence on different life cycles of the HIV. PI drugs have primarily been designed to inhibit the replication of HIV by binding the viral protease with high affinity. HIV-1 protease is an indispensable enzyme in the viral replication process, cleaving the Gag and Gag-Pol polyprotein precursors to produce mature virions [4]. PI effectively bind to the active site of the protease, preventing the approach and hydrolysis of substrates, thus inhibiting viral maturation and replication [5]. However, during viral replication, mutations can occur that may alter the amino acid sequence of the protease, reducing the binding affinity of PI and mutations near the active site of the protease are particularly problematic. Therefore, first-generation PIs are prone to developing drug resistance [4].

1.2 Association between Mutations and Drug Resistance

HIV-1 protease mutations can be classified into primary and secondary mutations. Primary mutations occur at the active site of the protease and are directly involved in substrate binding, reducing the affinity for binding with PI. Secondary mutations are located outside the active site and are typically compensatory, mitigating the adverse effects of primary mutations on the binding of the protease to its natural substrates [6, 7]. Therefore, developing high-level resistance to protease inhibitors involves a gradual accumulation of both primary and secondary mutations [6].

are these positions comparable?
mapping to structure would have been nice

Drug Classes	Drug Resistance Mutation Positions
PI	30, 32, 47, 48, 50, 54, 76, 82, 84, 88
NRTI	41, 65, 70, 74, 75, 151, 184, 210, 215
NNRTI	100I, 101P, 103N, 106A/M, 181C/I/A, 188C/L/H, 190A/E/S/Q, 230L
INI	66A/I/K, 92Q, 118R, 143, 148H/R/K, 155H, 263K

Table 2: Major drug resistance mutation positions by drug classes [1].

Darunavir (DRV), previously known as TMC114, is a second-generation PI, approved in 2006 for the treatment of HIV-1. It was specifically designed to inhibit drug-resistant strains of the virus. Compared to other PIs, DRV demonstrates potent inhibitory effects against a broad range of viral strains, including multiple drug-resistant HIV strains. It shows strong efficacy against both wild-type and various resistant variants. Moreover, DRV possesses a higher genetic barrier, which means that the virus faces greater difficulty in developing resistance to this drug [8].

In first-generation PI, there is a phenomenon called low genetic barrier, where HIV can develop drug resistance through a single mutation. In contrast, those requiring multiple mutations to exhibit resistance, known as having a higher genetic barrier like second-generation PI, typically demonstrate more complex resistance patterns [9]. Therefore, studying these multiple mutations is crucial for devising more effective therapeutic strategies and enhancing our understanding of HIV's resistance mechanisms.

good bio intro, well-written

1.3 Correlation between ΔΔG and Multiple Mutation

$\Delta\Delta G$ is calculated as the difference between the binding free energies of a mutant and the wild-type protein, expressed as:

$$\Delta\Delta G = \Delta G_{\text{Mut}} - \Delta G_{\text{WT}}$$

where ΔG_{Mut} is the binding free energy of the mutation and ΔG_{WT} is the binding free energy of the wild-type. A positive $\Delta\Delta G$ indicates that the mutation leads to less favorable binding, making the drug interaction energetically unfavorable. Conversely, a negative $\Delta\Delta G$ suggests that the mutation enhances binding affinity.

In HIV therapy, mutations in HIV protease can reduce the binding affinity to PI, leading to therapeutic failure. For second-generation PI like DRV, when the number of mutations increases, their contribution to drug resistance becomes more significant [8]. Such multiple mutations typically need to accumulate gradually, with each mutation potentially affecting the structure and function of the protein, ultimately leading to significant changes in the efficiency of drug binding.

what does that mean?

Therefore, we can use $\Delta\Delta G$ to quantify the effect of mutation, understanding the changes in $\Delta\Delta G$ and their relationship with protein mutations is crucial for the forecast alteration in drug resistance and optimization of therapeutic strategies.

2 Methods

2.1 Dataset

From the PDB dataset, we selected the HIV-1 protease structure bound with DRV, with PDB ID 1T3R (https://www.rcsb.org/structure/1t3r)

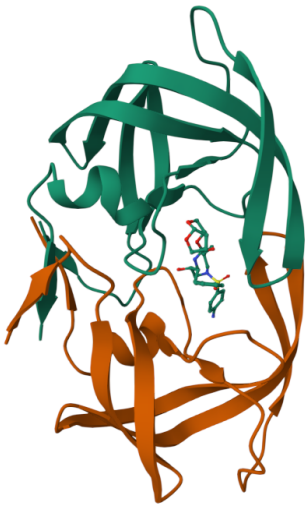


Figure 1: 3D structure of HIV-1 protease with bound DRV drug (PDB ID 1T3R), no water and ions are shown.

Eight mutations were modified to construct a wild-type structure suitable for this dataset in the context of the multiple mutation dataset.

Chain	Mutation
A	R 14 K, V 32 I, K 41 R, I 47 V, P 63 L, V 64 I, L 76 M, V 82 I
B	R 14 K, V 32 I, K 41 R, I 47 V, P 63 L, V 64 I, L 76 M, V 82 I

Table 3: Mutation List of transform Wild type PDB

ask into dimer nature and whether there can be differences between the sequences

The mutation dataset is from the HIV Drug Resistance Database (HIVDB) and focuses on multiple mutations (number of mutations greater than or equal to 2) associated with DRV drug resistance in PI drugs.

The PI dataset consists of isolates tested for in vitro sensitivity using the PhenoSense assay. This dataset has undergone high-quality filtering to exclude redundant and mixed viruses, thereby reducing bias and improving data reliability [1]. The PI dataset comprises

a total of 2395 isolates, including 32 with single mutations and 2363 with multiple mutations.

The DRV dataset is a subset of the PI dataset and includes only samples related to DRV drug resistance. After removing NA values, the DRV dataset contains a total of 1115 isolates, including 8 with single mutations and 1107 with multiple mutations.

In this project, I selected the DRV dataset to calculate $\Delta\Delta G$ for 2, 3, 4, and 5 multiple mutation. The 2 multiple mutation dataset refers to samples with 2 mutations and originally included 27 samples before filtering. The 3 multiple mutation dataset refers to samples with 3 mutations and originally included 34 samples before filtering. The 4 multiple mutation dataset refers to samples with exactly 4 mutations and originally included 65 samples before filtering. The 5 multiple mutation dataset refers to samples with exactly 5 mutations and originally included 83 samples before filtering.

Multiple Mutation	**Example**	**Datasat size**
2 Multiple Mutation	R57G, L63P	27
3 Multiple Mutation	I15V, M36I, I72V	34
4 Multiple Mutation	I32V, V47I, M76L, I82V	65
5 Multiple Mutation	L19I, R57K, L63S, I64V, E65D	83

Table 4: Multiple mutation dataset size before filtering

2.2 Data Selection

The data selection process primarily focuses on mutations. In the DRV dataset, there are instances such as R41RK and H69HY, where two or more amino acid codes indicate a mixture. This occurrence can potentially confound the analysis of genotype-phenotype correlations. Therefore, removing such instances helps maintain the reliability of the dataset for subsequent analyses. The size of the dataset after removal is as follows:

Multiple Mutation	Datasat size
2 Multiple Mutation	21
3 Multiple Mutation	24
4 Multiple Mutation	30
5 Multiple Mutation	33

Table 5: Multiple mutation dataset size after removing mixture

During the calculation of $\Delta\Delta G$, there are some outlier and missing value in the 2, 4, and 5 multiple mutation datasets, hence, additional data filtering and processing were performed to solve these issues.

