Project 6: Modelling mutational effects on protein structure

Computational biology and Systems biology 3

Qichen Fu

Word count (including figure legends and references): 2649

Modelling mutational effects on protein structure

Abstract

The mutations can affect protein properties including stability, dimerization and DNA binding and results in protein function changes which eventually lead to phenotypic changes. Predicting the mutational effects on protein structure is essential for genotype-phenotype mapping. Tetracycline repressor (TetR) is a transcriptional repressor that plays an essential role in tetracycline (TC) resistance in E. coli. TetR forms a homodimer and binds tet operators to repress transcription. Since no previous study systematically assesses the mutational effects on TetR, here I apply in-silico deep mutational scanning (DMS) using FoldX. The stability, dimerization and DNA binding affinity of TetR dimer upon mutations were calculated and analyzed. The results provide new insights into how different sites/mutations contribute to changes in TetR protein properties. It also showed that in-silico mutagenesis is a reliable and convenient way to assess mutational effects.

Background, motivation and aims

Mutations in the amino acids can affect protein structures and functions and lead to phenotypic changes. Specifically, mutations can affect protein stability, binding between the subunits and binding with DNA, ligands, peptides etc. Predicting mutational effects is essential for making accurate genetic predictions, interpreting personal genomes, and understanding evolution.

DMS is a new approach that introduces saturation mutation to a protein. Large-scale mutational data provides insights into the mutational effect systematically and genetic interactions (epistasis)[1]. This information can be used to assess how mutations affect protein and predict protein structures[1]. DMS have various advantages compared with traditional protein structure determination techniques such as Cryo-EM and X-ray diffraction since it can be used on all protein regions including disordered regions and is relatively less expensive.

A traditional DMS experiment is performed by first generating large mutation libraries and applying an appropriate selection method based on interested protein properties (e.g. binding, stability)[1,2]. The input and output libraries will be sequenced and the changes in the variant frequency will be analyzed and interpreted[1,2]. However, due to the available software that can directly calculate the total energy and interaction energy of a protein[3], it is possible to perform in-silico mutagenesis.

TetR is a transcription repressor that mediates the TC resistance in E. coli[4]. TetR has a DNA binding domain and an effector binding domain. The DNA binding domain binds with the Tet operator (promoter) and allows TetR to inhibit the expression of tetracycline activator (TetA) and tetracycline repressor (TetR)[4]. The invaded TC can bind with the effector binding domain and prevent TetR from binding with the promoter. This process activates the expression of TetR and TetA. The TetA protein is inserted into

the plasma membrane and can export invaded TC out of the cells[4]. Understanding how mutations affect TetR protein functions will provide potential solutions to antibiotic resistance which is a common issue in healthcare. Although some previous studies assessed the mutational effect on TetR, no systematic mutagenesis has been performed on TetR before.

In this project, I performed in-silico DMS on the TetR homodimer to understand mutational effects. This includes how mutations might affect stability, dimerization, DNA binding on different sites and how they relate to TetR functions.

Materials and methods

PDB files

All PDB files were downloaded from the RCSB database (TetR dimer: 3fk6[5]; TetR-DNA complex: 1qpi[6]). All PDB files were first repaired using the RepairPDB function in FoldX5[3]. TetR dimer from 3fk6 entry contains three mutations and was corrected to wild type using the BuildModel function. Edited PDB files for 3fk6 and 1qpi were available at: https://github.com/QichenFU/CBSB3_miniproject/PDB_files.

In silico deep mutagenesis

In silico deep mutagenesis was performed using the BuildModel function of FoldX5[3]. Each amino acid site was mutated to all 20 types of amino acids on both chains. Energy terms lists and PDB files of all mutants were generated. PDB files were used for subsequent binding energy analysis.

Total energy, binding energy, Van der Waals clashes

The total energy of the mutated TetR dimer was obtained in the BuildModel process. The binding energy between the two subunits of the TetR dimer was analyzed using the AnalyzeComplex function, while binding energy between the TetR dimer and DNA was obtained using the same function with complexwithDNA = true. ΔΔG was calculated as follows:

$$\Delta\Delta G = \Delta G_{Mutant} - \Delta G_{Wild\text{-}type}$$

Van der Waals clashes of mutants were included in the output of the BuildModel function.

Selection of thresholds for ΔΔG (both total and interaction energy)

The ΔΔG predicted by FoldX could be quite high (tens of kcal/mol), making the visualization of full-length mutational scanning difficult since large values mask the smaller but meaningful ΔΔGs. Moreover, proteins are generally considered destabilized if ΔΔG is larger than 10 and stabilized if ΔΔG is smaller than -10. Due to the lack of experimental ΔΔG of TetR in common databases like ProTherm[7], all ΔΔG have presented with an arbitrary threshold of [-10,10]. Note that in some correlation plots, no threshold was set for ΔΔG to avoid points accumulating at 10 or -10 and showing weird patterns.