In the 2 multiple mutation dataset, the outlier occurred when the 2 mutations are identical, the drug fold values differed. In such cases, the drug fold values take the average.

2 Multiple Mutation	Drug Fold	Average
E35D, L63P	1.4, 0.9, 0.7	1.00
I13V, I64V	0.5, 0.6	0.55
L63P, I64L	0.6, 0.7	0.65

how about the error?

Table 6: Outliers in the dataset of 2 multiple mutation

The missing value when $\Delta\Delta G$ results were missing due to multiple mutations occurring at the same position, resulting in only one calculated value. For example, mutations like R57K L63Q, R57G L63Q, and R57G L63P produced only one $\Delta\Delta G$ value, namely R57G L63Q.

huh, that seems wrong to me - why is it not specific to the target mutation?

After addressing these anomalies and missing values, the size of the 2 multiple mutation dataset was reduced to 14.

2 Multiple Mutation
R57G, L63P
R57K, L63Q
R57K, V77I

Table 7: Missing value of 2 multiple mutation

how does this happen? As in, how can those seqences be in the dataset but no value? Or are there only values for other inhibitors?

In the 4 multiple mutation dataset, one missing value was observed.

4 Multiple Mutation
L10I, I15V, L63T, I93L

Table 8: Missing value of 4 multiple mutation

Additionally, inconsistencies were observed in the wild mutation and the drug fold value was 0. Consequently, these data points were removed, resulting in the final dataset size of 27.

4 Multiple Mutation	Drug Flod
I32V, V47I, M76L, I82V	0
I15V, M36I, I64V, V82L	1.1

not sure I understand this… is I not the WT?

Table 9: Outliers in the dataset of 4 multiple mutation

Outliers in the 5 multiple mutation dataset were identified as mutated variants that did not correspond to the 20 canonical amino acids. The presence of 'X' could potentially indicate a sequencing error in the HIVDB, although no specific description of 'X' was found in the database.

5 Multiple Mutation
L19T, N37D, K43R, R57G, L63X
I13V, L19I, E35D, N37X, L63P

Table 10: Outliers in the dataset of 5 multiple mutation

Therefore, these outliers were also removed, reducing the dataset size to 31.

2.3 Workflow

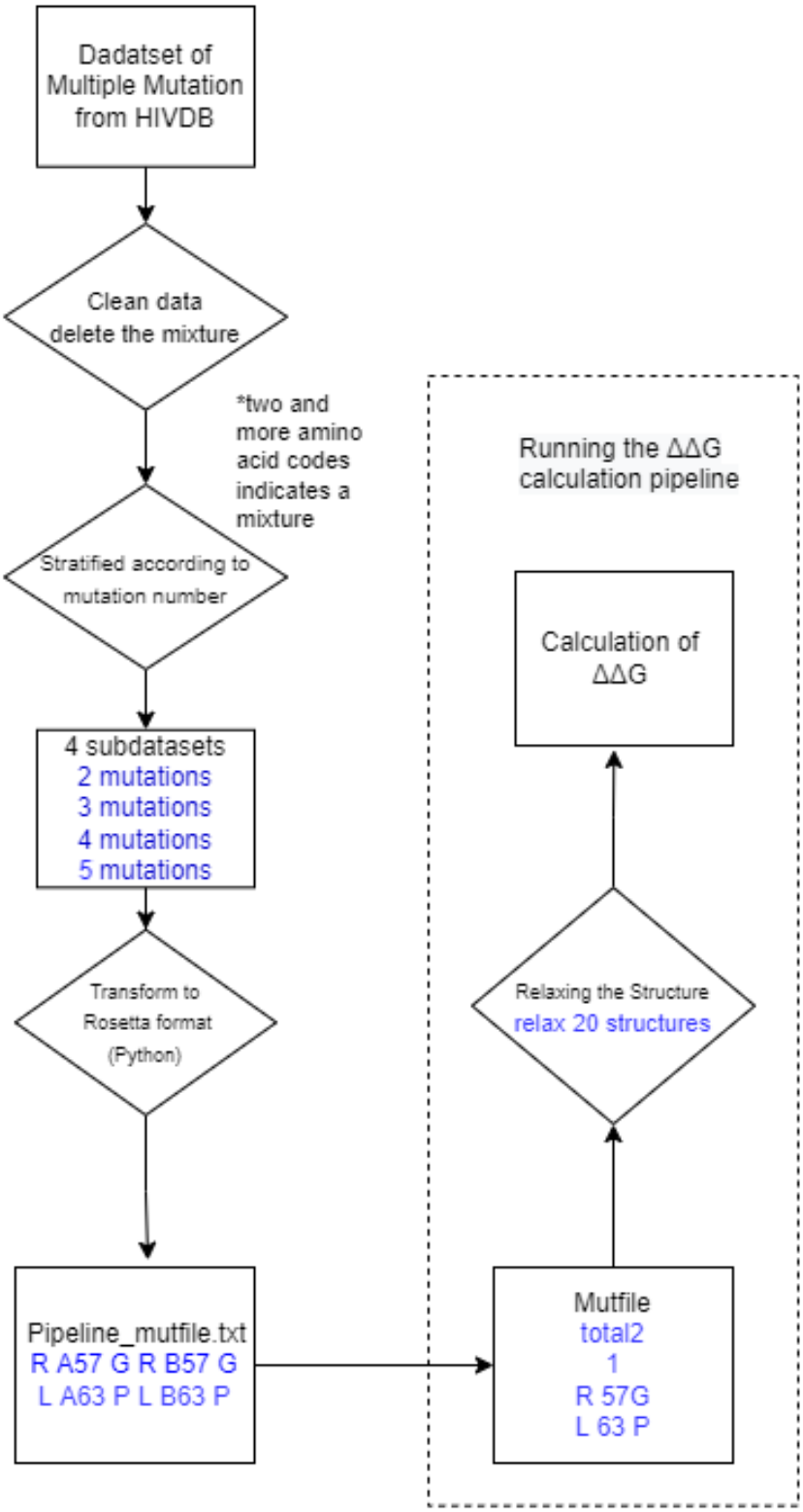


Figure 2: Flowchart of ΔΔG Calculation

2.4 Tools

The mutation selection step was finished in Python version 3.9.7. The code of data filtering and distribution can be downloaded directly from GitHub, and the download link is provided in the Data Availability section.

In this project, the ΔΔG calculation step was performed on the Danish e-Infrastructure Cooperation (DeiC) High-Performance Computing (HPC) cluster and utilizing the Rosetta software.

DeiC is an organization that coordinates the use of national supercomputers by Danish researchers. It is responsible for managing and coordinating the supercomputing resources operated and developed by universities within Denmark.

where from? Differs from other numberings I know

Rosetta is software library used for macromolecular modeling. The version of Rosetta used in this project is v0.2.6-14-gc174a69. Protocols are a collection of algorithms within Rosetta designed to accomplish specific tasks in biomolecular modeling. These protocols utilize the functionality provided by Rosetta to carry out tasks such as prediction, design, and analysis of the structures and interactions within biological systems [10].

The pipeline used to compute ΔΔG is derived from the Rosetta software suite, specifically the rosetta_ddG_pipeline, which is executed through the run_pipeline.py script. After activating the Conda environment named py3_ros_ddG_env, the pipeline generates a mut-file based on the input files(1t3r.pdb), performs relaxation operations, and subsequently calculates ΔΔG. For a detailed workflow, refer to section 2.3.

Command line for the pipeline:

```
python3 run_pipeline.py -s 1t3r.pdb -o Output -i proceed -mm mut_file
-m 1t3r_mutfile.txt --chainid AB --run_struc AB --overwrite_path True
--slurm_partition sbinlab_ib --ligand True --dump_pdb True
```

Flags for the pipeline:

-s: Specify the input structure file (PDB format)

-o: Specify the output directory

-i: Specify the initialization mode

-mm: Specify the type of mutfile to generate (mut_file indicates a mutation file)

-m: Specify the specific mutation file to use (1t3r_mutfile.txt)

–chainid AB: Indicate chains A and B

–run_struc: Specify the chains to use for running the structure

–overwrite_path: Specify whether to overwrite existing files

–slurm_partition: Specify the SLURM partition for job scheduling

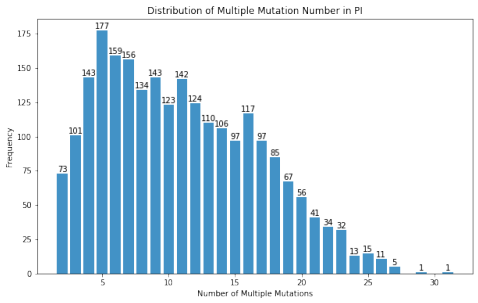
–ligand: Specify whether to consider ligands.

3 Results and Discussion

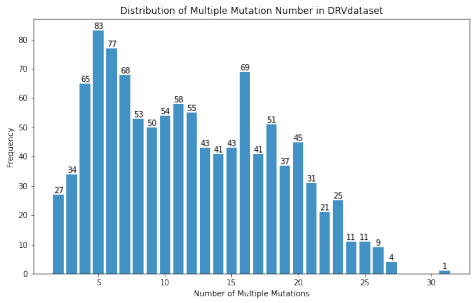
3.1 Data Distribution

The DRV subset follows the trend of the PI dataset, where the distribution of multiple mutations is similar, with the peak occurring at 5 multiple mutations. In the PI dataset, the samples are concentrated on the left side of the distribution, indicating that drug resistance in PI may mainly be associated with a lower number of multiple mutations. In the DRV subset, the sample distribution is more uniform, leaning towards a symmetric distribution. This suggests that resistance to DRV drugs may not be strongly correlated with a specific number of mutations.

Multiple mutations consisting of 2, 3, 4, and 5 mutations account for 20.63% and 18.74% of the total samples in the PI dataset (total 2395 samples) and the DRV subset (total 1115 samples). It is noteworthy that a peak is observed at 16 multiple mutations in the DRV subset, which may require additional attention and investigation.



(a) Distribution of Multiple Mutation Number in PI



(b) Distribution of Multiple Mutation Number in DRVdataset

Figure 3: Distribution of Multiple Mutation number

In the PI dataset, the amino acid ILE is most likely to mutate to VAL, followed by LEU mutating to PRO and MET mutating to ILE. In the DRV dataset, the mutation probabilities approximate those in the PI dataset.

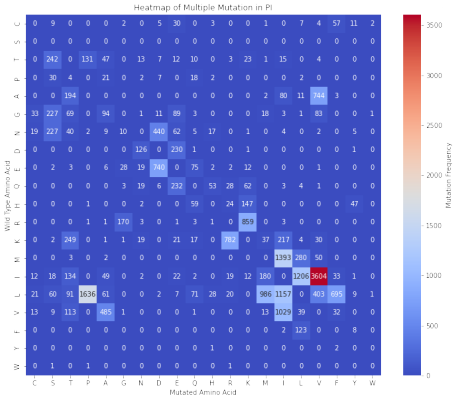
how do you explain these patterns?

However, in the dataset with 2 multiple mutations, the results show significant differences. The highest probability mutation is LEU mutating to PRO (19.5%), followed by

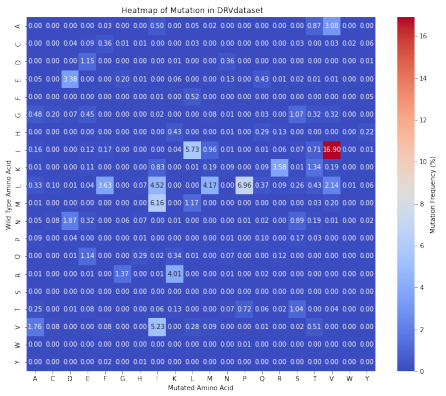
ILE mutating to VAL (11.9%). In the dataset with 3 multiple mutations, the amino acid mutation results are similar to those in the dataset with 2 multiple mutations. Specifically, 20.83% of mutations involve LEU mutating to PRO, and 19.33% involve ILE mutating to VAL. For both the 4 and 5 multiple mutation datasets, the results are similar. The highest probability mutation is ILE mutating to VAL, with probabilities of 18.33% and 14.73%, respectively.

figure? - should have been referenced

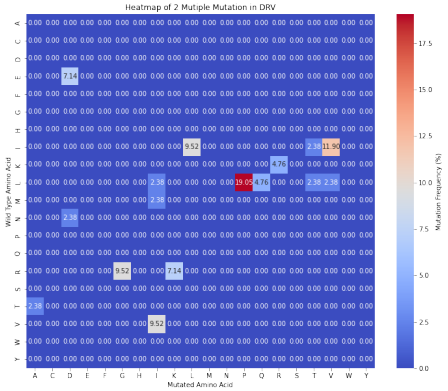
These heatmaps suggest that certain amino acid mutations, particularly ILE mutating to VAL and LEU mutating to PRO, play a consistently significant role across different datasets. The high probability of these mutations implies their potential functional importance in drug resistance.



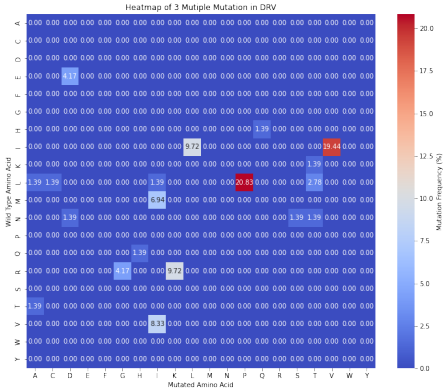
(a) Heatmap of Multiple Mutation in PI



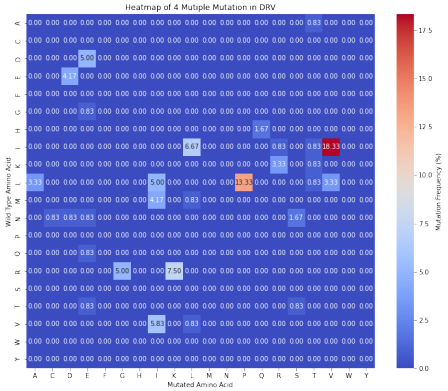
(b) Heatmap of Multiple Mutation in DRV



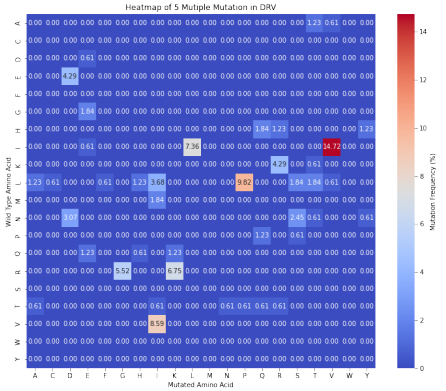
(c) Heatmap of 2 Mutations in DRV subset