Discovery of potential dominant mutations in TetR dimer

The mutagenesis was performed on a single chain. All other methods are the same as

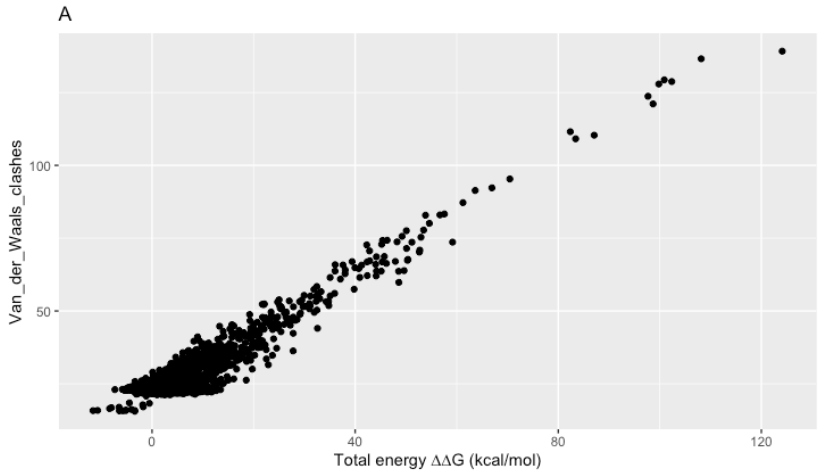
homozygous mutagenesis.

Results

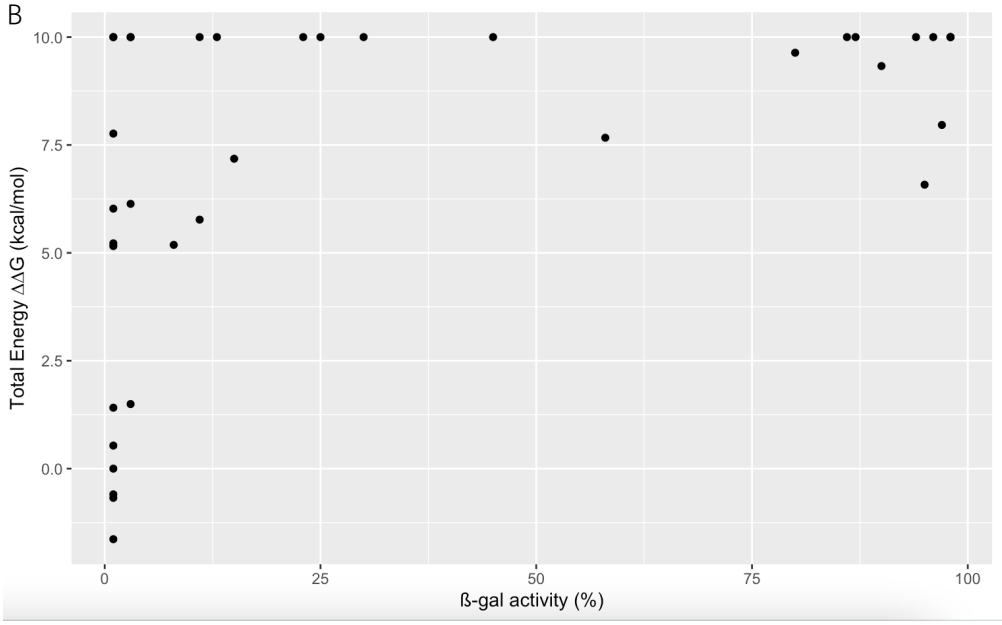
The assessment of the reliability of ΔΔG generated by FoldX

Homozygous mutations were applied to both subunits of TetR dimer. From visual inspection, the mutational effects showed a linear correlation with Van der Waals clashes (Fig. 1A) without obvious outliner, which indicated there was no abnormal and unreasonable mutagenesis.