(d) Heatmap of 3 Mutations in DRV subset



(e) Heatmap of 4 Mutations in DRV subset

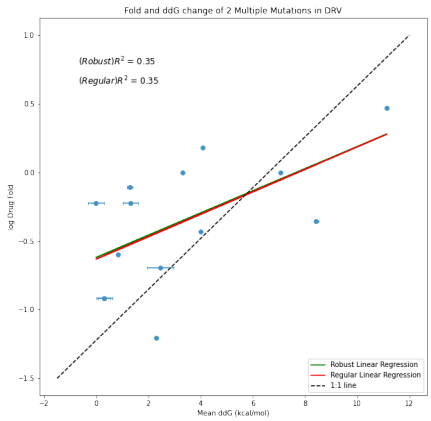


(f) Heatmap of 5 Mutations in DRV subset

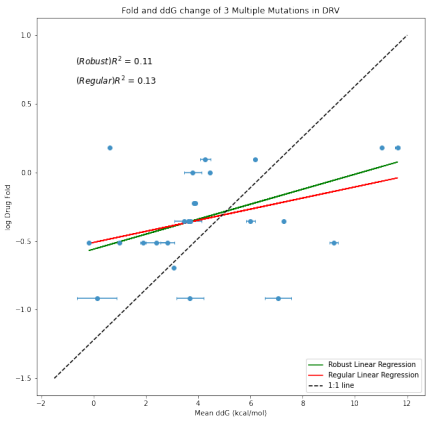
Figure 4: Heatmap of Multiple Mutation

3.2 Drug Fold and ΔΔG change of Mutiple Mutation in DRV Dataset

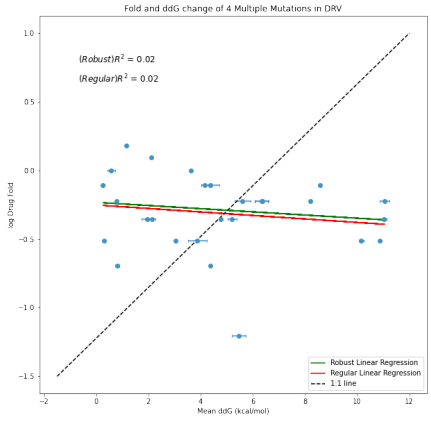
The results from datasets with 2 and 3 mutations demonstrate a positive correlation between ΔΔG and drug fold, indicating a significant impact of these mutations on binding affinity and consequent increase in drug resistance. However, the trend in the 4 and 5 multiple mutation datasets is flat.



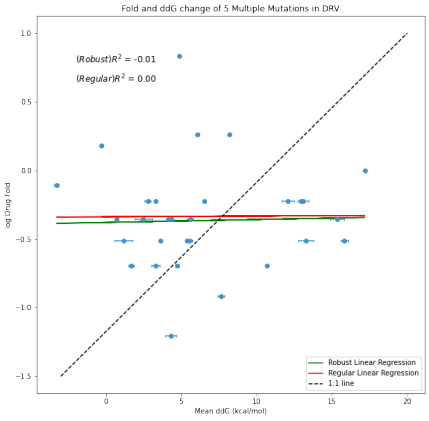
(a) Drug Fold and ΔΔG change of 2 Mutiple Mutation in DRV



(b) Drug Fold and ΔΔG change of 3 Mutiple Mutation in DRV



(c) Drug Fold and ΔΔG change of 4 Mutiple Mutation in DRV



(d) Drug Fold and ΔΔG change of 5 Mutiple Mutation in DRV

Figure 5: Drug Fold and ΔΔG change of Mutiple Mutation in DRV

Lower ΔΔG values for 2 mutations suggest a relatively minor effect on binding affinity. Conversely, the higher ΔΔG values and significant drug fold changes observed in the 5 mutations dataset indicate the structural changes have an impact on binding affinity, leading to increased drug resistance.

Furthermore, the majority of ΔΔG in the four groups of results are positive, indicating that mutations lead to the loss of stability. Considering the results from four datasets, the ΔΔG increase with the number of mutations, this is due to the additive effect of ΔΔG, no significant mutation effects (ΔΔG values greater than 1 kcal/mol) were observed.

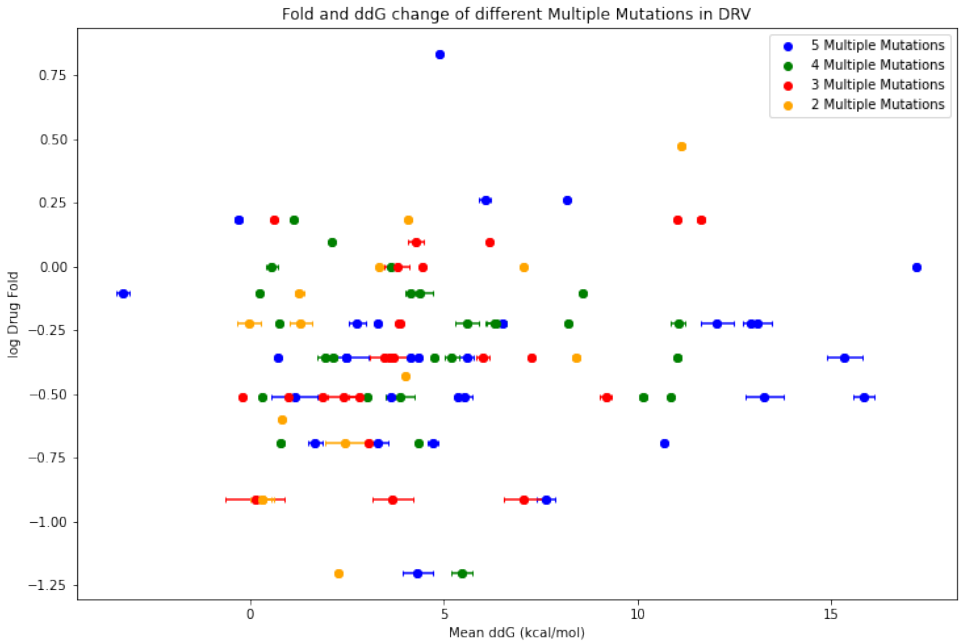


Figure 6: Drug Fold and ΔΔG change of different Multiple Mutations in DRV

4 Conclusion

This project examines the impact of multiple mutations on DRV drug resistance through $\Delta\Delta$G calculations, also uncovering the relationship between mutation count and $\Delta\Delta$G variations.

The PI dataset displays a right-skewed distribution, suggesting a potential link between drug resistance and fewer mutations. The DRV subset exhibits a more uniform distribution, indicating a weaker correlation between mutation count and drug resistance. Heatmap analyses consistently highlight mutations such as ILE to VAL and LEU to PRO across datasets, underscoring their critical roles in drug resistance mechanisms. Further analyses correlating drug fold with $\Delta\Delta$G reveal that $\Delta\Delta$G increases with the number of mutations. The datasets with 2 and 3 mutations demonstrate a positive correlation between $\Delta\Delta$G and drug fold, indicating significant impacts on binding affinity and subsequent drug resistance. In contrast, $\Delta\Delta$G changes in datasets with 4 and 5 mutations show slower variations.

Due to time constraints, this project encountered several limitations, such as issues with the backmutation process and the absence of comparison with other tools. Addressing these shortcomings would significantly enhance the reliability of the result.

5 Perspectives for Further Research

5.1 Issues in the Backmutation

Constructing the ”pseudo” wild-type PDB is one of the most critical steps in computing $\Delta\Delta G$ because all subsequent calculations rely on it. However, several key aspects need to be considered to ensure accuracy and reproducibility.

Firstly, when performing backmutations to create a ”pseudo” wild-type structure, it's crucial to carefully select the method. Backmutation involves replacing amino acid residues in the PDB with corresponding residues from the reference sequence to ensure consistency between the PDB and the reference sequence. This step typically requires the use of molecular simulation software, such as Rosetta, to ensure accurate backmutations without introducing unintended structural changes.

However, it's essential to carefully consider the methodology used for backmutations, While the rosetta_ddG_pipeline may seem convenient, it carries the risk of introducing unintended structural alterations. In contrast, protocols like relax with more global changes provide a more comprehensive approach, allowing for improving the reliability of $\Delta\Delta G$ calculations.

Secondly, when selecting which mutations to backmutate, it is important to consider drug fold data and additional datasets. In this project, I modified eight mutations but encountered an issue at position 82, where two different wild-type mutations were possible. I mistakenly chose the wild-type mutation as I, however, subsequent analysis revealed that the drug fold was 0 for the wild-type mutation I. Furthermore, further analysis of the 6 multiple mutation dataset indicated that the wild-type mutation at this position should be V.

5.2 Lack of comparison with other tools

Due to time constraints, this project lack of comparison with other tools. By comparing different tools, it is possible to find the highest accuracy and the most convenient way to calculate $\Delta\Delta G$ of multiple mutation.

6 Data Availability

The data is from HIVDB and can be downloaded directly via this link https://hivdb.stanford.edu/.

Code in this paper are available in my Github repository.

You can use this link https://github.com/yinghanJ/MutipleProtein to downloaded it.

7 References

[1] Stanford University, "Hiv drug resistance database," 2024, accessed: 2024-06-03. [Online]. Available: https://hivdb.stanford.edu/pages/genopheno.dataset.html

[2] Centers for Disease Control and Prevention, "Aids," National Center for Health Statistics, 2023, available online: https://www.cdc.gov/nchs/hus/sources-definitions/aids.htm.

[3] T. C. Quinn, "Population migration and the spread of types 1 and 2 human immunodeficiency viruses." Proceedings of the National Academy of Sciences, vol. 91, no. 7, pp. 2407–2414, 1994.

[4] A. K. Ghosh, H. L. Osswald, and G. Prato, "Recent progress in the development of hiv-1 protease inhibitors for the treatment of hiv/aids," Journal of medicinal chemistry, vol. 59, no. 11, pp. 5172–5208, 2016.

[5] A. Brik and C.-H. Wong, "Hiv-1 protease: mechanism and drug discovery," Organic & biomolecular chemistry, vol. 1, no. 1, pp. 5–14, 2003.

[6] H. Mitsuya, K. Maeda, D. Das, and A. K. Ghosh, "Development of protease inhibitors and the fight with drug-resistant hiv-1 variants," Advances in pharmacology, vol. 56, pp. 169–197, 2008.

[7] A. M. Wensing, N. M. van Maarseveen, and M. Nijhuis, "Fifteen years of hiv protease inhibitors: raising the barrier to resistance," Antiviral research, vol. 85, no. 1, pp. 59–74, 2010.

[8] S. De Meyer, H. Azijn, D. Surleraux, D. Jochmans, A. Tahri, R. Pauwels, P. Wigerinck, and M.-P. de Béthune, "Tmc114, a novel human immunodeficiency virus type 1 protease inhibitor active against protease inhibitor-resistant viruses, including a broad range of clinical isolates," Antimicrobial agents and chemotherapy, vol. 49, no. 6, pp. 2314–2321, 2005.

[9] C. V. Fletcher, J. Bartlett, P. Sax, and J. Mitty, "Overview of antiretroviral agents used to treat hiv," UpToDate https://www. uptodate. com/contents/overview-of-antiretroviral-agents-used-to-treat-hiv, 2018.

[10] Rosetta commons. [Online]. Available: https://www.rosettacommons.org/demos/latest/Home

A Appendices

Table 11: Original data of 2 mutations

Variant	Mean ΔΔG	Std ΔΔG
L10=:R41=:L109=:R140=	5.2273000742767785e-14	0.09550015753831494
L10V:R41K:L109V:R140K	-0.01183908045972538	0.3069826819600345
T12A:I64V:T111A:I163V	2.2940229885057044	0.0026858210219753847
T12=:I64=:T111=:I163=	-2.6134874068125164e-14	0.000905058376301795
I13=:I64=:I112=:I163=	-2.613491020819758e-14	0.0016009641697890713
I13V:I64V:I112V:I163V	0.8304597701149059	0.0012902266850988587
L19I:I93L:L118I:I192L	0.30781609195403575	0.2989366807872354
L19=:I93=:L118=:I192=	0.0	0.04283302141585795
E35D:L63P:E134D:L162P	3.3194252873563896	0.00756003846088205
E35=:L63=:E134=:L162=	5.226953129581583e-14	0.0039046638509164138
M36I:N37D:M135I:N136D	0.30781609195395737	0.23927843262220314
M36=:N37=:M135=:N136=	-2.613491020819758e-14	0.00028155054517284257
R57G:L63Q:R156G:L162Q	11.136666666666656	0.08745366627089042
R57G:K70R:R156G:K169R	7.063793103448268	0.005194079782850203
R57G:V77I:R156G:V176I	8.422183908045993	0.09151749834668949
R57=:L63=:R156=:L162=	0.0	0.1806872422406194
R57=:K70=:R156=:K169=	0.0	0.004680292660102364
R57=:V77=:R156=:V176=	2.6131874442114622e-14	0.17876338624744462
L63=:I64=:L162=:I163=	0.0	0.0
L63=:K70=:L162=:K169=	2.6134765647907916e-14	0.005004951967159089
L63=:I72=:L162=:I171=	2.6134982488342413e-14	0.0023612228257850957
L63=:V77=:L162=:V176=	2.613491020819758e-14	0.00225240436115868
L63P:I64L:L162P:I163L	4.011149425287322	0.0011721883939372253
L63P:K70R:L162P:K169R	1.2781609195402173	0.11200805889885382
L63P:I72T:L162P:I171T	4.087241379310393	0.0038225952065164904
L63T:V77I:L162T:V176I	2.4612643678161192	0.5078776441365234
V77I:I93L:V176I:I192L	1.307816091954015	0.2879755626660746
V77=:I93=:V176=:I192=	7.228014483236696e-20	0.002361222825748201

Table 12: Original data of 3 mutations (a)