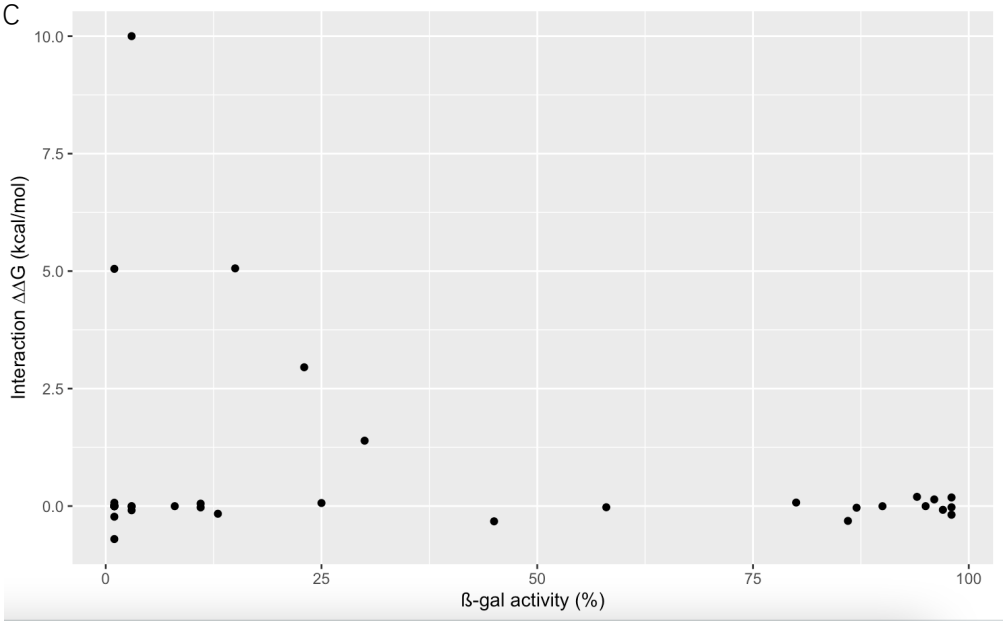
The ΔΔG of both stability and binding obtained from FoldX was compared with the ß-galactosidase activity data from a previous publication[8]. In this study, the authors used ß-galactosidase assays to test the effect of some mutations on TetR function. The stability ΔΔG is positively correlated (Spearman) with ß-galactosidase activity (Fig. 1B, 1D). This showed that stability ΔΔG generated by FoldX can reflect the experimental data. However, there is no correlation between DNA binding affinity and ß-galactosidase activity (Fig. 1C, 1D). This may be because the data only has one mutation on the DNA binding domain (site 17). Other mutations are on the effector binding domain and are less associated with DNA binding, which biases the analysis.



B



C



D

```
> res_spearman$P
                          gal delta_delta_G DNA_binding
gal                        NA  0.0001102747   0.6969832
delta_delta_G 0.0001102747             NA   0.1969182
DNA_binding   0.6969832170  0.1969182069          NA
```

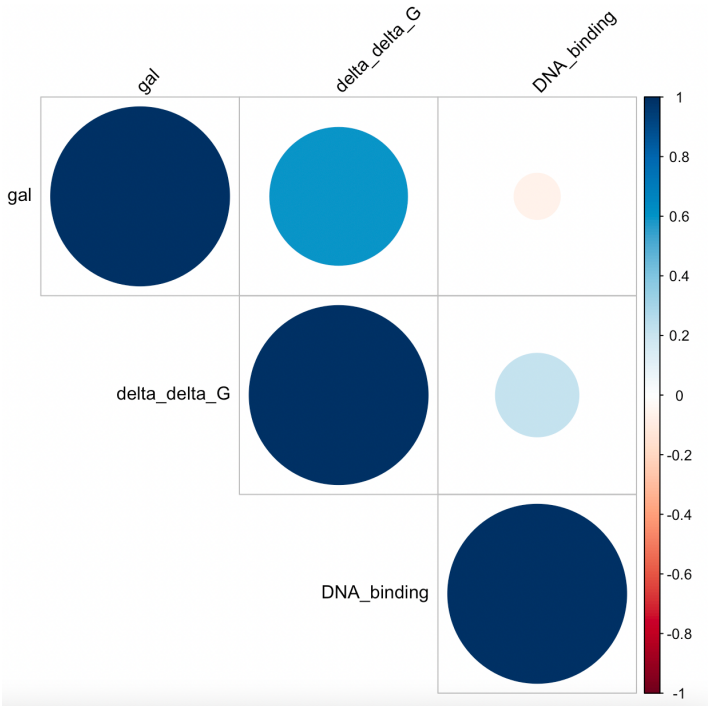


Figure 1. Van der Waals clashes and the correlation between FoldX stability/DNA binding ΔΔG and ß-galactosidase activity. **A**, The correlation between total energy ΔΔG and Van der Waals clashes. **B**, The dot plot of ß-galactosidase activity and total energy ΔΔG. **C**, The dot plot of ß-galactosidase activity and DNA binding ΔΔG. **D**, p-value (upper) and the matrix of correlations (lower) of Spearman correlation.