Variant	Mean ΔΔG	Std ΔΔG
L10I:L63P:I93L:L109I:L162P:I192L	4.276091954022962	0.20173818597315413
L10=:L63=:I93=:L109=:L162=:I192=	-2.6134765647907916e-14	0.004789118771508887
T12A:I13V:L63P:T111A:I112V:L162P	4.453448275862004	0.003164574689168524
T12=:I13=:L63=:T111=:I112=:L162=	-2.613491020819758e-14	0.0015506594900268108
I13=:Q18=:L63=:I112=:Q117=:L162=	2.613491020819758e-14	0.0018318824655636121
I13=:R57=:L63=:I112=:R156=:L162=	-5.2263748884229244e-14	0.27799819501720074
I13=:I64=:I72=:I112=:I163=:I171=	5.226982041639516e-14	0.0026858210219686367
I13V:Q18H:L63P:I112V:Q117H:L162P	3.456321839080479	0.0031140729158852104
I13V:R57G:L63P:I112V:R156G:L162P	11.019999999999976	0.037043963768532986
I13V:I64V:I72V:I112V:I163V:I171V	1.0001149425287903	0.0011721883939375447
I15=:E35=:R57=:I114=:E134=:R156=	-2.613418740674926e-14	0.07559199575114886
I15=:M36=:R41=:I114=:M135=:R140=	0.0	0.0074827497925172804
I15=:M36=:L63=:I114=:M135=:L162=	-2.6134765647907916e-14	0.01481644417868893
I15=:M36=:I72=:I114=:M135=:I171=	2.6135343889066576e-14	0.018506104087357162
I15=:I62=:L63=:I114=:I161=:L162=	5.2269964976684826e-14	0.003681751121398508
I15V:E35D:R57G:I114V:E134D:R156G	9.18724137931033	0.15907306806292587
I15V:M36I:R41K:I114V:M135I:R140K	-0.17333333333331158	0.0042637326416022185
I15V:M36I:L63C:I114V:M135I:L162C	3.7035632183907787	0.016322648514794678
I15V:M36I:I72V:I114V:M135I:I171V	0.6103448275862268	0.01113578974233223
I15V:I62V:L63P:I114V:I161V:L162P	3.892298850574745	0.017898585434307224
L19=:L63=:V77=:L118=:L162=:V176=	0.0	0.009002282413436907
L19T:L63P:V77I:L118T:L162P:V176I	3.8482758620689936	0.0030367459341676816
E35D:M36I:L63P:E134D:M135I:L162P	2.826206896551746	0.2593508976197864
E35=:M36=:L63=:E134=:M135=:L162=	2.613491020819758e-14	0.0013004262642573312
M36I:R41K:I93L:M135I:R140K:I192L	0.1374712643678583	0.7567847341453363
M36=:R41=:I93=:M135=:R140=:I192=	5.2275313707402425e-14	0.17419319331013608
N37D:L63P:I93L:N136D:L162P:I192L	3.7920689655172652	0.33584653051613444
N37=:R57=:L63=:N136=:R156=:L162=	0.0	0.004364790240463836
N37=:L63=:V77=:N136=:L162=:V176=	0.0	0.0056310109029447135
N37=:L63=:I93=:N136=:L162=:I192=	2.613483792805275e-14	0.0013888558590226462
N37S:L63T:V77I:N136S:L162T:V176I	3.0662068965517424	0.0029394739890794587
N37T:R57G:L63P:N136T:R156G:L162P	11.62666666666667	0.09681777793595334

Table 13: Original data of 3 mutations (b)

Variant	**Mean ΔΔG**	**Std ΔΔG**
R41K:L63P:V77I:R140K:L162P:V176I	1.8929885057470799	0.10426155408740985
R41=:L63=:V77=:R140=:L162=:V176=	-2.6136500371383893e-14	0.18818328180050567
R57K:L63P:E65D:R156K:L162P:E164D	6.00609195402298	0.18038075728329456
R57K:L63P:H69Q:R156K:L162P:H168Q	6.189885057471301	0.005447354303923744
R57K:L63A:V77I:R156K:L162A:V176I	7.279080459770138	0.004226385358812595
R57K:L63P:I93L:R156K:L162P:I192L	7.061954022988515	0.49707191249724675
R57=:L63=:E65=:R156=:L162=:E164=	0.0	0.005533978444696717
R57=:L63=:H69=:R156=:L162=:H168=	5.226837481349852e-14	0.09807110483679683
R57=:L63=:V77=:R156=:L162=:V176=	2.6145752229922437e-14	0.18528785545125112
R57=:L63=:I93=:R156=:L162=:I192=	-1.4456028966473392e-19	0.002715175128965129
L63=:K70=:I93=:L162=:K169=:I192=	5.226953129581583e-14	0.004722445651533297
L63=:V77=:I93=:L162=:V176=:I192=	-5.2263748884229244e-14	0.20319648032922946
L63P:K70T:I93L:L162P:K169T:I192L	3.6893103448276103	0.5151744397729927
L63P:V77I:I93L:L162P:V176I:I192L	3.6180459770114783	0.5159748478355373
I64=:V77=:I93=:I163=:V176=:I192=	5.226982041639516e-14	0.0024545007475720803
I64V:V77I:I93L:I163V:V176I:I192L	2.4104597701149952	0.46227352937393046

Table 14: Original data of 4 mutations (a)