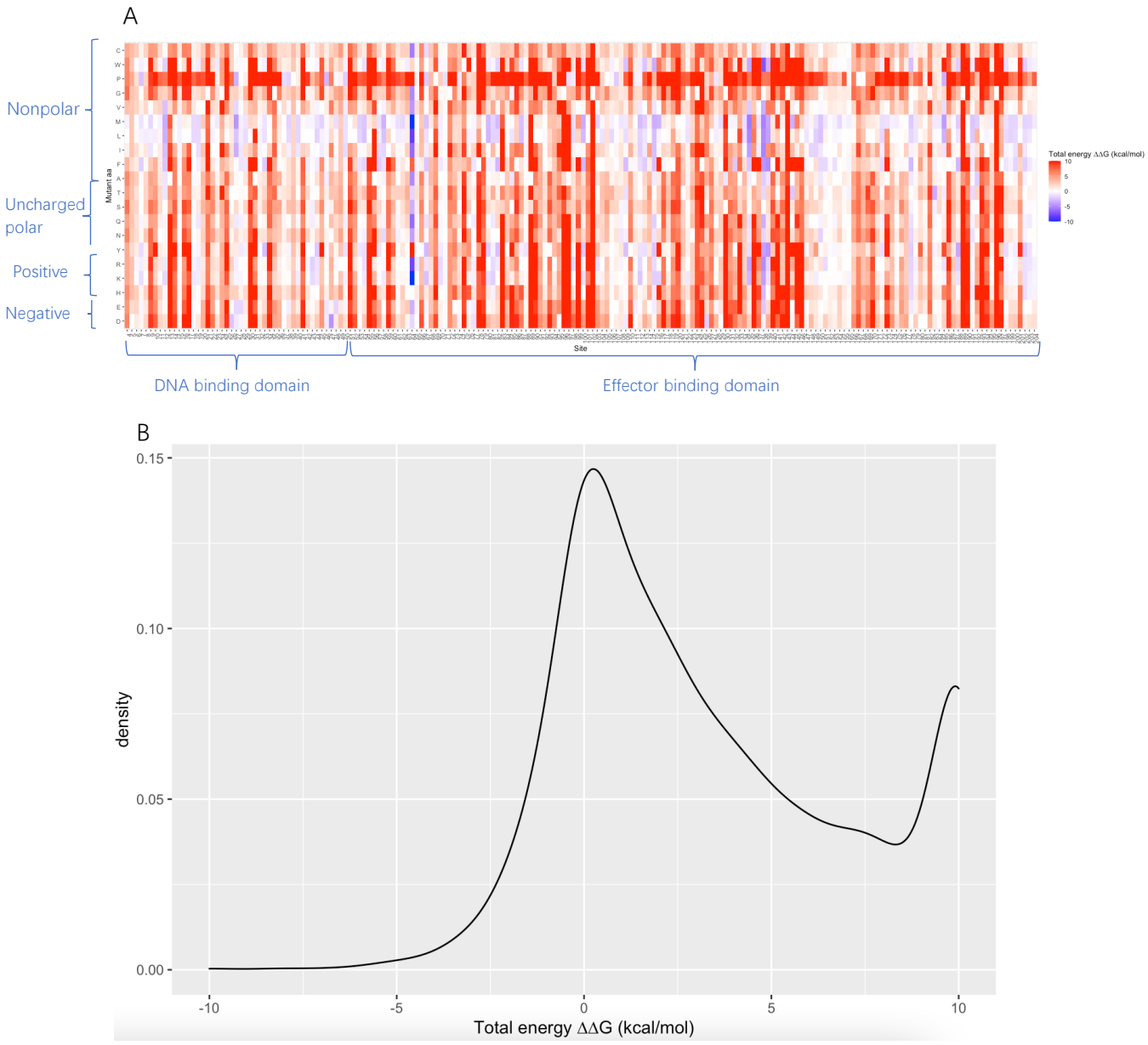
Mutational effects on TetR stability (total energy) predicted by FoldX

Most mutations caused the destabilizing effect on TetR dimer (Fig. 2A, 2B), which makes sense from the perspective of evolution. Mutations on effector binding domains have larger effects on the stability that the DNA binding domain (Fig. 2C). However, the effects are not uniformly distributed along the peptide chains and are also different depending on which amino acid that mutates to (Fig. 2A).

There are several scenarios when looking at the single mutation. Firstly, the ΔΔG was not affected by a mutation which indicates the mutation is not on the critical site for maintaining protein stability. They may locate in flexible regions or at the protein surface, such as sites 71, 72, 108 and 109 as labelled in the 3D structure (Fig. 2D). Secondly, the ΔΔG increases when a certain site is mutated to some amino acids but not others. Large ΔΔG is because of the drastic changes in the chemical properties such as polarity, residuals size etc. Site 116 (glutamine) displays a large ΔΔG when substituting the original side chains with large aromatic side chains (tyrosine, phenylalanine, histidine) (Fig. 2A). Thirdly, the ΔΔG increases regardless of which residual is mutated to. This shows that the sites have special roles in maintaining protein stability. For example, sites 102, 143 and 196 exhibit destabilizing effects on all mutations (Fig. 2A). As labelled in fig. 2D, these sites are located at the binding interfaces between two subunits which play critical roles in dimerization and stability. Fourthly, the mutation leads to weak stabilizing effects in a few sites which are not expected as the residuals are likely to be evolutionarily optimized. Weak stabilization

is observed in sites 64 and 138 (Fig. 2A). However, these effects are very weak (less than 5 kcal/mol). Moreover, FoldX may have a small bias on the results. Further experiments (both computational and experimental) should be done to confirm whether the stabilizing effects exist.

Depending on which amino acid the site is mutated to, it may result in different ΔΔGs. Mutating to proline generally caused destabilizing effects since proline can introduce special structural interactions (Fig. 2A). Proline can form cis bonds while other amino acids generally form planar and trans peptide bonds[9]. Proline can be a breaker of alpha-helix if there is no hydrogen bond donor[9]. Mutating to methionine or leucine generally leads to small destabilizing effects (Fig. 2A) and may be caused by the chemical properties of the residuals.



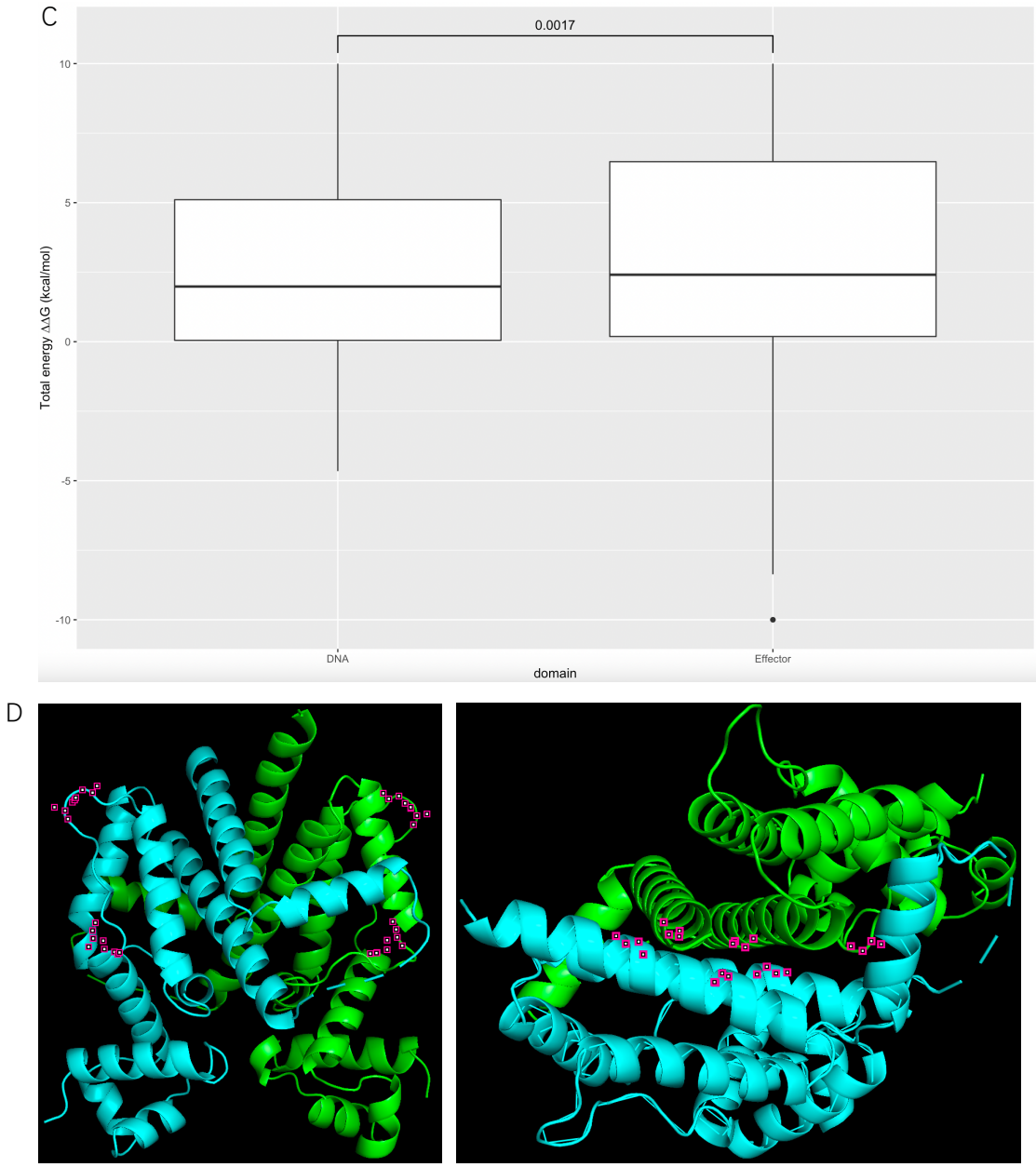


Figure 2. The mutational effects on the stability of TetR dimer. **A**, The heatmap of total energy ΔΔG on all TetR sites (x-axis) and amino acids that mutated to (y-axis). **B**, The accumulative distribution of all ΔΔG. **C**, The comparison of ΔΔG between DNA binding domain and effector binding domain (Wilcoxon Rank Sum test). **D**, Non-effective sites (left, 71, 72, 108 and 109) and destabilizing sites (right, 102, 143 and 196) on TetR dimer.