Variant	Mean ΔΔG	Std ΔΔG
L10I:I13V:R57G:I64V:L109I:I112V:R156G:I163V	8.577931034482724	0.03289485686052224
L10I:I15V:L63A:I93L:L109I:I114V:L162A:I192L	5.609080459770112	0.3008088360946854
L10=:I13=:R57=:I64=:L109=:I112=:R156=:I163=	-5.2263748884229244e-14	0.19464783353266002
L10=:I15=:L63=:I93=:L109=:I114=:L162=:I192=	2.6134765647907916e-14	0.006493998570961684
T12E:L19I:L63P:I64L:T111E:L118I:L162P:I163L	3.6394252873563233	0.024070513404113692
T12S:I15V:L19I:N37S:T111S:I114V:L118I:N136S	0.7728735632184093	0.015737475018642293
T12=:I15=:L19=:N37=:T111=:I114=:L118=:N136=	0.0	0.2231278744467597
T12=:L19=:L63=:I64=:T111=:L118=:L162=:I163=	-2.6135343889066576e-14	0.02869928736036943
I13=:R41=:D60=:L63=:I112=:R140=:D159=:L162=	-2.6135054768487247e-14	0.003843276666028681
I13V:R41K:D60E:L63P:I112V:R140K:D159E:L162P	2.0988505747126163	0.003985042717993427
K14=:I15=:R41=:L63=:K113=:I114=:R140=:L162=	2.6134765647907916e-14	0.005047011126451167
K14=:L33=:R41=:I64=:K113=:L132=:R140=:I163=	5.226953129581583e-14	0.006177024755194603
K14R:I15V:R41K:L63P:K113R:I114V:R140K:L162P	1.9478160919540652	0.2175357125469659
K14R:L33V:R41K:I64V:K113R:L132V:R140K:I163V	0.5725287356322356	0.1501778572987103
I15=:E35=:R57=:L63=:I114=:E134=:R156=:L162=	5.2273000742767785e-14	0.2615314107741374
I15=:L63=:V77=:I93=:I114=:L162=:V176=:I192=	2.6136500371383893e-14	0.09223018874362064
I15V:E35D:R57G:L63P:I114V:E134D:R156G:L162P	10.861149425287378	0.027704249361543767
I15V:L63A:V77I:I93L:I114V:L162A:V176I:I192L	6.354827586206927	0.25572576431510896
G16E:N37S:I62V:L63A:G115E:N136S:I161V:L162A	4.152413793103471	0.12624530956572363
G16=:N37=:I62=:L63=:G115=:N136=:I161=:L162=	2.3129646346357427e-18	0.06420916408537714
L19I:E35D:R57G:L63A:L118I:E134D:R156G:L162A	11.037011494252843	0.0846428227993725
L19=:E35=:R57=:L63=:L118=:E134=:R156=:L162=	-2.6134765647907916e-14	0.005163466079371176
I32=:V47=:M76=:I82=:I131=:V146=:M175=:I181=	-2.615500408846098e-14	0.4411556844895876
I32V:V47I:M76L:I82V:I131V:V146I:M175L:I181V	-1.3968965517241665	0.11839404079941371
L33=:N37=:R41=:L63=:L132=:N136=:R140=:L162=	0.0	0.0022757459624595695
L33=:R41=:D60=:I64=:L132=:R140=:D159=:I163=	5.226837481349852e-14	0.04754871086451429
L33=:D60=:I62=:L63=:L132=:D159=:I161=:L162=	5.226837481349852e-14	0.07177072610137591
L33V:N37C:R41K:L63P:L132V:N136C:R140K:L162P	4.771954022988514	0.019162681175864605
L33V:R41K:D60E:I64V:L132V:R140K:D159E:I163V	1.1451724137931516	0.0023045905351533067
L33V:D60E:I62V:L63P:L132V:D159E:I161V:L162P	4.373448275862119	0.34640665269533405
E35D:N37D:D60E:I72V:E134D:N136D:D159E:I171V	0.8013793103447732	0.08665584249001536
E35D:R57G:L63P:H69Q:E134D:R156G:L162P:H168Q	11.05482758620686	0.18358441351949967
E35D:R57G:L63P:I72V:E134D:R156G:L162P:I171V	10.150574712643646	0.0920202347826725
E35=:N37=:D60=:I72=:E134=:N136=:D159=:I171=	-2.6136500371383893e-14	0.09319450634390541
E35=:R57=:L63=:H69=:E134=:R156=:L162=:H168=	-5.227068777813315e-14	0.08737250303855114
E35=:R57=:L63=:I72=:E134=:R156=:L162=:I171=	-5.2273000742767785e-14	0.08904700892145087

Table 15: Original data of 4 mutations (b)

Variant	**Mean ΔΔG**	**Std ΔΔG**
M36I:R41K:R57G:I64V:M135I:R140K:R156G:I163V	8.200344827586216	0.004688753575740308
M36I:R41K:L63P:K70R:M135I:R140K:L162P:K169R	0.25850574712638724	0.07566746198864752
M36I:R57K:L63P:I93L:M135I:R156K:L162P:I192L	5.20827586206898	0.176555390886251
M36I:I62V:L63P:I93L:M135I:I161V:L162P:I192L	3.876321839080505	0.36336629463553954
M36=:R41=:R57=:I64=:M135=:R140=:R156=:I163=	3.469446951953614e-18	0.07058238754179857
M36=:R41=:L63=:K70=:M135=:R140=:L162=:K169=	-2.6131874442114622e-14	0.10111167006969017
M36=:R57=:L63=:I93=:M135=:R156=:L162=:I192=	0.0	0.0026611119316491106
M36=:I62=:L63=:I93=:M135=:I161=:L162=:I192=	2.6136500371383893e-14	0.06283457153169238
N37E:L63P:K70T:I72R:N136E:L162P:K169T:I171R	4.363678160919534	0.0062472083869744115
N37=:L63=:K70=:I72=:N136=:L162=:K169=:I171=	-2.6135054768487247e-14	0.01130534418935838
R41K:D60E:I64V:V77I:R140K:D159E:I163V:V176I	0.30862068965516615	0.004424914150561701
R41=:D60=:I64=:V77=:R140=:D159=:I163=:V176=	2.6136500371383893e-14	0.05609218956044877
D60=:Q61=:L63=:V77=:D159=:Q160=:L162=:V176=	0.0	0.06811506278572294
D60E:Q61E:L63P:V77I:D159E:Q160E:L162P:V176I	3.0412643678160727	0.0026160475119061492
L63=:I64=:A71=:V77=:L162=:I163=:A170=:V176=	-5.226989269654e-14	0.0017126051064045374
L63=:H69=:K70=:V77=:L162=:H168=:K169=:V176=	-2.613491020819758e-14	0.0011378729811928243
L63=:I72=:V77=:I93=:L162=:I171=:V176=:I192=	7.840458606430308e-14	0.006360384060818432
L63P:I64L:A71T:V77I:L162P:I163L:A170T:V176I	6.331609195402251	0.26119550953265697
L63P:H69Q:K70R:V77I:L162P:H168Q:K169R:V176I	2.1448275862068535	0.11201124360828474
L63P:I72T:V77I:I93L:L162P:I171T:V176I:I192L	5.465172413793211	0.2589579920441616

Table 16: Original data of 5 mutations (a)