Mutations affect the interaction between two subunits of TetR dimer

The interaction between two subunits is critical for the TetR dimer to exert its function[4]. To understand the specific sites involved in the interaction, the interaction energy between two subunits was calculated. From visual inspection, the mutations mainly caused effects on the effector binding domains and weaker effects on DNA binding domains (Fig. 2A) which is similar to stability (Fig. 1C). This indicates the effector binding domain is more important than the DNA binding domain for the interaction of

tetR dimer.

From visual inspection of the heatmap (Fig. 1A, 2A) and dot plot (Fig. 2B), the stability and interaction energy changes upon mutations have similar patterns and are mostly positively correlated. The sites that exhibit high ΔΔG are similar which include sites 102, 143 and 196. This shows that generally, sites that are important for stability are also critical for dimerization. Stability and dimerization are likely to be interrelated.

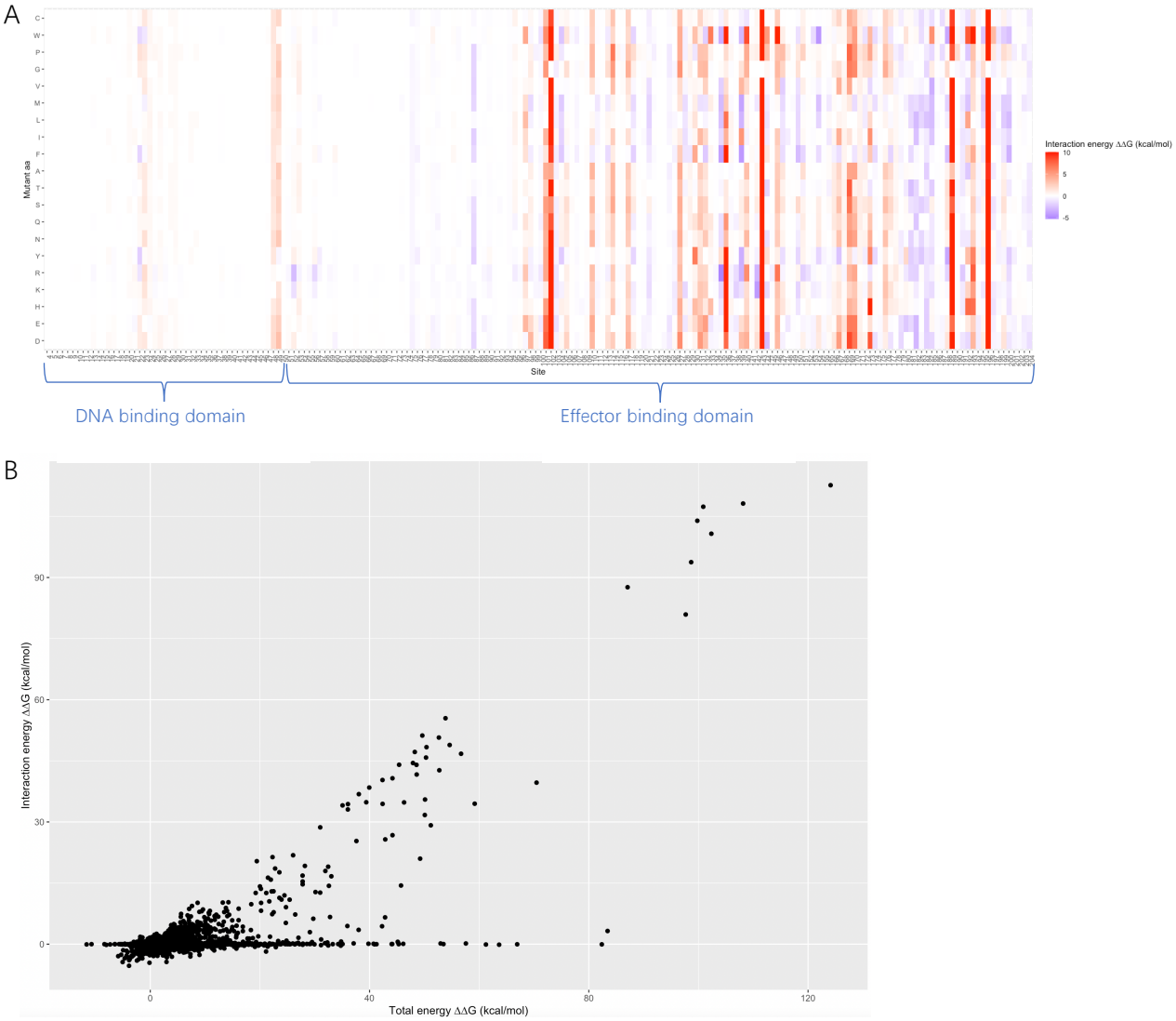


Figure 3. The mutational effects on TetR interaction energy. **A**, The heatmap of interaction ΔΔG on all TetR sites (x-axis) and amino acids that mutated to (y-axis). **B**, The correlation between total energy ΔΔG and interaction energy ΔΔG (ΔΔG not truncated).

Mutations affect TetR binding with DNA

TetR dimer binds with the 15-mer tet operator. The binding is stabilized by the interactions (e.g. hydrogen bonds, van der Waals interactions) between the DNA binding domain (specifically sites 27, 28, 38, 39, 40 and 42) and tet operator[10]. The mutations on the critical site for protein-DNA interaction are expected to greatly decrease the binding affinity between TetR dimer and DNA. In-silico mutagenesis data

shows that there are large increases in DNA binding ΔΔGs in the sites discussed in previous studies[11] (Fig. 4).

Figure 4. Mutational effects on the interaction energy between TetR dimer and DNA

FoldX predicts potential dominant mutations in TetR dimer

Since bacteria only have one copy of the genome, both subunits of TetR dimer carry the same mutations. However, the human genome has two alleles and may have different genotypes which produce heterozygous dimer. It will be interesting to see whether the same mutation has different effects on two chains. Fig. 5A shows that the effects of mutations on chain A are generally consistent with effects on chain B. However, there are a few outliers including sites 89, 102 and 145. These sites are all destabilizing sites as discussed previously or shown in the heatmap (Fig. 2A). This may indicate that mutations in one subunit may have dominant effects on the stability if the mutation is in the critical site. However, this contradicts the fact the mutational effects of these special mutations are not additive (Fig. 5B). The homozygous mutation ΔΔG is substantially lower or higher than the sum of heterozygous mutation ΔΔG, as shown in the green box of fig. 5B. This makes it difficult to determine exactly which allele contributes to the destabilization. It is also possible that the effect is technical. As previously discussed, the ΔΔG larger than 10 kcal/mol indicates the proteins are destabilized. It remains to be investigated whether the high ΔΔG predicted by FoldX indeed reflect that the mutations on these sites are more deleterious.