Variant	Mean ΔΔG	Std ΔΔG
T12A:I13V:L63P:I64V:V77I:T111A:I112V:L162P:I163V:V176I	5.551379310344882	0.18043641403873756
T12I:G16E:P39S:L63T:V77I:T111I:G115E:P138S:L162T:V176I	1.6744827586207922	0.1862120759887022
T12N:L63P:H69Q:V77I:I93L:T111N:L162P:H168Q:V176I:I192L	5.588275862068985	0.1961003757742027
T12P:K14R:L63P:G68E:A71T:T111P:K113R:L162P:G167E:A170T	6.067586206896579	0.14415387911553193
T12Q:L19I:M36I:R41K:L63V:T111Q:L118I:M135I:R140K:L162V	-0.2978160919540631	0.11085561852969923
T12R:I15V:L19I:R41K:H69Y:T111R:I114V:L118I:R140K:H168Y	-3.273563218390833	0.1541645429468869
T12=:I13=:L63=:I64=:V77=:T111=:I112=:L162=:I163=:V176=	2.6134982488342413e-14	0.004102659378753184
T12=:K14=:L63=:G68=:A71=:T111=:K113=:L162=:G167=:A170=	2.6134765647907916e-14	0.003516565181802356
T12=:I15=:L19=:R41=:H69=:T111=:I114=:L118=:R140=:H168=	-2.6145752229922437e-14	0.2131990882609573
T12=:G16=:P39=:L63=:V77=:T111=:G115=:P138=:L162=:V176=	7.840458606430308e-14	0.0014448051827571045
T12=:L19=:M36=:R41=:L63=:T111=:L118=:M135=:R140=:L162=	-2.6136500371383893e-14	0.17876139075948805
T12=:L63=:H69=:V77=:I93=:T111=:L162=:H168=:V176=:I192=	-9.25185853854297e-18	0.11081950107243121
I13=:K14=:I15=:I62=:L63=:I112=:K113=:I114=:I161=:L162=	5.2273000742767785e-14	0.22880106497825786
I13=:E35=:N37=:L63=:I64=:I112=:E134=:N136=:L162=:I163=	-5.229150445984487e-14	0.4038766105561881
I13=:N37=:R57=:L63=:H69=:I112=:N136=:R156=:L162=:H168=	-9.25185853854297e-18	0.16328774631362328
I13=:P39=:I62=:L63=:V77=:I112=:P138=:I161=:L162=:V176=	-2.6135054768487247e-14	0.004309961668802623
I13=:R41=:L63=:H69=:I93=:I112=:R140=:L162=:H168=:I192=	-5.2273000742767785e-14	0.1158199841465563
I13V:K14R:I15V:I62V:L63H:I112V:K113R:I114V:I161V:L162H	4.136781609195456	0.0711172006574636
I13V:E35D:N37T:L63P:I64V:I112V:E134D:N136T:L162P:I163V	4.348275862068931	0.003201928339604272
I13V:N37D:R57G:L63S:H69R:I112V:N136D:R156G:L162S:H168R	10.691494252873598	0.002131866320826621
I13V:P39Q:I62V:L63P:V77I:I112V:P138Q:I161V:L162P:V176I	4.732298850574693	0.14205808461848823
I13V:R41K:L63P:H69R:I93L:I112V:R140K:L162P:H168R:I192L	2.4810344827585986	0.5812973585904594
K14=:I15=:G17=:R57=:L63=:K113=:I114=:G116=:R156=:L162=	-2.6134765647907916e-14	0.00804433480775949
K14=:N37=:R41=:I64=:Q92=:K113=:N136=:R140=:I163=:Q191=	0.0	0.003955092076804902
K14R:I15V:G17E:R57K:L63P:K113R:I114V:G116E:R156K:L162P	7.638965517241387	0.25209299590368567
K14R:N37S:R41K:I64V:Q92K:K113R:N136S:R140K:I163V:Q191K	5.378275862068958	0.09292717810578843
I15=:N37=:K70=:V77=:I93=:I114=:N136=:K169=:V176=:I192=	2.6145752229922437e-14	0.20120807026880666
I15=:R41=:L63=:I64=:K70=:I114=:R140=:L162=:I163=:K169=	-7.840950111415168e-14	0.2925146666896562
I15V:N37D:K70R:V77I:I93L:I114V:N136D:K169R:V176I:I192L	1.159195402298881	0.5921848896971807
I15V:R41K:L63A:I64V:K70R:I114V:R140K:L162A:I163V:K169R	3.283678160919434	0.026093219879173398
Q18H:N37S:A71V:V77I:I93L:Q117H:N136S:A170V:V176I:I192L	2.7691954022988092	0.22358026705671186
Q18=:N37=:A71=:V77=:I93=:Q117=:N136=:A170=:V176=:I192=	-2.6135054768487247e-14	0.004959873335752662
L19I:R57K:L63S:I64V:E65D:L118I:R156K:L162S:I163V:E164D	6.510689655172392	0.09793804950319283
L19I:L63S:E65D:I72V:V77I:L118I:L162S:E164D:I171V:V176I	3.2998850574712453	0.26679242971810857
L19=:R57=:L63=:I64=:E65=:L118=:R156=:L162=:I163=:E164=	0.0	0.19012487986368462
L19=:L63=:E65=:I72=:V77=:L118=:L162=:E164=:I171=:V176=	-2.6145752229922437e-14	0.2257284562825066

Table 17: Original data of 5 mutations (b)

Variant	Mean ΔΔG	Std ΔΔG
L33F:K43T:L63C:H69Q:V77I:L132F:K142T:L162C:H168Q:V176I	4.889080459770046	0.07032811010130717
L33=:K43=:L63=:H69=:V77=:L132=:K142=:L162=:H168=:V176=	-2.6145752229922437e-14	0.36567556030577514
E35D:M36I:R57G:L63P:I64L:E134D:M135I:R156G:L162P:I163L	12.937356321839113	0.00652646876624959
E35D:N37D:L63P:A71T:I93L:E134D:N136D:L162P:A170T:I192L	8.182988505747076	0.056787658479933494
E35=:M36=:R57=:L63=:I64=:E134=:M135=:R156=:L162=:I163=	2.8912057932946783e-19	0.0035278182191121224
E35=:N37=:L63=:A71=:I93=:E134=:N136=:L162=:A170=:I192=	-5.226982041639516e-14	0.00399497636709117
M36I:D60E:I62V:L63P:H69Q:M135I:D159E:I161V:L162P:H168Q	2.471264367816118	0.04087916127318032
M36=:D60=:I62=:L63=:H69=:M135=:D159=:I161=:L162=:H168=	5.2269964976684826e-14	0.006978441028353832
N37D:R57G:L63P:V77I:I93L:N136D:R156G:L162P:V176I:I192L	12.057471264367749	0.4155779831916333
N37=:R41=:R57=:L63=:Q92=:N136=:R140=:R156=:L162=:Q191=	-5.2273000742767785e-14	0.09294125221942244
N37=:R41=:R57=:V77=:I93=:N136=:R140=:R156=:V176=:I192=	2.6136500371383893e-14	0.150491070916767
N37=:R57=:Q61=:L63=:I93=:N136=:R156=:Q160=:L162=:I192=	5.2268953054657175e-14	0.040715298108528586
N37=:R57=:L63=:V77=:I93=:N136=:R156=:L162=:V176=:I192=	-2.6131874442114622e-14	0.16775522435752294
N37S:R41K:R57G:L63A:Q92K:N136S:R140K:R156G:L162A:Q191K	17.18620689655167	0.012045406672474926
N37S:R57G:Q61E:L63P:I93L:N136S:R156G:Q160E:L162P:I192L	13.275057471264413	0.49969289022580493
N37Y:R41K:R57K:V77I:I93L:N136Y:R140K:R156K:V176I:I192L	3.6265517241379492	0.12484377871732284
P39=:R41=:I62=:L63=:H69=:P138=:R140=:I161=:L162=:H168=	7.840465834444792e-14	0.0010151449271401153
P39Q:R41K:I62V:L63P:H69Y:P138Q:R140K:I161V:L162P:H168Y	0.7197701149426269	0.007686553332105844
R57G:I62V:L63P:V77I:I93L:R156G:I161V:L162P:V176I:I192L	13.089885057471244	0.3755991206484464
R57G:L63P:I72E:V77I:I93L:R156G:L162P:I171E:V176I:I192L	15.831494252873528	0.2684567838366835
R57G:L63H:V77I:L89I:I93L:R156G:L162H:V176I:L188I:I192L	15.343793103448379	0.4599709906010201
R57=:I62=:L63=:V77=:I93=:R156=:I161=:L162=:V176=:I192=	7.228014483236696e-20	0.0031352141832042394
R57=:L63=:I72=:V77=:I93=:R156=:L162=:I171=:V176=:I192=	0.0	0.15759697977044868
R57=:L63=:V77=:L89=:I93=:R156=:L162=:V176=:L188=:I192=	7.840516430546174e-14	0.006520392921239166
Q61E:I62V:L63T:E65D:V77I:Q160E:I161V:L162T:E164D:V176I	4.322068965517304	0.3845531437387451
Q61=:I62=:L63=:E65=:V77=:Q160=:I161=:L162=:E164=:V176=	5.227068777813315e-14	0.020812534207291705