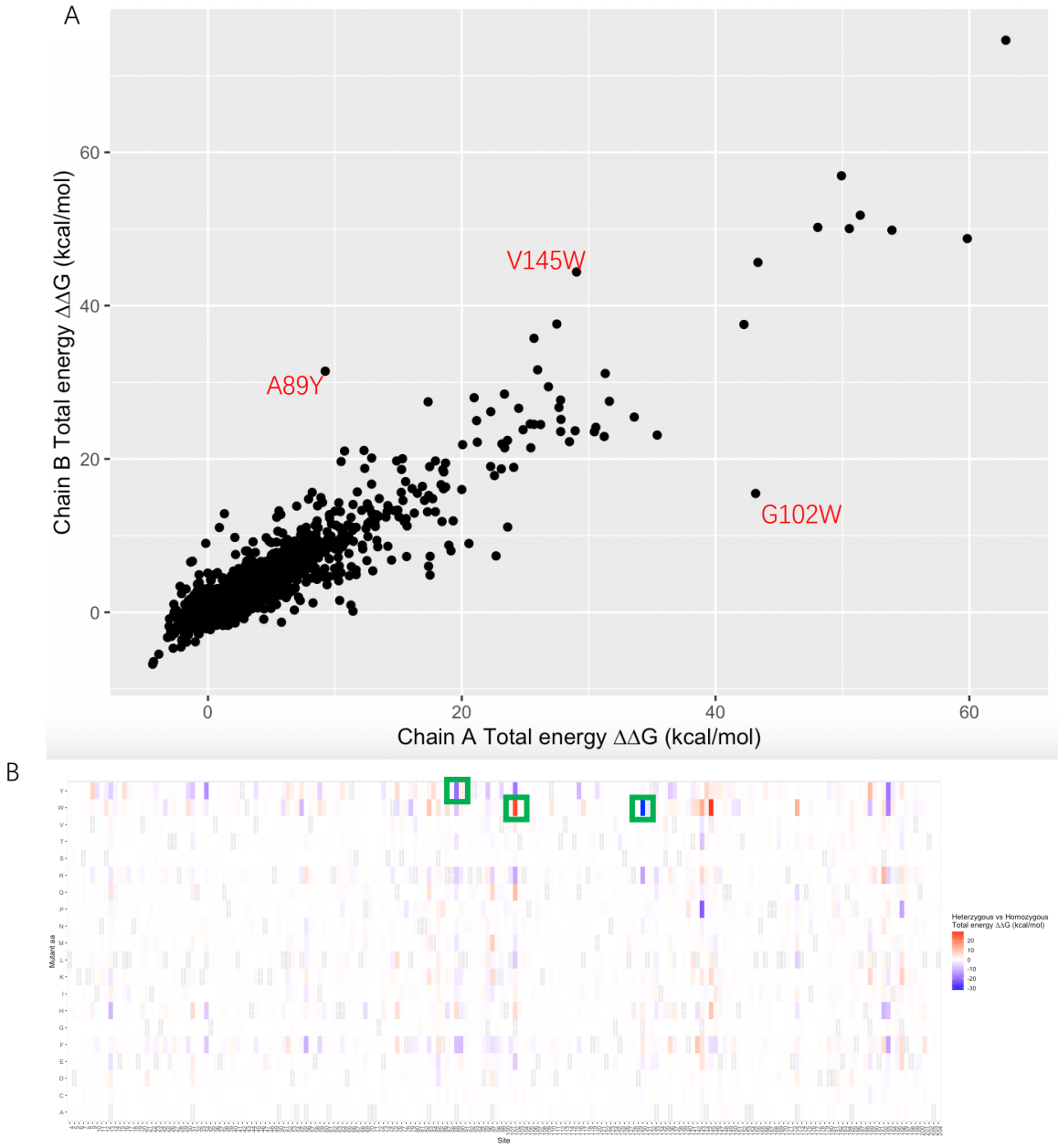


Figure 5. The potential dominant mutations in FoldX. **A**, The correlation between total energy of heterozygous mutation on chain A and chain B (ΔΔG not truncated). **B**, The heatmap of the sum of heterozygous mutation ΔΔG subtracted by homozygous mutation ΔΔG. The green box shows the outlier mutations (ΔΔG not truncated).

Discussion and critical evaluation

In this project, I assessed how mutations affect TetR stability, dimerization and DNA binding using in-silico deep mutagenesis. Most results agree with previous studies and consensus. FoldX can serve as a great tool for mutational effect prediction. The data also provided new insights into TetR protein structure. This includes different susceptibility to destabilizing or de-dimerizing proteins on different sites and domains and the possible reasons as well as the positive correlation between dimerization and stability. It also provides some ambiguous insights into dominant mutations in TetR

dimer.

However, the lack of experimental ΔΔGs affects the assessment of the accuracy of the FoldX. It would be better to test the correlation between experimental ΔΔG (e.g. temperature melt-based method) and FoldX ΔΔG rather than the ß-galactosidase activity which can be affected by both stability and DNA binding. This will make the FoldX data more confident. The experimental ΔΔGs will also provide more information on the distribution of ΔΔG which will help set more accurate thresholds for ΔΔGs. The biological significance of the extremely high ΔΔG produced by FoldX also remains to be investigated. Moreover, since the BuildModel process is stochastic and generates different ΔΔG in different runs, it will be more reliable to perform the mutagenesis several times and take the mean value. However, due to the time limitation, I only performed the mutagenesis once.

Further studies such as DMS experiments can be done to confirm and expand the computational analysis results. It will also be interesting to test the difference between TetR isoforms[12]. In this project, 3fk6 (fusion between the DNA binding domain of TetR(B) and effector binding domain of TetR(D)[5]) is used to perform in-silico deep mutagenesis. It will also be interesting to explore other TetR variants to see whether the results are similar or different.

Author contributions

Qichen Fu performed all computational analysis and data interpretations with the guidance of Dr Xianghua Li. Classmate Yuyun Wu had a few discussions with Qichen Fu about the project.

References

1. Fowler, D. M. & Fields, S. Deep mutational scanning: A new style of protein science. *Nature Methods* vol. 11 801–807 (2014).
2. Diss, G. & Lehner, B. The genetic landscape of a physical interaction. (2018) doi:10.7554/eLife.32472.001.
3. Delgado, J., Radusky, L. G., Cianferoni, D. & Serrano, L. FoldX 5.0: Working with RNA, small molecules and a new graphical interface. *Bioinformatics* **35**, 4168–4169 (2019).
4. Saenger, W., Orth, P., Kisker, C., Hillen, W. & Hinrichs, W. The Tetracycline Repressor—A Paradigm for a Biological Switch. *Angewandte Chemie International Edition* **39**, 2042–2052 (2000).
5. Klieber, M. A. *et al.* Structural origins for selectivity and specificity in an engineered bacterial repressor-inducer pair. *FEBS Journal* **276**, 5610–5621 (2009).
6. Nikam, R., Kulandaisamy, A., Harini, K., Sharma, D. & Michael Gromiha, M. ProThermDB: Thermodynamic database for proteins and mutants revisited after 15 years. *Nucleic Acids Research* **49**, D420–D424 (2021).
7. Scholz, O. *et al.* Activity reversal of Tet repressor caused by single amino acid exchanges. *Molecular Microbiology* **53**, 777–789 (2004).

8. Brandl, C. J. & Debertt, C. M. *Hypothesis about the function of membrane-buried proline residues in transport proteins*. *Proc. Nati. Acad. Sci. USA* vol. 83 https://www.pnas.org (1986).

9. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. *letters Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system*. http://structbio.nature.com (2000).

10. Hinrichs, W. *et al. Structure of the Tet Repressor-Tetracycline Complex and Regulation of Antibiotic Resistance*.

11. Yu, Z., Reichheld, S. E., Savchenko, A., Parkinson, J. & Davidson, A. R. A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *Journal of Molecular Biology* **400**, 847–864 (2010).

Code and data availability

All edited PDB files, code and data for reproducing the figures were deposited in the following GitHub repository:

https://github.com/QichenFU/CBSB3_miniproject